Revisiting Asymptotic Theory for Principal Component Estimators of Approximate Factor Models

Peiyun Jiang[†], Yoshimasa Uematsu[*], and Takashi Yamagata[‡]

[†]*Graduate School of Management, Tokyo Metropolitan University*
[*]*Department of Social Data Science, Hitotsubashi University*
[‡]*Department of Economics and Related Studies, University of York*
[‡]*Institute of Social and Economic Research, Osaka University*

April 30, 2024

Abstract

It is well-known that the approximate factor models have the rotation indeterminacy. It has been considered that the principal component (PC) estimators estimate some rotations of the true factors and factor loadings, but the rotation matrix commonly used in the literature depends on the PC estimator itself. This raises a question: what does the PC estimator consistently estimate? This paper aims to explore the answer. We first show that, assuming a quite general weak factor model with the r signal eigenvalues diverging possibly at different rates, there always exists a unique rotation matrix composed only of the true factors and loadings, such that it rotates the true model to the identifiable model satisfying the standard r^2 restrictions. We call the rotated factors and loadings the pseudo-true parameters. We next establish the consistency and asymptotic normality of the PC estimator for this pseudo-true parameter. The results give an answer for the question: the PC estimator consistently estimates the pseudo-true parameter. We also investigate similar problems in the factor augmented regression. Finite sample experiments confirm the excellent approximation of the theoretical results.

Keywords. Weak factor model, Rotation matrix, Consistency and asymptotic normality, Factor augmented regression.

1 Introduction

High-dimensional factor models have become an increasingly important analytical tool for psychology, finance, economics, biology, and so on. They are quite useful in reducing the high dimensionality of data by the low-rank approximation. For example, Chamberlain and Rothschild (1983) first introduce the approximate factor model to finance, and estimation and inferential methods are developed by Connor and Korajczyk (1986, 1993), Stock and Watson (2002a,b), Bai and Ng (2002), and Bai (2003), Fan et al. (2013), among many others.

[*]Correspondence: Yoshimasa Uematsu, Department of Social Data Science, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan (E-mail: yoshimasa.uematsu@r.hit-u.ac.jp).

1.1 Factor model and estimator

Consider the high-dimensional data matrix $\mathbf{X} \in \mathbb{R}^{T\times N}$ generated by the latent factor model:

$$\mathbf{X} = \mathbf{F}^*\mathbf{B}^{*\prime} + \mathbf{E}, \tag{1}$$

where $\mathbf{F}^* = (\mathbf{f}_1^*, \ldots, \mathbf{f}_T^*)'$ with $\mathbf{f}_t^* \in \mathbb{R}^r$ is a matrix of latent factors, $\mathbf{B}^{*\prime} = (\mathbf{b}_1^*, \ldots, \mathbf{b}_N^*)$ with $\mathbf{b}_i^* \in \mathbb{R}^r$ is a matrix of factor loadings, $\mathbf{E} \in \mathbb{R}^{T\times N}$ is an idiosyncratic error matrix with $\mathbb{E}[\mathbf{E}] = \mathbf{0}$, and $N, T \to \infty$ while r is fixed. Throughout the paper, we suppose that $\mathbf{B}^*$ and $\mathbf{F}^*$ are of full column rank and the largest eigenvalue of $\mathbb{E}[T^{-1}\mathbf{E}'\mathbf{E}]$ is uniformly bounded in N and T, but we do not require any specific structure in $(\mathbf{F}^*, \mathbf{B}^*)$, such as diagonality of $\mathbf{B}^{*\prime}\mathbf{B}^*$ and/or $\mathbf{F}^{*\prime}\mathbf{F}^*$, or sparseness of $\mathbf{B}^*$. For the later use, define λ_k as the kth largest eigenvalue of the signal part, $T^{-1}\mathbf{F}^*\mathbf{B}^{*\prime}\mathbf{B}^*\mathbf{F}^{*\prime}$, and set $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$.

The *principal component* (PC) estimator, $(\hat{\mathbf{F}}, \hat{\mathbf{B}})$, is defined as a minimizer of $\|\mathbf{X} - \mathbf{F}\mathbf{B}'\|_{\mathrm{F}}^2$ subject to the r^2 restrictions, $T^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_r$ and $\mathbf{B}'\mathbf{B} \in \mathcal{D}(r)$, where $\mathcal{D}(r)$ is a set of all the $r \times r$ diagonal matrices with r positive diagonal entries in non-increasing order. The constraint minimization problem reduces to the eigenvalue problem of $T^{-1}\mathbf{X}\mathbf{X}'$; the factor estimator $\hat{\mathbf{F}} \in \mathbb{R}^{T\times r}$ is obtained as $\sqrt{T}$ times the r eigenvectors associated with the r largest eigenvalues of $T^{-1}\mathbf{X}\mathbf{X}'$, and the loading estimator $\hat{\mathbf{B}} \in \mathbb{R}^{N\times r}$ is given by $\hat{\mathbf{B}} = T^{-1}\mathbf{X}'\hat{\mathbf{F}}$. By the construction, we can easily check $T^{-1}\hat{\mathbf{F}}'\hat{\mathbf{F}} = \mathbf{I}_r$ and $\hat{\mathbf{B}}'\hat{\mathbf{B}} \in \mathcal{D}(r)$ with its kth diagonal element equal to $\hat{\lambda}_k$, the kth largest eigenvalue of $T^{-1}\mathbf{X}\mathbf{X}'$. Setting $\hat{\mathbf{\Lambda}} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_r)$, we can write $\hat{\mathbf{B}}'\hat{\mathbf{B}} = \hat{\mathbf{\Lambda}}$.

1.2 Problem statement: What does the PC estimator estimate?

The large body of influential literature, including Bai (2003), Bai and Ng (2002, 2006, 2023), analyses the asymptotic properties of the PC estimator $(\hat{\mathbf{F}}, \hat{\mathbf{B}})$ relative to the "rotated parameter" $(\mathbf{F}^*\hat{\mathbf{H}}, \mathbf{B}^*\hat{\mathbf{H}}'^{-1})$, where[1]

$$\hat{\mathbf{H}} = \mathbf{B}^{*\prime}\mathbf{B}^*(T^{-1}\mathbf{F}^{*\prime}\hat{\mathbf{F}})\hat{\mathbf{\Lambda}}^{-1}. \tag{2}$$

Typically, they discuss the asymptotic approximations,

$$\hat{\mathbf{f}}_t = \hat{\mathbf{H}}'\mathbf{f}_t^* + o_p(1), \qquad \hat{\mathbf{b}}_i = \hat{\mathbf{H}}^{-1}\mathbf{b}_i^* + o_p(1), \tag{3}$$

and their associated asymptotic normality. However, it is questionable whether the results in (3) really establish the *consistency* of the PC estimator, because the "rotated parameter" $(\hat{\mathbf{H}}'\mathbf{f}_t^*, \hat{\mathbf{H}}^{-1}\mathbf{b}_i^*)$ depends on the PC estimator itself (and thus depends on data $\mathbf{X}$) through $\hat{\mathbf{H}}$. Approximation (3) literary says "each of $\hat{\mathbf{F}}$ and $\hat{\mathbf{B}}$, which is a function of $\mathbf{X}$, approximates another function of $\mathbf{X}$." This does not reveal which parameters in model (1) the PC estimators are actually estimating!

In view of this problem, we want a new rotation matrix that does not depend on data and that rotates the true parameter to an identifiable one. Precisely, we want to construct rotation matrix $\mathbf{H}$ composed only of the true parameter $(\mathbf{F}^*, \mathbf{B}^*)$ such that the rotated

[1] The data-dependent rotation matrix $\hat{\mathbf{H}}$ defined by (2) is usually denoted without the 'hat' in the literature; see Bai (2003) and Bai and Ng (2023), for example.

parameters, $\mathbf{F}^0 := \mathbf{F}^*\mathbf{H}$ and $\mathbf{B}^0 := \mathbf{B}^*\mathbf{H}'^{-1}$, satisfy the r^2 restrictions,

$$\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{F}^0 = \mathbf{I}_r \quad \text{and} \quad \mathbf{B}^{0\prime}\mathbf{B}^0 \in \mathcal{D}(r). \tag{4}$$

Thus the *pseudo-true* parameters $\mathbf{F}^0$ and $\mathbf{B}^0$ are separately identifiable. Then it is natural to investigate the asymptotic behavior of $(\hat{\mathbf{F}}, \hat{\mathbf{B}})$ relative to $(\mathbf{F}^0, \mathbf{B}^0)$ since the PC estimator is obtained under the same restrictions. This approach is employed by Uematsu and Yamagata (2023a,b), but they do not specify such matrix $\mathbf{H}$. The first contribution of this article is, under very mild conditions on model (1), to show the existence and its uniqueness (up to sign indeterminacy) of $\mathbf{H}$ that rotates true parameter $(\mathbf{F}^*, \mathbf{B}^*)$ to $(\mathbf{F}^0, \mathbf{B}^0)$ satisfying (4). This means that we can *always* rotate model (1) to the identifiable pseudo-true model,

$$\mathbf{X} = \mathbf{F}^0\mathbf{B}^{0\prime} + \mathbf{E}, \tag{5}$$

satisfying (4). As the second contribution, we will prove consistency and asymptotic normality of the PC estimator for the pseudo-true parameter. In light of the new asymptotic results, we can say that the PC estimator consistently estimates the pseudo-true model in (5) with (4). This is the first article to establish the asymptotic theory of the PC estimator for the pseudo-true parameter that does not depend on data.

There are two remarks on our approach. First, Bai and Ng (2013) directly impose the r^2 restrictions on the structural parameters: $T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^* = \mathbf{I}_r$ and $\mathbf{B}^{*\prime}\mathbf{B}^* \in \mathcal{D}(r)$, and prove $\hat{\mathbf{H}} \to_p \mathbf{I}_r$. The same strategy is taken by many authors, including Freyaldenhoven (2022). This approach is clearly different from ours, and seems restrictive since there is no evidence that the true structural model satisfies such artificial restrictions. Second, our new asymptotic results can be used for inference. Thanks to the obtained asymptotic normality of the PC estimator, testing for general statistical hypotheses for $(\mathbf{F}^0, \mathbf{B}^0)$ becomes feasible. Apparently, as long as the asymptotic normality is considered in relation to data-dependent rotations, we may encounter difficulties with hypothesis tests.

1.3 Weak factor models: Allowing different rates for signal eigenvalues

The majority of the literature on latent factor models, including Stock and Watson (2002a,b), Bai (2003), Bai and Ng (2002, 2006, 2013), have employed the *strong factor* (SF) models, where the signal eigenvalues satisfy $\lambda_k \asymp N$ for all $k = 1, \dots, r$. This requirement is somewhat strong in view of real data. Therefore, in the present study, the new asymptotic properties of the PC estimator are derived for the *weak factor* (WF) models that have possibly different divergence rates for the signal eigenvalues: $\lambda_k \asymp N^{\alpha_k}$ with $0 < \alpha_k \leq 1$ for $k = 1, \dots, r$. Observe that $\alpha_r = 1$ reduces to the SF model.

Recently, a growing body of literature has turned its attention to the WF models and has provided their empirical support. Uematsu and Yamagata (2023a,b) and Wei and Zhang (2023) consider the WF models induced by sparse factor loadings while in this article we allow the WF models with non-sparse loadings. Onatski (2010) and Freyaldenhoven (2022) propose methods to determine the number of factors for the WF models. Bai and Ng (2023, Section 5) consider a model with "weaker loadings," essentially assuming that $\mathbf{B}^{*\prime}\mathbf{B}^*$ is asymptotically diagonal with the elements possibly diverging at different rates. The rates coincide the signal eigenvalues if $\|T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^*\|_2$ is assumed to be bounded, and the model reduces to the WF model in such a case. However, the assumption is stringent since it directly imposes the restriction on the true parameter, $(\mathbf{F}^*, \mathbf{B}^*)$. They consider the PC estimator in relation to various data-dependent rotation matrices, but are silent about asymptotic normality of $\hat{\mathbf{b}}_i$ and $\hat{\mathbf{f}}_t'\hat{\mathbf{b}}_i$, and factor augmented regressions under (3), though they derive the asymptotic normality of some rotation of $\hat{\mathbf{b}}_i$. This article fill this important gap in the literature by providing these results.

1.4 Factor-augmented regression

Our approach can apply to the factor augmented regressions, considered in Bai and Ng (2006), Stock and Watson (2002a,b) and Ludvigson and Ng (2009), among others. They are widely used in situations with a large number of predictors. Assuming that certain unobserved common factors influence the movement of the predictors, the factors extracted from the predictors can be used to forecast a particular series. To illustrate, consider a simple factor augmented predictive regression model, $y_{t+1} = \boldsymbol{\gamma}^{*\prime}\mathbf{f}_t^* + \epsilon_{t+1}$, where y_t is a variable of interest and ϵ_t is an error term. As the predictive factor $\mathbf{f}_t^*$ is not observable, it is usually replaced by the PC estimator $\hat{\mathbf{f}}_t$ extracted from a larger set of predictors, $\mathbf{X}$. Bai and Ng (2006) employ the approximation (3) so that $\boldsymbol{\gamma}^{*\prime}\mathbf{f}_t^* = \boldsymbol{\gamma}_{\hat{\mathbf{H}}}'\hat{\mathbf{f}}_t + o_p(1)$, where $\boldsymbol{\gamma}_{\hat{\mathbf{H}}} = \hat{\mathbf{H}}^{-1}\boldsymbol{\gamma}^*$. Although this result is theoretically important and interesting, this approximation literally tells that the coefficient parameter $\boldsymbol{\gamma}_{\hat{\mathbf{H}}}$ on the PC estimator $\hat{\mathbf{f}}_t$ is viewed as a function of the PC estimator $(\hat{\mathbf{F}}, \hat{\mathbf{B}})$ itself, through $\hat{\mathbf{H}}$.

A t-test for the significance of the kth PC factor in the regression is routinely reported in empirical studies. However, due to the problem mentioned above, it does not seem to be asymptotically justified; see Ludvigson and Ng (2009, Table 2) as an example. Testing which factors are significant in factor augmented regressions is of great importance. This is because the extracted PC factors are ordered by their importance in the covariability of predictors, which may not correspond to their forecast power for the particular series of interest; see further discussions in Bai and Ng (2008, 2009) and Cheng and Hansen (2015). It is worth noting that using our approximation of the PC estimator to $(\mathbf{F}^0, \mathbf{B}^0)$, such t-tests are asymptotically justified because $\boldsymbol{\gamma}^{0\prime}\hat{\mathbf{f}}_t = \boldsymbol{\gamma}^{0\prime}\mathbf{f}_t^0 + o_p(1)$ with $\boldsymbol{\gamma}^0 = \mathbf{H}^{-1}\boldsymbol{\gamma}^*$, which is a function of the true parameters $(\mathbf{F}^*, \mathbf{B}^*, \boldsymbol{\gamma}^*)$. We provide a formal analysis of this, which is new to the literature.

1.5 Organization and notations

The rest of the paper is structured as follows. In section 2 the rotation matrix $\mathbf{H}$, which is purely a function of signals $(\mathbf{F}^*, \mathbf{B}^*)$, is derived and its uniqueness is proved. In section 3 the weak factor model is formally introduced and assumptions are made, then consistency of the PC factor is shown in sections 4 and its asymptotic normality is proved in section 5. Section 6 discusses the asymptotic properties of the estimators of factor augmented prediction regressions. Section 7 discuss the experimental design and summarize the results. Section 8 contains some concluding remarks.

Denote by $\lambda_k[\mathbf{A}]$ the kth largest eigenvalue of a square matrix $\mathbf{A}$. For any matrix $\mathbf{M} = (m_{ti}) \in \mathbb{R}^{T\times N}$, we define the Frobenius norm and ℓ_2-induced (spectral) norm as $\|\mathbf{M}\|_{\mathrm{F}} = (\sum_{t,i} m_{ti}^2)^{1/2}$ and $\|\mathbf{M}\|_2 = \lambda_1^{1/2}(\mathbf{M}'\mathbf{M})$, respectively. We denote by $\mathbf{I}_N$ and $\mathbf{0}_{T\times N}$ the $N\times N$ identity matrix and $T\times N$ zero matrix, respectively. We use $\lesssim$ ($\gtrsim$) to represent $\leq$ ($\geq$) up to a positive constant factor. For any positive sequences a_n and b_n, we write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $a_n \gtrsim b_n$. All asymptotic results are for cases where $N, T \to \infty$, and we do not specifically mention it.

2 Derivation of the Rotation Matrix

We derive the explicit form of rotation matrix $\mathbf{H}$ that rotates (1) to (5) with (4) under a quite general assumption. For this purpose, we consider the eigenvalue problem of the $r \times r$ matrix $\mathbf{B}^{*\prime}\mathbf{B}^{*}\left(T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^{*}\right)$. Let $\boldsymbol{\Lambda}$ and $\mathbf{P}$ denote the $r \times r$ diagonal matrix containing the eigenvalues of $\mathbf{B}^{*\prime}\mathbf{B}^{*}\left(T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^{*}\right)$ in descending order and the $r \times r$ matrix whose columns are composed of the corresponding normalized eigenvectors, respectively. Then, we may write

$$\mathbf{B}^{*\prime}\mathbf{B}^{*}\left(T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^{*}\right)\mathbf{P} = \mathbf{P}\boldsymbol{\Lambda}. \tag{6}$$

To progress further, we impose minimal conditions on (6).

Assumption 1. (i) The smallest eigenvalues of $\mathbf{B}^{*\prime}\mathbf{B}^{*}$ and $T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^{*}$ are bounded away from zero;
(ii) All the r diagonal elements of $\boldsymbol{\Lambda}$ are distinct.

Assumption 1 is very mild. Under Assumption 1, all the diagonal elements in $\boldsymbol{\Lambda}$ are positive, bounded away from zero, and distinct. Moreover, Assumption 1(ii) implies the linear independence of the column vectors in $\mathbf{P}$, so that $\mathbf{P}^{-1}$ is well-defined. Let $\mathbf{U} = \mathbf{P}^{-1}\mathbf{B}^{*\prime}\mathbf{B}^{*}\mathbf{P}^{-1\prime}$ and $\mathbf{V} = \mathbf{P}'\left(T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^{*}\right)\mathbf{P}$. Then (6) immediately implies $\mathbf{U}\mathbf{V} = \boldsymbol{\Lambda}$. The first result is obtained as follows.

Lemma 1. *If Assumption 1 is true, then $\mathbf{U}$ and $\mathbf{V}$ are invertible diagonal matrices.*

Theorem 1. *If Assumption 1 is true, then the rotation matrix that rotates (1) to (5) satisfying (4) uniquely exists up to column sign change, and is explicitly given by*

$$\mathbf{H} = \mathbf{P}\mathbf{V}^{-1/2}. \tag{7}$$

In particular, we have $\mathbf{B}^{0\prime}\mathbf{B}^{0} = \boldsymbol{\Lambda}$ in (4).

We emphasize that $\mathbf{H}$ consists only with the true parameters, $\mathbf{F}^{*}$ and $\mathbf{B}^{*}$, while $\hat{\mathbf{H}}$ defined in (2) depends on the PC estimators, $\hat{\mathbf{F}}$ and $\hat{\mathbf{B}}$. By (7) and Lemma 1, $\mathbf{H}$ is viewed as a matrix of eigenvectors associated with $\boldsymbol{\Lambda}$. Thus, it can be written as

$$\mathbf{H} = \mathbf{B}^{*\prime}\mathbf{B}^{*}(T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^{0})\boldsymbol{\Lambda}^{-1},$$

which can be regarded as a population version of $\hat{\mathbf{H}}$.

Theorem 1 indicates that, under a very mild assumption, the true structural model in (1) can *always* be rotated to the pseudo-true model (5) satisfying the conditions in (4). These two models, (1) and (5) with (4), are observationally equivalent. Thus, considering model (5) with (4) is equivalent to choosing the specific rotation matrix $\mathbf{H}$ defined by (7) with implicitly assuming the true model in (1).

3 Weak Factor Models

We formulate the WF models in preparation for deriving the asymptotic theory. We start with introducing the assumptions on the idiosyncratic error term.

Assumption 2 (Idiosyncratic errors)**.** For some constant $M < \infty$ that does not depending on N and T, we have:
(i) $\mathbb{E}[e_{t,i}|\mathbf{b}_i^{*}, \mathbf{f}_t^{*}] = 0$ and $\mathbb{E}[e_{t,i}^4] \leq M$ for all i and t;

(ii) $\mathbb{E}[\{N^{-\frac{1}{2}}\sum_{i=1}^{N}(e_{t,i}e_{s,i}-\mathbb{E}[e_{t,i}e_{s,i}])\}^2]\leq M$ for all t and s;
(iii) For all i, $\|\mathbb{E}[e_{s,i}e_{t,i}]\|\leq|\gamma_{s,t}|$ for some $\gamma_{s,t}$ such that $\sum_{t=1}^{T}|\gamma_{s,t}|\leq M$;
(iv) For all t, $|\mathbb{E}[e_{t,i}e_{t,j}]|\leq|\tau_{i,j}|$ for some $\tau_{i,j}$ such that $\sum_{j=1}^{N}|\tau_{i,j}|\leq M$;
(v) $\|\mathbf{E}\|_2^2=O_p(\max\{N,T\})$;
(vi) The minimum eigenvalue of $\mathbf{\Sigma}_e=\mathbb{E}[\mathbf{e}_t\mathbf{e}_t']$ is bounded away from zero.

The weak cross-sectional and serial correlations in Assumption 2(ii), (iii) and (iv) are similar to Bai (2003, Assumption C). Assumption 2(i), (v), and (vi) are also frequently imposed. We then strengthen Assumption 1 to explicitly characterize the WF models. Recall $\mathbf{\Lambda}=\operatorname{diag}(\lambda_1,\ldots,\lambda_r)$ with λ_k the kth largest eigenvalue of $T^{-1}\mathbf{F}^*\mathbf{B}^{*\prime}\mathbf{B}^*\mathbf{F}^{*\prime}$. The next assumption is crucial to characterize the WF models.

Assumption 3 (Signal strength)**.** There exist random or non-random variables $d_1,\ldots,d_r>0$ and constants $0<\alpha_r\leq\cdots\leq\alpha_1\leq 1$ such that $\lambda_k=d_k^2N^{\alpha_k}$ for $k=1,\ldots,r$ with ordered $0<\lambda_r<\cdots<\lambda_1$ for large N. If d_k's are random, we have $\mathbb{E}[d_k^4]\leq M$ for all k.

Hereafter, denote $\mathbf{N}=\operatorname{diag}(N^{\alpha_1},\ldots,N^{\alpha_r})$ and $\mathbf{D}=\operatorname{diag}(d_1,\ldots,d_r)$, so that we can write $\mathbf{\Lambda}=\mathbf{D}^2\mathbf{N}$. Under this condition, all the signal eigenvalues, $\lambda_1,\ldots,\lambda_r$, are distinct and $\lambda_k\asymp N^{\alpha_k}$. If $\alpha_r=1$, this reduces to the SF model. Compared with Assumption 3, Bai and Ng (2023, Section 5) and Freyaldenhoven (2022) impose the conditions that $\mathbf{N}^{-1/2}\mathbf{B}^{*\prime}\mathbf{B}^*\mathbf{N}^{-1/2}$ tends to a diagonal matrix. They additionally suppose that $T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^*$ converges (in probability) to a positive definite matrix and $\mathbf{I}_r$, respectively. These conditions lead to $\lambda_k\asymp N^{\alpha_k}$, and thus their models are examples of our WF models. However, their assumptions are much stronger than ours because they directly restrict the structure of $(\mathbf{F}^*,\mathbf{B}^*)$ in data generating process (1). They are unobserved and it does not seem possible to identify them without extra exogenous information. On the contrary, our strategy is to interpret the PC estimators $(\hat{\mathbf{F}},\hat{\mathbf{B}})$ as the consistent estimators of $\mathbf{F}^0=\mathbf{F}^*\mathbf{H}$ and $\mathbf{B}^0=\mathbf{B}^*\mathbf{H}'^{-1}$ in the pseudo-true model (5) with (4). This interpretation is always possible under Assumption 1 as considered in Section 2. If necessary, we rotate back $\mathbf{F}^0$ to identify $\mathbf{F}^*$ when sufficient exogenous information exists for the purpose.

4 Consistency

We reconsider statistical consistency of the PC estimator. As noted in the introduction, the majority of the existing results have considered "consistency" for the rotated parameter by $\hat{\mathbf{H}}$, which is a function of the PC estimator. In this section, we derive consistency for the pseudo-true parameters obtained by $\mathbf{H}$ in (5). Accordingly, we will make further assumptions on this pseudo-true model for investigating the asymptotics.

Write $\mathbf{B}^*=(\mathbf{b}_1^*,\ldots,\mathbf{b}_N^*)'=(\mathbf{B}_1^*,\ldots,\mathbf{B}_r^*)$ and $\mathbf{F}^*=(\mathbf{f}_1^*,\ldots,\mathbf{f}_T^*)'=(\mathbf{F}_1^*,\ldots,\mathbf{F}_r^*)$; the same notational rule applies to the other matrices.

Assumption 4 (Factors and Loadings)**.** For some constant $M<\infty$ that does not depending on N and T, we have:
(i) $\mathbb{E}\|\mathbf{f}_t^0\|_2^4\leq M$ and $\mathbb{E}\|\mathbf{b}_i^0\|_2^4\leq M$;
(ii) $\mathbb{E}\|\mathbf{N}^{-\frac{1}{2}}\sum_{i=1}^{N}\mathbf{b}_i^0e_{t,i}\|_2^2\leq M$ for each t;
(iii) $\mathbb{E}\|T^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{f}_t^0e_{t,i}\|_2^2\leq M$ for each t;
(iv) $\mathbb{E}\|T^{-\frac{1}{2}}\mathbf{N}^{-\frac{1}{2}}\sum_{t=1}^{T}\sum_{j=1}^{N}\mathbf{b}_j^0[e_{t,i}e_{t,j}-\mathbb{E}(e_{t,i}e_{t,j})]\|_2^2\leq M$ for each i;
(v) $\mathbb{E}\|(NT)^{-\frac{1}{2}}\sum_{s=1}^{T}\sum_{i=1}^{N}\mathbf{f}_t^0[e_{s,i}e_{t,i}-\mathbb{E}(e_{s,i}e_{t,i})]\|_2^2\leq M$ for each t;
(vi) the $r\times r$ matrix satisfies $\mathbb{E}\|T^{-\frac{1}{2}}\mathbf{N}^{-\frac{1}{2}}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{b}_i^0e_{i,t}\mathbf{f}_t^{0\prime}\|_2^2\leq M$.

The moment restrictions in Assumption 4(iii),(v),(vi) are essentially the same as Assumptions D, F1 and F2 in Bai (2003), and Assumption 4(ii) and (iv) are similar moment restrictions related to $\mathbf{b}_i^0$.

Following the analysis conducted by Bai and Ng (2023), we use other data-dependent rotations of $\mathbf{F}^*$ and $\mathbf{B}^*$ than $\hat{\mathbf{H}}$, such as

$$\hat{\mathbf{H}}_4 = \mathbf{B}^{*\prime}\hat{\mathbf{B}}\hat{\mathbf{\Lambda}}^{-1}, \quad \tilde{\mathbf{H}}_4 = \mathbf{B}^{0\prime}\hat{\mathbf{B}}\hat{\mathbf{\Lambda}}^{-1}, \quad \tilde{\mathbf{H}} = \mathbf{B}^{0\prime}\mathbf{B}^0(T^{-1}\mathbf{F}^{0\prime}\hat{\mathbf{F}})\hat{\mathbf{\Lambda}}^{-1},$$
$$\hat{\mathbf{Q}} = T^{-1}\hat{\mathbf{F}}'\mathbf{F}^*, \quad \tilde{\mathbf{Q}} = T^{-1}\hat{\mathbf{F}}'\mathbf{F}^0,$$

and others appear in the appendix. These rotation matrices are theoretically important since the discrepancy between the PC estimator and the parameter rotated by such matrices can converge faster. By the definition, we immediately obtain $\mathbf{F}^*\hat{\mathbf{H}}_4 = \mathbf{F}^0\tilde{\mathbf{H}}_4$, $\mathbf{B}^*\hat{\mathbf{Q}}' = \mathbf{B}^0\tilde{\mathbf{Q}}'$, and $\mathbf{F}^*\hat{\mathbf{H}} = \mathbf{F}^0\tilde{\mathbf{H}}$. Employing these rotation matrices, we achieve the "consistency" results for the data-dependent counterparts:

Lemma 2. *Suppose that Assumptions 1–4 hold. If $\frac{N^{1-\alpha_r}}{T} \to 0$, then we have*

(i) $$\frac{1}{\sqrt{T}}\|\hat{\mathbf{F}} - \mathbf{F}^*\hat{\mathbf{H}}_4\|_{\mathrm{F}} = \frac{1}{\sqrt{T}}\|\hat{\mathbf{F}} - \mathbf{F}^0\tilde{\mathbf{H}}_4\|_{\mathrm{F}} = O_p\left(\frac{N^{1-\alpha_r}}{T}\right) + O_p\left(N^{-\frac{1}{2}\alpha_r}\right),$$

(ii) $$\frac{1}{\sqrt{N}}\|\hat{\mathbf{B}} - \mathbf{B}^*\hat{\mathbf{Q}}'\|_{\mathrm{F}} = \frac{1}{\sqrt{N}}\|\hat{\mathbf{B}} - \mathbf{B}^0\tilde{\mathbf{Q}}'\|_{\mathrm{F}} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(N^{-\frac{1}{2}-\frac{1}{2}\alpha_r}\right),$$

(iii) $\tilde{\mathbf{H}}_4, \tilde{\mathbf{Q}}, \tilde{\mathbf{H}}$ are $O_p(1)$. If additionally $\alpha_1 < 2\alpha_r$, then we have $\tilde{\mathbf{H}}_4, \tilde{\mathbf{Q}}, \tilde{\mathbf{H}} \xrightarrow{p} \mathbf{I}_r$.

Lemma 2(i) is in line with the result in Bai and Ng (2023, Proposition 6(i)). Lemma 2(ii) differs from the corresponding results in Bai and Ng (2023, Proposition 6(ii)), but related to Wei and Zhang (2023, Proposition 3.4). Interestingly, Lemma 2(i)(ii) do not depend on the value of α_1, which may lead to the fast convergence rates. The boundedness of the rotation matrices with "tilde" implied by Lemma 2(iii) is key to deriving the asymptotic normality for the pseudo-true parameters in the next section. In contrast, the rotation matrices with "hat" are not necessarily bounded unless additional conditions are imposed on $\mathbf{F}^*$.

Theorem 2. *Suppose that Assumptions 1–4 hold. If $\alpha_1 < 2\alpha_r$ and $\frac{N^{1-\alpha_r}}{T} \to 0$, then we have*

(i) $$\frac{1}{\sqrt{T}}\|\hat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}} = O_p\left(\frac{N^{1-\alpha_r}}{T}\right) + O_p\left(N^{-\frac{1}{2}\alpha_r}\right),$$

(ii) $$\frac{1}{\sqrt{N}}\|\hat{\mathbf{B}} - \mathbf{B}^0\|_{\mathrm{F}} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(N^{\alpha_1-\frac{3}{2}\alpha_r-\frac{1}{2}}\right) + o_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\right),$$

(iii) $$\frac{1}{\sqrt{NT}}\|\hat{\mathbf{C}} - \mathbf{C}^*\|_{\mathrm{F}} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r-\frac{1}{2}}\right) + o_p\left(\frac{N^{1-\alpha_r}}{T}\right),$$

where $\hat{\mathbf{C}} = \hat{\mathbf{F}}\hat{\mathbf{B}}'$ and $\mathbf{C}^* = \mathbf{F}^*\mathbf{B}^{*\prime} = \mathbf{F}^0\mathbf{B}^{0\prime}$.

This consistency theorem is new as it involves the pseudo-true parameter $(\mathbf{F}^0, \mathbf{B}^0)$ rather than the data-dependent parameter as in Lemma 2 and Bai and Ng (2023). This means that Theorem 2 achieves the *consistency* of the PC estimator for the pseudo-true parameter. Compared to Lemma 2, the additional condition $\alpha_1 < 2\alpha_r$ is imposed, but it can be interpreted as a cost for separate identification of the kth factor and the kth loading for each $k = 1, \ldots, r$. The same interpretation can also be applied to the comparison of Lemma 2(ii) and Theorem 2(ii), the latter of which depends on both α_1 and α_r. As a result, the convergence rate of $\hat{\mathbf{B}}$ in Theorem 2(ii) cannot be faster than that in Lemma 2(ii). Meanwhile, the convergence speed of $\hat{\mathbf{F}}$ in Theorem 2(i) are not sacrificed. In fact, we can show that Theorem 2(i) is as fast as that of $T^{-1/2}\|\hat{\mathbf{F}} - \mathbf{F}^*\hat{\mathbf{H}}\|_{\mathrm{F}}$ under the same conditions.

In the case of SF models (i.e., $\alpha_r = 1$) in Theorem 2, the rates in (i) and (ii) become $\frac{1}{T} + \frac{1}{\sqrt{N}}$ and $\frac{1}{\sqrt{T}} + \frac{1}{N}$, respectively. They are the same as those in Lemma 2. This is remarkable because Theorem 2 reveals the consistency of the PC estimators in SF models, whereas Lemma 2 or Bai and Ng (2002) does not. Moreover, these rates are smaller than those in Bai and Ng (2002) for the SF models, which are $\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N}}$.

Regarding the common component estimator, $\hat{\mathbf{C}}$, the rate given by Theorem 2(iii) is faster than that shown by Wei and Zhang (2023, Proposition 3.6); their rate is equivalent to replacing $N^{\frac{1}{2}\alpha_1 - \frac{1}{2}\alpha_r - \frac{1}{2}}$ in (iii) with $N^{\alpha_1 - \alpha_r - \frac{1}{2}}$.

5 Asymptotic Normality

We derive the asymptotic normality of the PC estimator centered by the pseudo-true parameter. The significance of this normality lies in justification of testing for restrictions on the pseudo-true parameters that are not data-dependent. We emphasize that testing for a general restriction, excluding a zero restriction, on the true parameter rotated by a data-dependent rotation is formally impossible. To facilitate the discussion, we suppose a central limit theorem.

Assumption 5 (CLT)**.** The following holds for each i and t:

$$\mathbf{D}^{-1}\mathbf{N}^{-\frac{1}{2}}\sum_{i=1}^{N}\mathbf{b}_i^0 e_{t,i} \xrightarrow{d} N(\mathbf{0}, \mathbf{\Gamma}_t), \qquad \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{f}_t^0 e_{t,i} \xrightarrow{d} N(\mathbf{0}, \mathbf{\Phi}_i),$$

where

$$\mathbf{\Gamma}_t = \lim_{N\to\infty}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[\mathbf{D}^{-1}\mathbf{N}^{-1/2}\mathbf{b}_i^0 e_{t,i}e_{t,j}\mathbf{b}_j^{0\prime}\mathbf{N}^{-1/2}\mathbf{D}^{-1}],$$

$$\mathbf{\Phi}_i = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[\mathbf{f}_t^0 e_{t,i}e_{s,i}\mathbf{f}_s^{0\prime}].$$

They are bounded and positive definite for all t and i.

We first derive the asymptotic normality of the PC estimator with respect to the rotated parameters by data-dependent rotation matrices.

Lemma 3. *Suppose that Assumptions 1–5 hold.*
(i) If $\alpha_r > 1/2$ and $\frac{N^{\frac{3}{2}-\alpha_r}}{T} \to 0$, we have

$$\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{f}}_t - \hat{\mathbf{H}}_4'\mathbf{f}_t^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Gamma}_t).$$

(ii) If $\alpha_1 < 2\alpha_r$, $\frac{N^{1-\alpha_r}}{\sqrt{T}} \to 0$ and $\frac{\sqrt{T}}{N^{\alpha_r}} \to 0$, we have

$$\sqrt{T}(\hat{\mathbf{b}}_i - \hat{\mathbf{Q}}\mathbf{b}_i^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Phi}_i).$$

Lemma 3(i) is similar to Bai and Ng (2023, Proposition 7(i)). Lemma 3(ii) is new and theoretically useful in comparison to the existing results, such as Bai and Ng (2023, Proposition 7(ii)). Lemma 3(ii) is also similar to the result of Wei and Zhang (2023), but they impose a sparsity assumption to $\mathbf{B}^*$, which is more restrictive than ours. In the case where $N \asymp T$, the conditions for Lemma 3(i) and (ii) are simplified to $\alpha_r > 1/2$. The lemma leads to the next theorem:

Theorem 3. *Suppose that Assumptions 1–5 hold.*
(i) If $\alpha_r > 1/2$, $2\alpha_1 < 3\alpha_r$, $\frac{N^{\frac{3}{2}-\alpha_r}}{T} \to 0$, and $\frac{N^{\alpha_1-\alpha_r}}{\sqrt{T}} \to 0$, we have

$$\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{f}}_t - \mathbf{f}_t^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Gamma}_t).$$

(ii) If $\alpha_1 < 2\alpha_r$, $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r} \to 0$, and $\frac{N^{1-\alpha_r}}{\sqrt{T}} \to 0$, we have

$$\sqrt{T}(\hat{\mathbf{b}}_i - \mathbf{b}_i^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Phi}_i).$$

(iii) If $\alpha_r > 1/2$, $\frac{N^{\frac{3}{2}-\alpha_r}}{T} \to 0$, and $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r} \to 0$, we have

$$\frac{\hat{c}_{t,i} - c_{t,i}^*}{\sqrt{V_{t,i} + U_{t,i}}} \xrightarrow{d} N(0, 1),$$

where $V_{t,i} = \mathbf{b}_i^{0\prime}\mathbf{D}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{\Gamma}_t\mathbf{D}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{b}_i^0$ and $U_{t,i} = T^{-1}\mathbf{f}_t^{0\prime}\mathbf{\Phi}_i\mathbf{f}_t^0$.

Theorem 3 allows us to construct confidence intervals of the PC estimators for each of r (pseudo-true) factors and factor loadings. On the other hand, Lemma 3, Bai (2003), and Wei and Zhang (2023), among others are not generally applicable as the rotated parameters depend on the estimator itself; the only exception is the test for the (true) factors or factor loadings are jointly equal to zero.

In comparison to Lemma 3(i) and (ii), Theorem 3(i) and (ii) require the additional conditions. The conditions for Theorem 3(iii) are weaker than those for Theorem 3(i) and (ii) because the estimation of $\mathbf{C}^*$ does not require separate identification of the factors and factor loadings, and Lemma 3 applies directly. When $N \asymp T$, the conditions for Theorem 3(i) are simplified to $\alpha_r > 1/2$ and $2\alpha_1 < 3\alpha_r$, those for Theorem 3(ii) and (iii) reduce to $\alpha_r > 1/2$ and $\alpha_1 < 3\alpha_r - 1$. When the SF model (i.e., $\alpha_r = 1$) is considered, the conditions for Theorem 3(i) and (ii) reduce to $\sqrt{N}/T \to 0$ and $\sqrt{T}/N \to 0$, respectively, which are identical to those in Bai (2003, Theorems 1&2) for the case with the data-dependent rotation.

To conduct statistical inference based on Theorem 3, the unknowns should be replaced with estimators. In Theorem 3(i), the scaling matrix $\mathbf{D}\mathbf{N}^{1/2}$ can be consistently estimated by $(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{1/2}$ because Assumption 3 gives $(\mathbf{B}^{0\prime}\mathbf{B}^0)^{1/2} = \mathbf{D}\mathbf{N}^{1/2}$. As for $\mathbf{\Gamma}_t$ and $\mathbf{\Phi}_i$, the choice of the estimators depend on the dependence structure of $e_{t,i}$. In general, we may use $\hat{\mathbf{\Gamma}}_t = (\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1/2}\hat{\mathbf{\Omega}}_t(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1/2}$ with $\hat{\mathbf{\Omega}}_t$ appropriately chosen. For example, when cross-sectional independence and time-series heteroskedasticity are allowed, choose $\hat{\mathbf{\Omega}}_t = \sum_{i=1}^N \hat{\mathbf{b}}_i\hat{e}_{t,i}^2\hat{\mathbf{b}}_i'$, where $\hat{\mathbf{E}} = \mathbf{X} - \hat{\mathbf{F}}\hat{\mathbf{B}}' = (\hat{e}_{t,i})$. For $\mathbf{\Phi}_i$, when $e_{t,i}$ is serially correlated and heteroskedastic over i, the HAC estimator of Newey and West (1987) can be employed: $\hat{\mathbf{\Phi}}_i = \hat{\mathbf{V}}_{0,i} + \sum_{\ell=1}^L(1 - \ell/(L+1))(\hat{\mathbf{V}}_{\ell,i} + \hat{\mathbf{V}}_{\ell,i}')$, where $L > 0$ is a slowly diverging sequence and $\hat{\mathbf{V}}_{\ell,i} = T^{-1}\sum_{t=\ell+1}^T \hat{\mathbf{f}}_t\hat{e}_{t,i}\hat{e}_{t-\ell,i}\hat{\mathbf{f}}_{t-\ell}'$. Finally, for Theorem 3(iii), we may construct $\hat{V}_{t,i} = \hat{\mathbf{b}}_i'(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1/2}\hat{\mathbf{\Gamma}}_t(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1/2}\hat{\mathbf{b}}_i$ and $\hat{U}_{t,i} = T^{-1}\hat{\mathbf{f}}_t'\hat{\mathbf{\Phi}}_i\hat{\mathbf{f}}_t$.

6 Factor-Augmented Regression

We consider factor augmented regression models. Assuming that certain unobserved common factors influence the predictors, the factors extracted from the predictors can be used to forecast a particular series; see Stock and Watson (2002a). In such an analysis, significance tests of each factor is important because the estimated factors ordered by the magnitude of the associated eigenvalues may not correspond to their forecast power for a particular series of interest; see further discussions in Bai and Ng (2008, 2009) and Cheng and Hansen (2015).

Consider the factor augmented regression model

$$y_{t+h} = \boldsymbol{\gamma}^{*\prime}\mathbf{f}_t^* + \boldsymbol{\beta}'\mathbf{w}_t + \epsilon_{t+h}, \tag{8}$$

where $\mathbf{f}_t^*$ is a vector of latent factors, $\mathbf{w}_t$ is a vector of exogenous variables, ϵ_t is an error term, and $(\boldsymbol{\gamma}^{*\prime}, \boldsymbol{\beta}')'$ is a coefficient vector. Since $\mathbf{f}_t^*$ is unobservable, it is replaced by the PC estimator $\hat{\mathbf{f}}_t$ estimated from a set of many predictors, $\mathbf{X}$. Following Bai and Ng (2006, 2023), we may choose the data-dependent rotation matrix $\hat{\mathbf{H}}$. Then, under approximation (3) and some boundedness condition on $\boldsymbol{\gamma}_{\hat{\mathbf{H}}} := \hat{\mathbf{H}}^{-1}\boldsymbol{\gamma}^*$, we have

$$\boldsymbol{\gamma}^{*\prime}\mathbf{f}_t^* = \boldsymbol{\gamma}_{\hat{\mathbf{H}}}'\hat{\mathbf{f}}_t - \boldsymbol{\gamma}_{\hat{\mathbf{H}}}'(\hat{\mathbf{f}}_t - \hat{\mathbf{H}}'\mathbf{f}_t^*) = \boldsymbol{\gamma}_{\hat{\mathbf{H}}}'\hat{\mathbf{f}}_t + o_p(1).$$

Since $\hat{\mathbf{f}}_t$ is observed, we can estimate its coefficient vector, and denote it as $\hat{\boldsymbol{\gamma}}$. We may derive the asymptotic normality for the discrepancy, $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_{\hat{\mathbf{H}}}$, and construct the associated t-ratio. However, it is questionable whether using such statistics is justified as the "parameter" $\boldsymbol{\gamma}_{\hat{\mathbf{H}}}$ depends on the PC estimator through $\hat{\mathbf{H}}$. Despite the problem, such tests are routinely reported in empirical studies; see Ludvigson and Ng (2009) among many others.

As in the previous section, we consider $\hat{\mathbf{f}}_t$ as the estimator of $\mathbf{f}_t^0$. To facilitate this perspective, we introduce $\mathbf{H}$ and rewrite model (8) to the pseudo-true model:

$$y_{t+h} = \boldsymbol{\gamma}^{0\prime}\mathbf{f}_t^0 + \boldsymbol{\beta}'\mathbf{w}_t + \epsilon_{t+h}, \tag{9}$$

where $\boldsymbol{\gamma}^0 := \mathbf{H}^{-1}\boldsymbol{\gamma}^*$ and $\mathbf{f}_t^0 = \mathbf{H}'\mathbf{f}_t^*$. Observe that $\boldsymbol{\gamma}^0$ is a function of the parameters, $\mathbf{F}^*$, $\mathbf{B}^*$, and $\boldsymbol{\gamma}^*$. Therefore, tests for general parameter restrictions on $\boldsymbol{\gamma}^0$ seem justified. Rewrite (9) to the matrix form:

$$\mathbf{Y} = \mathbf{F}^0\boldsymbol{\gamma}^0 + \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \mathbf{Z}^0\boldsymbol{\delta}^0 + \boldsymbol{\epsilon}, \tag{10}$$

where $\mathbf{Y} = (y_{1+h}, \ldots, y_{T+h})'$, $\boldsymbol{\epsilon} = (\epsilon_{1+h}, \ldots, \epsilon_{T+h})'$, $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_T)'$, $\boldsymbol{\delta}^0 = (\boldsymbol{\gamma}^{0\prime}, \boldsymbol{\beta}')'$, and $\mathbf{Z}^0 = (\mathbf{F}^0, \mathbf{W}) = (\mathbf{z}_1^0, \ldots, \mathbf{z}_T^0)'$. We make the assumptions to derive the asymptotic theory.

Assumption 6. (i) $\|\boldsymbol{\gamma}^0\|_2 \leq M$ and $\|\boldsymbol{\beta}\|_2 \leq M$;
(ii) $\mathbb{E}(\epsilon_{t+h}|y_t, \mathbf{z}_t^0, y_{t-1}, \mathbf{z}_{t-1}^0, \cdots) = 0$ for any $h > 0$;
(iii) $T^{-1/2}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{z^0\epsilon})$, $T^{-1}\mathbf{Z}^{0\prime}\mathbf{Z}^0 \xrightarrow{p} \boldsymbol{\Sigma}_{z^0}$, where $\boldsymbol{\Sigma}_{z^0\epsilon}$ and $\boldsymbol{\Sigma}_{z^0}$ are positive definite and bounded;
(iv) $\mathbb{E}\,\|T^{-1/2}\mathbf{N}^{-\frac{1}{2}}\sum_{t=1}^T\sum_{i=1}^N \mathbf{b}_i^0 e_{i,t}\mathbf{w}_t'\|_2^2 \leq M$ and $\mathbb{E}\,\|T^{-1/2}\mathbf{N}^{-\frac{1}{2}}\sum_{t=1}^T\sum_{j=1}^N \mathbf{b}_j^0[\epsilon_{t+h}e_{t,j} - \mathbb{E}(\epsilon_{t+h}e_{t,j})]\|_2^2 \leq M$;
(v) For all t, $|\mathbb{E}\,(\epsilon_{t+h}e_{t,i})| \leq |\tau_i|$ for some τ_i such that $\sum_{i=1}^N |\tau_i| \leq M$.

Assumption 6(i) and (ii) are standard. Assumption 6(iii) is for identification of $\boldsymbol{\delta}^0$ and asymptotic normality of its estimator. Assumption 6(iv) and (v) permit weak correlations between the errors in the factor model and in the augmented regression model. They imply $\mathbf{W}'\mathbf{E}\mathbf{B}^0 = O_p(\sqrt{TN^{\alpha_1}})$ and $\frac{1}{T}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}'\boldsymbol{\epsilon} = O_p\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N^{\alpha_r}}}\right)$.

6.1 Asymptotic normality

In model (10), unknown $\mathbf{f}_t^0$ is replaced with the estimated factor $\hat{\mathbf{f}}_t$. Then (10) is written as

$$\mathbf{Y} = \hat{\mathbf{F}}\gamma^0 + \mathbf{W}\boldsymbol{\beta} + \mathbf{u} = \hat{\mathbf{Z}}\delta^0 + \mathbf{u}, \tag{11}$$

where $\mathbf{u} = \boldsymbol{\epsilon} - (\hat{\mathbf{F}} - \mathbf{F}^0)\gamma^0$ and $\hat{\mathbf{Z}} = (\hat{\mathbf{F}}, \mathbf{W}) = (\hat{\mathbf{z}}_1, \dots, \hat{\mathbf{z}}_T)'$. The estimator of $\boldsymbol{\delta}^0$ is obtained as $\hat{\boldsymbol{\delta}} = (\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}'\mathbf{Y}$. Using (11), we have

$$\sqrt{T}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0) = (T^{-1}\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}T^{-\frac{1}{2}}\hat{\mathbf{Z}}'\boldsymbol{\epsilon} - (T^{-1}\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}T^{-\frac{1}{2}}\hat{\mathbf{Z}}'(\hat{\mathbf{F}} - \mathbf{F}^0)\boldsymbol{\gamma}^0. \tag{12}$$

In the right-hand side of (12), the first term tends to a normal distribution with mean zero, and the second term causes potential bias in $\hat{\boldsymbol{\delta}}$.

The following theorem ensures that the effect of the replacement by the PC estimator is asymptotically negligible and a test for a general restriction on $\boldsymbol{\gamma}^0$ is asymptotically valid.

Theorem 4. *Suppose that Assumptions 1–6 hold. If $\alpha_1 < 2\alpha_r$, $\frac{N^{1-\alpha_r}}{\sqrt{T}} \to 0$, and $\sqrt{T}N^{\frac{1}{2}\alpha_1 - \frac{3}{2}\alpha_r} \to 0$, we have*

$$\sqrt{T}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\delta^0}),$$

where $\boldsymbol{\Sigma}_{\delta^0} = \boldsymbol{\Sigma}_{z^0}^{-1}\boldsymbol{\Sigma}_{z^0\epsilon}\boldsymbol{\Sigma}_{z^0}^{-1}$.

In general, $\boldsymbol{\Sigma}_{\delta^0}$ can be estimated by $\hat{\boldsymbol{\Sigma}}_{\delta^0} = (T^{-1}\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}\hat{\boldsymbol{\Sigma}}_{z^0\epsilon}(T^{-1}\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}$, where $\hat{\boldsymbol{\Sigma}}_{z^0\epsilon}$ is a consistent estimator. For example, when ϵ_t is heteroskedastic, we may use $\hat{\boldsymbol{\Sigma}}_{z^0\epsilon} = T^{-1}\sum_{t=1+h}^{T+h}\hat{\mathbf{z}}_t\hat{\epsilon}_t^2\hat{\mathbf{z}}_t'$, where $\hat{\epsilon}_{t+h} = y_{t+h} - \hat{\boldsymbol{\delta}}'\hat{\mathbf{z}}_t$.

The rate conditions for Theorem 4 (and Theorem 5 below) are identical to the corresponding results for approximations with data-dependent rotations, such as $(\hat{\mathbf{F}}, \hat{\mathbf{B}}) \approx (\mathbf{F}^*\hat{\mathbf{H}}_4, \mathbf{B}^*\hat{\mathbf{Q}}')$, where $\hat{\mathbf{H}}_4\hat{\mathbf{Q}} \xrightarrow{p} \mathbf{I}_r$ by Bai and Ng (2023, Lemma 3). They imply the associated approximations $\hat{\gamma} \approx \hat{\mathbf{H}}_4^{-1}\gamma^*$ and $\hat{\gamma} \approx \hat{\mathbf{Q}}\gamma^*$, respectively; see Lemma B.9. However, for such approximations the "parameter" estimated by $\hat{\gamma}$ depends on $\mathbf{X}$, and thus the only justifiable inference is about the joint significance, $H_0 : \boldsymbol{\gamma}^* = \mathbf{0}$.

6.2 Forecasting

We consider the h-step ahead forecast. From (9), the expectation of y_{T+h} conditional on $\{\mathbf{z}_T^0, \dots, \mathbf{z}_1^0\}$ is computed as $y_{T+h|T} := \mathbb{E}(y_{T+h} \mid \mathbf{z}_T^0, \dots, \mathbf{z}_1^0) = \boldsymbol{\delta}^{0\prime}\mathbf{z}_T^0$, which is the infeasible h-step ahead forecast of y at time T. After estimating the forecast regression (11), $y_{T+h|T}$ is estimated by $\hat{y}_{T+h|T} = \hat{\boldsymbol{\delta}}'\hat{\mathbf{z}}_T$. Thus the estimation error, $\hat{y}_{T+h|T} - y_{T+h|T}$, is derived as

$$\hat{y}_{T+h|T} - y_{T+h|T} = (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0)'\hat{\mathbf{z}}_T + \boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1\prime}(\hat{\mathbf{f}}_T - \mathbf{H}'\mathbf{f}_T^*).$$

Theorem 5. *Suppose that Assumptions 1–6 hold. If $\sqrt{T}N^{\frac{1}{2}\alpha_1 - \frac{3}{2}\alpha_r} \to 0$, $\frac{1}{2} < \alpha_r$, and*

$\frac{N^{\frac{3}{2}-\alpha_r}}{T} \to 0$, *we have*

$$\frac{\hat{y}_{T+h|T} - y_{T+h|T}}{\sigma_{T+h|T}} \xrightarrow{d} N(0,1),$$

where $\sigma^2_{T+h|T} = T^{-1}\mathbf{z}^{0\prime}_T \mathbf{\Sigma}_{\delta^0} \mathbf{z}^0_T + \boldsymbol{\gamma}^{0\prime}\mathbf{D}^{-1}\mathbf{N}^{-1/2}\mathbf{\Gamma}_T\mathbf{N}^{-1/2}\mathbf{D}^{-1}\boldsymbol{\gamma}^0$.

We may estimate $\sigma^2_{T+h|T}$ by $\hat{\sigma}^2_{T+h|T} = T^{-1}\hat{\mathbf{z}}'_T\hat{\mathbf{\Sigma}}_{\delta^0}\hat{\mathbf{z}}_T + \hat{\boldsymbol{\gamma}}'(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1/2}\hat{\mathbf{\Gamma}}_T(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1/2}\hat{\boldsymbol{\gamma}}$ for some consistent estimators $\hat{\mathbf{\Sigma}}_{\delta^0}$ and $\hat{\mathbf{\Gamma}}_T$. The rate conditions for Theorem 5 are identical to the corresponding results with data-dependent rotations; see Lemma B.10.

Finally, we consider the out-of-sample forecast error. By $y_{T+h} = y_{T+h|T} + \epsilon_{T+h}$, we define the out-of-sample forecast error as

$$v_{T+h|T} = \hat{y}_{T+h|T} - y_{T+h} = \hat{y}_{T+h|T} - y_{T+h|T} - \epsilon_{T+h}.$$

Assuming that $\epsilon_t \sim$ i.i.d.$N\left(0, \sigma^2_\epsilon\right)$, we have $v_{T+h|T} \sim N(0, \sigma^2_\epsilon + \sigma^2_{T+h|T})$. The variance of $v_{T+h|T}$ can be estimated by $\hat{\sigma}^2_\epsilon + \hat{\sigma}^2_{T+h|T}$, where $\hat{\sigma}^2_\epsilon = T^{-1}\sum_{t=1+h}^{T+h}\hat{\epsilon}^2_t$ and $\hat{\sigma}^2_{T+h|T}$ is given above. This result yields the confidence band for $\hat{y}_{T+h|T}$. If normality assumption is not plausible, bootstrap confidence interval should be computed; see Gonçalves and Perron (2014).

7 Monte Carlo Experiments

We check finite sample performance of the PC estimator relative to the pseudo-true parameter in comparison to that relative to the "parameter" rotated by other data-dependent rotations. Section 7.1 treats the approximate factor models. Section 7.2 considers the factor-augmented regression models.

7.1 Approximate factor models

We discuss estimation accuracy of the PC estimator $(\hat{\mathbf{F}}, \hat{\mathbf{B}})$ against the pseudo-true parameter $(\mathbf{F}^0, \mathbf{B}^0)$ and other existing data-dependent rotations, $(\mathbf{F}^*\hat{\mathbf{H}}, \mathbf{B}^{*\prime}\hat{\mathbf{H}}^{-1\prime})$ and $(\mathbf{F}^*\hat{\mathbf{H}}_4, \mathbf{B}^{*\prime}\mathbf{Q}')$. We also examine the multivariate normal approximation for $\hat{\mathbf{f}}_t$ and $\hat{\mathbf{b}}_i$. Furthermore, we investigate the size of tests, such as $H_0 : f_{t,k} = f^0_{t,k}$ and $H_0 : b_{i,k} = b^0_{i,k}$, which are only justified in our proposed approach.

7.1.1 Experimental design

We choose $r = 2$ throughout, and generate $f_{t,k} \sim$ i.i.d.$U(\mu_k - \sqrt{3}, \mu_k + \sqrt{3})$ with $\mu_k = 1$ for $t = 1, \dots, T$ and $k = 1, 2$, to form $\mathbf{F} = (f_{t,k})$. Applying the Gram-Schmidt procedure to $\mathbf{F}$, we obtain $\mathbf{F}^0$ such that $\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_2$. We next generate two types of loading matrices:

1. Non-sparse factor loadings: $\mathbf{B}^0 = \left(\mathbf{b}^0_1, \mathbf{b}^0_2\right) = \left(N^{(\alpha_1-1)/2}g_{i,1}, N^{(\alpha_2-1)/2}g_{i,2}\right)$, where $g_{i,1}$ is 2 for $i = 1, \dots, N/2$ and 1 for the rest, and $g_{i,2}$ is 1/2 for $i = 1, \dots, N/2$ and -1 for the rest.

2. Sparse factor loadings: $\mathbf{B}^0 = \left(\mathbf{b}^0_1, \mathbf{b}^0_2\right)$, where $b^0_{i,1}$ is 2 for $i = 1, \dots, N_1$ and 0 for the rest, and $b^0_{i,2}$ is 1 for $i = 1, \dots, N_2/2$, -1 for $i = N_2/2 + 1, \dots, N_2$, and zero for the rest, where N_k are the closest even number to N^{α_k} for $k = 1, 2$.

These designs ensure that $\mathbf{b}_k^{0\prime}\mathbf{b}_k^0 \asymp N^{\alpha_k}$. We construct $\mathbf{F}^* = \mathbf{F}^0\mathbf{H}^{-1}$ and $\mathbf{B}^* = \mathbf{B}^0\mathbf{H}'$, where $\mathbf{H} = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 2 \end{smallmatrix}\right)$. Then the data is generated by $\mathbf{X} = \mathbf{F}^*\mathbf{B}^{*\prime} + \mathbf{E}$, where $\mathbf{E} = (e_{t,i})$ with $e_{t,i} \sim$ i.i.d.$N(0, \sigma_e^2)$ and $\sigma_e^2 = 0.5$. We consider six models: $(\alpha_1, \alpha_2) = (1.0, 1.0)$, $(1.0, 0.9)$, $(1.0, 0.8)$, $(0.9, 0.7)$, $(0.8, 0.6)$, and $(0.7, 0.5)$, where $(1.0, 1.0)$ is the SF model. We consider sample sizes, $T = N = 50, 100, 200, 500$. All the results are based on 50,000 replications.

The PC estimate, $\hat{\mathbf{F}}$, is computed as $\sqrt{T}$ times r eigenvectors of $T^{-1}\mathbf{X}\mathbf{X}'$ corresponding to its first r largest eigenvalues. The loading estimator is computed as $\hat{\mathbf{B}} = \mathbf{X}'\hat{\mathbf{F}}/T$. Using the correlation between $\mathbf{F}^0$ and $\hat{\mathbf{F}}$, the signs and the orders of the r columns of $\hat{\mathbf{F}}$ and $\hat{\mathbf{B}}$ are adjusted to fit to those of $\mathbf{F}^0$, when necessary, so that $\hat{\mathbf{F}}$ and $\hat{\mathbf{B}}$ can be regarded as the estimates of $\mathbf{F}^0$ and $\mathbf{B}^0$ (i.e. no sign indeterminacy). This adjustment is for experimental purposes only; this may be irrelevant in practice.

To save space, we only report the results for the models with non-sparse loadings. The results with sparse loadings are very similar and are available in the online appendix.

7.1.2 Estimation accuracy and normal approximation

To assess the convergence results for the PC estimator $(\hat{\mathbf{F}}, \hat{\mathbf{B}})$ and their product, Figure 1 summarizes the averages over the replications of the relevant norm losses. Figure 1(i) reports $T^{-\frac{1}{2}}\|\hat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}}$, $T^{-\frac{1}{2}}\|\hat{\mathbf{F}} - \mathbf{F}^*\hat{\mathbf{H}}_4\|_{\mathrm{F}}$, and $T^{-\frac{1}{2}}\|\hat{\mathbf{F}} - \mathbf{F}^*\hat{\mathbf{H}}\|_{\mathrm{F}}$ while Figure 1(ii) compares $N^{-\frac{1}{2}}\|\hat{\mathbf{B}} - \mathbf{B}^0\|_{\mathrm{F}}$, $N^{-\frac{1}{2}}\|\hat{\mathbf{B}} - \mathbf{B}^*\hat{\mathbf{Q}}\|_{\mathrm{F}}$, and $N^{-\frac{1}{2}}\|\hat{\mathbf{B}} - \mathbf{B}^*\hat{\mathbf{H}}^{-1}\|_{\mathrm{F}}$. Figure 1(iii) shows $(NT)^{-\frac{1}{2}}\|\hat{\mathbf{C}} - \mathbf{C}^*\|_{\mathrm{F}}$.

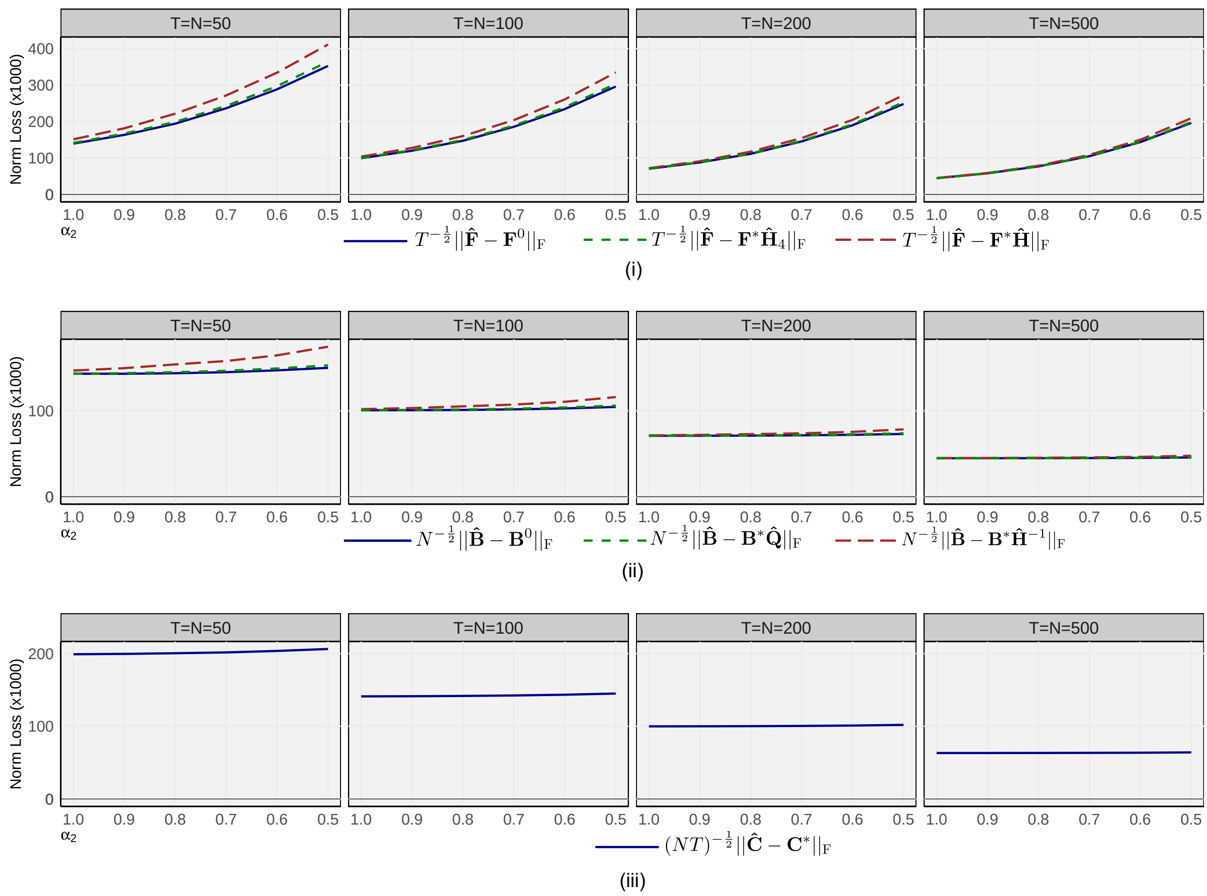

Figure 1: Plots of (i) $T^{-\frac{1}{2}}\|\hat{\mathbf{F}} - \mathbf{F}\|_{\mathrm{F}}$; (ii) $N^{-\frac{1}{2}}\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}$; (iii) $(NT)^{-\frac{1}{2}}\|\hat{\mathbf{C}} - \mathbf{C}^*\|_{\mathrm{F}}$

Each block in the figure is the result for a particular sample size. Within the block, the vertical axis shows the magnitude of the average norm loss, and the horizontal axis indicates the value of α_2, which identifies one of the six models considered, hence, the weakness of the model – the smaller the value, the weaker the factor model.

It is clear from Figure 1(i) and (ii) that the norm loss of the PC estimator $(\hat{\mathbf{F}}, \hat{\mathbf{B}})$ against $(\mathbf{F}^0, \mathbf{B}^0)$ is the smallest compared with those against $(\mathbf{F}^*\hat{\mathbf{H}}_4, \mathbf{B}^*\hat{\mathbf{Q}}')$ and $(\mathbf{F}^*\hat{\mathbf{H}}, \mathbf{B}^*\hat{\mathbf{H}}')$, closely followed by the former approximation. The approximation by $(\mathbf{F}^0, \mathbf{B}^0)$ improves as the model weakens especially when the sample size is small. This superiority eventually disappears as the sample size increases. It is interesting that $(\mathbf{F}^0, \mathbf{B}^0)$ achieves the best performance because the theoretical results of Lemma 2 and Theorem 2 suggest that the $(\mathbf{F}^*\hat{\mathbf{H}}_4, \mathbf{B}^*\hat{\mathbf{Q}}')$ approximation may be more accurate. A similar comment applies to (iii).

Next, we check the quality of the joint normal approximations of the PC estimators, by comparing the following statistics with the χ^2_r distribution:

$$Q^2_f(\Delta\hat{\mathbf{f}}_t) := (\mathbf{B}^{0\prime}\mathbf{B}^0)^{\frac{1}{2}}(\Delta\hat{\mathbf{f}}_t)'\mathbf{\Gamma}_t^{-1}(\Delta\hat{\mathbf{f}}_t)(\mathbf{B}^{0\prime}\mathbf{B}^0)^{\frac{1}{2}}, \quad Q^2_b(\Delta\hat{\mathbf{b}}_i) := T(\Delta\hat{\mathbf{b}}_i)'\mathbf{\Phi}_i^{-1}(\Delta\hat{\mathbf{b}}_i)$$

with $\Delta\hat{\mathbf{f}}_t = \hat{\mathbf{f}}_t - \mathbf{f}^0_t$, $\hat{\mathbf{f}}_t - \hat{\mathbf{H}}_4'\mathbf{f}^*_t$, $\hat{\mathbf{f}}_t - \hat{\mathbf{H}}'\mathbf{f}^*_t$, $\mathbf{\Gamma}_t = \sigma^2_e\mathbf{I}_r$, $\Delta\hat{\mathbf{b}}_i = \hat{\mathbf{b}}_i - \mathbf{b}^0_i$, $\hat{\mathbf{b}}_i - \hat{\mathbf{Q}}\mathbf{b}^*_i$, $\hat{\mathbf{b}}_i - \hat{\mathbf{H}}^{-1}\mathbf{b}^*_i$, and $\mathbf{\Phi}_i = \sigma^2_e\mathbf{I}_r$. We assess the accuracy of the tail behavior by computing the frequencies of $Q^2_\cdot > \chi^2_{r,0.95}$ over the replications, where $\chi^2_{r,0.95}$ is the 95-percentile of a χ^2_r distribution. The closer the size (frequency) is to 5%, the more accurate the approximation by the statistics.

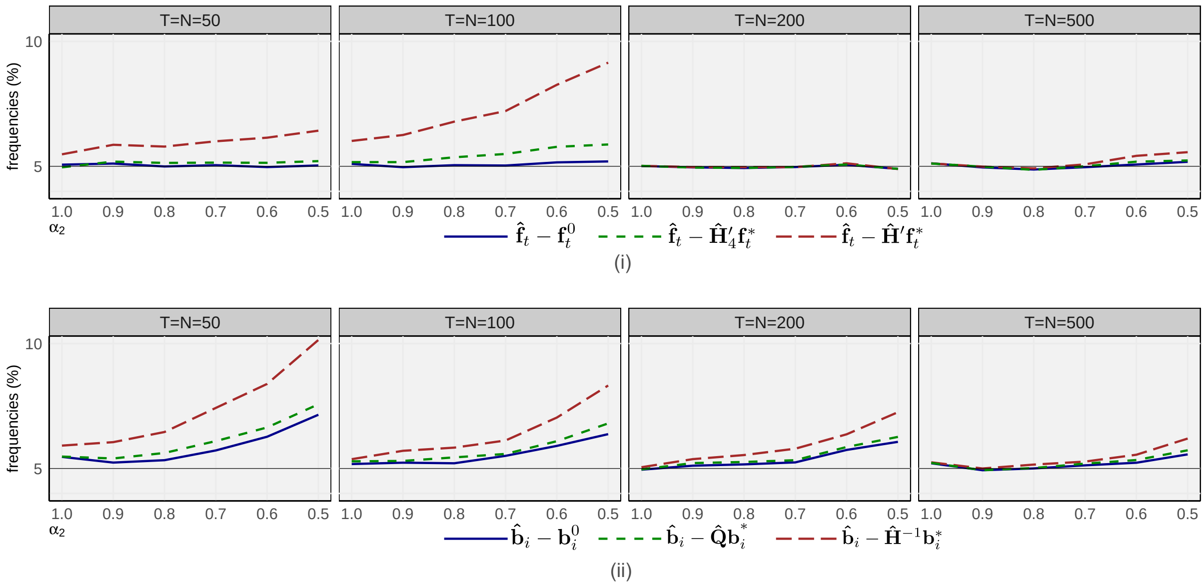

Figure 2: Plots of frequencies of $Q^2_\cdot > \chi^2_{r,0.95}$ for (i) $Q^2_f(\Delta\hat{\mathbf{f}}_t)$; (ii) $Q^2_b(\Delta\hat{\mathbf{b}}_i)$

Figure 2 shows the frequencies. The PC estimator performs the best in the case when centered at $(\mathbf{f}^0_t, \mathbf{b}^0_i)$ compared to those when centered at the data-dependent statistics, $(\hat{\mathbf{H}}_4'\mathbf{f}^*_t, \hat{\mathbf{Q}}\mathbf{b}^*_i)$ and $(\hat{\mathbf{H}}'\mathbf{f}^*_t, \hat{\mathbf{H}}\mathbf{b}^*_i)$. The result is interesting because Lemma 3 and Theorem 3 suggest that centering by $(\hat{\mathbf{H}}_4'\mathbf{f}^*_t, \hat{\mathbf{Q}}\mathbf{b}^*_i)$ may provide a better approximation with a finite sample. We note that there are sudden performance improvements for all the statistics for $N = T = 200$ and 500, which we believe to be a finite sample phenomenon.

7.1.3 Test of linear restrictions

We report the performance of testing for $H_0 : f_{1,k} = f^0_{1,k}$, $H_0 : b_{1,k} = b^0_{1,k}$, and $H_0 : c_{1,1} = c^*_{1,1}$ using the statistics,

$$z_{f,k} = \sqrt{N^{\alpha_k}\gamma_t^{(k,k)}}(\hat{f}_{1,k} - f^0_{1,k}), \quad z_{b,k} = \sqrt{T\phi_1^{(k,k)}}(\hat{b}_{1,k} - b^0_{1,k}), \quad z_c = \frac{\hat{c}_{1,1} - c^*_{1,1}}{\sigma_c},$$

respectively, where $\gamma_1^{(k,k)} = (\mathbf{\Gamma}_1^{-1})_{k,k}$, $\phi_1^{(k,k)} = (\mathbf{\Phi}_1^{-1})_{k,k}$, and $\sigma_c^2 = \sigma_e^2[\mathbf{b}_1^{0\prime}(\mathbf{B}^{0\prime}\mathbf{B}^0)^{-1}\mathbf{b}_1^0 + T^{-1}\mathbf{f}_1^{0\prime}\mathbf{f}_1^0]$. We check the empirical size of the 5% test, i.e., by computing the frequencies of their exceeding 1.96 in modulus over the replications. Again, the results are the first to asymptotically justify such tests in approximate factor models.

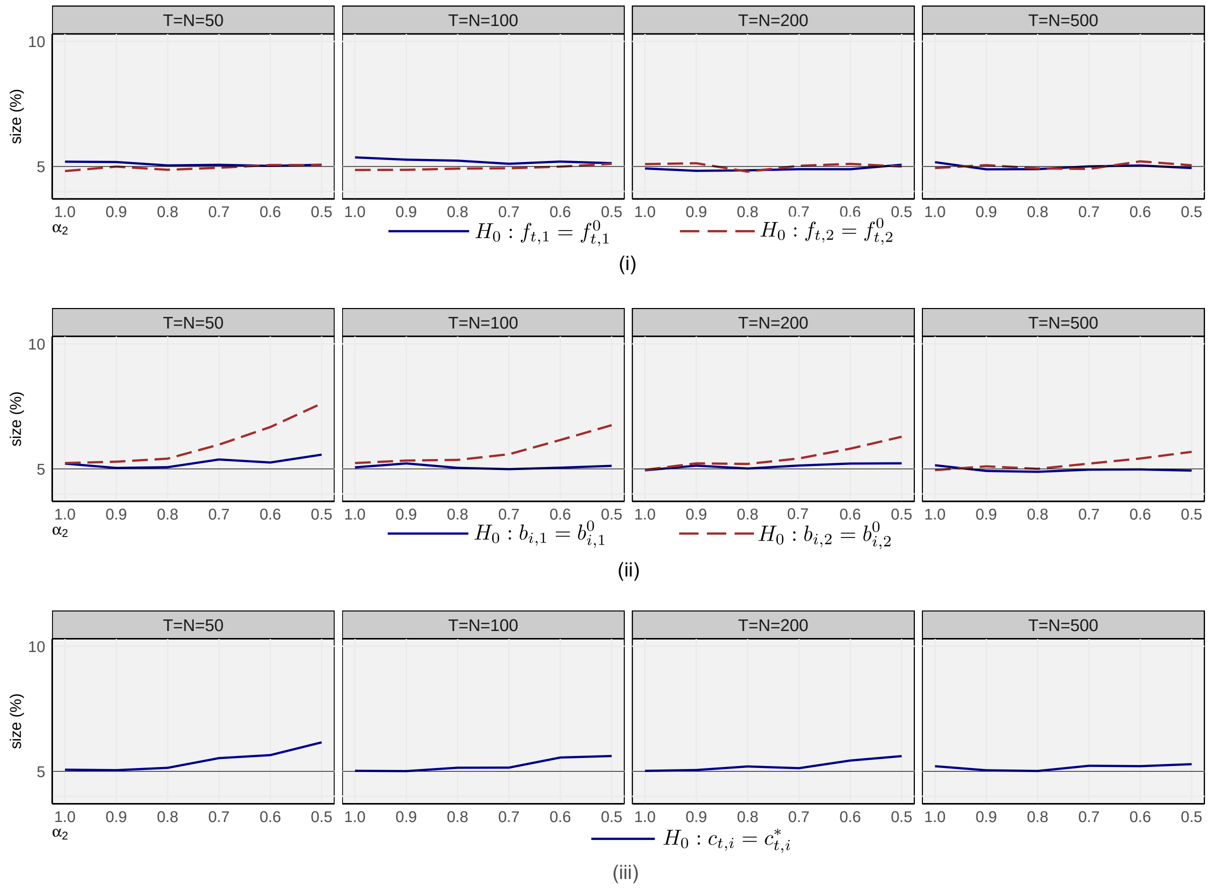

Figure 3: Size of the 5% level tests for (i) $H_0 : f_{1,k} = f^0_{1,k}$; (ii) $H_0 : b_{1,k} = b^0_{1,k}$; (iii) $H_0 : c_{1,1} = c^*_{1,1}$.

Figure 3(i) shows that our proposed test has an empirical size very close to the nominal level for all the cases. Figure 3(ii) indicates a moderate over-rejection in the very weak models, but it decreases as the sample size increases. A similar comment applies to (iii).

7.2 Factor augmented regressions

We examine the normal approximation of the coefficient estimates and the size of the tests for, e.g., $H_0 : \gamma_k = \gamma_k^0$. Again, this is only justified in our proposed approach. Finally, we investigate the empirical coverage of the confidence interval for the h-step ahead forecast, $\hat{y}_{T+h|T}$, against both the conditional mean $(y_{T+h|T})$ and the actual value (y_{T+h}).

7.2.1 Experimental design

Using $\mathbf{f}_t^0$ and $\mathbf{H}$ in Section 7.1.1, we generate

$$y_{t+h} = \mathbf{f}_t^{0\prime}\boldsymbol{\gamma}^0 + \mathbf{w}_t^\prime\boldsymbol{\beta} + \epsilon_{t+h}, \quad t = 1, \ldots, T, \tag{13}$$

where $\boldsymbol{\gamma}^0 = (1, \ldots, 1)^\prime \in \mathbb{R}^r$, $\boldsymbol{\beta} = (1, \ldots, 1)^\prime \in \mathbb{R}^L$, $\epsilon_t \sim$ i.i.d.$N\left(0, \sigma_\epsilon^2\right)$, and $\mathbf{w}_t = (w_{t,1}, \ldots, w_{t,L})^\prime$ with $w_{t,\ell} = \sum_{k=1}^r \rho_\ell[f_{t,k}^0 - \mathbb{E}\left(f_k^0\right)] + \varepsilon_{w,t,\ell}$, where $\varepsilon_{w,t,\ell} \sim$ i.i.d.$N(0, \sigma_{w,\ell}^2)$, $\ell = 1, \ldots, L$, and $\sigma_{w,\ell}$ is chosen such that $\mathrm{Cov}(f_{t,k}^0, w_{t,\ell}) =: \rho_\ell$ for all k given ℓ. Note that $\mathbf{f}_t^{0\prime}\boldsymbol{\gamma}^0 = \mathbf{f}_t^{*\prime}\boldsymbol{\gamma}^*$ with $\boldsymbol{\gamma}^* = \mathbf{H}\boldsymbol{\gamma}^0$ and $\mathbf{f}_t^{*\prime} = \mathbf{f}_t^{0\prime}\mathbf{H}^{-1}$ in (13). We choose $L = 1$, $\sigma_{w,\ell} = 1$, $\rho_\ell = 0.5$ for all ℓ, and $\sigma_\epsilon = 1$, so that the population R^2 of the augmented regression (13) is $3/4$. We have chosen $\alpha = 0.05$ and $h = 1$. All the results are obtained by 50,000 replications as in Section 7.1.1.

The infeasible estimator of $\boldsymbol{\gamma}^0$ is defined as $\hat{\boldsymbol{\gamma}}^0 = (\sum_{t=1}^{T-h} \tilde{\mathbf{f}}_t^0\tilde{\mathbf{f}}_t^{0\prime})^{-1} \sum_{t=1}^{T-h} \tilde{\mathbf{f}}_t^0 y_{t+h}$, where $\tilde{\mathbf{f}}_t^{0\prime}$ is tth rows of the $T\times r$ matrix $\mathbf{M}_w\mathbf{F}^0$ with $\mathbf{M}_w = \mathbf{I}_T - \mathbf{W}(\mathbf{W}^\prime\mathbf{W})^{-1}\mathbf{W}^\prime$ and $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_T)^\prime$. The feasible estimator of $\boldsymbol{\gamma}^0$ is defined by replacing $\mathbf{F}^0$ with $\hat{\mathbf{F}}$ in $\hat{\boldsymbol{\gamma}}^0$, and denoted by $\hat{\boldsymbol{\gamma}}$.

7.2.2 Estimation accuracy and normal approximation

Figure 4 reports the norm losses, $\|\hat{\boldsymbol{\gamma}}^0 - \boldsymbol{\gamma}^0\|_{\mathrm{F}}$, $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\|_{\mathrm{F}}$ and $\|\hat{\boldsymbol{\gamma}} - \hat{\mathbf{H}}^{-1}\boldsymbol{\gamma}^*\|_{\mathrm{F}}$. Replacing $\mathbf{F}^0$ with the *consistent* estimator $\hat{\mathbf{F}}$ has no significant effect, but approximating $\hat{\mathbf{F}}$ with $\mathbf{F}^*\hat{\mathbf{H}}$ leads to much worse performance, the inaccuracy of which is exaggerated as the underlining factor model becomes weaker. As the sample size increases, all norm losses decrease.

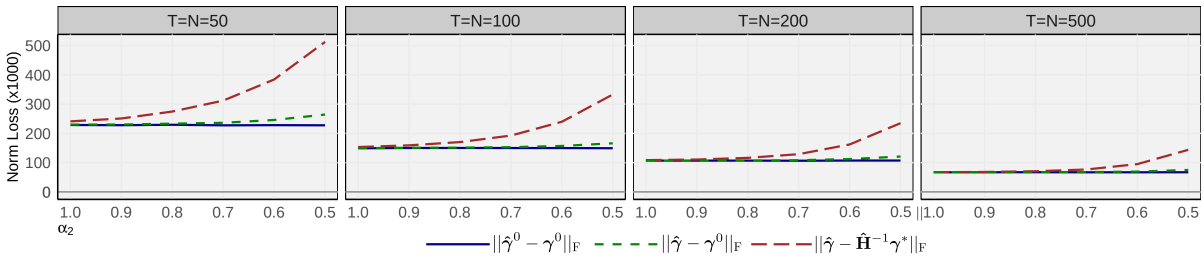

Figure 4: Plots of $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_{\mathrm{F}}$.

Next, we consider approximations of the three statistics to the χ_r^2 distribution:

$$Q_\gamma^2\left(\Delta\hat{\boldsymbol{\gamma}}\right) = T\left(\Delta\hat{\boldsymbol{\gamma}}\right)^\prime \boldsymbol{\Sigma}_{\gamma^0}^{-1}\left(\Delta\hat{\boldsymbol{\gamma}}\right),$$

where $\Delta\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}^0 - \boldsymbol{\gamma}^0, \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0, \hat{\boldsymbol{\gamma}} - \hat{\mathbf{H}}^{-1}\boldsymbol{\gamma}^*$, and $\boldsymbol{\Sigma}_{\gamma^0} = \sigma_\epsilon^2\left(\mathbf{F}^{0\prime}\mathbf{M}_w\mathbf{F}^0\right)^{-1}$. We assess the accuracy of the tail behavior of the distribution by computing the frequencies of $Q_\gamma^2(\Delta\hat{\boldsymbol{\gamma}}) > \chi_{r,0.95}^2$ over the replications. The closer the rejection frequency is to 5%, the more accurate the approximation by the statistics. Figure 5 summarizes the frequencies. The approximation of

$\hat{\mathbf{F}}$ to $\mathbf{F}^*\hat{\mathbf{H}}$ gives about 60% of rejection frequency for the weakest model, grossly exceeding 5%. In contrast, our approximation relative to the pseudo true parameter looks very accurate, apart from the weakest model. Actually, this weakest model with $(\alpha_1, \alpha_2) = (0.7, 0.5)$ is weak enough to violate the conditions in Theorem 4, $\alpha_r > 0.5$ and $\alpha_1 < 3\alpha_r - 1$ for $N \asymp T$.

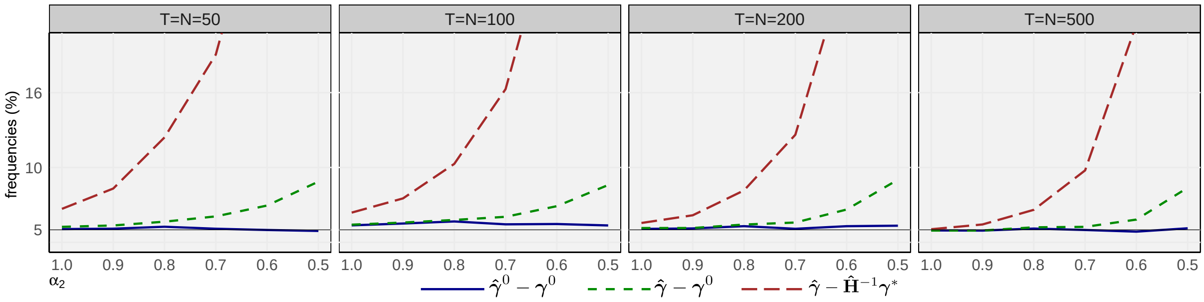

Figure 5: Plots of frequencies of $Q^2\left(\Delta\hat{\gamma}\right) > \chi^2_{r,0.95}$

7.2.3 Test of linear restrictions

We report the size of the 5% level test for $H_0 : \gamma_k = \gamma_k^0$ vs. $H_1 : \gamma_k \neq \gamma_k^0$ for $k = 1, \ldots, r$ using the statistics $z\left(\Delta\gamma_k\right) = [T\sigma_{\gamma^0}^{(k,k)}]^{\frac{1}{2}}\Delta\gamma_k$ for $\Delta\gamma_k = \hat{\gamma}_k^0 - \gamma_k^0$ and $\hat{\gamma}_k - \gamma_k^0$, where $\sigma_{\gamma^0}^{(k,k)}$ is the kth diagonal element of $\mathbf{\Sigma}_{\gamma^0}^{-1}$. The rejection frequencies are computed in terms of $N(0,1)$. Figure 6 confirms to our theory; the tests have the correct size except the weakest model, but it does not satisfy the conditions in Theorem 4.

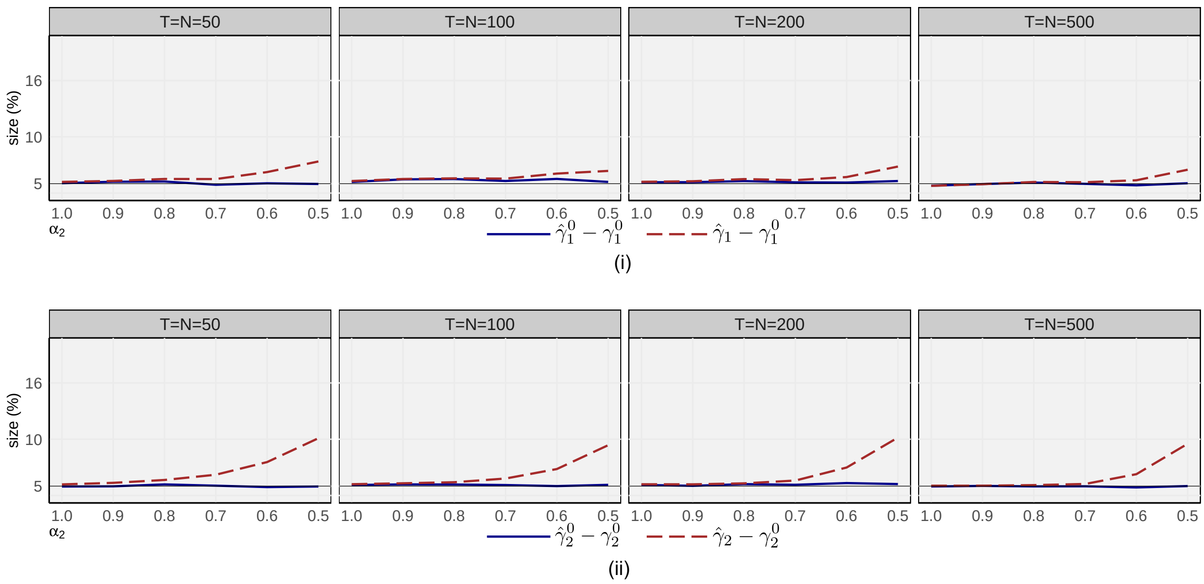

Figure 6: Size of the 5% level tests for (i) $H_0 : \gamma_1 = \gamma_1^0$; (ii) $H_0 : \gamma_2 = \gamma_2^0$.

7.2.4 Coverage by the confidence interval

We examine the empirical coverage of the $1-\alpha$ confidence interval for the h-step ahead forecast, $\hat{y}_{T+h|T} = \hat{\mathbf{f}}_T'\hat{\gamma} + \mathbf{w}_T'\hat{\boldsymbol{\beta}}$, against the conditional mean, $y_{T+h|T} = \mathbf{f}_T^{0\prime}\gamma^0 + \mathbf{w}_T'\boldsymbol{\beta}$, as well as the actual value, y_{T+h}. Denoting by $y_{T+h|T}^{(n)}$, $y_{T+h}^{(n)}$, and $\hat{y}_{T+h|T}^{(n)}$, the values for the nth replication for $n = 1, \ldots, R$, we compute the average coverages of $y_{T+h|T}$ and y_{T+h} by

$$CI_\alpha(y) = R^{-1}\sum_{n=1}^{R} 1\{y^{(n)} \in [LB_{\alpha,y}^{(n)}, UB_{\alpha,y}^{(n)}]\}, \qquad y = y_{T+h|T},\ y_{T+h},$$

where $UB_{\alpha,y_{T+h|T}}^{(n)}, LB_{\alpha,y_{T+h|T}}^{(n)} = \hat{y}_{T+h|T}^{(n)} \pm z_{1-\alpha/2}\sigma_{T+h|T}$ and $UB_{\alpha,y_{T+h}}^{(n)}, LB_{\alpha,y_{T+h}}^{(n)} = \hat{y}_{T+h|T}^{(n)} \pm z_{1-\alpha/2}(\sigma_{T+h|T}^2 + \sigma_\epsilon^2)^{\frac{1}{2}}$ with $z_{1-\alpha/2}$ defined as $\mathbb{P}(|Z| \leq z_{1-\alpha/2}) = 1-\alpha$ for $Z \sim N(0,1)$ and $\sigma_{T+h|T}^2 = \sigma_\epsilon^2 \mathbf{z}_T^{0\prime}\left(\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\mathbf{z}_T^0 + \sigma_e^2\gamma^{0\prime}(\mathbf{B}^{0\prime}\mathbf{B}^0)^{-1}\gamma^0$, where $\mathbf{Z}^0 = (\mathbf{F}^0, \mathbf{W}) = (\mathbf{z}_1^0, \ldots, \mathbf{z}_T^0)'$. We consider $\alpha = 0.9, 0.95$. Figure 7 shows that the empirical coverages are very close to the nominal level for all the models with the coverage accuracy improving as the sample size increases.

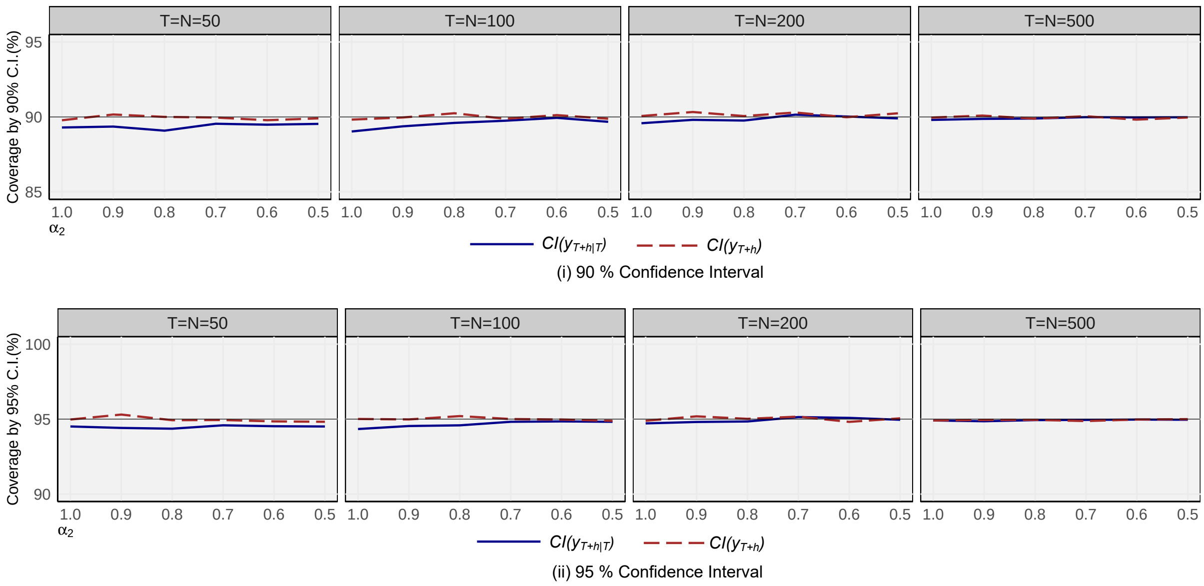

Figure 7: Empirical converges of (i) 90%; (ii) 95% confidence intervals for $y_{T+h|T}$ and y_{T+h}.

8 Conclusion

In the literature, including (Bai, 2003) and (Bai and Ng, 2023), the PC estimators have been considered as estimating $(\mathbf{F}^*\hat{\mathbf{H}}, \mathbf{B}^*\hat{\mathbf{H}}^{-1\prime})$, where $\mathbf{F}^*$ and $\mathbf{B}^*$ are the true parameters in model (1) and $\hat{\mathbf{H}}$ defined in (2). Since the "rotated parameters" depend on the PC estimators via $\hat{\mathbf{H}}$, however, this does not mean establishing the *consistency* of the PC estimator. A natural question is what the PC estimator estimates. To answer the question, we have achieved some theoretical results. First, under a quite general condition, we have proved the existence and uniqueness of rotation matrix $\mathbf{H}$ that depends only on the true parameters, $(\mathbf{F}^*, \mathbf{B}^*)$, and that rotates the true model in (1) to the pseudo-true model in (5) satisfying (4) (Theorem

1). Next, thanks to the identifiability of the obtained pseudo-parameters, $(\mathbf{F}^0, \mathbf{B}^0)$, we have proved the consistency and asymptotic normality of the PC estimators for them (Theorems 2 and 3). From these observations, we can say that the PC estimator consistently estimates the pseudo-true parameter that is uniquely determined by the true model in (1).

Another significant aspect of this paper is that for asymptotic theory we have considered the WF models, which allow the r largest eigenvalues of $\mathbf{B}^*\mathbf{F}^{*\prime}\mathbf{F}^*\mathbf{B}^{*\prime}$ to diverge at possibly different rates N^{α_k} for $\alpha_1 \geq \cdots \geq \alpha_r > 0$. This modeling framework is important in view of real data analyses, but it makes the theory difficult. Remarkably, the asymptotic normality of the PC estimator (Theorem 3) can be used for statistical inference for each of r (pseudo-true) factors and factor loadings. On the other hand, the normal approximation with data-dependent rotation matrices is not generally applicable to inference.

We have considered a similar problem in the factor augmented regression; using the approximation with the data-dependent rotation matrix $\hat{\mathbf{F}} \approx \mathbf{F}^*\hat{\mathbf{H}}$, the model slope coefficients on the regressor $\hat{\mathbf{F}}$ depend on $\hat{\mathbf{H}}^{-1}$, which is a function of the regressor $\hat{\mathbf{F}}$ itself. Thus, a t-test for the significance of the kth PC factor in the regression does not seem to be asymptotically justified though it is routinely reported in empirical studies; for an example, see (Ludvigson and Ng, 2009, Table 2). Using our approximation $\hat{\mathbf{F}} \approx \mathbf{F}^0$, such a test is asymptotically justified. We have established the consistency and the asymptotic normality of the least squares estimator of factor augmented regression coefficients allowing for WF models.

We have carried out extensive finite sample experiments for different divergence rates of the factor strength. The results show that the accuracy of the approximation of the PC estimators by the pseudo-true parameters is almost always better than the data-dependent ones. Importantly, the size of the t-tests of $H_0 : f_{t,k} = f^0_{t,k}$ and $H_0 : b_{i,k} = b^0_{i,k}$ are very close to the significance level. We note that such t-tests are only justified by our approach. Similarly, the accuracy of the least squares estimators and the joint normality approximation for the factor augmented regression with our rotation is shown to be almost always better than those with the data-dependent rotations, such as in Bai and Ng (2006). The size of the t-test for significance of each regression coefficient is shown to be very close to the level of significance. Again, such t-tests are only asymptotically justified by our article.

Finally, we comment on the estimation of the structural parameters $(\mathbf{F}^*, \mathbf{B}^*)$. As discussed in Uematsu and Yamagata (2023a), to directly identify the true loading matrix $\mathbf{B}^*$, r^2 (or more) constraints should be imposed along with the cross-sectional ordering of $x_{t,i}$. Such restrictions are informed exogenously. One way to identify structural parameters is to look for constraints, guided by economic and financial theory; see discussions in Stock and Watson (2016). Another is to use exogenous shocks that result in structural breaks in the statistical model. Recently, Yamamoto and Hara (2022) proposed a method to identify factor augmented regression models using changes in unconditional shock variances. Extending our approach to these methods would be a useful direction for future research.

Acknowledgment

We are grateful to Naoko Hara, Kazuhiko Hayakawa, and Yohei Yamamoto for helpful discussions and useful comments.

Funding

This work was supported by JSPS KAKENHI (grant numbers 20H01484, 21H00700, 21H04397 and 23H00804).

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–171.

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

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference with factor-augmented regressions. *Econometrica 74*, 1133–1150.

Bai, J. and S. Ng (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*(2), 304–317.

Bai, J. and S. Ng (2009). Boosting diffusion indices. *Journal of Applied Econometrics 24*(4), 607–629.

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of Econometrics 176*, 18–29.

Bai, J. and S. Ng (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica 51*, 1281–1304.

Cheng, X. and B. E. Hansen (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics 186*(2), 280–293.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15*, 373–394.

Connor, G. and R. A. Korajczyk (1993). A test for the number of factors in an approximate factor modela test for the number of factors in an approximate factor model. *Journal of Finance 48*, 1263–1291.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B 75*, 603–680.

Freyaldenhoven (2022). Factor models with local factors - determining the number of relevant factors. *Journal of Econometrics 229*, 80–102.

Gonçalves, S. and B. Perron (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics 182*(1), 156–173.

Ludvigson, C. S. and S. Ng (2009). Macro factors in bond risk premia. *Review of Financial Studies 22*, 5027–5067.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics 92*, 1004–1016.

Stock, J. and M. Watson (2016). Chapter 8 - dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. Volume 2 of *Handbook of Macroeconomics*, pp. 415–525. Elsevier.

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

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 30*, 147–162.

Uematsu, Y. and T. Yamagata (2023a). Estimation of sparsity-induced weak factor models. *Journal of Business & Economic Statistics 41*, 213–227.

Uematsu, Y. and T. Yamagata (2023b). Inference in sparsity-induced weak factor models. *Journal of Business & Economic Statistics 41*, 126–139.

Wei, J. and Y. Zhang (2023). Does principal component analysis preserve the sparsity in sparse weak factor models? *arXiv:2305.05934*.

Yamamoto, Y. and N. Hara (2022). Identifying factor-augmented vector autoregression models via changes in shock variances. *Journal of Applied Econometrics 37*(4), 722–745.

Supplementary Material for

Revisiting Asymptotic Theory for Principal Component Estimators of Approximate Factor Models

Peiyun Jiang[†], Yoshimasa Uematsu[*], Takashi Yamagata[‡]

[†] *Graduate School of Management, Tokyo Metropolitan University*
[*] *Department of Social Data Science, Hitotsubashi University*
[‡] *Department of Economics and Related Studies, University of York*
[‡] *Institute of Social Economic Research, Osaka University*

A Proofs of the Main Results

A.1 Proofs for the results in Section 2

Proof of Lemma 1. The eigen-decomposition of $\mathbf{B}^{*\prime}\mathbf{B}^{*}\left(T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^{*}\right)$ yields

$$\mathbf{U}\mathbf{V} = \mathbf{\Lambda} = \mathbf{\Lambda}' = \mathbf{V}'\mathbf{U}' = \mathbf{V}\mathbf{U}.$$

Since $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Lambda}$ are invertible under Assumption 1, we obtain

$$\mathbf{\Lambda}^{-1}\mathbf{U} = \mathbf{V}^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1} \quad \text{and} \quad \mathbf{V}\mathbf{\Lambda}^{-1} = \mathbf{U}^{-1} = \mathbf{\Lambda}^{-1}\mathbf{V}. \tag{A.1}$$

From the first equation in (A.1), we have $(\mathbf{\Lambda}^{-1}\mathbf{U})_{ij} = (\mathbf{U}\mathbf{\Lambda}^{-1})_{ij}$ for any $i, j \in \{1, \dots, r : i \neq j\}$, which entails $(1/\lambda_i - 1/\lambda_j)u_{ij} = 0$. Because λ_i's are distinct and bounded away from zero by Assumption 1, it must reduce $u_{ij} = 0$; that is, $\mathbf{U}$ is a diagonal matrix. In the same way, $\mathbf{V}$ is also a diagonal matrix from the second equation in (A.1). This completes the proof. □

Proof of Theorem 1. For $\mathbf{F}^0 = \mathbf{F}^*\mathbf{H}$ and $\mathbf{B}^0 = \mathbf{B}^*\mathbf{H}^{-1\prime}$ with $\mathbf{H} = \mathbf{P}\mathbf{V}^{-1/2}$, Lemma 1 gives

$$T^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0 = \mathbf{V}^{-1/2}\mathbf{P}'\left(T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^*\right)\mathbf{P}\mathbf{V}^{-1/2} = \mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2} = \mathbf{I}$$

and

$$\mathbf{B}^{0\prime}\mathbf{B}^0 = \mathbf{V}^{1/2}\mathbf{P}^{-1}\mathbf{B}^{*\prime}\mathbf{B}^*\mathbf{P}^{-1\prime}\mathbf{V}^{1/2} = \mathbf{V}^{1/2}\mathbf{U}\mathbf{V}^{1/2} = \mathbf{U}\mathbf{V} = \mathbf{\Lambda} \quad \text{(diagonal)}.$$

Therefore (4) holds.

Finally, we prove the uniqueness of $\mathbf{H}$. Suppose there exists another rotation matrix $\bar{\mathbf{H}}$ such that $T^{-1}\bar{\mathbf{H}}'\mathbf{F}^{*\prime}\mathbf{F}^*\bar{\mathbf{H}} = \mathbf{I}$ and $\bar{\mathbf{H}}^{-1}\mathbf{B}^{*\prime}\mathbf{B}^*\bar{\mathbf{H}}^{-1\prime}$ becomes diagonal (with elements ordered decreasingly). Then we must have

$$\bar{\mathbf{H}}^{-1\prime}\bar{\mathbf{H}}^{-1} = T^{-1}\mathbf{F}^{*\prime}\mathbf{F}^* = \mathbf{P}^{-1\prime}\mathbf{V}\mathbf{P}^{-1} = \mathbf{H}'^{-1}\mathbf{H}^{-1},$$

which is equivalent to $\bar{\mathbf{H}}\bar{\mathbf{H}}' = \mathbf{H}\mathbf{H}'$. Thus it is written as $\bar{\mathbf{H}} = \mathbf{H}\mathbf{Q}$ for some orthogonal matrix $\mathbf{Q}$. We further have

$$\bar{\mathbf{H}}^{-1}\mathbf{B}^{*\prime}\mathbf{B}^{*}\bar{\mathbf{H}}^{-1\prime} = \mathbf{Q}^{\prime}\mathbf{H}^{-1}\mathbf{B}^{*\prime}\mathbf{B}^{*}\mathbf{H}^{-1\prime}\mathbf{Q} = \mathbf{Q}^{\prime}\mathbf{\Lambda}\mathbf{Q},$$

which must be diagonalized. Since $\mathbf{\Lambda}$ is a diagonal matrix with distinct elements in descending order and $\mathbf{Q}$ is orthogonal, the only choice of $\mathbf{Q}$ is $\mathbf{Q} = \mathbf{I}$. Hence, the construction of $\mathbf{H}$ is unique. This completes the proof. □

A.2 Proofs for the results in Section 4

Proof of Lemma 2. (i) Since $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{F}}$ are the eigenvalues and eigenvectors of $\mathbf{X}\mathbf{X}^{\prime}/T$, respectively, we have

$$T^{-1}\mathbf{X}\mathbf{X}^{\prime}\hat{\mathbf{F}} = \hat{\mathbf{F}}\hat{\mathbf{\Lambda}}.$$

Meanwhile, $\mathbf{X}\mathbf{X}^{\prime}$ can be expanded as

$$\mathbf{X}\mathbf{X}^{\prime} = \mathbf{F}^{0}\mathbf{B}^{0\prime}\mathbf{B}^{0}\mathbf{F}^{0\prime} + \mathbf{F}^{0}\mathbf{B}^{0\prime}\mathbf{E}^{\prime} + \mathbf{E}\mathbf{B}^{0}\mathbf{F}^{0\prime} + \mathbf{E}\mathbf{E}^{\prime}.$$

Thus, we obtain

$$\hat{\mathbf{F}} = \frac{1}{T}\left(\mathbf{F}^{0}\mathbf{B}^{0\prime}\mathbf{B}^{0}\mathbf{F}^{0\prime} + \mathbf{F}^{0}\mathbf{B}^{0\prime}\mathbf{E}^{\prime} + \mathbf{E}\mathbf{B}^{0}\mathbf{F}^{0\prime} + \mathbf{E}\mathbf{E}^{\prime}\right)\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}.$$

From $\hat{\mathbf{B}} = \mathbf{X}^{\prime}\hat{\mathbf{F}}/T$, we can see that

$$\begin{aligned}
\hat{\mathbf{B}} &= \frac{1}{T}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}} + \frac{1}{T}\mathbf{E}^{\prime}\hat{\mathbf{F}},\\
\mathbf{B}^{0\prime}\hat{\mathbf{B}} &= \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}} + \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\hat{\mathbf{F}},\\
\mathbf{B}^{0\prime}\hat{\mathbf{B}}\hat{\mathbf{\Lambda}}^{-1} &= \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1} + \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1},\\
\tilde{\mathbf{H}}_{4} &= \tilde{\mathbf{H}} + \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1},
\end{aligned}$$

where $\tilde{\mathbf{H}} = \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}$. Then, we have

$$\hat{\mathbf{F}} - \mathbf{F}^{*}\hat{\mathbf{H}}_{4} = \hat{\mathbf{F}} - \mathbf{F}^{0}\tilde{\mathbf{H}}_{4} = \left(\frac{1}{T}\mathbf{E}\mathbf{E}^{\prime}\hat{\mathbf{F}} + \frac{1}{T}\mathbf{E}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\right)\hat{\mathbf{\Lambda}}^{-1}.$$

Lemma B.1(iii) yields $\left\|\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\right\|_{\mathrm{F}} = O_p(\sqrt{TN^{\alpha_1}})$. Thus the inequality $\|\mathbf{A}\mathbf{B}\|_{\mathrm{F}} \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_{\mathrm{F}}$ gives

$$\left\|\mathbf{E}\mathbf{E}^{\prime}\hat{\mathbf{F}}\right\|_{\mathrm{F}} \leq \lambda_1[\mathbf{E}\mathbf{E}^{\prime}]\|\hat{\mathbf{F}}\|_{\mathrm{F}} = O_p\left(N+T\right)\sqrt{T}.$$

Therefore, we obtain the first result:

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{F}}-\mathbf{F}^{*}\hat{\mathbf{H}}_4\right\|_{\mathrm{F}} \\
&\le \left\|\frac{1}{T^{3/2}}\mathbf{E}\mathbf{E}'\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}+\left\|\frac{1}{T^{3/2}}\mathbf{E}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}} \\
&\le \frac{\left\|\mathbf{E}\mathbf{E}'\hat{\mathbf{F}}\right\|_{\mathrm{F}}}{T^{3/2}}\left\|\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}+\left\|\frac{1}{T^{3/2}}\left(\mathbf{E}\mathbf{B}^{0}\mathbf{N}^{-\frac{1}{2}}\right)\left(\mathbf{N}^{\frac{1}{2}}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\mathbf{N}^{-\frac{1}{2}}\right)\mathbf{N}^{\frac{1}{2}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}} \\
&= O_p\left(\left(\frac{N}{T}+1\right)N^{-\alpha_r}\right)+O_p\left(N^{-\frac{1}{2}\alpha_r}\right) \\
&= O_p\left(\frac{N^{1-\alpha_r}}{T}\right)+O_p\left(N^{-\frac{1}{2}\alpha_r}\right).
\end{aligned}$$

(ii) By the definition,

$$\hat{\mathbf{B}} = \frac{1}{T}\mathbf{X}'\hat{\mathbf{F}}=\frac{1}{T}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}}+\frac{1}{T}\mathbf{E}'\hat{\mathbf{F}},$$

implies

$$\hat{\mathbf{B}}-\mathbf{B}^{0}\tilde{\mathbf{Q}}'=\frac{1}{T}\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F}^{0}\tilde{\mathbf{H}}_4)+\frac{1}{T}\mathbf{E}'\mathbf{F}^{0}\tilde{\mathbf{H}}_4, \tag{A.2}$$

where $\tilde{\mathbf{Q}}'=T^{-1}\mathbf{F}^{0\prime}\hat{\mathbf{F}}$. Using $\tilde{\mathbf{H}}_4=O_p(1)$ in Lemma 2(iii), Lemma B.3(i) and

$$\left\|\frac{1}{T}\mathbf{E}'\mathbf{F}^{0}\tilde{\mathbf{H}}_4\right\|_{\mathrm{F}} = O_p\left(\frac{1}{T}\sqrt{NT}\right)=O_p\left(\sqrt{\frac{N}{T}}\right),$$

we have

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\left\|\hat{\mathbf{B}}-\mathbf{B}^{0}\tilde{\mathbf{Q}}'\right\|_{\mathrm{F}} \\
&=\frac{1}{\sqrt{N}}\left\|\frac{1}{T}\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F}^{0}\tilde{\mathbf{H}}_4)+\frac{1}{T}\mathbf{E}'\mathbf{F}^{0}\tilde{\mathbf{H}}_4\right\|_{F} \\
&\le \frac{1}{\sqrt{N}}\left[O_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right)+O_p\left(\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right)+O_p\left(N^{-\frac{1}{2}\alpha_r}\right)\right]+\frac{1}{\sqrt{N}}O_p\left(\sqrt{\frac{N}{T}}\right) \\
&=O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(N^{-\frac{1}{2}-\frac{1}{2}\alpha_r}\right),
\end{aligned}$$

if $\frac{N^{1-\alpha_r}}{T}\to 0$.

(iii) Let $\tilde{\mathbf{H}}_2=(\mathbf{F}^{0\prime}\mathbf{F}^{0})^{-1}\mathbf{F}^{0\prime}\hat{\mathbf{F}}=\mathbf{F}^{0\prime}\hat{\mathbf{F}}/T=\tilde{\mathbf{Q}}'$. A comparable result to Lemma 2(i) is

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{F}}-\mathbf{F}^{0}\tilde{\mathbf{H}}_2\right\|_{\mathrm{F}} \\
&\le \frac{1}{\sqrt{T}}\left\|\hat{\mathbf{F}}-\mathbf{F}^{0}\tilde{\mathbf{H}}_4\right\|_{\mathrm{F}}+\frac{1}{\sqrt{T}}\left\|\mathbf{F}^{0}(\tilde{\mathbf{H}}_2-\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}} \\
&= O_p\left(\frac{N^{1-\alpha_r}}{T}\right)+O_p\left(N^{-\frac{1}{2}\alpha_r}\right)+O_p(\Delta_1),
\end{aligned}$$

where we have used Lemmas B.5(ii), B.6(i), and Δ_1 is defined by (B.10). Suppose that $\frac{N^{1-\alpha_r}}{T} \to 0$ holds, $\frac{1}{\sqrt{T}} \left\| \hat{\mathbf{F}} - \mathbf{F}^0 \tilde{\mathbf{H}}_2 \right\|_{\mathrm{F}} = o_p(1)$ implies that $\tilde{\mathbf{H}}_2 = O_p(1)$. If $\|\tilde{\mathbf{H}}_2\|_{\mathrm{F}} \to \infty$, then $\left\| \frac{1}{\sqrt{T}} (\hat{\mathbf{F}} - \mathbf{F}^0 \tilde{\mathbf{H}}_2) \right\|_{\mathrm{F}}^2 = tr(\mathbf{I}_r - \tilde{\mathbf{H}}_2' \tilde{\mathbf{H}}_2)$ will diverge, which contradicts the prior result. If $\|\tilde{\mathbf{H}}_4\|_{\mathrm{F}} \to \infty$, then $\left\| \frac{1}{\sqrt{T}} (\hat{\mathbf{F}} - \mathbf{F}^0 \tilde{\mathbf{H}}_4) \right\|_{\mathrm{F}}^2 = tr(\mathbf{I}_r - \tilde{\mathbf{H}}_4' \tilde{\mathbf{H}}_2 - \tilde{\mathbf{H}}_2' \tilde{\mathbf{H}}_4 + \tilde{\mathbf{H}}_4' \tilde{\mathbf{H}}_4)$ will diverge. This contradiction implies that $\tilde{\mathbf{H}}_4$ is $O_p(1)$. Similarly, $\frac{1}{\sqrt{T}} \left\| \hat{\mathbf{F}} - \mathbf{F}^0 \tilde{\mathbf{H}} \right\|_{\mathrm{F}} = o_p(1)$ implies that $\tilde{\mathbf{H}} = O_p(1)$, which is indicated by Wei and Zhang (2023, Proof of Lemma A.4).

Next, we show the probability limit of the rotation matrices assuming that $\frac{N^{1-\alpha_r}}{T} \to 0$ and $\alpha_1 < 2\alpha_r$. Given Lemma B.4(i), we obtain $\tilde{\mathbf{Q}} \xrightarrow{p} \mathbf{I}_r$. Lemmas B.5(ii), B.6(i), and B.3(iii) imply

$$\left\| \tilde{\mathbf{Q}}' - \tilde{\mathbf{H}}_4 \right\|_{\mathrm{F}} = O_p(\Delta_1), \quad \left\| \tilde{\mathbf{Q}}' - \tilde{\mathbf{H}} \right\|_{\mathrm{F}} = O_p(\Delta_{NT}),$$

which are $o_p(1)$ and Δ_1, Δ_{NT} are defined as (B.10), (B.1). Thus, we complete the proof. □

Proof of Theorem 2. (i) Lemma B.3(iii) shows that $\left\| \tilde{\mathbf{H}}' - \frac{\hat{\mathbf{F}}' \mathbf{F}^0}{T} \right\|_{\mathrm{F}} = O_p\left(\Delta_{NT}\right)$, where Δ_{NT} is defined by (B.1). From Lemma B.4(i), we have

$$\left\| \tilde{\mathbf{H}} - \mathbf{I}_r \right\|_{\mathrm{F}} \leq \left\| \tilde{\mathbf{H}} - \frac{\mathbf{F}^{0\prime} \hat{\mathbf{F}}}{T} \right\|_{\mathrm{F}} + \left\| \frac{\mathbf{F}^{0\prime} \hat{\mathbf{F}}}{T} - \mathbf{I}_r \right\|_{\mathrm{F}} = O_p\left(\Delta_{NT}\right).$$

By the definition of $\hat{\mathbf{B}}$ and expanding X, we obtain

$$\hat{\mathbf{F}} = \frac{1}{T} \mathbf{X} \mathbf{X}' \hat{\mathbf{F}} \hat{\mathbf{\Lambda}}^{-1} = \mathbf{X} \hat{\mathbf{B}} \hat{\mathbf{\Lambda}}^{-1} = \mathbf{F}^0 \mathbf{B}^{0\prime} \hat{\mathbf{B}} \hat{\mathbf{\Lambda}}^{-1} + \mathbf{E} \hat{\mathbf{B}} \hat{\mathbf{\Lambda}}^{-1}.$$

Then, $\frac{1}{T} \mathbf{F}^{0\prime} \hat{\mathbf{F}} = \tilde{\mathbf{H}}_4 + \frac{1}{T} \mathbf{F}^{0\prime} \mathbf{E} \hat{\mathbf{B}} \hat{\mathbf{\Lambda}}^{-1}$. Applying Lemmas 2(i), B.6(i), and B.4(i), we show the first result:

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \left\| \hat{\mathbf{F}} - \mathbf{F}^0 \right\|_{\mathrm{F}} \\
&\leq \frac{1}{\sqrt{T}} \left\| \hat{\mathbf{F}} - \mathbf{F}^0 \tilde{\mathbf{H}}_2 \right\|_{\mathrm{F}} + \frac{1}{\sqrt{T}} \left\| \mathbf{F}^0 (\tilde{\mathbf{H}}_2 - \mathbf{I}_r) \right\|_{\mathrm{F}} \\
&= \frac{1}{\sqrt{T}} \left\| \frac{1}{T} \mathbf{E} \mathbf{E}' \hat{\mathbf{F}} \hat{\mathbf{\Lambda}}^{-1} + \frac{1}{T} \mathbf{E} \mathbf{B}^0 \mathbf{F}^{0\prime} \hat{\mathbf{F}} \hat{\mathbf{\Lambda}}^{-1} - \frac{1}{T} \mathbf{F}^0 \mathbf{F}^{0\prime} \mathbf{E} \hat{\mathbf{B}} \hat{\mathbf{\Lambda}}^{-1} \right\|_{\mathrm{F}} + O_p\left(\Delta_{NT}\right) \\
&\leq O_p\left(\left(\frac{N}{T} + 1 \right) N^{-\alpha_r} \right) + O_p\left(N^{-\frac{1}{2}\alpha_r} \right) + O_p\left(\Delta_1\right) + O_p\left(\Delta_{NT}\right) \\
&= O_p\left(\left(\frac{N}{T} + 1 \right) N^{-\alpha_r} \right) + O_p\left(N^{-\frac{1}{2}\alpha_r} \right) \\
&\quad + \left[\left(\frac{N^{1-\alpha_r}}{T} \right)^2 + \left(\frac{N^{1-\alpha_r}}{T} \right)^{\frac{3}{2}} + \sqrt{\frac{N^{1-\alpha_r}}{T}} N^{-\alpha_r} + \frac{N^{1-\alpha_r}}{T} + \frac{1}{\sqrt{T N^{\alpha_r}}} \right] O_p\left(1\right) \\
&\quad + \left(\frac{N^{1-\alpha_r}}{T} + N^{\frac{1}{2}\alpha_1 - \alpha_r} \frac{N^{1-\alpha_r}}{T} + N^{\frac{1}{2}\alpha_1 - \frac{3}{2}\alpha_r} + \frac{N^{\frac{1}{2}\alpha_1 - \alpha_r}}{\sqrt{T}} \right) O_p(1) \\
&= O_p\left(\frac{N^{1-\alpha_r}}{T} \right) + O_p\left(N^{-\frac{1}{2}\alpha_r} \right),
\end{aligned}$$

where Δ_1 and Δ_{NT} are defined by (B.10) and (B.1), if $\frac{N^{1-\alpha_r}}{T} \to 0$ and $\frac{1}{2}\alpha_1 < \alpha_r$. Note that

$$N^{\frac{1}{2}\alpha_1 - \frac{3}{2}\alpha_r} = N^{-\frac{1}{2}\alpha_r} N^{\frac{1}{2}\alpha_1 - \alpha_r} \lesssim N^{-\frac{1}{2}\alpha_r},$$

$$\frac{N^{\frac{1}{2}\alpha_1 - \alpha_r}}{\sqrt{T}} = \frac{N^{\frac{1}{2}\alpha_1 - \frac{1}{2}\alpha_r}}{\sqrt{T}} N^{-\frac{1}{2}\alpha_r} \lesssim N^{-\frac{1}{2}\alpha_r},$$

$$N^{\frac{1}{2}\alpha_1 - \alpha_r} \frac{N^{1-\alpha_r}}{T} \lesssim \frac{N^{1-\alpha_r}}{T}.$$

(ii) The proof of Lemma 2(ii) and Lemma B.4(i) imply

$$\begin{aligned}
&\frac{1}{\sqrt{N}} \left\| \hat{\mathbf{B}} - \mathbf{B}^0 \right\|_{\mathrm{F}} \\
&\leq \frac{1}{\sqrt{N}} \left\| \hat{\mathbf{B}} - \mathbf{B}^0 \tilde{\mathbf{Q}}' \right\|_{\mathrm{F}} + \frac{1}{\sqrt{N}} \left\| \mathbf{B}^0 \tilde{\mathbf{Q}}' - \mathbf{B}^0 \right\|_{\mathrm{F}} \\
&= O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(N^{-\frac{1}{2}-\frac{1}{2}\alpha_r}\right) + O_p\left(N^{\frac{1}{2}\alpha_1 - \frac{1}{2}}\right) O_p\left(\Delta_{NT}\right) \\
&= O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(N^{-\frac{1}{2}-\frac{1}{2}\alpha_r}\right) + \left(\frac{N^{\frac{1}{2}+\frac{1}{2}\alpha_1-\alpha_r}}{T} + N^{\alpha_1 - \frac{3}{2}\alpha_r - \frac{1}{2}} + \frac{N^{\frac{1}{2}+\alpha_1-2\alpha_r}}{T} + \frac{N^{\alpha_1-\alpha_r-\frac{1}{2}}}{\sqrt{T}}\right) O_p(1) \\
&= O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(N^{\alpha_1 - \frac{3}{2}\alpha_r - \frac{1}{2}}\right) + o_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\right),
\end{aligned}$$

where the final equality is because

$$\frac{N^{\frac{1}{2}+\frac{1}{2}\alpha_1-\alpha_r}}{T} = \frac{N^{1-\alpha_r}}{T} N^{\frac{1}{2}\alpha_1 - \frac{1}{2}} = o\left(\frac{N^{1-\alpha_r}}{T}\right),$$

$$\frac{N^{\frac{1}{2}+\alpha_1-2\alpha_r}}{T} = \frac{N^{\frac{1}{2}+\frac{1}{2}\alpha_1-\alpha_r}}{T} N^{\frac{1}{2}\alpha_1 - \alpha_r} = o\left(\frac{N^{1-\alpha_r}}{T}\right),$$

$$\frac{N^{\alpha_1-\alpha_r-\frac{1}{2}}}{\sqrt{T}} = \sqrt{\frac{N^{\alpha_1-\alpha_r}}{T}} N^{\frac{1}{2}\alpha_1 - \frac{1}{2} - \frac{1}{2}\alpha_r} = o\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\right),$$

if $\frac{N^{1-\alpha_r}}{T} \to 0$ and $\alpha_1 < 2\alpha_r$.

(iii) Applying Lemmas 2 and B.5(vi), we obtain

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}}\left\|\hat{\mathbf{C}}-\mathbf{C}^*\right\|_{\mathrm{F}} \\
&\le \frac{1}{\sqrt{NT}}\left\|\hat{\mathbf{F}}\hat{\mathbf{B}}'-\mathbf{F}^0\tilde{\mathbf{H}}_4\hat{\mathbf{B}}'\right\|_{\mathrm{F}}+\frac{1}{\sqrt{NT}}\left\|\mathbf{F}^0\tilde{\mathbf{H}}_4\hat{\mathbf{B}}'-\mathbf{F}^0\tilde{\mathbf{H}}_4\tilde{\mathbf{H}}_4^{-1}\mathbf{B}^{0\prime}\right\|_{\mathrm{F}} \\
&= \frac{1}{\sqrt{NT}}\left\|(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\hat{\mathbf{B}}'\right\|_{\mathrm{F}}+\frac{1}{\sqrt{NT}}\left\|\mathbf{F}^0\tilde{\mathbf{H}}_4(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{H}}_2+\mathbf{B}^0\tilde{\mathbf{H}}_2-\mathbf{B}^0\tilde{\mathbf{H}}_4^{\prime -1})'\right\|_{\mathrm{F}} \\
&\le \frac{1}{\sqrt{NT}}\left\|(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\hat{\mathbf{B}}'\right\|_{\mathrm{F}}+\frac{1}{\sqrt{NT}}\left\|\mathbf{F}^0\tilde{\mathbf{H}}_4\right\|_{\mathrm{F}}\left\|\frac{1}{T}\mathbf{E}'\hat{\mathbf{F}}-\mathbf{B}^0\frac{1}{T}(\hat{\mathbf{B}}'\mathbf{B}^0)^{-1}\hat{\mathbf{B}}'\mathbf{E}'\hat{\mathbf{F}}\right\|_{\mathrm{F}} \\
&\le \frac{1}{\sqrt{NT}}\left\|(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\hat{\mathbf{B}}'\right\|_{\mathrm{F}}+\frac{1}{\sqrt{NT}}\left\|\mathbf{F}^0\tilde{\mathbf{H}}_4\right\|_{\mathrm{F}}\left\|\frac{1}{T}\mathbf{E}'\hat{\mathbf{F}}\right\|_{\mathrm{F}} \\
&= O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}}\right)\left[O_p\left(\frac{N^{1-\alpha_r}}{T}\right)+O_p\left(N^{-\frac{1}{2}\alpha_r}\right)\right]+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(N^{-\frac{1}{2}-\frac{1}{2}\alpha_r}\right) \\
&= O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r-\frac{1}{2}}\right)+o_p\left(\frac{N^{1-\alpha_r}}{T}\right).
\end{aligned}
$$

□

A.3 Proofs for the results in Section 5

Proof of Lemma 3. (i) We multiply $\hat{\mathbf{B}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}$ to the both side of $\mathbf{X}=\mathbf{F}^0\mathbf{B}^{0\prime}+\mathbf{E}$, then, we have,

$$
\begin{aligned}
\mathbf{X}\hat{\mathbf{B}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1} &= \mathbf{F}^0\mathbf{B}^{0\prime}\hat{\mathbf{B}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}+\mathbf{E}\hat{\mathbf{B}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1} \\
\hat{\mathbf{F}} &= \mathbf{F}^0\tilde{\mathbf{H}}_4+\mathbf{E}\mathbf{B}^0\tilde{\mathbf{Q}}'(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}+\mathbf{E}(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{Q}}')(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}.
\end{aligned}
$$

Since $\mathbf{F}^*\hat{\mathbf{H}}_4=\mathbf{F}^0\tilde{\mathbf{H}}_4$, the t-th row of $\hat{\mathbf{F}}$ is given by

$$
\hat{\mathbf{f}}_t'-\mathbf{f}_t^{*\prime}\hat{\mathbf{H}}_4=\hat{\mathbf{f}}_t'-\mathbf{f}_t^{0\prime}\tilde{\mathbf{H}}_4=\mathbf{e}_t'\mathbf{B}^0\tilde{\mathbf{Q}}'(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}+\mathbf{e}_t'(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{Q}}')(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}.
$$

That is

$$
\hat{\mathbf{f}}_t-\hat{\mathbf{H}}_4'\mathbf{f}_t^*=\hat{\mathbf{f}}_t-\tilde{\mathbf{H}}_4'\mathbf{f}_t^0 \quad = \quad (\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\tilde{\mathbf{Q}}\mathbf{B}^{0\prime}\mathbf{e}_t+(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{Q}}')'\mathbf{e}_t.
$$

Then, we have

$$
\begin{aligned}
&\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{f}}_t-\hat{\mathbf{H}}_4'\mathbf{f}_t^*) \\
&= \mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{f}}_t-\tilde{\mathbf{H}}_4'\mathbf{f}_t^0) \\
&= \mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\tilde{\mathbf{Q}}\mathbf{B}^{0\prime}\mathbf{e}_t+\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\left(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{Q}}'\right)'\mathbf{e}_t.
\end{aligned}
\tag{A.3}
$$

We first consider the first term on the right-hand side of the above equation. By the order (B.3), $\mathbf{N}^{-\frac{1}{2}}\hat{\mathbf{B}}'\hat{\mathbf{B}}\mathbf{N}^{-\frac{1}{2}}-\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}}=O_p(\Delta_{NT})=o_p(1)$ if $\frac{N^{1-\alpha_r}}{T}\to 0$ and $\alpha_1<2\alpha_r$. It implies that

$$
\mathbf{D}^{-1}\mathbf{N}^{-\frac{1}{2}}\hat{\mathbf{B}}'\hat{\mathbf{B}}\mathbf{N}^{-\frac{1}{2}}\mathbf{D}^{-1}\xrightarrow{p}\mathbf{D}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}}\mathbf{D}^{-1}=\mathbf{I}_r.
$$

Now $\mathbf{D}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{e}_t \xrightarrow{d} N(0, \boldsymbol{\Gamma}_t)$ by Assumption 5 and $\tilde{\mathbf{Q}} - \mathbf{I}_r = O_p(\Delta_{NT})$. The first term on the right-hand side of (A.3) is thus asymptotically normal, and we obtain

$$\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\tilde{\mathbf{Q}}\mathbf{B}^{0\prime}\mathbf{e}_t = \mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\mathbf{B}^{0\prime}\mathbf{e}_t(\mathbf{I}_r + o_p(1)),$$

whose asymptotic distribution is

$$\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\mathbf{N}^{\frac{1}{2}}\mathbf{D}\mathbf{D}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{e}_t \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_t).$$

Next, the second term on the right-hand side of (A.3) is $o_p(1)$ by Lemmas B.7 if $\alpha_r > 1/2$ and $\frac{N^{\frac{3}{2}-\alpha_r}}{T} \to 0$. Collecting these results, we obtain (i):

$$\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{f}}_t - \hat{\mathbf{H}}_4'\mathbf{f}_t^*) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_t).$$

(ii) Recall (A.2)

$$\hat{\mathbf{B}} - \mathbf{B}^*\hat{\mathbf{Q}}' = \hat{\mathbf{B}} - \mathbf{B}^0\tilde{\mathbf{Q}}' = \frac{1}{T}\mathbf{E}'(\hat{\mathbf{F}} - \mathbf{F}^0\tilde{\mathbf{H}}_4) + \frac{1}{T}\mathbf{E}'\mathbf{F}^0\tilde{\mathbf{H}}_4.$$

The i-th row is given by

$$\hat{\mathbf{b}}_i' - \mathbf{b}_i^{*\prime}\hat{\mathbf{Q}}' = \hat{\mathbf{b}}_i' - \mathbf{b}_i^{0\prime}\tilde{\mathbf{Q}}' = \frac{1}{T}\mathbf{e}_i'(\hat{\mathbf{F}} - \mathbf{F}^0\tilde{\mathbf{H}}_4) + \frac{1}{T}\mathbf{e}_i'\mathbf{F}^0\tilde{\mathbf{H}}_4$$
$$\hat{\mathbf{b}}_i - \hat{\mathbf{Q}}\mathbf{b}_i^* = \hat{\mathbf{b}}_i - \tilde{\mathbf{Q}}\mathbf{b}_i^0 = \frac{1}{T}(\hat{\mathbf{F}} - \mathbf{F}^0\tilde{\mathbf{H}}_4)'\mathbf{e}_i + \frac{1}{T}\tilde{\mathbf{H}}_4'\mathbf{F}^{0\prime}\mathbf{e}_i.$$

We have

$$\sqrt{T}(\hat{\mathbf{b}}_i - \hat{\mathbf{Q}}\mathbf{b}_i^*) = \sqrt{T}(\hat{\mathbf{b}}_i - \tilde{\mathbf{Q}}\mathbf{b}_i^0) = \frac{1}{\sqrt{T}}\tilde{\mathbf{H}}_4'\mathbf{F}^{0\prime}\mathbf{e}_i + \frac{1}{\sqrt{T}}(\hat{\mathbf{F}} - \mathbf{F}^0\tilde{\mathbf{H}}_4)'\mathbf{e}_i. \tag{A.4}$$

By Lemmas B.5(ii), B.6(i) and B.4(i),

$$\tilde{\mathbf{H}}_4' - \mathbf{I}_r = \tilde{\mathbf{H}}_4' - \tilde{\mathbf{H}}_2' + \tilde{\mathbf{H}}_2' - \mathbf{I}_r = O_p(\Delta_1) + O_p(\Delta_{NT}) = o_p(1), \text{ if } \frac{N^{1-\alpha_r}}{T} \to 0 \text{ and } \alpha_1 < 2\alpha_r.$$

Under Assumption 5, the first term on the right-hand side of (A.4) is thus asymptotically normal. That is

$$\frac{1}{\sqrt{T}}\tilde{\mathbf{H}}_4'\mathbf{F}^{0\prime}\mathbf{e}_i = \frac{1}{\sqrt{T}}\mathbf{F}^{0\prime}\mathbf{e}_i + (\tilde{\mathbf{H}}_4' - \mathbf{I}_r)\frac{1}{\sqrt{T}}\mathbf{F}^{0\prime}\mathbf{e}_i = \frac{1}{\sqrt{T}}\mathbf{F}^{0\prime}\mathbf{e}_i + O_p(\Delta_{NT}),$$

whose asymptotic distribution is $N(\mathbf{0}, \boldsymbol{\Phi}_i)$.

Next, the second term on the right-hand side of (A.4) is bounded by

$$\left\|\frac{1}{\sqrt{T}}\mathbf{e}_i{}'(\hat{\mathbf{F}} - \mathbf{F}^0\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}} = O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right) + O_p\left(\sqrt{T}N^{-\alpha_r}\right) + O_p\left(N^{-\frac{1}{2}\alpha_r}\right),$$

from Lemma B.3(iv). It is negligible if $\frac{N^{1-\alpha_r}}{\sqrt{T}} \to 0$ and $\frac{\sqrt{T}}{N^{\alpha_r}} \to 0$ are satisfied. Collecting these results, we obtain (ii)

$$\sqrt{T}(\hat{\mathbf{b}}_i - \hat{\mathbf{Q}}\mathbf{b}_i^*) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Phi}_i).$$

if $\alpha_1 < 2\alpha_r$, $\frac{N^{1-\alpha_r}}{\sqrt{T}} \to 0$, and $\frac{\sqrt{T}}{N^{\alpha_r}} \to 0$. $\square$

Proof of Theorem 3. (i)

$$\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{f}}_t - \mathbf{f}_t^0) = \mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{f}}_t - \tilde{\mathbf{H}}_4'\mathbf{f}_t^0) + \mathbf{D}\mathbf{N}^{\frac{1}{2}}(\tilde{\mathbf{H}}_4' - \mathbf{I}_r)\mathbf{f}_t^0.$$

We only consider the second term on the right-hand side of the above equation. Lemmas B.5(ii), B.6(i), and B.4(i) imply

$$\mathbf{N}^{\frac{1}{2}}(\tilde{\mathbf{H}}_4' - \mathbf{I}_r) = \mathbf{N}^{\frac{1}{2}}\left(\tilde{\mathbf{H}}_4' - \tilde{\mathbf{H}}_2' + \tilde{\mathbf{H}}_2' - \mathbf{I}_r\right) \leq N^{\frac{1}{2}\alpha_1}O_p(\Delta_1) + N^{\frac{1}{2}\alpha_1}O_p(\Delta_{NT}) = o_p(1),$$

if $\frac{N^{\frac{3}{2}-\alpha_r}}{T} \to 0$, $N^{\alpha_1 - \frac{3}{2}\alpha_r} \to 0$, and $\frac{N^{\alpha_1-\alpha_r}}{\sqrt{T}} \to 0$. Note that

$$\frac{N^{1+\frac{1}{2}\alpha_1-\alpha_r}}{T} = \frac{N^{\frac{3}{2}-\alpha_r}}{T}N^{\frac{1}{2}\alpha_1-\frac{1}{2}} \to 0.$$

Using the result in Lemma 3, we obtain

$$\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{f}}_t - \mathbf{f}_t^0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_t).$$

(ii)

$$\sqrt{T}(\hat{\mathbf{b}}_i - \mathbf{b}_i^0) = \sqrt{T}(\hat{\mathbf{b}}_i - \tilde{\mathbf{Q}}\mathbf{b}_i^0) + \sqrt{T}(\tilde{\mathbf{Q}} - \mathbf{I}_r)\mathbf{b}_i^0.$$

The second term on the right-hand side of the above equation is bounded by

$$\left\|\sqrt{T}(\tilde{\mathbf{Q}} - \mathbf{I}_r)\mathbf{b}_i^0\right\|_{\mathrm{F}} = \sqrt{T}O_p\left(\Delta_{NT}\right) = o_p(1),$$

if $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r} \to 0$, $\frac{N^{1-\alpha_r}}{\sqrt{T}} \to 0$, and $\alpha_1 < 2\alpha_r$ hold. Thus, we have

$$\sqrt{T}(\hat{\mathbf{b}}_i - \mathbf{b}_i^0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Phi}_i).$$

(iii) By the definition of $\hat{c}_{t,i}$,

$$\begin{aligned}
&\hat{c}_{t,i} - c_{t,i}^* \\
&= \hat{\mathbf{b}}_i'\hat{\mathbf{f}}_t - \mathbf{b}_i^{0\prime}\mathbf{f}_t^0 \\
&= \hat{\mathbf{b}}_i'\hat{\mathbf{f}}_t - (\tilde{\mathbf{H}}_4^{-1}\mathbf{b}_i^0)'\tilde{\mathbf{H}}_4'\mathbf{f}_t^0 \\
&= (\tilde{\mathbf{H}}_4^{-1}\mathbf{b}_i^0)'(\hat{\mathbf{f}}_t - \tilde{\mathbf{H}}_4'\mathbf{f}_t^0) + (\hat{\mathbf{b}}_i - \tilde{\mathbf{H}}_4^{-1}\mathbf{b}_i^0)'\tilde{\mathbf{H}}_4'\mathbf{f}_t^0 + (\hat{\mathbf{b}}_i - \tilde{\mathbf{H}}_4^{-1}\mathbf{b}_i^0)'(\hat{\mathbf{f}}_t - \tilde{\mathbf{H}}_4'\mathbf{f}_t^0).
\end{aligned} \tag{A.5}$$

The first term on the right-hand side of (A.5) is

$$(\tilde{\mathbf{H}}_4^{-1}\mathbf{b}_i^0)'(\hat{\mathbf{f}}_t - \tilde{\mathbf{H}}_4'\mathbf{f}_t^0) = (\tilde{\mathbf{H}}_4^{-1}\mathbf{b}_i^0)'(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\tilde{\mathbf{Q}}\mathbf{B}^{0\prime}\mathbf{e}_t + (\tilde{\mathbf{H}}_4^{-1}\mathbf{b}_i^0)'(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\left(\hat{\mathbf{B}} - \mathbf{B}^0\tilde{\mathbf{Q}}'\right)'\mathbf{e}_t.$$

By Lemma 3(i), we have, if $\alpha_r > 1/2$ and $\frac{N^{\frac{3}{2}-\alpha_r}}{T} \to 0$,

$$\mathbf{b}_i^{0\prime}\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{f}}_t - \tilde{\mathbf{H}}_4'\mathbf{f}_t^0) \quad \xrightarrow{d} \quad N(0, \mathbf{b}_i^{0\prime}\boldsymbol{\Gamma}_t\mathbf{b}_i^0).$$

Next, consider the second term on the right-hand side of (A.5),

$$
\begin{aligned}
&\sqrt{T}(\hat{\mathbf{b}}_i - \tilde{\mathbf{H}}_4^{-1}\mathbf{b}_i^0)'\tilde{\mathbf{H}}_4'\mathbf{f}_t^0 \\
&= \sqrt{T}(\hat{\mathbf{b}}_i - \tilde{\mathbf{Q}}\mathbf{b}_i^0)'\tilde{\mathbf{H}}_4'\mathbf{f}_t^0 + \sqrt{T}(\tilde{\mathbf{H}}_4^{-1}\mathbf{b}_i^0 - \tilde{\mathbf{Q}}\mathbf{b}_i^0)'\tilde{\mathbf{H}}_4'\mathbf{f}_t^0 \\
&= \sqrt{T}(\hat{\mathbf{b}}_i - \tilde{\mathbf{Q}}\mathbf{b}_i^0)'\tilde{\mathbf{H}}_4'\mathbf{f}_t^0 + O_p(\sqrt{T}\Delta_{NT}),
\end{aligned}
$$

where $\tilde{\mathbf{H}}_4^{-1} - \tilde{\mathbf{Q}} = O_p(\Delta_{NT})$ by Lemmas B.5(vi) and B.6(iii). Applying the result in Lemma 3(ii), if $\alpha_1 < 2\alpha_r$, $\frac{N^{1-\alpha_r}}{\sqrt{T}} \to 0$, and $\sqrt{T}N^{\frac{1}{2}\alpha_1 - \frac{3}{2}\alpha_r} \to 0$, we have

$$
\sqrt{T}(\hat{\mathbf{b}}_i - \tilde{\mathbf{H}}_4^{-1}\mathbf{b}_i^0)'\tilde{\mathbf{H}}_4'\mathbf{f}_t^0 \xrightarrow{d} N(\mathbf{0}, \mathbf{f}_t^{0\prime}\mathbf{\Phi}_i\mathbf{f}_t^0).
$$

The third term on the right-hand side of (A.5) is dominated by the first and the second terms. Note that $\frac{N^{1-\alpha_r}}{\sqrt{T}} = (\frac{N^{\frac{3}{2}-\alpha_r}}{T})^{\frac{1}{2}}N^{\frac{1}{4}-\frac{1}{2}\alpha_r} \to 0$ if $\alpha_r > 1/2$ and $\frac{N^{\frac{3}{2}-\alpha_r}}{T} \to 0$ hold. Therefore, we come to that if $\alpha_r > 1/2$, $\frac{N^{\frac{3}{2}-\alpha_r}}{T} \to 0$, and $\sqrt{T}N^{\frac{1}{2}\alpha_1 - \frac{3}{2}\alpha_r} \to 0$, we have

$$
\frac{\hat{c}_{t,i} - c_{t,i}^*}{\sigma_{c(t,i)}} \xrightarrow{d} N(0,1), \text{ with } \sigma_{c(t,i)}^2 = V_{t,i} + U_{t,i},
$$

where $V_{t,i} = \mathbf{b}_i^{0\prime}\mathbf{D}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{\Gamma}_t\mathbf{D}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{b}_i^0$, $U_{t,i} = T^{-1}\mathbf{f}_t^{0\prime}\mathbf{\Phi}_i\mathbf{f}_t^0$. Thus, we complete the proof. □

A.4 Proofs for the results in Section 6

Proof of Theorem 4. By the definition of $\hat{\boldsymbol{\delta}}$, we have the following decomposition:

$$
\begin{aligned}
&\sqrt{T}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0) \\
&= \left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'\boldsymbol{\epsilon} + \left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'(\mathbf{F}^0 - \hat{\mathbf{F}})\boldsymbol{\gamma}^0 \\
&= \left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon} + \left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}(\hat{\mathbf{Z}} - \mathbf{Z}^0)'\boldsymbol{\epsilon} + \left[\left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1} - \left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\right]\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'\boldsymbol{\epsilon} \\
&\quad + \left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'(\mathbf{F}^0 - \hat{\mathbf{F}})\boldsymbol{\gamma}^0.
\end{aligned}
$$

The third term is bounded by $O_p\left(\Delta_{NT}\right)$ because

$$
\begin{aligned}
&\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}} \\
&= \frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0 + \frac{1}{T}(\hat{\mathbf{Z}} - \mathbf{Z}^0)'\hat{\mathbf{Z}} + \frac{1}{T}\mathbf{Z}^{0\prime}(\hat{\mathbf{Z}} - \mathbf{Z}^0) \\
&= \frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0 + \frac{1}{T}(\hat{\mathbf{F}} - \mathbf{F}^0)'\hat{\mathbf{F}} + \frac{1}{T}\mathbf{F}^{0\prime}(\hat{\mathbf{F}} - \mathbf{F}^0) \\
&= \frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0 + O_p\left(\Delta_{NT}\right),
\end{aligned}
$$

by Lemma B.4(i). The second and fourth terms are bounded by

$$\begin{aligned}
&\left\|\left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}(\mathbf{Z}^0-\hat{\mathbf{Z}})'\boldsymbol{\epsilon}\right\|_{\mathrm{F}} \\
&= O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(\frac{\sqrt{T}}{N^{\alpha_r}}\right)+o_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\right), \\
&\left\|\left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'(\mathbf{F}^0-\hat{\mathbf{F}})\boldsymbol{\gamma}^0\right\|_{\mathrm{F}} \\
&= O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\right),
\end{aligned}$$

where we have used Lemmas B.8(ii) and (i). Collecting these non-dominating terms,

$$\begin{aligned}
\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0) &= \left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon} \\
&+O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\right).
\end{aligned}$$

If $\alpha_1<2\alpha_r$, $\frac{N^{1-\alpha_r}}{\sqrt{T}}\to 0$, $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\to 0$, and under Assumption 6(iii)

$$\left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon}\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Sigma}_{\delta^0}),$$

thus, we complete the proof. □

Proof of Theorem 5. We start with the decomposition using the rotation matrix $\mathbf{H}$

$$\begin{aligned}
&\hat{y}_{T+h|T}-y_{T+h|T} \\
&= \hat{\boldsymbol{\gamma}}'\hat{\mathbf{f}}_T+\hat{\boldsymbol{\beta}}'\mathbf{w}_T-\boldsymbol{\gamma}^{*\prime}\mathbf{f}_T-\boldsymbol{\beta}'\mathbf{w}_T \\
&= \left(\hat{\boldsymbol{\gamma}}-\mathbf{H}^{-1\prime}\boldsymbol{\gamma}^*\right)'\hat{\mathbf{f}}_T+\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\left(\hat{\mathbf{f}}_T-\mathbf{H}\mathbf{f}_T\right)+(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{w}_T \\
&= \hat{\mathbf{z}}_T'(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)+\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\left(\hat{\mathbf{f}}_T-\mathbf{H}\mathbf{f}_T\right) \\
&= T^{-1/2}\hat{\mathbf{z}}_T'[\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)]+\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-1/2}\left[\mathbf{N}^{1/2}\left(\hat{\mathbf{f}}_T-\mathbf{H}\mathbf{f}_T\right)\right] \\
&= T^{-1/2}\mathbf{z}_T^{0\prime}[\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)]+T^{-1/2}(\hat{\mathbf{z}}_T-\mathbf{z}_T^0)'[\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)] \\
&\quad+\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-1/2}\left[\mathbf{N}^{1/2}\left(\hat{\mathbf{f}}_T-\tilde{\mathbf{H}}_4\mathbf{f}_T^0\right)\right]+\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\left(\tilde{\mathbf{H}}_4-\mathbf{I}\right)\mathbf{f}_T^0. \qquad \text{(A.6)}
\end{aligned}$$

Consider the first term on the right-hand side of the above equation,

$$
\begin{aligned}
&T^{-1/2}\mathbf{z}_T^{0\prime}\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)\\
&=T^{-1/2}\mathbf{z}_T^{0\prime}\left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'\boldsymbol{\epsilon}+T^{-1/2}\mathbf{z}_T^{0\prime}\left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'(\mathbf{F}^0-\hat{\mathbf{F}})\mathbf{H}^{-1}\boldsymbol{\gamma}^*\\
&=T^{-1/2}\mathbf{z}_T^{0\prime}\left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon}\\
&\quad+\frac{1}{\sqrt{T}}\left[O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\right)\right],
\end{aligned}
$$

where the term in the bracket is dominated by the first one if $\frac{N^{1-\alpha_r}}{\sqrt{T}}\to 0$, $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}$, and $\frac{1}{2}\alpha_1<\alpha_r$. The second term on the right-hand side of (A.6) is dominated by the first one, thus, we ignore it.

Next, consider the third term in (A.6) and Lemma B.7 implies

$$
\begin{aligned}
&\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-\frac{1}{2}}\left[\mathbf{N}^{1/2}\left(\hat{\mathbf{f}}_T-\tilde{\mathbf{H}}_4\mathbf{f}_T^0\right)\right]\\
&=\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{D}^{-1}\left[\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\tilde{\mathbf{Q}}\mathbf{B}^{0\prime}\mathbf{e}_t+\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\left(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{Q}}'\right)'\mathbf{e}_t\right]\\
&=\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{D}^{-1}\left[\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\tilde{\mathbf{Q}}\mathbf{B}^{0\prime}\mathbf{e}_t\right]+\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-\frac{1}{2}}\left[\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\left(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{Q}}'\right)'\mathbf{e}_t\right]\\
&=\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{D}^{-1}\left[\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\mathbf{B}^{0\prime}\mathbf{e}_t\right]+O_p(\|\mathbf{N}^{-\frac{1}{2}}\|_F)O_p(\Delta_{NT})\\
&\quad+O_p(\|\mathbf{N}^{-\frac{1}{2}}\|_F)\left[O_p(N^{\frac{1}{2}-\alpha_r})+O_p\left(\frac{N^{\frac{3}{2}-\alpha_r}}{T}\right)+O_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\right)\right],
\end{aligned}
$$

where the last two terms are dominated by the first one if $\frac{N^{\frac{3}{2}-\alpha_r}}{T}\to 0$ and $\frac{1}{2}<\alpha_r$.

The fourth term in (A.6) becomes

$$
\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\left(\tilde{\mathbf{H}}_4-\mathbf{I}\right)\mathbf{f}_T^0=\frac{1}{\sqrt{T}}O_p(\sqrt{T}\Delta_{NT}),
$$

which is dominated by the first term in (A.6) if $\frac{N^{1-\alpha_r}}{\sqrt{T}}\to 0$, $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}$, and $\frac{1}{2}\alpha_1<\alpha_r$. Collecting these terms, if $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}$, $\frac{1}{2}<\alpha_r$, and $\frac{N^{\frac{3}{2}-\alpha_r}}{T}\to 0$, we obtain

$$
\frac{\left(\hat{y}_{T+h|T}-y_{T+h|T}\right)}{\sigma_{T+h|T}}\xrightarrow{d}N(0,1),
$$

where $\sigma^2_{T+h|T}=T^{-1}\mathbf{z}_T^{0\prime}\boldsymbol{\Sigma}_{\delta^0}\mathbf{z}_T^0+\boldsymbol{\gamma}^{0\prime}\mathbf{D}^{-1}\mathbf{N}^{-1/2}\boldsymbol{\Gamma}_T\mathbf{N}^{-1/2}\mathbf{D}^{-1}\boldsymbol{\gamma}^0$.

□

B Related Lemmas and their Proofs

Lemma B.1. *The followings hold under Assumptions 1–4:*

(i) $\frac{1}{NT}\mathbf{e}_t'\mathbf{E}'\mathbf{F}^0 = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right)$,

(ii) $\frac{1}{T}\mathbf{e}_i'\mathbf{E}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}} = O_p\left(\frac{1}{\sqrt{N^{\alpha_r}}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$,

(iii) $\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}' = O_p(\sqrt{T})$,

(iv) $\frac{\mathbf{N}^{-\frac{1}{2}}\hat{\mathbf{F}}'\mathbf{F}^0\mathbf{N}^{\frac{1}{2}}}{T} = O_p(1)$.

Proof of Lemma B.1. (i)

$$
\begin{aligned}
&\frac{1}{NT}\mathbf{e}_t'\mathbf{E}'\mathbf{F}^0 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T} e_{t,i}e_{s,i}\mathbf{f}_s^{0\prime} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T} \mathbf{f}_s^{0\prime}[e_{t,i}e_{s,i} - \mathbb{E}(e_{t,i}e_{s,i})] + \frac{1}{T}\left(\frac{1}{N}\sum_{s=1}^{T}\mathbf{f}_s^{0\prime}\sum_{i=1}^{N}\mathbb{E}(e_{t,i}e_{s,i})\right) \\
&= O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right).
\end{aligned}
$$

Because for all i, $|\mathbb{E}(e_{s,i}e_{t,i})| \le |\gamma_{s,t}|$ for some $\gamma_{s,t}$ such that $\sum_{t=1}^{T}|\gamma_{s,t}| \le M$ by Assumption 2(iii), and $\|\mathbf{f}_s^0\|_2$ is bounded by $\mathbb{E}\|\mathbf{f}_t^0\|^4 \le M$ in Assumption 4(i).
(ii)

$$
\begin{aligned}
&\frac{1}{T}\mathbf{e}_i'\mathbf{E}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}} \\
&= \frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T} e_{t,i}e_{t,j}\mathbf{b}_j^{0\prime}\mathbf{N}^{-\frac{1}{2}} \\
&= \frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T} \mathbf{b}_j^{0\prime}\mathbf{N}^{-\frac{1}{2}}[e_{t,i}e_{t,j} - \mathbb{E}(e_{t,i}e_{t,j})] + \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N}\mathbf{b}_j^{0\prime}\mathbf{N}^{-\frac{1}{2}}\mathbb{E}(e_{t,i}e_{t,j}) \\
&= O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N^{\alpha_r}}}\right).
\end{aligned}
$$

Under weak cross-sectional dependence as in Assumption 2(iv), $(1/TN)\sum_{t=1}^{T}\sum_{j=1}^{N}|\mathbb{E}(e_{t,i}e_{t,j})| \le (1/N)\sum_{j=1}^{N}|\tau_{i,j}| = O\left(N^{-1}\right)$. By $\mathbb{E}\|\mathbf{b}_i^0\|^4 \le M$ in Assumption 4(i), $\|\mathbf{b}_i^0\|_2$ is bounded and $||\mathbf{b}_j^{0\prime}\mathbf{N}^{-\frac{1}{2}}||_F \le ||\mathbf{b}_j^{0\prime}||_F||\mathbf{N}^{-\frac{1}{2}}||_F \le \frac{1}{\sqrt{N^{\alpha_r}}}O_p(1)$. The upper bounds in (i) and (ii) are not consistent with Bai and Ng (2023, Assumption A3'), because we impose a moment restriction related to $\mathbf{b}_i^0$ in Assumption 4(iv).

(iii) Assumption 4(ii) implies

$$\frac{\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}'\mathbf{E}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}}}{T} = \frac{1}{T}\sum_{t=1}^{T}\left[\left(\mathbf{N}^{-\frac{1}{2}}\sum_{i=1}^{N}\mathbf{b}_i^0 e_{t,i}\right)\left(\mathbf{N}^{-\frac{1}{2}}\sum_{i=1}^{N}\mathbf{b}_i^0 e_{t,i}\right)'\right] = O_p(1),$$

where $\mathbf{B}^0 = (\mathbf{b}_1^0, \cdots, \mathbf{b}_N^0)'$. Thus,

$$\left\|\mathbf{E}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}}\right\|_{\mathrm{F}}^2 = tr(\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}'\mathbf{E}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}}) = O_p(T),$$

$\left\|\mathbf{E}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}}\right\|_{\mathrm{F}} = O_p(\sqrt{T})$, and $\left\|\mathbf{E}\mathbf{B}^0\right\|_{\mathrm{F}} = O_p(\sqrt{TN^{\alpha_1}})$.

(iv) A comparable result to Bai and Ng (2023, Lemma 1) is

$$\mathbf{N}^{-1}\hat{\mathbf{\Lambda}} \xrightarrow{p} \mathbf{N}^{-1}\mathbf{\Lambda},$$

which can be obtained by (B.3). The equations $\frac{1}{T}\mathbf{X}\mathbf{X}'\hat{\mathbf{F}} = \hat{\mathbf{F}}\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{F}}'\hat{\mathbf{F}}/T = \mathbf{I}_r$ implies

$$\frac{1}{T^2}\mathbf{N}^{-\frac{1}{2}}\hat{\mathbf{F}}'\mathbf{X}\mathbf{X}'\hat{\mathbf{F}}\mathbf{N}^{-\frac{1}{2}} = \mathbf{N}^{-1}\hat{\mathbf{\Lambda}}.$$

The dominating term of the left-hand side is

$$\frac{\mathbf{N}^{-\frac{1}{2}}\hat{\mathbf{F}}'\mathbf{F}^0\mathbf{N}^{\frac{1}{2}}}{T}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}}\frac{\mathbf{N}^{\frac{1}{2}}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\mathbf{N}^{-\frac{1}{2}}}{T} \xrightarrow{p} \mathbf{N}^{-1}\mathbf{\Lambda}.$$

Thus, $\frac{\mathbf{N}^{-\frac{1}{2}}\hat{\mathbf{F}}'\mathbf{F}^0\mathbf{N}^{\frac{1}{2}}}{T} = O_p(1)$. $\square$

Lemma B.2. *Suppose that Assumption 2 holds. If* $\frac{N^{1-\alpha_r}}{T} \to 0$, *we have* $\hat{\lambda}_k = \lambda_k\left[T^{-1}\mathbf{X}\mathbf{X}'\right] \asymp N^{\alpha_k}$ *with high probability.*

Proof of Lemma B.2. Let $\sigma_k[\mathbf{A}]$ be k-th largest singular value of matrix $\mathbf{A}$. By the definition of WF models, we have

$$\lambda_k = \lambda_k\left[\frac{\mathbf{B}^*\mathbf{F}^{*\prime}\mathbf{F}^*\mathbf{B}^{*\prime}}{T}\right] = \lambda_k\left[\mathbf{B}^{*\prime}\mathbf{B}^*\frac{\mathbf{F}^{*\prime}\mathbf{F}^*}{T}\right] \asymp N^{\alpha_k}.$$

By the singular value version of Weyl's inequalities,

$$\sigma_{k+l-1}[A+B] \le \sigma_k[A] + \sigma_l[B], \quad 1 \le k, l \le \min(N, T).$$

We first show the upper bound for $\hat{\lambda}_k$:

$$\begin{aligned}
\sigma_k[\mathbf{X}] &\le \sigma_k[\mathbf{F}^*\mathbf{B}^{*\prime}] + \sigma_1[\mathbf{E}] \\
\sqrt{\lambda_k\left[\frac{\mathbf{X}\mathbf{X}'}{T}\right]} &\le \sqrt{\lambda_k\left[\frac{\mathbf{F}^*\mathbf{B}^{*\prime}\mathbf{B}^*\mathbf{F}^{*\prime}}{T}\right]} + \sqrt{\lambda_1\left[\frac{\mathbf{E}\mathbf{E}'}{T}\right]} \\
\lambda_k\left[\frac{\mathbf{X}\mathbf{X}'}{T}\right] &\le \lambda_k\left[\frac{\mathbf{F}^*\mathbf{B}^{*\prime}\mathbf{B}^*\mathbf{F}^{*\prime}}{T}\right] + \lambda_1\left[\frac{\mathbf{E}\mathbf{E}'}{T}\right] + 2\sqrt{\lambda_k\left[\frac{\mathbf{F}^*\mathbf{B}^{*\prime}\mathbf{B}^*\mathbf{F}^{*\prime}}{T}\right]\lambda_1\left[\frac{\mathbf{E}\mathbf{E}'}{T}\right]} \\
\lambda_k\left[\frac{\mathbf{X}\mathbf{X}'}{T}\right] &\le N^{\alpha_k} + \left(\frac{N}{T}+1\right) + \sqrt{N^{\alpha_k}\left(\frac{N}{T}+1\right)}.
\end{aligned}$$

Next, the lower bound for $\hat{\lambda}_k$ becomes

$$\begin{aligned}
\sigma_k[\mathbf{X}-\mathbf{E}] &\leq \sigma_k[\mathbf{X}]+\sigma_1[\mathbf{E}] \\
\sigma_k[\mathbf{F}^*\mathbf{B}^{*\prime}] &\leq \sigma_k[\mathbf{X}]+\sigma_1[\mathbf{E}] \\
\sqrt{\lambda_k\left[\frac{\mathbf{F}^*\mathbf{B}^{*\prime}\mathbf{B}^*\mathbf{F}^{*\prime}}{T}\right]} &\leq \sqrt{\lambda_k\left[\frac{\mathbf{X}\mathbf{X}'}{T}\right]}+\sqrt{\lambda_1\left[\frac{\mathbf{E}\mathbf{E}'}{T}\right]} \\
\sqrt{\lambda_k\left[\frac{\mathbf{X}\mathbf{X}'}{T}\right]} &\geq \sqrt{\lambda_k\left[\frac{\mathbf{F}^*\mathbf{B}^{*\prime}\mathbf{B}^*\mathbf{F}^{*\prime}}{T}\right]}-\sqrt{\lambda_1\left[\frac{\mathbf{E}\mathbf{E}'}{T}\right]}.
\end{aligned}$$

Thus, $\hat{\lambda}_k=\lambda_k\left[\frac{\mathbf{X}\mathbf{X}'}{T}\right]\asymp N^{\alpha_k}$. Note that N^{α_k} dominates $\frac{N}{T}$ if $\frac{N^{1-\alpha_r}}{T}\to 0$.

□

Lemma B.3. *Define*

$$\Delta_{NT}=\frac{N^{1-\alpha_r}}{T}+N^{\frac{1}{2}\alpha_1-\alpha_r}\frac{N^{1-\alpha_r}}{T}+N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}+\frac{N^{\frac{1}{2}\alpha_1-\alpha_r}}{\sqrt{T}}. \tag{B.1}$$

Suppose that Assumptions 1–4 hold. Then, we have

(i) $\left\|\frac{1}{T}\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}}=O_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right)+O_p\left(\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right)+O_p\left(N^{-\frac{1}{2}\alpha_r}\right),$

(ii) $\left\|\frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}}=O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right),$

$\left\|\frac{1}{T}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}}=O_p\left(N^{-\frac{1}{2}\alpha_r}\right)+O_p\left(\frac{N^{1-\alpha_r}}{T}\right),$

(iii) $\left\|\frac{1}{T}\mathbf{F}^{0\prime}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}})\right\|_{\mathrm{F}}=O_p\left(\Delta_{NT}\right),$

(iv) $\left\|\frac{1}{T}\mathbf{e}_i{}'(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}}=O_p\left(\frac{N^{1-\alpha_r}}{T}\right)+O_p\left(N^{-\alpha_r}\right)+O_p\left(\frac{1}{\sqrt{TN^{\alpha_r}}}\right).$

Proof of Lemma B.3. (i) By the definition of $\hat{\mathbf{F}}$, we have the following decomposition:

$$\begin{aligned}
&\left\|\frac{1}{T}\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}} \\
&=\left\|\frac{1}{T^2}\mathbf{E}'\mathbf{E}\mathbf{E}'\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}+\frac{1}{T^2}\mathbf{E}'\mathbf{E}\mathbf{B}^0\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}} \\
&\leq\left\|\frac{1}{T^2}\mathbf{E}'\mathbf{E}\mathbf{E}'\right\|_F\left\|\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_F+\left\|\frac{1}{T}\mathbf{E}'\mathbf{E}\right\|_F\left\|\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}}\right\|_F\left\|\mathbf{N}^{\frac{1}{2}}\frac{1}{T}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\mathbf{N}^{-\frac{1}{2}}\mathbf{N}^{\frac{1}{2}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}} \\
&=\frac{1}{T^2}O_p\left(N^{3/2}+T^{3/2}\right)O_p\left(\sqrt{T}N^{-\alpha_r}\right)+O_p\left(\frac{N}{T}+1\right)O_p\left(N^{-\frac{1}{2}\alpha_r}\right) \\
&=O_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right)+O_p\left(\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right)+O_p\left(N^{-\frac{1}{2}\alpha_r}\right).
\end{aligned}$$

Consider (ii)

$$
\begin{aligned}
&\left\|\frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}(\hat{\mathbf{F}}-\mathbf{F}^{0}\tilde{\mathbf{H}}_{4})\right\|_{\mathrm{F}}\\
&=\left\|\frac{1}{T^{2}}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\mathbf{E}\mathbf{E}^{\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}+\frac{1}{T^{2}}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\mathbf{E}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}\\
&\leq\left\|\frac{1}{T^{2}}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\right\|_{F}\left\|\mathbf{E}\mathbf{E}^{\prime}\right\|_{F}\left\|\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{F}+\left\|\mathbf{N}^{\frac{1}{2}}\frac{1}{T}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\mathbf{E}\mathbf{B}^{0}\mathbf{N}^{-\frac{1}{2}}\right\|_{F}\left\|\mathbf{N}^{\frac{1}{2}}\frac{1}{T}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\mathbf{N}^{-\frac{1}{2}}\mathbf{N}^{\frac{1}{2}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}\\
&=O_{p}\left(N^{\frac{1}{2}\alpha_{1}-\frac{1}{2}\alpha_{r}}\left(\frac{N^{1-\frac{1}{2}\alpha_{r}}}{T}+N^{-\frac{1}{2}\alpha_{r}}\right)\right)+N^{\frac{1}{2}\alpha_{1}}O_{p}\left(N^{-\frac{1}{2}\alpha_{r}}\right)\\
&=O_{p}\left(N^{\frac{1}{2}\alpha_{1}-\frac{1}{2}\alpha_{r}}\frac{N^{1-\frac{1}{2}\alpha_{r}}}{T}\right)+O_{p}\left(N^{\frac{1}{2}\alpha_{1}-\frac{1}{2}\alpha_{r}}\right).
\end{aligned}
$$

$$
\begin{aligned}
&\left\|\frac{1}{T}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}(\hat{\mathbf{F}}-\mathbf{F}^{0}\tilde{\mathbf{H}}_{4})\right\|_{\mathrm{F}}\\
&=\left\|\frac{1}{T^{2}}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\mathbf{E}\mathbf{E}^{\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}+\frac{1}{T^{2}}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\mathbf{E}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}\\
&\leq\left\|\frac{1}{T^{2}}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\right\|_{F}\left\|\mathbf{E}\mathbf{E}^{\prime}\right\|_{F}\left\|\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{F}+\left\|\frac{1}{T}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\mathbf{E}\mathbf{B}^{0}\mathbf{N}^{-\frac{1}{2}}\right\|_{F}\left\|\mathbf{N}^{\frac{1}{2}}\frac{1}{T}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\mathbf{N}^{-\frac{1}{2}}\mathbf{N}^{\frac{1}{2}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}\\
&=O_{p}\left(\frac{N^{1-\alpha_{r}}}{T}\right)+O_{p}\left(N^{-\alpha_{r}}\right)+O_{p}\left(N^{-\frac{1}{2}\alpha_{r}}\right)\\
&=O_{p}\left(\frac{N^{1-\alpha_{r}}}{T}\right)+O_{p}\left(N^{-\frac{1}{2}\alpha_{r}}\right).
\end{aligned}
$$

Consider (iii)

$$
\begin{aligned}
&\frac{1}{T}\mathbf{F}^{0\prime}(\hat{\mathbf{F}}-\mathbf{F}^{0}\tilde{\mathbf{H}})\\
&=\frac{1}{T}\mathbf{F}^{0\prime}\left(\frac{1}{T}\mathbf{E}\mathbf{E}^{\prime}\hat{\mathbf{F}}+\frac{1}{T}\mathbf{F}^{0}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\hat{\mathbf{F}}+\frac{1}{T}\mathbf{E}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\right)\hat{\mathbf{\Lambda}}^{-1}\\
&=\frac{1}{T^{2}}\mathbf{F}^{0\prime}\mathbf{E}\mathbf{E}^{\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}+\frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}^{\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}+\frac{1}{T^{2}}\mathbf{F}^{0\prime}\mathbf{E}\mathbf{B}^{0}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}.
\end{aligned}
$$

The first term on the right-hand side of the above equation is bounded by

$$
\left\|\frac{1}{T^{2}}\mathbf{F}^{0\prime}\mathbf{E}\mathbf{E}^{\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}\leq O_{p}\left(\left(\frac{N}{T}+1\right)N^{-\alpha_{r}}\right).
$$

Next, we consider the second term:

$$
\begin{aligned}
&\left\|\frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'\hat{\mathbf{F}}\hat{\boldsymbol{\Lambda}}^{-1}\right\|_{\mathrm{F}} \\
&\le \left\|\frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\hat{\boldsymbol{\Lambda}}^{-1}\right\|_{\mathrm{F}} + \left\|\frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'\mathbf{F}^0\tilde{\mathbf{H}}_4\hat{\boldsymbol{\Lambda}}^{-1}\right\|_{\mathrm{F}} \\
&\le \left\|\frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}}\left\|\hat{\boldsymbol{\Lambda}}^{-1}\right\|_{\mathrm{F}} + \left\|\frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'\mathbf{F}^0\right\|_{\mathrm{F}}\left\|\tilde{\mathbf{H}}_4\hat{\boldsymbol{\Lambda}}^{-1}\right\|_{\mathrm{F}} \\
&= O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\frac{N^{1-\alpha_r}}{T}\right) + O_p\left(N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\right) + O_p\left(\frac{\sqrt{TN^{\alpha_1}}}{T}N^{-\alpha_r}\right),
\end{aligned}
$$

where we have used Lemma B.3 (ii) and Assumption 4(vi). The third term is bounded by

$$
\left\|\frac{1}{T^2}\mathbf{F}^{0\prime}\mathbf{E}\mathbf{B}^0\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\boldsymbol{\Lambda}}^{-1}\right\|_{\mathrm{F}} \le \left\|\frac{1}{T^2}\mathbf{F}^{0\prime}\mathbf{E}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}}\right\|_{\mathrm{F}}\left\|\mathbf{N}^{\frac{1}{2}}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\mathbf{N}^{-\frac{1}{2}}\mathbf{N}^{\frac{1}{2}}\hat{\boldsymbol{\Lambda}}^{-1}\right\|_{\mathrm{F}} = O_p\left(\frac{1}{\sqrt{TN^{\alpha_r}}}\right).
$$

Collecting terms, we obtain

$$
\begin{aligned}
&\left\|\frac{1}{T}\mathbf{F}^{0\prime}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}})\right\|_{\mathrm{F}} \\
&\le O_p\left(\left(\frac{N}{T}+1\right)N^{-\alpha_r}\right) + O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\frac{N^{1-\alpha_r}}{T}\right) + O_p\left(N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\right) + O_p\left(\frac{N^{\frac{1}{2}\alpha_1-\alpha_r}}{\sqrt{T}}\right) \\
&= O_p\left(\frac{N^{1-\alpha_r}}{T}\right) + O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\frac{N^{1-\alpha_r}}{T}\right) + O_p\left(N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\right) + O_p\left(\frac{N^{\frac{1}{2}\alpha_1-\alpha_r}}{\sqrt{T}}\right),
\end{aligned}
$$

which is $o_p(1)$ if $\frac{N^{1-\alpha_r}}{T} \to 0$ and $\frac{1}{2}\alpha_1 < \alpha_r$.

(iv)

$$
\begin{aligned}
&\left\|\frac{1}{T}\mathbf{e}_i'(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}} \\
&= \left\|\frac{1}{T}\left(\frac{1}{T}\mathbf{e}_i'\mathbf{E}\mathbf{E}'\hat{\mathbf{F}} + \frac{1}{T}\mathbf{e}_i'\mathbf{E}\mathbf{B}^0\mathbf{F}^{0\prime}\hat{\mathbf{F}}\right)\hat{\boldsymbol{\Lambda}}^{-1}\right\|_{\mathrm{F}} \\
&\le O_p\left(\left(\frac{N}{T}+1\right)N^{-\alpha_r}\right) + \left\|\frac{1}{T}\mathbf{e}_i'\mathbf{E}\mathbf{B}^0\mathbf{N}^{-\frac{1}{2}}\right\|_{\mathrm{F}}\left\|\frac{1}{T}\mathbf{N}^{\frac{1}{2}}\mathbf{F}^{0\prime}\hat{\mathbf{F}}\mathbf{N}^{-\frac{1}{2}}\mathbf{N}^{\frac{1}{2}}\hat{\boldsymbol{\Lambda}}^{-1}\right\|_{\mathrm{F}} \\
&= O_p\left(\left(\frac{N}{T}+1\right)N^{-\alpha_r}\right) + \left[O_p\left(\frac{1}{\sqrt{N^{\alpha_r}}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)\right]O_p\left(N^{-\frac{1}{2}\alpha_r}\right) \\
&= O_p\left(\frac{N^{1-\alpha_r}}{T}\right) + O_p\left(N^{-\alpha_r}\right) + O_p\left(\frac{1}{\sqrt{TN^{\alpha_r}}}\right),
\end{aligned}
$$

since $\frac{1}{T}\mathbf{e}_i'\mathbf{E}\mathbf{B}^*\mathbf{N}^{-\frac{1}{2}} = O_p\left(\frac{1}{\sqrt{N^{\alpha_r}}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$ by Lemma B.1(ii). □

Lemma B.4. *Suppose that Assumptions 1–4 hold. Then, we have*

(i) $\left\|\frac{\hat{\mathbf{F}}'\mathbf{F}^0}{T} - \mathbf{I}_r\right\|_{\mathrm{F}} = O_p(\Delta_{NT})$,

(ii) $\left\|\mathbf{N}^{\frac{1}{2}}\left(\frac{\hat{\mathbf{F}}'\mathbf{F}^0}{T} - \mathbf{I}_r\right)\right\|_{\mathrm{F}} = O_p(N^{\frac{1}{2}\alpha_1}\Delta_{NT})$.

The result in the case of strong factors has been given by Bai and Ng (2013). For weak factors, as $N \asymp T$, the convergence rate in (i) is reduced to $N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}+N^{\frac{1}{2}\alpha_1-\alpha_r-\frac{1}{2}}$, which is faster than the convergence rate $N^{\frac{1}{4}\alpha_1-\frac{1}{2}\alpha_r}+N^{\frac{1}{2}-\alpha_r}$ in Freyaldenhoven (2022) Lemma 3.

Proof of Lemma B.4. (i) Denote $\hat{\mathbf{F}}=(\hat{\mathbf{f}}_1,\cdots,\hat{\mathbf{f}}_T)'=(\hat{\mathbf{F}}_1,\cdots,\hat{\mathbf{F}}_r)$ and $\mathbf{F}^0=(\mathbf{f}_1^0,\cdots,\mathbf{f}_T^0)'=(\mathbf{F}_1^0,\cdots,\mathbf{F}_r^0)$. From the equation

$$\begin{aligned}\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0&=\frac{1}{T}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}})'\mathbf{F}^0+\tilde{\mathbf{H}}'\\&=\frac{1}{T}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}})'\mathbf{F}^0+\hat{\boldsymbol{\Lambda}}^{-1}\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\boldsymbol{\Lambda},\end{aligned}$$

and Lemma B.3(iii),

$$\begin{aligned}\left\|\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0-\hat{\boldsymbol{\Lambda}}^{-1}\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\boldsymbol{\Lambda}\right\|_{\max}&\le\left\|\frac{1}{T}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}})'\mathbf{F}^0\right\|_{\mathrm{F}}=O_p\left(\Delta_{NT}\right)\\\max_{k,l}\left|\frac{\hat{\mathbf{F}}_k'\mathbf{F}_l^0}{T}-\frac{\lambda_l}{\hat{\lambda}_k}\frac{\hat{\mathbf{F}}_k'\mathbf{F}_l^0}{T}\right|&\le O_p\left(\Delta_{NT}\right)\\\max_{k,l}\left|\frac{\hat{\mathbf{F}}_k'\mathbf{F}_l^0}{T}\left(1-\frac{\lambda_l}{\hat{\lambda}_k}\right)\right|&\le O_p\left(\Delta_{NT}\right).\end{aligned}\tag{B.2}$$

Lemmas 7 and 8 of Freyaldenhoven (2022) showed that

$$\frac{\hat{\mathbf{F}}_k'\mathbf{F}_l^0}{T}\xrightarrow{p}0\ (k\neq l),\ \frac{\hat{\mathbf{F}}_k'\mathbf{F}_k^0}{T}\xrightarrow{p}1,\ k=1,\cdots,r,\ l=1,\cdots,r.$$

It implies that, for the diagonal elements,

$$\begin{aligned}\max_k\left|1-\frac{\lambda_k}{\hat{\lambda}_k}\right|&\le O_p\left(\Delta_{NT}\right)\\\left|\hat{\lambda}_k-\lambda_k\right|&\le O_p\left(\Delta_{NT}\right).\end{aligned}\tag{B.3}$$

If $k>l$, then $\alpha_k<\alpha_l$ and $\frac{\lambda_l}{\hat{\lambda}_k}$ will diverge. The order (B.2) implies that

$$\left|\frac{\hat{\mathbf{F}}_k'\mathbf{F}_l^0}{T}\right|\le O_p\left(\Delta_{NT}\right)\left|\frac{\hat{\lambda}_k}{\hat{\lambda}_k-\lambda_l}\right|=o_p(\Delta_{NT})\tag{B.4}$$

and

$$N^{\frac{1}{2}\alpha_k}\frac{\hat{\mathbf{F}}_k'\mathbf{F}_l^0}{T}=N^{\frac{1}{2}\alpha_k}o_p(\Delta_{NT}).$$

If $k<l$, then $\alpha_k>\alpha_l$ and $\frac{\lambda_l}{\hat{\lambda}_k}\to 0$. The order (B.2) implies that

$$\left|\frac{\hat{\mathbf{F}}_k'\mathbf{F}_l^0}{T}\right|\le O_p\left(\Delta_{NT}\right)\tag{B.5}$$

and

$$N^{\frac{1}{2}\alpha_k}\frac{\hat{\mathbf{F}}_k'\mathbf{F}_l^0}{T}=N^{\frac{1}{2}\alpha_k}O_p(\Delta_{NT}).$$

Now, we consider the case $k=l$ and follow the Lemmas 7 and 8 of Freyaldenhoven (2022) to find the upper bound of the term $\left|N^{\frac{1}{2}\alpha_k}\frac{\hat{\mathbf{F}}_k'\mathbf{F}_k^0}{T}\right|$. We first consider the upper bound of the matrix $\frac{1}{T}\hat{\mathbf{F}}'\left(\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}-\frac{\mathbf{X}\mathbf{X}'}{T}\right)\hat{\mathbf{F}}$.

$$\begin{aligned}
&\left\|\frac{1}{T}\hat{\mathbf{F}}'\left(\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}-\frac{\mathbf{X}\mathbf{X}'}{T}\right)\hat{\mathbf{F}}\right\|_{\mathrm{F}}\\
&\leq\left\|\frac{1}{T}\hat{\mathbf{F}}'\left(\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{E}'}{T}+\frac{\mathbf{E}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}+\frac{\mathbf{E}\mathbf{E}'}{T}\right)\hat{\mathbf{F}}\right\|_{\mathrm{F}}\\
&\leq\left\|\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{E}'}{T}\hat{\mathbf{F}}\right\|_{\mathrm{F}}+\left\|\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{E}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}\hat{\mathbf{F}}\right\|_{\mathrm{F}}+\left\|\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{E}\mathbf{E}'}{T}\hat{\mathbf{F}}\right\|_{\mathrm{F}}.
\end{aligned}\tag{B.6}$$

Using $\mathbf{B}^{0\prime}\mathbf{E}'\mathbf{F}^0=O_p(\sqrt{TN^{\alpha_1}})$ and Lemma B.3 (ii), the first term on the right-hand side of (B.6) becomes

$$\begin{aligned}
&\left\|\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{E}'}{T}\hat{\mathbf{F}}\right\|_{\mathrm{F}}\\
&\leq\left\|\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\frac{\mathbf{B}^{0\prime}\mathbf{E}'}{T}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}}+\left\|\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\frac{\mathbf{B}^{0\prime}\mathbf{E}'}{T}\mathbf{F}^0\tilde{\mathbf{H}}_4\right\|_{\mathrm{F}}\\
&\leq\left\|\frac{\hat{\mathbf{F}}'\mathbf{F}^0}{T}\frac{\mathbf{B}^{0\prime}\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)}{T}\right\|_{\mathrm{F}}+\left\|\mathbf{N}^{\frac{1}{2}}\mathbf{N}^{-\frac{1}{2}}\frac{\hat{\mathbf{F}}'\mathbf{F}^0}{T}\mathbf{N}^{\frac{1}{2}}\mathbf{N}^{-\frac{1}{2}}\frac{\mathbf{B}^{0\prime}\mathbf{E}'\mathbf{F}^0}{T}\tilde{\mathbf{H}}_4\right\|_{\mathrm{F}}\\
&=O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\right)+N^{\frac{1}{2}\alpha_1}O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

For the second and third terms on the right-hand side of (B.6),

$$\begin{aligned}
&\left\|\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{E}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}\hat{\mathbf{F}}\right\|_{\mathrm{F}}+\left\|\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{E}\mathbf{E}'}{T}\hat{\mathbf{F}}\right\|_{\mathrm{F}}\\
&\leq\left\|\frac{1}{T}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)'\mathbf{E}\mathbf{B}^0\frac{\mathbf{F}^{0\prime}\hat{\mathbf{F}}}{T}\right\|_{\mathrm{F}}+\left\|\tilde{\mathbf{H}}_4'\frac{\mathbf{F}^{0\prime}\mathbf{E}\mathbf{B}^0}{T}\frac{\mathbf{F}^{0\prime}\hat{\mathbf{F}}}{T}\right\|_{\mathrm{F}}+\left\|\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{E}\mathbf{E}'}{T}\hat{\mathbf{F}}\right\|_{\mathrm{F}}\\
&\leq O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\right)+O_p\left(\frac{N^{\frac{1}{2}\alpha_1}}{\sqrt{T}}\right)+O_p\left(\frac{N}{T}+1\right).
\end{aligned}$$

Thus,

$$
\begin{aligned}
&\left\|\frac{1}{T}\hat{\mathbf{F}}'\left(\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}-\frac{\mathbf{X}\mathbf{X}'}{T}\right)\hat{\mathbf{F}}\right\|_{\max}\\
&\leq\left\|\frac{1}{T}\hat{\mathbf{F}}'\left(\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}-\frac{\mathbf{X}\mathbf{X}'}{T}\right)\hat{\mathbf{F}}\right\|_{\mathrm{F}}\\
&\leq O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\right)+O_p\left(\frac{N^{\frac{1}{2}\alpha_1}}{\sqrt{T}}\right)+O_p\left(\frac{N}{T}\right).
\end{aligned}
\tag{B.7}
$$

Next, using the following decomposition

$$
\begin{aligned}
&\frac{1}{T}\hat{\mathbf{F}}'\left(\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}-\frac{\mathbf{X}\mathbf{X}'}{T}\right)\hat{\mathbf{F}}\\
&=\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}\hat{\mathbf{F}}-\frac{1}{T}\mathbf{F}^{0\prime}\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}\mathbf{F}^0+\left(\frac{1}{T}\mathbf{F}^{0\prime}\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}\mathbf{F}^0-\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{X}\mathbf{X}'}{T}\hat{\mathbf{F}}\right)\\
&=\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}\hat{\mathbf{F}}-\mathbf{\Lambda}+\left(\mathbf{\Lambda}-\hat{\mathbf{\Lambda}}\right),
\end{aligned}
$$

we get

$$
\frac{1}{T}\hat{\mathbf{F}}'\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}\hat{\mathbf{F}}-\mathbf{\Lambda}=\frac{1}{T}\hat{\mathbf{F}}'\left(\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}-\frac{\mathbf{X}\mathbf{X}'}{T}\right)\hat{\mathbf{F}}+\left(\hat{\mathbf{\Lambda}}-\mathbf{\Lambda}\right).
\tag{B.8}
$$

Collecting (B.8), (B.7), and (B.3), we obtain

$$
\begin{aligned}
&\frac{1}{T}\hat{\mathbf{F}}_k'\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}\hat{\mathbf{F}}_k-\lambda_k\leq\left\|\frac{1}{T}\hat{\mathbf{F}}'\left(\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}-\frac{\mathbf{X}\mathbf{X}'}{T}\right)\hat{\mathbf{F}}\right\|_{\max}+\left|\hat{\lambda}_k-\lambda_k\right|\\
&\leq O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\right)+O_p\left(\frac{N^{\frac{1}{2}\alpha_1}}{\sqrt{T}}\right)+O_p\left(\frac{N}{T}\right)+O_p\left(\Delta_{NT}N^{\alpha_k}\right).
\end{aligned}
$$

Then, from the equation

$$
\frac{1}{T}\hat{\mathbf{F}}_k'\frac{\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}}{T}\hat{\mathbf{F}}_k-\lambda_k=\left[\left(\frac{\hat{\mathbf{F}}_k'\mathbf{F}_k^0}{T}\right)^2-1\right]\lambda_k+\sum_{l\neq k}^{r}\left[\left(\frac{\hat{\mathbf{F}}_k'\mathbf{F}_l^0}{T}\right)^2\lambda_l\right]
$$

and its upper bound, we have

$$
\begin{aligned}
&\left(\frac{\hat{\mathbf{F}}_k'\mathbf{F}_k^0}{T}\right)^2-1\leq O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}+N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}+\frac{N^{\frac{1}{2}\alpha_1}}{\sqrt{T}}+\frac{N}{T}+\Delta_{NT}N^{\alpha_k}\right)O_p\left(N^{-\alpha_k}\right),\\
&\frac{\hat{\mathbf{F}}_k'\mathbf{F}_k^0}{T}-1\leq O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}+N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}+\frac{N^{\frac{1}{2}\alpha_1}}{\sqrt{T}}+\frac{N}{T}+\Delta_{NT}N^{\alpha_k}\right)O_p\left(N^{-\alpha_k}\right).
\end{aligned}
\tag{B.9}
$$

From orders (B.4), (B.5) and (B.9), we have

$$\begin{aligned}
&\left\|\frac{\hat{\mathbf{F}}'\mathbf{F}^0}{T}-\mathbf{I}_r\right\|_{\mathrm{F}} \le r\left\|\frac{\hat{\mathbf{F}}'\mathbf{F}^0}{T}-\mathbf{I}_r\right\|_{\max} \\
&\le \max_k\left\{\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\frac{N^{1-\frac{1}{2}\alpha_r}}{T}+N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}+\frac{N^{\frac{1}{2}\alpha_1}}{\sqrt{T}}+\frac{N}{T}\right)N^{-\alpha_k}+\Delta_{NT}\right\}O_p(1) \\
&= O_p(\Delta_{NT}) = o_p(1),
\end{aligned}$$

if $\frac{N^{1-\alpha_r}}{T}\to 0$ and $\frac{1}{2}\alpha_1<\alpha_r$.

(ii) From (i)

$$\left\|\mathbf{N}^{\frac{1}{2}}\left(\frac{\hat{\mathbf{F}}'\mathbf{F}^0}{T}-\mathbf{I}_r\right)\right\|_{\mathrm{F}} \le r\left\|\mathbf{N}^{\frac{1}{2}}\left(\frac{\hat{\mathbf{F}}'\mathbf{F}^0}{T}-\mathbf{I}_r\right)\right\|_{\max} \le N^{\frac{1}{2}\alpha_1}O_p(\Delta_{NT}) = o_p(1),$$

if $\frac{N^{\frac{3}{2}-\alpha_r}}{T}\to 0, N^{\alpha_1-\frac{3}{2}\alpha_r}\to 0$, and $\frac{N^{\alpha_1-\alpha_r}}{\sqrt{T}}\to 0$. Because

$$\frac{N}{T}N^{\frac{1}{2}\alpha_1-\alpha_r}=\frac{N^{\frac{3}{2}-\alpha_r}}{T}N^{\frac{1}{2}\alpha_1-\frac{1}{2}}\to 0, \quad \frac{N^{1-\frac{1}{2}\alpha_r}}{T}=\frac{N^{\frac{3}{2}-\alpha_r}}{T}N^{\frac{1}{2}\alpha_r-\frac{1}{2}}\to 0.$$

□

Define

$$\begin{aligned}
&\tilde{\mathbf{H}}_1=(\mathbf{B}^{0\prime}\mathbf{B}^0)(\hat{\mathbf{B}}'\mathbf{B}^0)^{-1}, \quad \tilde{\mathbf{H}}_2=(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\hat{\mathbf{F}}, \\
&\tilde{\mathbf{H}}_3=(\hat{\mathbf{F}}'\mathbf{F}^0)^{-1}\hat{\mathbf{F}}'\hat{\mathbf{F}}, \quad \tilde{\mathbf{H}}_4=(\mathbf{B}^{0\prime}\hat{\mathbf{B}})(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}, \\
&\tilde{\mathbf{H}}=\mathbf{B}^{0\prime}\mathbf{B}^0\frac{\mathbf{F}^{0\prime}\hat{\mathbf{F}}}{T}\hat{\mathbf{\Lambda}}^{-1}, \quad \tilde{\mathbf{Q}}=\frac{\hat{\mathbf{F}}'\mathbf{F}^0}{T}.
\end{aligned}$$

Lemma B.5 gives equivalence among the rotation matrices.

Lemma B.5. *Suppose that Assumptions 1–4 hold. Then, we have,*

(i) $\tilde{\mathbf{H}}_4-\tilde{\mathbf{H}}=\frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}$,

(ii) $\tilde{\mathbf{H}}_2-\tilde{\mathbf{H}}_4=\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\hat{\mathbf{B}}\mathbf{\Lambda}^{-1}$,

(iii) $\tilde{\mathbf{H}}_3-\tilde{\mathbf{H}}_1=\left(\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\right)^{-1}\frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\mathbf{B}^0(\hat{\mathbf{B}}'\mathbf{B}^0)^{-1}$,

(iv) $\tilde{\mathbf{H}}_3-\tilde{\mathbf{H}}_4=\left(\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\right)^{-1}\frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\hat{\mathbf{B}}\hat{\mathbf{\Lambda}}^{-1}$,

(v) $\tilde{\mathbf{H}}_1'^{-1}-\tilde{\mathbf{H}}_2=\frac{1}{T}\mathbf{\Lambda}^{-1}\mathbf{B}^{0\prime}\mathbf{E}'\hat{\mathbf{F}}$,

(vi) $\tilde{\mathbf{H}}_4'^{-1}-\tilde{\mathbf{H}}_2=\frac{1}{T}(\hat{\mathbf{B}}'\mathbf{B}^0)^{-1}\hat{\mathbf{B}}'\mathbf{E}'\hat{\mathbf{F}}$,

Proof of Lemma B.5. (i) We proceed by left multiplying $\mathbf{B}^{0\prime}$ and right multiplying $\hat{\mathbf{\Lambda}}^{-1}$ to the first equation, to get

$$\begin{aligned}
\hat{\mathbf{B}} &= \frac{1}{T}\mathbf{X}'\hat{\mathbf{F}} = \frac{1}{T}\mathbf{B}^0\mathbf{F}^{0\prime}\hat{\mathbf{F}} + \frac{1}{T}\mathbf{E}'\hat{\mathbf{F}} \\
\mathbf{B}^{0\prime}\hat{\mathbf{B}} &= \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}\hat{\mathbf{F}} + \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'\hat{\mathbf{F}} \\
\mathbf{B}^{0\prime}\hat{\mathbf{B}}\hat{\mathbf{\Lambda}}^{-1} &= \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{B}^0\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1} + \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1} \\
\tilde{\mathbf{H}}_4 &= \tilde{\mathbf{H}} + \frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}.
\end{aligned}$$

(ii) By the definition of $\hat{\mathbf{B}}$ and expanding $\mathbf{X}$, we obtain

$$\begin{aligned}
\hat{\mathbf{F}} &= \frac{1}{T}\mathbf{X}\mathbf{X}'\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1} = \mathbf{X}\hat{\mathbf{B}}\hat{\mathbf{\Lambda}}^{-1} = \mathbf{F}^0\mathbf{B}^{0\prime}\hat{\mathbf{B}}\hat{\mathbf{\Lambda}}^{-1} + \mathbf{E}\hat{\mathbf{B}}\hat{\mathbf{\Lambda}}^{-1} \\
\frac{1}{T}\mathbf{F}^{0\prime}\hat{\mathbf{F}} &= \mathbf{B}^{0\prime}\hat{\mathbf{B}}\hat{\mathbf{\Lambda}}^{-1} + \frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\hat{\mathbf{B}}\hat{\mathbf{\Lambda}}^{-1}.
\end{aligned}$$

(iii) $\mathbf{B}^{0\prime}\hat{\mathbf{B}}$ in (i) implies

$$\begin{aligned}
\hat{\mathbf{B}}'\mathbf{B}^0 &= \frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0 + \frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\mathbf{B}^0 \\
\tilde{\mathbf{H}}_3 &= \tilde{\mathbf{H}}_1 + \left(\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\right)^{-1}\frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\mathbf{B}^0\left(\hat{\mathbf{B}}'\mathbf{B}^0\right)^{-1}.
\end{aligned}$$

(iv) Multiplying $\hat{\mathbf{B}}'$ to the first equation in (i),

$$\begin{aligned}
\hat{\mathbf{B}}'\hat{\mathbf{B}} &= \frac{1}{T}\hat{\mathbf{B}}'\mathbf{B}^0\mathbf{F}^{0\prime}\hat{\mathbf{F}} + \frac{1}{T}\hat{\mathbf{B}}'\mathbf{E}'\hat{\mathbf{F}} \\
\hat{\mathbf{B}}'\hat{\mathbf{B}} &= \frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\mathbf{B}^{0\prime}\hat{\mathbf{B}} + \frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\hat{\mathbf{B}} \\
\tilde{\mathbf{H}}_3 &= \tilde{\mathbf{H}}_4 + \left(\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\right)^{-1}\frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\hat{\mathbf{B}}\left(\hat{\mathbf{B}}'\hat{\mathbf{B}}\right)^{-1}.
\end{aligned}$$

(v) Post-multiplying $\mathbf{B}^0$ to the transpose of the first equation in (i),

$$\begin{aligned}
\hat{\mathbf{B}}'\mathbf{B}^0 &= \frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{B}^0 + \frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\mathbf{B}^0 \\
\tilde{\mathbf{H}}_1^{-1} &= \tilde{\mathbf{H}}_2' + \frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\mathbf{B}^0\left(\mathbf{B}^{0\prime}\mathbf{B}^0\right)^{-1} \\
\tilde{\mathbf{H}}_1'^{-1} &= \tilde{\mathbf{H}}_2 + \left(\mathbf{B}^{0\prime}\mathbf{B}^0\right)^{-1}\frac{1}{T}\mathbf{B}^{0\prime}\mathbf{E}'\hat{\mathbf{F}}.
\end{aligned}$$

(vi) The first equation in (iv) implies

$$\tilde{\mathbf{H}}_4'^{-1} = \tilde{\mathbf{H}}_2 + \left(\hat{\mathbf{B}}'\mathbf{B}^0\right)^{-1}\frac{1}{T}\hat{\mathbf{B}}'\mathbf{E}'\hat{\mathbf{F}}.$$

□

Lemma B.6. *Suppose that Assumptions 1–4 hold. Then, we have,*

(i) $\left\|\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\hat{\mathbf{B}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\right\|_{\mathrm{F}} = O_p(\Delta_1)$,

(ii) $\left\|\frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\mathbf{B}^0(\hat{\mathbf{B}}'\mathbf{B}^0)^{-1}\right\|_{\mathrm{F}} = O_p(\Delta_2)$,

(iii) $\left\|\frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\hat{\mathbf{B}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\right\|_{\mathrm{F}} = O_p(\Delta_1+\Delta_2)$.

where

$$\Delta_1 = \left(\frac{N^{1-\alpha_r}}{T}\right)^2 + \left(\frac{N^{1-\alpha_r}}{T}\right)^{\frac{3}{2}} + \sqrt{\frac{N^{1-\alpha_r}}{T}}N^{-\alpha_r} + \frac{N^{1-\alpha_r}}{T} + \frac{1}{\sqrt{TN^{\alpha_r}}}, \tag{B.10}$$
$$\Delta_2 = N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r} + N^{\frac{1}{2}\alpha_1-\alpha_r}\frac{N^{1-\alpha_r}}{T} + \frac{N^{\frac{1}{2}\alpha_1-\alpha_r}}{\sqrt{T}}.$$

Proof of Lemma B.6. (i) Using the results in the proof of Lemma 2, we show that

$$\begin{aligned}
&\left\|\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\hat{\mathbf{B}}(\hat{\mathbf{B}}\hat{\mathbf{B}})^{-1}\right\|_{\mathrm{F}} \\
&\leq \left\|\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\left(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{Q}}'\right)\left(\hat{\mathbf{B}}'\hat{\mathbf{B}}\right)^{-1}\right\|_{\mathrm{F}} + \left\|\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\mathbf{B}^0\tilde{\mathbf{Q}}'\left(\hat{\mathbf{B}}'\hat{\mathbf{B}}\right)^{-1}\right\|_{\mathrm{F}} \\
&\leq \frac{1}{T}O_p(\sqrt{NT})O_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\frac{N^{1-\frac{1}{2}\alpha_r}}{T} + \frac{N^{1-\frac{1}{2}\alpha_r}}{T} + N^{-\frac{1}{2}\alpha_r} + \sqrt{\frac{N}{T}}\right)O_p(N^{-\alpha_r}) + O_p\left(\frac{1}{\sqrt{TN^{\alpha_r}}}\right) \\
&= O_p\left(\left(\frac{N^{1-\alpha_r}}{T}\right)^2\right) + O_p\left(\left(\frac{N^{1-\alpha_r}}{T}\right)^{\frac{3}{2}}\right) + O_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}N^{-\alpha_r}\right) + O_p\left(\frac{N^{1-\alpha_r}}{T}\right) + O_p\left(\frac{1}{\sqrt{TN^{\alpha_r}}}\right).
\end{aligned}$$

(ii) Lemma B.3(ii) implies

$$\begin{aligned}
&\left\|\frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\mathbf{B}^0\left(\hat{\mathbf{B}}'\mathbf{B}^0\right)^{-1}\right\|_{\mathrm{F}} \\
&\leq \left\|\frac{1}{T}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)'\mathbf{E}\mathbf{B}^0\left(\hat{\mathbf{B}}'\mathbf{B}^0\right)^{-1}\right\|_{\mathrm{F}} + \left\|\frac{1}{T}\tilde{\mathbf{H}}_4'\left(\mathbf{F}^{0\prime}\mathbf{E}\mathbf{B}^0\right)\left(\hat{\mathbf{B}}'\mathbf{B}^0\right)^{-1}\right\|_{\mathrm{F}} \\
&\leq O_p\left(N^{\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_r}\left(\frac{N^{1-\frac{1}{2}\alpha_r}}{T}+1\right)N^{-\alpha_r}\right) + O_p\left(\frac{\sqrt{TN^{\alpha_1}}}{T}N^{-\alpha_r}\right) \\
&= O_p\left(N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\right) + O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\frac{N^{1-\alpha_r}}{T}\right) + O_p\left(\frac{N^{\frac{1}{2}\alpha_1-\alpha_r}}{\sqrt{T}}\right).
\end{aligned}$$

(iii) Lemma B.3(i) implies

$$
\begin{aligned}
&\left\|\frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\hat{\mathbf{B}}\left(\hat{\mathbf{B}}'\hat{\mathbf{B}}\right)^{-1}\right\|_{\mathrm{F}}\\
&\leq \left\|\frac{1}{T}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)'\mathbf{E}\hat{\mathbf{B}}\left(\hat{\mathbf{B}}'\hat{\mathbf{B}}\right)^{-1}\right\|_{\mathrm{F}} + \left\|\frac{1}{T}\tilde{\mathbf{H}}_4'\left(\mathbf{F}^{0\prime}\mathbf{E}\hat{\mathbf{B}}\right)\left(\hat{\mathbf{B}}'\hat{\mathbf{B}}\right)^{-1}\right\|_{\mathrm{F}}\\
&\leq \left\|\frac{1}{T}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)'\mathbf{E}(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{Q}})\left(\hat{\mathbf{B}}'\hat{\mathbf{B}}\right)^{-1} + \frac{1}{T}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)'\mathbf{E}\mathbf{B}^0\tilde{\mathbf{Q}}\left(\hat{\mathbf{B}}'\hat{\mathbf{B}}\right)^{-1}\right\|_{\mathrm{F}}\\
&\qquad + \left\|\frac{1}{T}\tilde{\mathbf{H}}_4'\left(\mathbf{F}^{0\prime}\mathbf{E}\hat{\mathbf{B}}\right)\left(\hat{\mathbf{B}}'\hat{\mathbf{B}}\right)^{-1}\right\|_{\mathrm{F}}\\
&\leq O_p(\Delta_1) + O_p(\Delta_2),
\end{aligned}
$$

where $\Delta_1 + \Delta_2 \lesssim \Delta_{NT}$ if $\frac{N^{1-\alpha_r}}{T} \to 0$. □

Lemma B.7. *Suppose that Assumptions 1–4 hold. If $\frac{N^{1-\alpha_r}}{T} \to 0$, then, we have*

$$
\left\|\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\left(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{Q}}'\right)'\mathbf{e}_t\right\|_{\mathrm{F}} = O_p(N^{\frac{1}{2}-\alpha_r}) + O_p\left(\frac{N^{\frac{3}{2}-\alpha_r}}{T}\right) + O_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\right).
$$

Proof of Lemma B.7. (A.2) implies

$$
\begin{aligned}
(\hat{\mathbf{B}}-\mathbf{B}^0\tilde{\mathbf{Q}}')'\mathbf{e}_t &= \frac{1}{T}(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)'\mathbf{E}\mathbf{e}_t + \tilde{\mathbf{H}}_4'\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\mathbf{e}_t\\
&= \frac{1}{T}\hat{\mathbf{\Lambda}}^{-1}\left(\frac{1}{T}\mathbf{E}'\mathbf{E}\mathbf{E}'\hat{\mathbf{F}} + \frac{1}{T}\mathbf{E}'\mathbf{E}\mathbf{B}^0\mathbf{F}^{0\prime}\hat{\mathbf{F}}\right)'\mathbf{e}_t + \tilde{\mathbf{H}}_4'\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\mathbf{e}_t.
\end{aligned}
$$

Consider the first term on the right-hand side of the above equation:

$$
\begin{aligned}
&\left\|\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\frac{1}{T^2}\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{F}}'\mathbf{E}\mathbf{E}'\mathbf{E}\mathbf{e}_t\right\|_{\mathrm{F}}\\
&= \left\|\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\mathbf{N}^{\frac{1}{2}}\mathbf{N}^{-\frac{1}{2}}\frac{1}{T^2}\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{F}}'\mathbf{E}\mathbf{E}'\mathbf{E}\mathbf{e}_t\right\|_{\mathrm{F}}\\
&\leq \left\|\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\mathbf{N}^{\frac{1}{2}}\mathbf{N}^{-\frac{1}{2}}\frac{1}{T^2}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}\left\|\hat{\mathbf{F}}'\right\|_{\mathrm{F}}\|\mathbf{E}\|_{sp}^3\|\mathbf{e}_t\|_{\mathrm{F}}\\
&\leq \frac{1}{T^2}N^{-\frac{3}{2}\alpha_r}\sqrt{T}(N^{3/2}+T^{3/2})\sqrt{N}O_p(1)\\
&= O_p\left(\frac{N^{\frac{3}{2}-\alpha_r}}{T}\sqrt{\frac{N^{1-\alpha_r}}{T}}\right) + O_p\left(N^{\frac{1}{2}-\frac{3}{2}\alpha_r}\right).
\end{aligned}
$$

Next, consider the upper bound of the second term.

$$\begin{aligned}
&\left\|\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\frac{1}{T^2}\hat{\boldsymbol{\Lambda}}^{-1}\hat{\mathbf{F}}'\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{E}'\mathbf{E}\mathbf{e}_t\right\|_{\mathrm{F}} \\
&\leq \left\|\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\mathbf{N}^{\frac{1}{2}}\frac{1}{T}\hat{\boldsymbol{\Lambda}}^{-1}\right\|_{\mathrm{F}}\left\|\mathbf{N}^{-\frac{1}{2}}\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0\mathbf{N}^{\frac{1}{2}}\right\|_{\mathrm{F}}\left\|\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^0\right\|_{\mathrm{F}}\lambda_1[\mathbf{E}'\mathbf{E}]\,\|\mathbf{e}_t\|_{\mathrm{F}} \\
&\leq \frac{1}{T}N^{-\alpha_r}(N+T)\sqrt{N}O_p(1) \\
&= O_p\left(\frac{N^{\frac{3}{2}-\alpha_r}}{T}\right)+O_p\left(N^{\frac{1}{2}-\alpha_r}\right).
\end{aligned}$$

Consider the third term and Lemma B.1(i) implies

$$\begin{aligned}
&\left\|\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\tilde{\mathbf{H}}_4'\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\mathbf{e}_t\right\|_{\mathrm{F}} \\
&\leq \left\|\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\tilde{\mathbf{H}}_4'\right\|_{\mathrm{F}}\left\|\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\mathbf{e}_t\right\|_{\mathrm{F}} \\
&\leq N^{-\frac{1}{2}\alpha_r}N\left(\frac{1}{\sqrt{NT}}+\frac{1}{T}\right)O_p(1) \\
&= O_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\right)+O_p\left(\frac{N^{1-\frac{1}{2}\alpha_r}}{T}\right).
\end{aligned}$$

If $\frac{N^{1-\alpha_r}}{T}\to 0$ holds, collecting these terms completes the proof. □

Lemma B.8. *Suppose that Assumptions 1–6 hold. If* $\frac{N^{1-\alpha_r}}{T}\to 0$*, then, we have*

(i) $\frac{1}{T}\|\hat{\mathbf{Z}}'(\mathbf{F}^0-\hat{\mathbf{F}})\|_{\mathrm{F}}=O_p\left(\Delta_{NT}\right)$,

(ii) $\frac{1}{\sqrt{T}}\|(\hat{\mathbf{F}}-\mathbf{F}^0)'\boldsymbol{\epsilon}\|_{\mathrm{F}}=O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(\frac{\sqrt{T}}{N^{\alpha_r}}\right)+o_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\right)$.

Proof of Lemma B.8. (i) Because $\hat{\mathbf{Z}}$ includes $\hat{\mathbf{F}}$ and $\mathbf{W}$, we have

$$\frac{1}{T}\left\|\hat{\mathbf{F}}'(\mathbf{F}^0-\hat{\mathbf{F}})\right\|_{\mathrm{F}} = \left\|\tilde{\mathbf{Q}}-\mathbf{I}_r\right\|_{\mathrm{F}}=O_p\left(\Delta_{NT}\right),$$

$$\begin{aligned}
&\frac{1}{T}\left\|\mathbf{W}'(\mathbf{F}^0-\hat{\mathbf{F}})\right\|_{\mathrm{F}} \\
&=\frac{1}{T}\left\|\mathbf{W}'(\mathbf{F}^0-\mathbf{F}^0\tilde{\mathbf{H}}_4+\mathbf{F}^0\tilde{\mathbf{H}}_4-\hat{\mathbf{F}})\right\|_{\mathrm{F}} \\
&\leq \frac{1}{T}\left\|\mathbf{W}'\mathbf{F}^0(\mathbf{I}_r-\tilde{\mathbf{H}}_4)\right\|_{\mathrm{F}}+\frac{1}{T}\left\|\mathbf{W}'(\mathbf{F}^0\tilde{\mathbf{H}}_4-\hat{\mathbf{F}})\right\|_{\mathrm{F}} \\
&= O_p\left(\Delta_{NT}\right)+O_p\left(\frac{N^{1-\alpha_r}}{T}\right)+O_p\left(\frac{1}{N^{\alpha_r}}\right)+O_p\left(\frac{1}{\sqrt{TN^{\alpha_r}}}\right).
\end{aligned}$$

The final inequality is from

$$
\begin{aligned}
&\frac{1}{T}\left\|\mathbf{W}'(\mathbf{F}^0\tilde{\mathbf{H}}_4-\hat{\mathbf{F}})\right\|_{\mathrm{F}}\\
&=\frac{1}{T}\left\|\mathbf{W}'\left(\frac{1}{T}\mathbf{E}\mathbf{E}'\hat{\mathbf{F}}+\frac{1}{T}\mathbf{E}\mathbf{B}^0\mathbf{F}^{0\prime}\hat{\mathbf{F}}\right)\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}\\
&\leq\frac{1}{T^2}\left\|\mathbf{W}'\mathbf{E}\mathbf{E}'\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}+\frac{1}{T^2}\left\|(\mathbf{W}'\mathbf{E}\mathbf{B}^0)\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\mathbf{\Lambda}}^{-1}\right\|_{\mathrm{F}}\\
&=O_p\left(\frac{N^{1-\alpha_r}}{T}+\frac{1}{N^{\alpha_r}}\right)+O_p\left(\frac{1}{\sqrt{TN^{\alpha_r}}}\right)\\
&=O_p\left(\frac{N^{1-\alpha_r}}{T}\right)+O_p\left(\frac{1}{N^{\alpha_r}}\right)+O_p\left(\frac{1}{\sqrt{TN^{\alpha_r}}}\right),
\end{aligned}
$$

which is less than $O_p(\Delta_{NT})$.

(ii) Using Lemmas B.5(ii), B.6(i), and B.4(i)

$$
\begin{aligned}
&\left\|\tilde{\mathbf{H}}_4-\mathbf{I}_r\right\|_{\mathrm{F}}\\
&=\left\|\tilde{\mathbf{H}}_4-\tilde{\mathbf{H}}_2+\tilde{\mathbf{H}}_2-\mathbf{I}_r\right\|_{\mathrm{F}}\\
&=\left\|-\frac{1}{T}\mathbf{F}^{0\prime}\mathbf{E}\hat{\mathbf{B}}\mathbf{\Lambda}^{-1}+\tilde{\mathbf{H}}_2-\mathbf{I}_r\right\|_{\mathrm{F}}\\
&\leq O_p(\Delta_1)+O_p(\Delta_{NT})=O_p(\Delta_{NT}),
\end{aligned}
$$

if $\frac{N^{1-\alpha_r}}{T}\to 0$. Then,

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\left\|(\hat{\mathbf{F}}-\mathbf{F}^0)'\boldsymbol{\epsilon}\right\|_{\mathrm{F}}\\
&\leq\frac{1}{\sqrt{T}}\left\|(\hat{\mathbf{F}}-\mathbf{F}^0\tilde{\mathbf{H}}_4)'\boldsymbol{\epsilon}\right\|_{\mathrm{F}}+\frac{1}{\sqrt{T}}\left\|(\tilde{\mathbf{H}}_4-\mathbf{I}_r)'\mathbf{F}^{0\prime}\boldsymbol{\epsilon}\right\|_{\mathrm{F}}\\
&\leq\frac{1}{T^{3/2}}\left\|\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{F}}'\mathbf{E}\mathbf{E}'\boldsymbol{\epsilon}\right\|_{\mathrm{F}}+\frac{1}{T^{3/2}}\left\|\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{F}}'\mathbf{F}^0\mathbf{B}^{0\prime}\mathbf{E}'\boldsymbol{\epsilon}\right\|_{\mathrm{F}}+\frac{1}{\sqrt{T}}\left\|(\tilde{\mathbf{H}}_4-\mathbf{I}_r)'\mathbf{F}^{0\prime}\boldsymbol{\epsilon}\right\|_{\mathrm{F}}\\
&\leq O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}+\frac{\sqrt{T}}{N^{\alpha_r}}\right)+O_p\left(\frac{\sqrt{T}}{N^{\alpha_r}}+\frac{1}{\sqrt{N^{\alpha_r}}}\right)+O_p(\Delta_{NT})\\
&=O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(\frac{\sqrt{T}}{N^{\alpha_r}}\right)+o_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\right),
\end{aligned}
$$

since $\frac{1}{T}\mathbf{N}^{-\frac{1}{2}}\mathbf{B}^{0\prime}\mathbf{E}'\boldsymbol{\epsilon}=O_p\left(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{N^{\alpha_r}}}\right)$ and $\frac{N^{1-\alpha_r}}{T}\to 0$. □

Lemma B.9. *Suppose that Assumptions 1–6 hold. If $\frac{1}{2}\alpha_1<\alpha_r$, $\frac{N^{1-\alpha_r}}{\sqrt{T}}\to 0$, and $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\to 0$ as $N,T\to\infty$, we have*

$$
\begin{aligned}
&\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_3)\xrightarrow{d}N(\mathbf{0},\mathbf{\Sigma}_{\delta^0}),\\
&\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_4)\xrightarrow{d}N(\mathbf{0},\mathbf{\Sigma}_{\delta^0}),
\end{aligned}
$$

where $\boldsymbol{\gamma}_j=\tilde{\mathbf{H}}_j^{-1}\boldsymbol{\gamma}^*$, $\boldsymbol{\delta}_j=(\boldsymbol{\gamma}_j',\boldsymbol{\beta}')'$, $j=3,4$, $\mathbf{\Sigma}_{\delta^0}=\mathbf{\Sigma}_{z^0}^{-1}\mathbf{\Sigma}_{z^0\epsilon}\mathbf{\Sigma}_{z^0}^{-1}$.

Proof of Lemma B.9. The augmented regression model is rewritten as

$$
\begin{aligned}
y_{t+h} &= \boldsymbol{\gamma}^{*\prime}\mathbf{f}_t^* + \boldsymbol{\beta}'\mathbf{w}_t + \epsilon_{t+h} \\
&= \boldsymbol{\gamma}^{*\prime}\tilde{\mathbf{H}}_j^{\prime -1}\hat{\mathbf{f}}_t + \boldsymbol{\beta}'\mathbf{w}_t + \epsilon_{t+h} + \boldsymbol{\gamma}^{*\prime}\tilde{\mathbf{H}}_j^{\prime -1}\tilde{\mathbf{H}}_j\mathbf{f}_t^* - \boldsymbol{\gamma}^{*\prime}\tilde{\mathbf{H}}_j^{\prime -1}\hat{\mathbf{f}}_t.
\end{aligned}
$$

(i) We start with $j = 3$.

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\hat{\mathbf{F}}'(\mathbf{F}^0\tilde{\mathbf{H}}_3 - \hat{\mathbf{F}})\tilde{\mathbf{H}}_3^{-1}\boldsymbol{\gamma}^* = 0 \\
&\frac{1}{\sqrt{T}}\mathbf{W}'(\mathbf{F}^0\tilde{\mathbf{H}}_3 - \hat{\mathbf{F}})\tilde{\mathbf{H}}_3^{-1}\boldsymbol{\gamma}^* \\
&= \frac{1}{\sqrt{T}}(\mathbf{W}'\mathbf{F}^0 - \mathbf{W}'\hat{\mathbf{F}}\tilde{\mathbf{H}}_3^{-1})\boldsymbol{\gamma}^* \\
&= \frac{1}{\sqrt{T}}(\mathbf{W}'\mathbf{F}^0 - \mathbf{W}'\mathbf{X}\hat{\mathbf{B}}\hat{\boldsymbol{\Lambda}}^{-1}\tilde{\mathbf{Q}})\boldsymbol{\gamma}^* \\
&= \frac{1}{\sqrt{T}}\mathbf{W}'\mathbf{F}^0(\mathbf{I}_r - \mathbf{B}^{0\prime}\hat{\mathbf{B}}\hat{\boldsymbol{\Lambda}}^{-1}\tilde{\mathbf{Q}}) - \frac{1}{\sqrt{T}}(\mathbf{W}'\mathbf{E}'\hat{\mathbf{B}}\hat{\boldsymbol{\Lambda}}^{-1}\tilde{\mathbf{Q}})\boldsymbol{\gamma}^* \\
&= \sqrt{T}O_p(\Delta_{NT}).
\end{aligned}
$$

By the proof of Theorem 4, we have

$$
\begin{aligned}
&\sqrt{T}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_3) \\
&= \left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'\boldsymbol{\epsilon} + \left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'(\mathbf{F}^0\tilde{\mathbf{H}}_3 - \hat{\mathbf{F}})\tilde{\mathbf{H}}_3^{-1}\boldsymbol{\gamma}^* \\
&= \left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon} + \left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}(\hat{\mathbf{Z}} - \mathbf{Z}^0)'\boldsymbol{\epsilon} + \left[\left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1} - \left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\right]\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'\boldsymbol{\epsilon} \\
&\quad + \left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'(\mathbf{F}^0\tilde{\mathbf{H}}_3 - \hat{\mathbf{F}})\tilde{\mathbf{H}}_3^{-1}\boldsymbol{\gamma}^* \\
&= \left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon} + O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right) + O_p\left(\sqrt{T}N^{\frac{1}{2}\alpha_1 - \frac{3}{2}\alpha_r}\right) + O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}N^{\frac{1}{2}\alpha_1 - \alpha_r}\right) \\
&\quad + O_p\left(N^{\frac{1}{2}\alpha_1 - \alpha_r}\right).
\end{aligned}
$$

Thus, $\sqrt{T}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_3) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\delta^0})$ if $\frac{1}{2}\alpha_1 < \alpha_r$, $\frac{N^{1-\alpha_r}}{\sqrt{T}} \to 0$, and $\sqrt{T}N^{\frac{1}{2}\alpha_1 - \frac{3}{2}\alpha_r} \to 0$.

(ii) Let $j = 4$,

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\hat{\mathbf{F}}'(\mathbf{F}^0\tilde{\mathbf{H}}_4 - \hat{\mathbf{F}})\tilde{\mathbf{H}}_4^{-1}\boldsymbol{\gamma}^* = \sqrt{T}\left(\frac{1}{T}\hat{\mathbf{F}}'\mathbf{F}^0 - \tilde{\mathbf{H}}_4^{-1}\right)\boldsymbol{\gamma}^* = \sqrt{T}\left[\frac{1}{T}\hat{\mathbf{F}}'\mathbf{E}\hat{\mathbf{B}}(\hat{\mathbf{B}}'\mathbf{B}^0)^{-1}\right]\boldsymbol{\gamma}^* = \sqrt{T}O_p(\Delta_1 + \Delta_2), \\
&\frac{1}{\sqrt{T}}\mathbf{W}'(\mathbf{F}^0\tilde{\mathbf{H}}_4 - \hat{\mathbf{F}})\tilde{\mathbf{H}}_4^{-1}\boldsymbol{\gamma}^* \\
&= \sqrt{T}\left(-\frac{1}{T^2}\mathbf{W}'\mathbf{E}\mathbf{B}^0\mathbf{F}^{0\prime}\hat{\mathbf{F}}\hat{\boldsymbol{\Lambda}}^{-1} - \frac{1}{T^2}\mathbf{W}'\mathbf{E}\mathbf{E}'\hat{\mathbf{F}}\hat{\boldsymbol{\Lambda}}^{-1}\right)\tilde{\mathbf{H}}_4^{-1}\boldsymbol{\gamma}^* \\
&= O_p\left(N^{\frac{1}{2}\alpha_1 - \alpha_r}\right) + O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}} + \sqrt{T}N^{-\alpha_r}\right).
\end{aligned}
$$

By the proof of Theorem 4, we have

$$
\begin{aligned}
&\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_4)\\
&=\left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'\boldsymbol{\epsilon}+\left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'(\mathbf{F}^0\tilde{\mathbf{H}}_4-\hat{\mathbf{F}})\tilde{\mathbf{H}}_4^{-1}\boldsymbol{\gamma}^*\\
&=\left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon}+\left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}(\hat{\mathbf{Z}}-\mathbf{Z}^0)'\boldsymbol{\epsilon}+\left[\left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}-\left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\right]\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'\boldsymbol{\epsilon}\\
&\quad+\left(\frac{1}{T}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{\sqrt{T}}\hat{\mathbf{Z}}'(\mathbf{F}^0\tilde{\mathbf{H}}_4-\hat{\mathbf{F}})\tilde{\mathbf{H}}_4^{-1}\boldsymbol{\gamma}^*\\
&=\left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon}+O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\right)+O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}N^{\frac{1}{2}\alpha_1-\alpha_r}\right)\\
&\quad+O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\right).
\end{aligned}
$$

Thus, $\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_4)\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Sigma}_{\delta^0})$ if $\frac{1}{2}\alpha_1<\alpha_r$, $\frac{N^{1-\alpha_r}}{\sqrt{T}}\to 0$, and $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\to 0$.

□

Lemma B.10. *Under the assumptions of Theorem 5, if we use rotation matrices $\tilde{\mathbf{H}}_3$ or $\tilde{\mathbf{H}}_4$, we obtain the same results as in Theorem 5.*

$$
\frac{\left(\hat{y}_{T+h|T}-y_{T+h|T}\right)}{\sigma_{T+h|T}}\xrightarrow{d}N(0,1),
$$

where $\sigma^2_{T+h|T}=T^{-1}\mathbf{z}_T^{0\prime}\boldsymbol{\Sigma}_{\delta^0}\mathbf{z}_T^0+\boldsymbol{\gamma}^{0\prime}\mathbf{D}^{-1}\mathbf{N}^{-1/2}\boldsymbol{\Gamma}_T\mathbf{N}^{-1/2}\mathbf{D}^{-1}\boldsymbol{\gamma}^0$.

Proof of Lemma B.10. Expand the term

$$
\begin{aligned}
&\hat{y}_{T+h|T}-y_{T+h|T}\\
&=\hat{\boldsymbol{\gamma}}'\hat{\mathbf{f}}_T+\hat{\boldsymbol{\beta}}'\mathbf{w}_T-\boldsymbol{\gamma}^{*\prime}\mathbf{f}_T-\boldsymbol{\beta}'\mathbf{w}_T\\
&=\left(\hat{\boldsymbol{\gamma}}-\hat{\mathbf{H}}_j^{-1\prime}\boldsymbol{\gamma}^*\right)'\hat{\mathbf{f}}_T+\boldsymbol{\gamma}^{*\prime}\hat{\mathbf{H}}_j^{-1}\left(\hat{\mathbf{f}}_T-\hat{\mathbf{H}}_j\mathbf{f}_T\right)+(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{w}_T\\
&=\hat{\mathbf{z}}_T'(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_j)+\boldsymbol{\gamma}^{*\prime}\hat{\mathbf{H}}_j^{-1}\left(\hat{\mathbf{f}}_T-\hat{\mathbf{H}}_j\mathbf{f}_T\right)\\
&=T^{-1/2}\hat{\mathbf{z}}_T'[\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_j)]+\boldsymbol{\gamma}^{*\prime}\hat{\mathbf{H}}_j^{-1}\mathbf{N}^{-1/2}\left[\mathbf{N}^{1/2}\left(\hat{\mathbf{f}}_T-\tilde{\mathbf{H}}_j\mathbf{f}_T^0\right)\right]\\
&=T^{-1/2}\mathbf{z}_T^{0\prime}[\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_j)]+T^{-1/2}(\hat{\mathbf{z}}_T-\mathbf{z}_T^0)'[\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_j)]\\
&\quad+\boldsymbol{\gamma}^{*\prime}\hat{\mathbf{H}}_j^{-1}\mathbf{N}^{-1/2}\left[\mathbf{N}^{1/2}\left(\hat{\mathbf{f}}_T-\tilde{\mathbf{H}}_4\mathbf{f}_T^0\right)\right]+\boldsymbol{\gamma}^{*\prime}\hat{\mathbf{H}}_j^{-1}\mathbf{N}^{-1/2}\left[\mathbf{N}^{1/2}\left(\tilde{\mathbf{H}}_4-\tilde{\mathbf{H}}_j\right)\mathbf{f}_T^0\right].
\end{aligned}
$$

(i) We start with $j=3$. The second term on the righat-hand side of the above equation is dominated by the first one, thus, we next focus on the first, third and fourth terms.

$$
\begin{aligned}
&T^{-1/2}\mathbf{z}_T^{0\prime}\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_3)\\
&=T^{-1/2}\mathbf{z}_T^{0\prime}\left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon}\\
&\quad+T^{-1/2}\left[O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\right)+O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}N^{\frac{1}{2}\alpha_1-\alpha_r}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\right)\right].
\end{aligned}
$$

$$\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-\frac{1}{2}}\left[\mathbf{N}^{1/2}\left(\hat{\mathbf{f}}_T-\tilde{\mathbf{H}}_4\mathbf{f}_T^0\right)\right]$$
$$=\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{D}^{-1}\left[\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\mathbf{B}^{0\prime}\mathbf{e}_t\right]+O_p(\|\mathbf{N}^{-\frac{1}{2}}\|_F)O_p(\Delta_{NT})$$
$$+O_p(\|\mathbf{N}^{-\frac{1}{2}}\|_F)\left[O_p(N^{\frac{1}{2}-\alpha_r})+O_p\left(\frac{N^{\frac{3}{2}-\alpha_r}}{T}\right)+O_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\right)\right].$$
$$\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\left(\tilde{\mathbf{H}}_3-\tilde{\mathbf{H}}_4\right)\mathbf{f}_T^0$$
$$=\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{N}^{\frac{1}{2}}\left(\tilde{\mathbf{H}}_3-\tilde{\mathbf{H}}_4\right)\mathbf{f}_T^0=O_p(\|\mathbf{N}^{-\frac{1}{2}}\|_F)O_p(N^{\frac{1}{2}\alpha_1}\Delta_{NT}),$$

where

$$N^{\frac{1}{2}\alpha_1}\Delta_{NT}=\frac{N^{1+\frac{1}{2}\alpha_1-\alpha_r}}{T}+N^{\frac{1}{2}\alpha_1-\alpha_r}\frac{N^{1+\frac{1}{2}\alpha_1-\alpha_r}}{T}+N^{\alpha_1-\frac{3}{2}\alpha_r}+\frac{N^{\alpha_1-\alpha_r}}{\sqrt{T}}.$$

Thus, if $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}$, $\frac{1}{2}<\alpha_r$, and $\frac{N^{\frac{3}{2}-\alpha_r}}{T}\to 0$, we have

$$\frac{\left(\hat{y}_{T+h|T}-y_{T+h|T}\right)}{\sigma_{T+h|T}}\xrightarrow{d}N(0,1),$$

where $\sigma^2_{T+h|T}=T^{-1}\mathbf{z}_T^{0\prime}\boldsymbol{\Sigma}_{\delta^0}\mathbf{z}_T^0+\boldsymbol{\gamma}^{0\prime}\mathbf{D}^{-1}\mathbf{N}^{-1/2}\boldsymbol{\Gamma}_T\mathbf{N}^{-1/2}\mathbf{D}^{-1}\boldsymbol{\gamma}^0$.

(ii) Let $j=4$,

$$T^{-1/2}\mathbf{z}_T^{0\prime}\sqrt{T}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_4)$$
$$=T^{-1/2}\mathbf{z}_T^{0\prime}\left(\frac{1}{T}\mathbf{Z}^{0\prime}\mathbf{Z}^0\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{Z}^{0\prime}\boldsymbol{\epsilon}$$
$$+T^{-1/2}\left[O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}\right)+O_p\left(\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}\right)+O_p\left(\frac{N^{1-\alpha_r}}{\sqrt{T}}N^{\frac{1}{2}\alpha_1-\alpha_r}\right)+O_p\left(N^{\frac{1}{2}\alpha_1-\alpha_r}\right)\right].$$
$$\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-\frac{1}{2}}\left[\mathbf{N}^{1/2}\left(\hat{\mathbf{f}}_T-\tilde{\mathbf{H}}_4\mathbf{f}_T^0\right)\right]$$
$$=\boldsymbol{\gamma}^{*\prime}\mathbf{H}^{-1}\mathbf{N}^{-\frac{1}{2}}\mathbf{D}^{-1}\left[\mathbf{D}\mathbf{N}^{\frac{1}{2}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1}\mathbf{B}^{0\prime}\mathbf{e}_t\right]+O_p(\|\mathbf{N}^{-\frac{1}{2}}\|_F)O_p(\Delta_{NT})$$
$$+O_p(\|\mathbf{N}^{-\frac{1}{2}}\|_F)\left[O_p(N^{\frac{1}{2}-\alpha_r})+O_p\left(\frac{N^{\frac{3}{2}-\alpha_r}}{T}\right)+O_p\left(\sqrt{\frac{N^{1-\alpha_r}}{T}}\right)\right].$$

Thus, if $\sqrt{T}N^{\frac{1}{2}\alpha_1-\frac{3}{2}\alpha_r}$, $\frac{1}{2}<\alpha_r$, and $\frac{N^{\frac{3}{2}-\alpha_r}}{T}\to 0$, we have

$$\frac{\left(\hat{y}_{T+h|T}-y_{T+h|T}\right)}{\sigma_{T+h|T}}\xrightarrow{d}N(0,1),$$

where $\sigma^2_{T+h|T}=T^{-1}\mathbf{z}_T^{0\prime}\boldsymbol{\Sigma}_{\delta^0}\mathbf{z}_T^0+\boldsymbol{\gamma}^{0\prime}\mathbf{D}^{-1}\mathbf{N}^{-1/2}\boldsymbol{\Gamma}_T\mathbf{N}^{-1/2}\mathbf{D}^{-1}\boldsymbol{\gamma}^0$. □