MONASH University

MONASH BUSINESS SCHOOL

DEPARTMENT OF ECONOMETRCS AND BUSINESS STATISTICS

ISSN 1440-771X

WORKING PAPER SERIES

Estimation and Inference for Three-Dimensional Panel Data Models

Guohua Feng, Jiti Gao, Fei Liu and Bin Peng

Working Paper No. 20/23
November 2023

Estimation and Inference for Three-Dimensional Panel Data Models

Guohua Feng*, Jiti Gao†, Fei Liu♯ and Bin Peng†

August 22, 2023

Abstract

In this paper, we develop estimation and inferential methods for three-dimensional (3D) panel data models with homogeneous/heterogeneous coefficients. Our 3D panel data models specify the nature of common shocks through the use of a hierarchical factor structure (i.e., global factors and sector factors). Accordingly, we develop an approach to estimating the hierarchy, thus enabling us to have a better understanding of the relative importance of the two types of unobservable shocks. Second, we propose bias corrected estimators, and give bootstrap procedures to construct the confidence intervals for the parameters of interest while allowing for correlation along three dimensions of idiosyncratic errors. We justify the theoretical findings using extensive simulations. In an empirical study, we examine the twin hypotheses of conditional and unconditional-convergence for manufacturing industries across countries.

Keywords: Asymptotic Theory, Bias Correction, Dependent Wild Bootstrap, Hierarchical Model

JEL classification: C23, O10, L60

*Department of Economics, University of North Texas, United States.
†Department of Econometrics and Business Statistics, Monash University, Australia.
♯School of Finance, Nankai University, China.

Gao and Peng would like to acknowledge the Australian Research Council Discovery Projects Program for its financial support under Grant Numbers: DP200102769 & DP210100476. Liu's research is financially supported by the National Natural Science Foundation of China under Grant No. 72203114.

1 Introduction

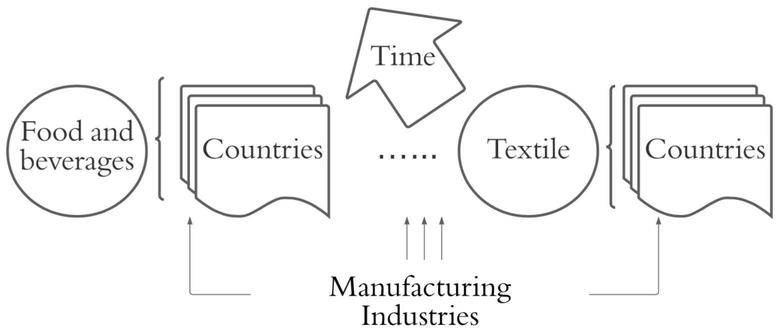
The development of panel data research has progressed extensively over the past thirty years. Particularly in applied works, panel data models are capable of providing accurate estimates by utilizing information along both of the cross-sectional and time dimensions. Early works specifically focus on large N but small T setting, where N and T represent the numbers of observations available over the cross-sectional and time dimensions respectively. Hsiao (2003) provides a comprehensive review on models of this kind. This line of research usually assumes independence among individual units (e.g., Arellano and Bond, 1991; Blundell and Bond, 1998). When making inferences, such as calculating confidence intervals, the correlation among different units is not well represented in regression results. As a consequence, some results may appear more significant than they actually are.

With the emergence of data rich environment, panel data research shifts its focus to large N and T framework. Recent seminal studies include Pesaran (2006) and Bai (2009). In particular, Bai (2009) draws attention to two types of cross-sectional dependence (CSD), namely strong and weak. In his paper, strong CSD is captured by the factor structure, which involves estimating a low-rank representation. On the other hand, weak CSD is measured by the correlation of idiosyncratic errors and is an issue of inference associated with the estimation of the asymptotic covariance matrix. To deal with weak CSD, Gonçalves (2011) proposes to use block bootstrap, while Bai et al. (2020) recently consider a combination of heteroskedasticity and autocorrelation consistent (HAC) covariance estimation approach and thresholding approach. In both papers, a panel data model with fixed effects is considered. To the best of our knowledge, however, two types of CSD have not been satisfactorily addressed within one framework.

On top of the aforementioned issues, hierarchical panel data models have recently received an increasing amount of attention in the general field of econometrics as they allow for simultaneous examination of unobservable shocks occurring at different levels of the hierarchy. Also, they have the advantage of accommodating a more flexible degree of CSD along different dimensions. To give a detailed example, Figure 1 below plots an empirical example to be studied in Section 5, in which we have 23 industry sectors, and each sector has variables from a large number of countries over a long time period. It then calls for a framework which can naturally accommodate the hierarchy of the dataset while allowing for different types of CSD. All things considered, how to infer the parameters of interest in practice becomes even more daunting.

That said, in this paper we shall work with a 3D panel data model, and, with moderate modifications, the techniques employed can be applied to a wide range

Figure 1: The Hierarchical Structure of Dataset



of applications such as those in Matyas et al. (2017). Up to this point, it is worth mentioning that existing studies on 3D panel data models suffer from two problems. First, many of them are simulation studies without providing associated asymptotic validations (see Breitung and Eickmeier, 2016 for an excellent review). A few exceptions are Jin et al. (2023), Lin (2023), Choi et al. (2023) and Kapetanios et al. (2021). We comment on each of them below. Jin et al. (2023) consider a three-dimensional latent factor model without regressors, and propose a two-step estimation procedure based on principal component analysis (PCA) to decompose the factor structure. Choi et al. (2023) consider a similar model, and assume one dimension is fixed. From the perspective of theoretical developments, the analysis of Choi et al. (2023) is similar to a two-dimensional factor model. Kapetanios et al. (2021) allow for regressors, and adopt the common correlated effects (CCE) estimation approach. However, to obtain inference (such as confidence interval), CCE approach usually turns to a random coefficient setting, which imposes independent and identically distributed (i.i.d.) conditions on individual coefficients. By doing so, the challenge raised with weak CSD is automatically bypassed. Very recently, Lin (2023) considers a model very similar to those to be studied in this paper but having one dimension fixed. Therefore, his study reduces to a typical two dimensional panel data model from the methodological perspective. Moreover, there is no inference provided in his paper, and the bias correction is done by assuming away all kinds of weak dependence.

Specifically, our study contributes to the hierarchical panel data literature in three ways.

1. We propose two 3D panel data models to cover the cases with homogeneous and heterogeneous coefficients in Sections 2 and 3 respectively. Our 3D panel data models not only contain regressors but also specify the nature of common shocks through the use of a hierarchical factor structure (i.e., global factors and sector factors). Accordingly, we develop an approach to estimating the

hierarchy, thus enabling us to have a better understanding of the relative importance of these two types of unobservable shocks in explaining the response variable.

2. For both models, we provide bias corrected estimators, and give bootstrap procedures to establish inference for the parameters of interest while allowing for both types of CSD and time series autocorrelation. It is worth pointing out that the asymptotic results established for the heterogeneous coefficients case also apply to for example Choi et al. (2023) and Lin (2023), although all results are established assuming the three dimensions to diverge. It is also worth stressing that the mean group estimator of the heterogeneous case suffers some serious asymptotic bias unless a specific structure on the regressors is imposed (such as that in Pesaran, 2006). We therefore provide a treatment in Section 3.

3. Third, we justify the theoretical findings using extensive simulated and real data examples. In the empirical study, we examine the twin hypotheses of conditional and unconditional-convergence for manufacturing industries across countries. The empirical results presented in this paper suggest that unconditional-convergence does not obtain. This finding is quite robust to the exclusion of OECD countries and to the use of different sample periods. On the other hand, there is strong and consistent evidence of convergence once factors that affect steady-state levels of labour productivity are controlled for.

The rest of the paper is organized as follows. Section 2 investigates the model with homogeneous coefficient and establishes the corresponding asymptotic properties. Section 3 extends the results of Section 2 to the case with heterogeneous coefficients. Section 4 conducts extensive simulation studies. Empirical results are presented in Section 5. Section 6 concludes the paper.

Before proceeding further, it is convenient to introduce some notations that will be repeatedly used throughout the article. We emphasize that vector and matrix are always written in bold font. For a matrix $\mathbf{A}$, $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_2$ denote the Frobenius norm and the spectral norm of $\mathbf{A}$, respectively, and $\mathbf{A}^\top$ stands for the transpose of $\mathbf{A}$. Provided that $\mathbf{A}$ has full column rank, let $\mathbf{M}_\mathbf{A} = \mathbf{I} - \mathbf{P}_\mathbf{A}$ with $\mathbf{P}_\mathbf{A} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$. For two scalars m and n, $m \wedge n = \min\{m, n\}$, $m \vee n = \max\{m, n\}$. For two random variables a and b, $a \asymp b$ stands for $a = O_P(b)$ and $b = O_P(a)$. For a positive integer L, let $[L]$ define the set $\{1, 2, \ldots, L\}$. $\mathbb{I}(\cdot)$ represents the conventional indicator function, and the mathematical symbols "$\to_P$" and "$\to_D$", stand for convergence in probability and convergence in distribution respectively;

$\mathbb{E}$ stands for the expectation operation; $\mathbb{E}^*$ and $\Pr^*$ stand for the expectation and probability induced by the bootstrap procedure.

2 A Model with Homogeneous Coefficients

As aforementioned, we consider the models with homogeneous coefficients and heterogeneous coefficients, respectively. Although the case with homogeneous coefficients sounds easy, it is actually more challenging to deal with. We shall be specific below. Therefore, we start with the homogeneous case. This section consists of five subsections: Section 2.1 presents the model, and the estimation approach; In Section 2.2, we derive the asymptotic distribution, while Section 2.3 provides a bootstrap procedure to establish confidence interval in practice; Section 2.4 decomposes the unobservable factor structure; Section 2.5 discusses the treatment for unbalanced data.

2.1 The Setup

The panel data model that we study can be written as

$$y_{ijt} = \mathbf{x}_{ijt}^\top \boldsymbol{\beta} + \boldsymbol{\gamma}_{ij}^{G\top} \mathbf{f}_t^G + \boldsymbol{\gamma}_{ij}^{S\top} \mathbf{f}_{it}^S + \varepsilon_{ijt}, \tag{2.1}$$

where $i \in [L]$ and $j \in [N]$ index the units in the first and second layers of the hierarchy respectively; $t \in [T]$ indexes the time periods; and $\mathbf{x}_{ijt}$ is a $d \times 1$ vector of observable explanatory variables. In (2.1), only y_{ijt} and $\mathbf{x}_{ijt}$ are observables, and ε_{ijt} is the idiosyncratic error that has weak correlation along three dimensions.

One main goal is inferring $\boldsymbol{\beta}$, which requires estimation and inference when $(L, N, T) \to (\infty, \infty, \infty)$. Equivalently important, we would like to decompose the two layers of unobservable shocks. Here, $\mathbf{f}_t^G$ is regarded as an $l^G \times 1$ vector of global shocks which has an impact on all units, while $\mathbf{f}_{it}^S$ is an $l_i^S \times 1$ vector of shocks from the i^{th} sector. We require d, l^G and $\max_{i \le L} l_i^S$ to be finite, and allow for the fact that there is no global factor, and that the factor structure does not necessarily exist for every single sector.

The two-layer factor structure is general and nests several widely-used fixed settings as special cases.

Example 1. *When $\boldsymbol{\gamma}_{ij}^G = (\alpha_i + \gamma_j, 1)^\top$ and $\mathbf{f}_t^G = (1, g_t)^\top$, the two-layer structure captures the three way fixed effects used in Abowd et al. (1999), who analyse the wage data of different individuals employed by different firms. The firms correspond to the sectors of the first layer.*

Example 2. *When $\boldsymbol{\gamma}_{ij}^G = D_j$, $\mathbf{f}_t^G = 1$, $\boldsymbol{\gamma}_{ij}^S = 1$, and $\mathbf{f}_{it}^S = D_{it}$, the two-layer structure captures the fixed effects employed in Rodrik (2013), who investigates unconditional convergence of manufacturing industries using a large number of countries. The industries correspond to the sectors of the first layer.*

More examples can be found in Matyas et al. (2017). We acknowledge the fact that it is impossible to cover all possible fixed effects structures in one framework, and refer the interested reader to Lu et al. (2021) for more discussions on this matter.

Using matrix notation, $\forall t \in [T]$, it is easy to see the two-layer unobservable shocks affect y_{ijt}'s through the following structure:

$$\begin{pmatrix} y_{11t} \\ \vdots \\ y_{1Nt} \\ \vdots \\ y_{L1t} \\ \vdots \\ y_{LNt} \end{pmatrix} \begin{matrix} \\ \Lleftarrow \\ \\ \Lleftarrow \\ \\ \Lleftarrow \\ \\ \end{matrix} \begin{pmatrix} \boldsymbol{\gamma}_{11}^{G\top} & \boldsymbol{\gamma}_{11}^{S\top} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\gamma}_{1N}^{G\top} & \boldsymbol{\gamma}_{1N}^{S\top} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \\ \boldsymbol{\gamma}_{L1}^{G\top} & \mathbf{0} & \cdots & \boldsymbol{\gamma}_{L1}^{S\top} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\gamma}_{LN}^{G\top} & \mathbf{0} & \cdots & \boldsymbol{\gamma}_{LN}^{S\top} \end{pmatrix} \begin{pmatrix} \mathbf{f}_t^G \\ \mathbf{f}_{1t}^S \\ \vdots \\ \mathbf{f}_{Lt}^S \end{pmatrix} = (\boldsymbol{\Gamma}^G, \boldsymbol{\Gamma}^S) \begin{pmatrix} \mathbf{f}_t^G \\ \mathbf{f}_{1t}^S \\ \vdots \\ \mathbf{f}_{Lt}^S \end{pmatrix}, \tag{2.2}$$

where the definitions of $\boldsymbol{\Gamma}^G$ and $\boldsymbol{\Gamma}^S$ are self evident. It is obvious that at each time period t, the global factor $\mathbf{f}_t^G$ affects all units, possibly to a different degree, whereas $\mathbf{f}_{it}^S$ affects only the units of the i^{th} sector. Notably, by design, $\boldsymbol{\Gamma}^S$ admits a sparse representation, which under some moderate conditions guarantees the following results[1], and is useful to separate two-layer unobservable shocks later on.

$$\|\boldsymbol{\Gamma}^G\|_2 \asymp \|\boldsymbol{\Gamma}^G\| \quad \text{and} \quad \|\boldsymbol{\Gamma}^S\|_2 = o_P(\|\boldsymbol{\Gamma}^S\|).$$

To proceed, it is convenient to rewrite (2.1) in a vector form:

$$\mathbf{Y}_{ij\cdot} = \mathbf{X}_{ij\cdot}\boldsymbol{\beta} + \mathbf{F}^G\boldsymbol{\gamma}_{ij}^G + \mathbf{F}_i^S\boldsymbol{\gamma}_{ij}^S + \mathbf{E}_{ij\cdot}, \tag{2.3}$$

where $\mathbf{E}_{ij\cdot} = (\varepsilon_{ij1}, \ldots, \varepsilon_{ijT})^\top$, and $\mathbf{Y}_{ij\cdot}$, $\mathbf{X}_{ij\cdot}$, $\mathbf{F}^G$, and $\mathbf{F}_i^S$ are defined accordingly. We then define the following objective function:

$$Q(\mathbf{b}, \mathbf{C}) = \sum_{i=1}^{L}\sum_{j=1}^{N}(\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\mathbf{b})^\top \mathbf{M}_{\mathbf{C}_i}(\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\mathbf{b}),$$

[1] See (A.36) of the Appendix for detailed development.

where $\mathbf{C} = (\mathbf{C}_1, \dots, \mathbf{C}_L)$ with each $\mathbf{C}_i$ being a $T \times d_{\max}$ matrix and $d_{\max}$ $(\geq l^G + \max_i l_i^S)$ being a user-specified fixed large integer.

The estimators are intuitively given by

$$(\widehat{\mathbf{b}}, \widehat{\mathbf{C}}) = \operatorname{argmin} Q(\mathbf{b}, \mathbf{C}), \tag{2.4}$$

where $\frac{1}{T}\mathbf{C}_i^\top \mathbf{C}_i = \mathbf{I}_{d_{\max}}$ for each $i \in [L]$ for the purpose of identification. Simple algebra shows that (2.4) admits the following expressions:

$$\widehat{\mathbf{b}} = \left(\sum_{i=1}^{L}\sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i} \mathbf{X}_{ij\cdot}\right)^{-1} \sum_{i=1}^{L}\sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i} \mathbf{Y}_{ij\cdot} \quad \text{and} \quad \widehat{\mathbf{C}}_i \widehat{\mathbf{V}}_i = \widehat{\mathbf{\Sigma}}_i \widehat{\mathbf{C}}_i, \tag{2.5}$$

where $\widehat{\mathbf{\Sigma}}_i = \frac{1}{NT}\sum_{j=1}^{N} (\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}})(\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}})^\top$, and $\widehat{\mathbf{V}}_i$ is a diagonal matrix including the largest $d_{\max}$ eigenvalues of $\widehat{\mathbf{\Sigma}}_i$.

In what follows we derive a lemma (Lemma 2.1), which will not only ensure the consistency of $\widehat{\mathbf{b}}$ under mild conditions, but also yield a rate of convergence.

Assumption 1.

1. *For the regressors and errors, let*

   (a) $\max_{i,j,t} \mathbb{E}\|\mathbf{x}_{ijt}\|^4 < \infty$*;*

   (b) $\max_i \|\mathbf{E}_{i\cdot\cdot}\|_2 = O_P(\sqrt{N} \vee \sqrt{T})$*, where* $\mathbf{E}_{i\cdot\cdot} = (\mathbf{E}_{i1\cdot}, \dots, \mathbf{E}_{iN\cdot})^\top$*;*

   (c) $\|\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{E}_{ij\cdot}\| = O_P(\frac{1}{\sqrt{LNT}})$*.*

2. *For the factors, let*

   (a) $\max_{i,t} \mathbb{E}(\|\mathbf{f}_t^G\|^4 + \|\mathbf{f}_{it}^S\|^4) < \infty$*;*

   (b) $\max_i \|\frac{1}{T}\mathbf{F}_i^\top \mathbf{F}_i - \mathbf{\Sigma}_{\mathbf{F}_i}\| \to_P 0$*, and* $0 < \min_i \lambda_{\min}\{\mathbf{\Sigma}_{\mathbf{F}_i}\} \leq \max_i \lambda_{\max}\{\mathbf{\Sigma}_{\mathbf{F}_i}\} < \infty$*, where* $\mathbf{F}_i = (\mathbf{F}^G, \mathbf{F}_i^S)$*.*

3. *For the loadings, let*

   (a) $\max_{i,j} \mathbb{E}(\|\boldsymbol{\gamma}_{ij}^G\|^4 + \|\boldsymbol{\gamma}_{ij}^S\|^4) < \infty$*;*

   (b) $\max_i \|\frac{1}{N}\mathbf{\Gamma}_{i\cdot}^\top \mathbf{\Gamma}_{i\cdot} - \mathbf{\Sigma}_{\mathbf{\Gamma}_{i\cdot}}\| \to_P 0$ *and* $0 < \min_i \lambda_{\min}\{\mathbf{\Sigma}_{\mathbf{\Gamma}_{i\cdot}}\} \leq \max_i \lambda_{\max}\{\mathbf{\Sigma}_{\mathbf{\Gamma}_{i\cdot}}\} < \infty$*, where* $\mathbf{\Gamma}_{i\cdot} = (\mathbf{\Gamma}_{i\cdot}^G, \mathbf{\Gamma}_{i\cdot}^S)$ *and* $\mathbf{\Gamma}_{i\cdot}^\star = (\mathbf{\Gamma}_{i1}^\star, \dots, \mathbf{\Gamma}_{iN}^\star)^\top$ *with* $\star \in \{G, S\}$*.*

4. *Suppose that* $\inf_{\mathbf{C}} \lambda_{\min}\{\frac{1}{LNT}\sum_{i=1}^{L} \mathbf{D}_i\} > 0$*, where* $\mathbf{D}_i = \mathbf{D}_{i,1} - \mathbf{D}_{i,2}^\top[(\mathbf{\Gamma}_{i\cdot}^\top \mathbf{\Gamma}_{i\cdot}) \otimes \mathbf{I}_T]^{-1}\mathbf{D}_{i,2}$ *with* $\mathbf{D}_{i,1} = \sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\mathbf{C}_i} \mathbf{X}_{ij\cdot}$ *and* $\mathbf{D}_{i,2} = \sum_{j=1}^{N} \boldsymbol{\gamma}_{ij} \otimes (\mathbf{M}_{\mathbf{C}_i} \mathbf{X}_{ij\cdot})$*.*

Assumption 1.1 requires the fourth moment of the regressors to be bounded, which is standard. The rest of Assumption 1.1 can be verified in the case where ε_{ijt} satisfies certain mixing conditions as in Assumption 2.1 below. For the time being, our focus is consistency only, so we do not impose more detailed structure. Assumptions 1.2-1.3 are pretty standard in the panel data literature, and therefore we do not discuss them here. Assumption 1.4 is an identification condition, which is similar to Assumption A of Bai (2009). Because of the additional dimension, we need to further average across i.

Lemma 2.1. *Under Assumption 1, as $(L, N, T) \to (\infty, \infty, \infty)$,*

1. $\widehat{\mathbf{b}} - \boldsymbol{\beta} = O_P(\frac{1}{\sqrt{T}} \vee \frac{1}{\sqrt{N}})$;
2. $\|\mathbf{P}_{\widehat{\mathbf{C}}_i} - \mathbf{P}_{\mathbf{F}_i}\|_2 = O_P(\|\widehat{\mathbf{b}} - \boldsymbol{\beta}\| + \frac{1}{\sqrt{N \wedge T}})$, *for each i.*

Without much information on the unknown factors, Lemma 2.1.(1) does not only achieve consistency for $\widehat{\mathbf{b}}$, but also provides a preliminary rate of convergence. The second result of Lemma 2.1 indicates that we can recover the space spanned by $\mathbf{F}_i$. However, this result does not enable us to separate the global and sector shocks. We will turn to the decomposition of a two-layer factor structure in Section 2.4. There are three questions that remain unresolved: (i) Asymptotic distribution; (ii) Valid inference procedure provided that ε_{ijt} has weak correlation along three dimensions; and (iii) Decomposition of the factor structure. We will deal with the three questions in the subsequent subsections.

2.2 Asymptotic Properties

In this subsection, we establish an asymptotic distribution for $\widehat{\mathbf{b}} - \boldsymbol{\beta}$ under the assumption that the number of global factors and the number of factors in different sectors (i.e., l^G and l_i^S's) are given. In this case the condition under (2.4) should be written as $\frac{1}{T}\mathbf{C}_i^\top \mathbf{C}_i = \mathbf{I}_{l^G + l_i^S}$ instead.

To proceed, we impose more structures, which are summarized in the following assumption.

Assumption 2.

1. *Define $\mathbf{E}_{\cdot\cdot t} = (\varepsilon_{11t}, \ldots, \varepsilon_{1Nt}, \ldots, \varepsilon_{L1t}, \ldots, \varepsilon_{LNt})^\top$. Suppose that the error terms are independent of the other variables. Let further $\{\mathbf{E}_{\cdot\cdot t} \,|\, t \in [T]\}$ be strictly stationary and α-mixing such that $\mathbb{E}[\mathbf{E}_{\cdot\cdot t}] = \mathbf{0}$, for some $\kappa > 0$, where the mixing coefficients satisfy $\sum_{t=1}^{\infty} [\alpha(t)]^{\kappa/(2+\kappa)} < \infty$ and $\max_{i,j} \mathbb{E}|\varepsilon_{ij1}|^{4+\kappa} < \infty$. Additionally, let*

   (a) $\sum_{i_1,i_2=1}^{L} \sum_{j_1,j_2=1}^{N} |\mathbb{E}[\varepsilon_{i_1 j_1 1} \varepsilon_{i_2 j_2 1}]| = O(LN)$;

(b) $\sum_{i_1,i_2=1}^{L}\sum_{j_1,j_2=1}^{N}\sum_{t_1,t_2=1}^{T}|\mathbb{E}[\varepsilon_{i_1j_1t_1}\varepsilon_{i_2j_2t_2}]| = O(LNT)$;

(c) $\sum_{i=1}^{L}\sum_{j_1,j_2,j_3,j_4=1}^{N}\sum_{t_1,t_2=1}^{T}|\mathrm{cov}(\varepsilon_{ij_1t_1}\varepsilon_{ij_2t_1},\varepsilon_{ij_3t_2}\varepsilon_{ij_4t_2})| = O(LN^2T)$;

(d) $\sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{t_1,t_2,t_3,t_4=1}^{T}|\mathrm{cov}(\varepsilon_{ijt_1}\varepsilon_{ijt_2},\varepsilon_{ijt_3}\varepsilon_{ijt_4})| = O(LNT^2)$;

(e) $\max_i\sum_{j_1,j_2=1}^{N}|E[\varepsilon_{ij_11}\varepsilon_{ij_2,1}]| = O(N)$;

(f) $\max_i\sum_{j=1}^{N}\sum_{t=1}^{\infty}|E[\varepsilon_{ij,1+t}\varepsilon_{ij1}]| = O(N)$.

2. *As* $(L,N,T)\to(\infty,\infty,\infty)$, *suppose that*

(a) $\frac{LN}{T^2}\to 0$, $\frac{LT}{N^2}\to 0$, *and* $\frac{L}{T\wedge N}\to 0$;

(b) $\frac{1}{\sqrt{LNT}}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{E}_{ij.}\to_D N(\mathbf{0},\mathbf{\Omega}_1)$, *where* $\mathbf{Z}_{ij.}$ *is defined as* $\mathbf{Z}_{ij.} = \mathbf{X}_{ij.}-\sum_{k=1}^{N}\mathbf{X}_{ik.}\boldsymbol{\gamma}_{ij}^\top(\mathbf{\Gamma}_{i.}^\top\mathbf{\Gamma}_{i.})^{-1}\boldsymbol{\gamma}_{ik}$;

(c) $\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{Z}_{ij.} = \mathbf{\Omega}_2+O_P(\frac{1}{\sqrt{N\wedge T}})$, *where* $\mathbf{\Omega}_2$ *is nonsingular.*

Assumption 2.1 imposes some specific restrictions on the error components. Specifically, it requires ε_{ijt} to behave as a stationary time series across t, while it allows ε_{ijt} to have weak cross-sectional dependence over both i and j. Based on the hierarchical tree structure considered in this paper, Assumption 2.2.a regulates (L,N,T), and basically requires each sector (indexed by i) to have a sufficiently large number of individuals (indexed by j) and time periods of observations (indexed by t), which can easily be realised in practice (e.g., the empirical study to be studied in Section 5). Suppose that $N\asymp T$, Assumption 2.2.a reduces to $\frac{L}{T\wedge N}\to 0$, which does not lose too much generality given the hierarchy design. Assumption 2.2.b assumes an asymptotic normality, which is quite standard in the relevant literature.

Lemma 2.2. *Let Assumptions 1-2 hold. As* $(L,N,T)\to(\infty,\infty,\infty)$,

$$\sqrt{LNT}\left(\widehat{\mathbf{b}}-\boldsymbol{\beta}-\frac{1}{N}\mathbf{a}_1-\frac{1}{T}\mathbf{a}_2\right)\to_D N(\mathbf{0},\mathbf{\Omega}_2^{-1}\mathbf{\Omega}_1\mathbf{\Omega}_2^{-1}),$$

where $\mathbf{a}_1$ *and* $\mathbf{a}_2$ *are defined as follows:*

$$\mathbf{a}_1 = -\mathbf{\Omega}_2^{-1}\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{\mathbf{Z}_{ij_1.}^\top\mathbf{F}_i}{T}\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i.}^\top\mathbf{\Gamma}_{i.}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij_2}\sigma_{i,j_1j_2},$$
$$\mathbf{a}_2 = -\mathbf{\Omega}_2^{-1}\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij.}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,i}\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i.}^\top\mathbf{\Gamma}_{i.}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij},$$

with $\sigma_{i,j_1j_2} = E[\varepsilon_{ij_1t}\varepsilon_{ij_2t}]$ *and* $\mathbf{\Sigma}_{\varepsilon,i} = N^{-1}\sum_{j=1}^{N}E[\mathbf{E}_{ij.}\mathbf{E}_{ij.}^\top]$.

Lemma 2.2 establishes the asymptotic distribution of $\widehat{\mathbf{b}}-\boldsymbol{\beta}$. It is noteworthy that $\widehat{\mathbf{b}}$ is asymptotically biased due to the fact that the bias terms $\mathbf{a}_1$ and $\mathbf{a}_2$ are not asymptotically negligible. Moreover, the two biases are actually dominating, bias correction therefore has to be conducted.

In what follows, we develop a method for bias correction, which has a similar spirit to these analytically bias-corrected estimators for two dimensional panel data model discussed by Chen et al. (2021). Given the extra dimension, the two types of CSD and time series autocorrelation involved, the development is more mathematically involved in this paper. With the the estimators $\widehat{\mathbf{C}}_i$ and $\widehat{\mathbf{b}}$, we can obtain the following loading estimators:

$$\widehat{\boldsymbol{\gamma}}_{ij}=\frac{1}{T}\widehat{\mathbf{C}}_i^\top(\mathbf{Y}_{ij\cdot}-\mathbf{X}_{ij\cdot}\widehat{\mathbf{b}})\quad\text{and}\quad\widehat{\boldsymbol{\Gamma}}_{i\cdot}=(\widehat{\boldsymbol{\gamma}}_{i1},\cdots,\widehat{\boldsymbol{\gamma}}_{iN})^\top.$$

Then, $\mathbf{Z}_{ij\cdot}$, $\boldsymbol{\Omega}_2$ and σ_{i,j_1j_2} in the bias term $\mathbf{a}_1$ can be estimated by

$$\widehat{\mathbf{Z}}_{ij\cdot}=\mathbf{X}_{ij\cdot}-\sum_{k=1}^{N}\mathbf{X}_{ik\cdot}\widehat{\boldsymbol{\gamma}}_{ij}^\top(\widehat{\boldsymbol{\Gamma}}_{i\cdot}^\top\widehat{\boldsymbol{\Gamma}}_{i\cdot})^{-1}\widehat{\boldsymbol{\gamma}}_{ik},\quad\widehat{\boldsymbol{\Omega}}_2=\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\widehat{\mathbf{Z}}_{ij\cdot},$$

and

$$\widehat{\sigma}_{i,j_1j_2}=\frac{1}{T}(\mathbf{Y}_{ij_1\cdot}-\mathbf{X}_{ij_1\cdot}\widehat{\mathbf{b}}-\widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_1})^\top(\mathbf{Y}_{ij_2\cdot}-\mathbf{X}_{ij_2\cdot}\widehat{\mathbf{b}}-\widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_2}),$$

respectively. Additionally, we can construct an estimator of $\boldsymbol{\Sigma}_{\varepsilon,i}$: $\widehat{\boldsymbol{\Sigma}}_{\varepsilon,i}$, with its t,s-th element being $\widehat{\sigma}_{\varepsilon,i,ts}=N^{-1}(\mathbf{Y}_{i\cdot t}-\mathbf{X}_{i\cdot t}\widehat{\mathbf{b}}-\widehat{\boldsymbol{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{it})^\top(\mathbf{Y}_{i\cdot s}-\mathbf{X}_{i\cdot s}\widehat{\mathbf{b}}-\widehat{\boldsymbol{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{is})$, where $\widehat{\mathbf{c}}_{it}$ is the t-th column of $\widehat{\mathbf{C}}_i^\top$. Sequentially, we define the following feasible estimators for $\mathbf{a}_1$ and $\mathbf{a}_2$ by replacing each element in them with its estimator:

$$\widehat{\mathbf{a}}_1=-\widehat{\boldsymbol{\Omega}}_2^{-1}\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{\widehat{\mathbf{Z}}_{ij_1\cdot}^\top\widehat{\mathbf{C}}_i}{T}\left(\frac{\widehat{\boldsymbol{\Gamma}}_{i\cdot}^\top\widehat{\boldsymbol{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widehat{\boldsymbol{\gamma}}_{ij_2}\widehat{\sigma}_{i,j_1j_2},$$

$$\widehat{\mathbf{a}}_2=-\widehat{\boldsymbol{\Omega}}_2^{-1}\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\widehat{\boldsymbol{\Sigma}}_{\varepsilon,i}\widehat{\mathbf{C}}_i\left(\frac{\widehat{\boldsymbol{\Gamma}}_{i\cdot}^\top\widehat{\boldsymbol{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widehat{\boldsymbol{\gamma}}_{ij}.$$

With $\widehat{\mathbf{a}}_1$ and $\widehat{\mathbf{a}}_2$, we are ready to propose a bias corrected estimator of the form:

$$\widehat{\mathbf{b}}_{bc}=\widehat{\mathbf{b}}-\frac{1}{N}\widehat{\mathbf{a}}_1-\frac{1}{T}\widehat{\mathbf{a}}_2,$$

which is asymptotically unbiased.

We now establish the asymptotic distribution of $\widehat{\mathbf{b}}_{bc}$ as the first main result of this paper.

Theorem 2.1. *Let Assumptions 1 and 2 hold. As* $(L, N, T) \to (\infty, \infty, \infty)$,

$$\sqrt{LNT}\,(\widehat{\mathbf{b}}_{bc} - \boldsymbol{\beta}) \to_D N(\mathbf{0}, \boldsymbol{\Omega}_2^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Omega}_2^{-1}).$$

To obtain valid inference in practice, it requires the feasibility and availability of $\boldsymbol{\Omega}_2^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Omega}_2^{-1}$, which however is not really straightforward in the presence of weak correlation over three dimensions of ε_{ijt}. In the following subsection, we tackle this issue by developing a bootstrap procedure.

2.3 Bootstrap Inference

We now derive a valid confidence interval for $N(\mathbf{0}, \boldsymbol{\Omega}_2^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Omega}_2^{-1})$ by proposing a dependent wild bootstrap (DWB) method.

Before proceeding further, it is worth briefly reviewing the relevant literature. The DWB method is initially introduced in Shao (2010), wherein a comprehensive comparison between the DWB and some existing bootstrap methods can be found. The moving block bootstrap (MBB) approach for the fixed effects panel data model of Gonçalves (2011) shares a similar motivation to what we will propose below. However, resampling the blocks along the time dimension can destroy the unobservable factor structure. As a consequence, it is not clear to us whether such a procedure can be applied when interactive fixed effects are involved.

We are now ready to proceed. The detailed steps are as follows.

1. Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_T)^\top$ be an ℓ-dependent time series for each bootstrap draw, and let $\boldsymbol{\xi}$ satisfy that

$$\mathbb{E}[\xi_t] = 0, \quad \mathbb{E}|\xi_t|^2 = 1, \quad \mathbb{E}|\xi_t|^4 < \infty, \quad \mathbb{E}[\xi_t\xi_s] = a\left(\frac{t-s}{\ell}\right),$$

   where $\ell \to \infty$ and $\frac{\ell}{\sqrt{T}} \to 0$, and $a(\cdot)$ is a symmetric kernel defined on $[-1, 1]$ satisfying that $a(0) = 1$ and $K_a(x) = \int_{\mathbb{R}} a(u)e^{-iux}du \geq 0$ for $x \in \mathbb{R}$.

2. Construct a new set of dependent variables by $\mathbf{Y}_{ij\cdot}^* = \widehat{\mathbf{Z}}_{ij\cdot}\widehat{\mathbf{b}}_{bc} + \widehat{\mathbf{U}}_{ij\cdot} \circ \boldsymbol{\xi}$, where $\circ$ defines the Hadamard product, and $\widehat{\mathbf{U}}_{ij\cdot} = (\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}}_{bc} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij})$. Accordingly, the new estimator of $\boldsymbol{\beta}$ is given by

$$\widehat{\mathbf{b}}^* = \left(\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\widehat{\mathbf{Z}}_{ij\cdot}\right)^{-1}\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{Y}_{ij\cdot}^*.$$

3. We repeat the above procedure $\mathcal{L}$ times.

The condition $K_a(x) = \int_{\mathbb{R}} a(u)e^{-iux}du \geq 0$ for $x \in \mathbb{R}$ essentially regulates the kernel function, which, together with other restrictions imposed on $a(\cdot)$, is satisfied by a few commonly used kernels, such as the Bartlett and Parzen kernels. More choices of the kernel function can be found in Andrews (1991) and Shao (2010) for example. In practice, we may use Bartlett kernel for simplicity.

When generating $\boldsymbol{\xi}$, one can always use a Normal distribution like $N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} = \{a\left(\frac{t-s}{\ell}\right)\}_{T\times T}$. Although normality is not really necessary in theory, it fulfils every aforementioned requirement.

For the above bootstrap procedure, we establish the asymptotic consistency in the following theorem.

Theorem 2.2. *Suppose that* $\mathbb{E}\|\frac{1}{\sqrt{LN}}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{z}_{ijt}\varepsilon_{ijt}\|^4 < \infty$, *where* $\mathbf{z}_{ijt}^\top$ *stands for the* t^{th} *row of* $\mathbf{Z}_{ij\cdot}$. *Under Assumptions 1 and 2, as* $(L, N, T) \to (\infty, \infty, \infty)$,

$$\sup_{\mathbf{w}}\left|\Pr{}^*(\sqrt{LNT}\,(\widehat{\mathbf{b}}^* - \widehat{\mathbf{b}}_{bc}) \leq \mathbf{w}) - \Pr(\sqrt{LNT}\,(\widehat{\mathbf{b}}_{bc} - \boldsymbol{\beta}) \leq \mathbf{w})\right| = o_P(1).$$

Up to this point, we have established the asymptotic distribution associated with (2.4), and have developed a bootstrap procedure to yield valid inference practically. We shall further examine our theoretical findings in the simulations of Section 4.

Note that the above results are obtained assuming l^G and l_i^S's are known. Therefore, it is imperative to develop a data-driven method which is capable of separating the two types of factors in the hierarchical factor structure and of estimating the number of factors of each type. This is exactly what we do in the following subsection.

2.4 Decomposition of the Factor Structure

To close our theoretical investigation, we estimate the numbers of factors (i.e., l^G and l_i^S's) in this subsection. Recall that in (2.2), the global factors have an impact on each individual unit over i and j, while the sector factors can only affect a subset of the cross-sectional units. From a signal-noise ratio point of view, it is convenient to identify the number of global factors (i.e., l^G) first. In addition, we will also utilize the rate of convergence achieved in Lemma 2.1, which does not require any prior knowledge on the numbers of factors. In other words, $\widehat{\mathbf{b}}$ used below is from (2.5).

For the global factors, we define the following covariance matrix:

$$\widehat{\boldsymbol{\Sigma}}^G = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}})(\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}})^\top, \tag{2.6}$$

and estimate the number of global factors, l^G, as follows:

$$\widehat{\ell}^G = \operatorname*{argmin}_{0\le \ell \le d_{\max}} \left\{ \frac{\widehat{\lambda}^G_{\ell+1}}{\widehat{\lambda}^G_{\ell}} \cdot \mathbb{I}(\widehat{\lambda}^G_{\ell} \ge \omega) + \mathbb{I}(\widehat{\lambda}^G_{\ell} < \omega) \right\}, \tag{2.7}$$

where $\omega = [\log(N \vee T)]^{-1}$, $\widehat{\lambda}^G_0 = 1$ is a mock eigenvalue, and $\widehat{\lambda}^G_\ell$ stands for the ℓ^{th} largest eigenvalue of $\widehat{\mathbf{\Sigma}}^G$.

Sequentially, we define the following covariance matrix for each sector i.

$$\widehat{\mathbf{\Sigma}}^S_i = \frac{1}{NT} \sum_{j=1}^{N} \mathbf{M}_{\widehat{\mathbf{C}}^G} (\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}})(\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}})^\top \mathbf{M}_{\widehat{\mathbf{C}}^G}, \tag{2.8}$$

where $\widehat{\mathbf{C}}^G$ represents the eigenvectors associated with the largest $\widehat{\ell}^G$ eigenvalues of $\widehat{\mathbf{\Sigma}}^G$, and $\frac{1}{T}\widehat{\mathbf{C}}^{G\top}\widehat{\mathbf{C}}^G = \mathbf{I}_{\widehat{\ell}^G}$. The construction of $\widehat{\mathbf{\Sigma}}^S_i$ allows us to identify the sector factors block by block. Accordingly, the numbers of sector factors can be estimated as follows:[2]

$$\widehat{\ell}^S_i = \operatorname*{argmin}_{\ell^S_i} \Big\{ \frac{\widehat{\lambda}^S_{i,\ell^S_i+1}}{\widehat{\lambda}^S_{i,\ell^S_i}} \cdot \mathbb{I}(\widehat{\lambda}^S_{i,\ell^S_i} \ge \omega) + \mathbb{I}(\widehat{\lambda}^S_{i,\ell^S_i} < \omega) \Big\}, \tag{2.9}$$

where $0 \le \ell^S_i \le d_{\max}$ for $\forall i \in [L]$, $\widehat{\lambda}^S_{i,0} \equiv 1$ is a mock eigenvalue, and $\widehat{\lambda}^S_{i,l^S_i}$ stands for the l^S_i-th largest eigenvalue of $\widehat{\mathbf{\Sigma}}^S_i$.

The estimators in (2.7) and (2.9) can be considered as extensions of Lam and Yao (2012), and Ahn and Horenstein (2013). However, as pointed out by Lam and Yao (2012), it remains unresolved to bound the ratio associated with the eigenvalues which converge to 0 from below. To bypass this unresolved issue, we introduce a tuning parameter ω. The idea behind involving the turning parameter is that although it is challenging to study a ratio with a denominator converging to 0, we can discard this ratio and construct a U-shape curve by employing the indicator function in (2.7) and (2.9), respectively.

We still need to make one more important assumption to separate the global and sector factors.

Assumption 3.

1. *Let* $\max_{i,j} \|\boldsymbol{\gamma}^S_{ij}\|^2 = O_P(\log(LN))$, $\|\mathbf{F}^S\|_2 = O_P(\sqrt{T} \vee \sqrt{L})$, *where* $\mathbf{F}^S = (\mathbf{F}^S_1, \ldots, \mathbf{F}^S_L)$.

[2]Lin (2023) considers the factor selection when L is fixed, and requires orthogonality between global and sector factors in his Assumption G. However, his estimation on the number of factors does not include any treatment for the eigenvalue ratio when both the indices of the numerator and denominator are greater than the true one. As a result, it is unclear whether his estimator yields a U-shape as claimed.

2. $\max_i \frac{1}{T}\|\mathbf{F}_i^{S\top}\mathbf{F}^G\|_2 = O_P(T^\nu)$ *with* $-1/2 \leq \nu < 0$.

Note that to separate global and sector factors, Assumption 3 requires much less restrictions than Assumption D of Choi et al. (2023). The conditions $\|\mathbf{F}^S\|_2 = O_P(\sqrt{T} \vee \sqrt{L})$ and $\max_i \frac{1}{T}\|\mathbf{F}_i^{S\top}\mathbf{F}^G\|_2 = O_P(T^\nu)$ can easily be fulfilled if $\mathbb{E}[\mathbf{f}_{it}^S] = \mathbf{0}$ over both i and t. The validity of these conditions can be justified as follows. For simplicity, we let $l^G = 1$, and $f_t^G \equiv 1$, and assume that $\mathbf{f}_{it}^S = \mathbf{f}^S + \boldsymbol{\xi}_{it}^S$, where $\mathbb{E}[\boldsymbol{\xi}_{it}^S] = \mathbf{0}$. Then we can write

$$\gamma_{ij}^G f_t + \boldsymbol{\gamma}_{ij}^{S\top}\mathbf{f}_{it}^S = \gamma_{ij}^{G*} + \boldsymbol{\gamma}_{ij}^{S\top}\boldsymbol{\xi}_{it}^S, \tag{2.10}$$

where $\gamma_{ij}^{G*} = \gamma_{ij}^G + \boldsymbol{\gamma}_{ij}^{S\top}\mathbf{f}^S$. Equation (2.10) suggests that when studying a hierarchical factor structure, for the purpose of identification at most one layer is allowed to have non-zero mean. Therefore, it is reasonable to assume that $\mathbb{E}[\mathbf{f}_{it}^S] = \mathbf{0}$, which is implicitly covered by Assumption 3.

Theorem 2.3. *Let Assumptions 1, 2.1, and 3 hold. As* $(L, N, T) \to (\infty, \infty, \infty)$, *we have* $\Pr(\widehat{\ell}^G = l^G, \widehat{\boldsymbol{\ell}}^S = \mathbf{l}^S) \to 1$, *where* $\widehat{\boldsymbol{\ell}}^S = (\widehat{\ell}_1^S, \ldots, \widehat{\ell}_L^S)$ *and* $\mathbf{l}^S = (l_1^S, \ldots, l_L^S)$.

This theorem implies that we can separate the global factors and the sector factors in (2.1). In addition, Theorem 2.3 implies that we can identify l^G and l_i^S's jointly although a sequential estimation procedure is proposed above. Finally, it is worth stressing again that the cases with 0 factor are allowed, and that Assumption 2.2 is not required for Theorem 2.3 to hold.

2.5 On Unbalanced Data

In practice, it is not guaranteed that each sector has the same number of individuals, so it brings certain unbalanceness to the data set. Formally, $\forall i \in [L]$, we may have $j \in [N_i]$, which is exactly the same as our empirical study of Section 5.

In this case we let $\mathbb{N} = \sum_{i=1}^L N_i$. The estimators can be modified accordingly:

$$\widehat{\mathbf{b}} = \left(\sum_{i=1}^L \sum_{j=1}^{N_i} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{X}_{ij\cdot}\right)^{-1} \sum_{i=1}^L \sum_{j=1}^{N_i} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{Y}_{ij\cdot} \quad \text{and} \quad \widehat{\mathbf{C}}_i\widehat{\mathbf{V}}_i = \widehat{\boldsymbol{\Sigma}}_i\widehat{\mathbf{C}}_i, \tag{2.11}$$

where $\widehat{\boldsymbol{\Sigma}}_i = \frac{1}{N_i T}\sum_{j=1}^{N_i}(\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}})(\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}})^\top$, and $\widehat{\mathbf{V}}_i$ is a diagonal matrix including the largest d_{max} eigenvalues of $\widehat{\boldsymbol{\Sigma}}_i$. Then, the bias correction estimator can be redefined as

$$\widehat{\mathbf{b}}_{bc} = \widehat{\mathbf{b}} - \frac{L}{\mathbb{N}}\widehat{\mathbf{a}}_1 - \frac{1}{T}\widehat{\mathbf{a}}_2,$$

where

$$\widehat{\mathbf{a}}_1 = -\widehat{\mathbf{\Omega}}_2^{-1} \frac{1}{L} \sum_{i=1}^{L} \frac{1}{N_i} \sum_{j_1,j_2=1}^{N_i} \frac{\widehat{\mathbf{Z}}_{ij_1\cdot}^\top \widehat{\mathbf{C}}_i}{T} \left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top \widehat{\mathbf{\Gamma}}_{i\cdot}}{N_i} \right)^{-1} \widehat{\gamma}_{ij_2} \widehat{\sigma}_{i,j_1 j_2},$$

$$\widehat{\mathbf{a}}_2 = -\widehat{\mathbf{\Omega}}_2^{-1} \frac{1}{LT} \sum_{i=1}^{L} \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i} \widehat{\mathbf{\Sigma}}_{\varepsilon,i} \widehat{\mathbf{C}}_i \left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top \widehat{\mathbf{\Gamma}}_{i\cdot}}{N_i} \right)^{-1} \widehat{\gamma}_{ij}.$$

After carefully checking the proofs, we can see that as long as $\min_i N_i \to \infty$ and $\frac{L^2}{\mathbb{N}} \to 0$, all the aforementioned results are still valid with trivial modifications. We will conduct simulations below to support this point.

3 A Model with Heterogeneous Coefficients

This section studies the following 3D panel data model with heterogeneous coefficients:

$$y_{ijt} = \mathbf{x}_{ijt}^\top \boldsymbol{\beta}_i + \boldsymbol{\gamma}_{ij}^{G\top} \mathbf{f}_t^G + \boldsymbol{\gamma}_{ij}^{S\top} \mathbf{f}_{it}^S + \varepsilon_{ijt}, \tag{3.1}$$

where $\boldsymbol{\beta}_i$ is a $d \times 1$ vector of sector-specific coefficients. In what follows, we consider the estimation and inference of $\boldsymbol{\beta}_i$. Specifically, the estimation can be achieved by

$$(\widehat{\mathbf{b}}_i, \widetilde{\mathbf{C}}_i) = \operatorname{argmin} Q_i(\mathbf{b}_i, \mathbf{C}_i), \text{ subject to } T^{-1}\mathbf{C}_i^\top \mathbf{C}_i = \mathbf{I}_{d_{\max}}, \tag{3.2}$$

where

$$Q_i(\mathbf{b}_i, \mathbf{C}_i) = \sum_{j=1}^{N} (\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\mathbf{b}_i)^\top \mathbf{M}_{\mathbf{C}_i} (\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\mathbf{b}_i).$$

for each $i \in [L]$. Simple algebra gives expressions for the theoretical solution of (3.2):

$$\widehat{\mathbf{b}}_i = \left(\sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widetilde{\mathbf{C}}_i} \mathbf{X}_{ij\cdot} \right)^{-1} \sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widetilde{\mathbf{C}}_i} \mathbf{Y}_{ij\cdot} \quad \text{and} \quad \widetilde{\mathbf{C}}_i \widetilde{\mathbf{V}}_i = \widetilde{\mathbf{\Sigma}}_i \widetilde{\mathbf{C}}_i,$$

where $\widetilde{\mathbf{\Sigma}}_i$ and $\widetilde{\mathbf{V}}_i$ are defined in the same manner as $\widehat{\mathbf{\Sigma}}_i$ and $\widehat{\mathbf{V}}_i$ in Section 2. For each $i \in [L]$, the asymptotic properties of $\widehat{\mathbf{b}}_i$ are identical to those of Bai (2009). Therefore, we do not pursue the results along this line. Instead, we aim to develop asymptotic properties for the mean group (MG) estimator:

$$\widehat{\mathbf{b}}_{mg} = L^{-1} \sum_{i=1}^{L} \widehat{\mathbf{b}}_i.$$

Accordingly, we define the average value of the true coefficients as: $\overline{\boldsymbol{\beta}} = L^{-1}\sum_{i=1}^{L}\boldsymbol{\beta}_i$. Before we establish the asymptotic distribution of $\widehat{\mathbf{b}}_{mg}$, we first introduce some additional conditions corresponding to those listed in Assumptions 1.1 and 1.4 as well as 2.2(b)(c).

Assumption 4.

1. $\min_i \inf_{\mathbf{C}_i} \lambda_{\min}\{\frac{1}{NT}\mathbf{D}_i\} > 0$, *where* $\mathbf{D}_i$ *is defined in Assumption 1.*

2. $\max_{i,j}\|\frac{1}{\sqrt{T}}\mathbf{X}_{ij.}\| = O_P(\log(LN))$, $\max_{i,j}\|\boldsymbol{\gamma}_{ij}\| = O_P(\log(LN))$ *and* $\max_i \|\frac{1}{\sqrt{T}}\mathbf{F}_i\| = O_P(\log L)$.

3. $\max_i \|\frac{1}{NT}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^{\top}\mathbf{M}_{\mathbf{F}_i}\mathbf{Z}_{ij.} - \boldsymbol{\Omega}_{2,i}\| = O_P(\frac{\log N}{\sqrt{N\wedge T}})$ *and* $0 < \min_i \lambda_{\min}\{\boldsymbol{\Omega}_{2,i}\} \leq \max_i \lambda_{\max}\{\boldsymbol{\Omega}_{2,i}\} < \infty$.

4. $\frac{1}{\sqrt{LNT}}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^{\top}\mathbf{M}_{\mathbf{F}_i}\mathbf{E}_{ij.} \to_D N(\mathbf{0}, \boldsymbol{\Omega}_{mg})$, *where* $\mathbf{Z}_{ij.}$ *is defined in Assumption 2.*

The asymptotic distribution of $\widehat{\mathbf{b}}_{mg}$ is then studied in the following lemma.

Lemma 3.1. *Let Assumptions 1, 2 and 4 hold. As* $(L, N, T) \to (\infty, \infty, \infty)$,

$$\sqrt{LNT}\left(\widehat{\mathbf{b}}_{mg} - \overline{\boldsymbol{\beta}} - \frac{1}{N}\mathbf{a}_{1,mg} - \frac{1}{T}\mathbf{a}_{2,mg}\right) \to_D N(\mathbf{0}, \boldsymbol{\Omega}_{mg}),$$

where the bias terms $\mathbf{a}_{1,mg}$ *and* $\mathbf{a}_{2,mg}$ *are defined as follows:*

$$\mathbf{a}_{1,mg} = -\frac{1}{LN}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\sum_{j_1,j_2=1}^{N}\frac{\mathbf{Z}_{ij_1.}^{\top}\mathbf{F}_i}{T}\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i.}^{\top}\boldsymbol{\Gamma}_{i.}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij_2}\sigma_{i,j_1j_2},$$

$$\mathbf{a}_{2,mg} = -\frac{1}{LNT}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{X}_{ij.}^{\top}\mathbf{M}_{\mathbf{F}_i}\boldsymbol{\Sigma}_{\varepsilon,i}\mathbf{F}_i\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i.}^{\top}\boldsymbol{\Gamma}_{i.}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij},$$

with σ_{i,j_1j_2} *and* $\boldsymbol{\Sigma}_{\varepsilon,i}$ *having been defined in Section 2.*

As an extension of Lemma 2.2, Lemma 3.1 establishes the asymptotic distribution of the mean group estimator, which is robust to the heterogeneity in regression coefficients. Similar to the homogeneous case, the mean group estimator also suffers from two bias terms arising from the cross-sectional and temporal dependence of random errors. To obtain an asymptotically unbiased estimator, we use the analytical method to conduct bias correction. Specifically, we can construct the bias corrected mean group estimator as follows:

$$\widehat{\mathbf{b}}_{bcmg} = \widehat{\mathbf{b}}_{mg} - \frac{1}{N}\widehat{\mathbf{a}}_{1,mg} - \frac{1}{T}\widehat{\mathbf{a}}_{2,mg}, \tag{3.3}$$

where $\widehat{\mathbf{a}}_{1,mg} = \frac{1}{L}\sum_{i=1}^{L}\widehat{\mathbf{a}}_{1,i}$ and $\widehat{\mathbf{a}}_{2,mg} = \frac{1}{L}\sum_{i=1}^{L}\widehat{\mathbf{a}}_{2,i}$ with

$$\widehat{\mathbf{a}}_{i,1} = -\frac{1}{N}\widetilde{\mathbf{\Omega}}_{2,i}^{-1}\sum_{j_1,j_2=1}^{N}\frac{\widetilde{\mathbf{Z}}_{ij_1\cdot}^{\top}\widetilde{\mathbf{C}}_i}{T}\left(\frac{\widetilde{\mathbf{\Gamma}}_{i\cdot}^{\top}\widetilde{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widetilde{\gamma}_{ij_2}\widetilde{\sigma}_{i,j_1j_2},$$

$$\widehat{\mathbf{a}}_{i,2} = -\frac{1}{NT}\widetilde{\mathbf{\Omega}}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^{\top}\mathbf{M}_{\widehat{\mathbf{C}}_i}\widetilde{\mathbf{\Sigma}}_{\varepsilon,i}\widetilde{\mathbf{C}}_i\left(\frac{\widetilde{\mathbf{\Gamma}}_{i\cdot}^{\top}\widetilde{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widetilde{\gamma}_{ij}, \tag{3.4}$$

and $\widetilde{\mathbf{\Omega}}_{2,i} = \frac{1}{NT}\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\cdot}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\widetilde{\mathbf{Z}}_{ij\cdot}$. All other estimators are constructed in an analogous way with those for the homogeneous model.

We establish the asymptotic distribution of the bias corrected mean group estimator in the following theorem.

Theorem 3.1. *Let Assumptions 1, 2 and 4 hold. As* $(L, N, T) \to (\infty, \infty, \infty)$,

$$\sqrt{LNT}\,(\widehat{\mathbf{b}}_{bcmg} - \overline{\boldsymbol{\beta}}) \to_D N(\mathbf{0}, \mathbf{\Omega}_{mg}).$$

For the convenience of inference, we modify the bootstrap approach in Section 2 to accommodate the heterogeneity of the coefficients. Before we present the details, we first introduce the sector-specific bias corrected estimator that will used in the bootstrap procedure: $\widehat{\mathbf{b}}_{i,bc} = \widehat{\mathbf{b}}_i - \frac{1}{N}\widehat{\mathbf{a}}_{i,1} - \frac{1}{T}\widehat{\mathbf{a}}_{i,2}$. Then, it is straightforward to see that $\widehat{\mathbf{b}}_{bcmg} = N^{-1}\sum_{i=1}^{L}\widehat{\mathbf{b}}_{i,bc}$. The bootstrap procedure for inference can be then summarized into the following steps:

1. Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_T)^{\top}$ be an ℓ-dependent time series for each bootstrap draw, and let $\boldsymbol{\xi}$ satisfy that

$$\mathbb{E}[\xi_t] = 0, \quad \mathbb{E}|\xi_t|^2 = 1, \quad \mathbb{E}|\xi_t|^4 < \infty, \quad \mathbb{E}[\xi_t\xi_s] = a\left(\frac{t-s}{\ell}\right),$$

   where $\ell \to \infty$ and $\frac{\ell}{\sqrt{T}} \to 0$, and $a(\cdot)$ is a symmetric kernel defined on $[-1, 1]$ satisfying that $a(0) = 1$ and $K_a(x) = \int_{\mathbb{R}} a(u)e^{-iux}du \geq 0$ for $x \in \mathbb{R}$.

2. Construct a new set of dependent variables by $\mathbf{Y}_{ij\cdot}^{*} = \widetilde{\mathbf{Z}}_{ij\cdot}\widehat{\mathbf{b}}_{i,bc} + \widetilde{\mathbf{U}}_{ij\cdot} \circ \boldsymbol{\xi}$, where $\widetilde{\mathbf{U}}_{ij\cdot} = (\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}}_{i,bc} - \widetilde{\mathbf{C}}_i\widetilde{\gamma}_{ij})$. Accordingly, the new mean group estimator can be obtained using $\widehat{\mathbf{b}}_{mg}^{*} = L^{-1}\sum_{i=1}^{L}\widehat{\mathbf{b}}_i^{*}$, where

$$\widehat{\mathbf{b}}_i^{*} = \left(\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\cdot}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\widetilde{\mathbf{Z}}_{ij\cdot}\right)^{-1}\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\cdot}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\mathbf{Y}_{ij\cdot}^{*}.$$

3. We repeat the above procedure $\mathcal{L}$ times.

In the following theorem, we establish the consistency of the bootstrap distribution.

Theorem 3.2. *Suppose that* $\mathbb{E}\|\frac{1}{\sqrt{LN}}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{z}_{ijt}\varepsilon_{ijt}\|^4<\infty$*, where* $\mathbf{z}_{ijt}^\top$ *stands for the* t^{th} *row of* $\mathbf{Z}_{ij\cdot}$*. Under Assumptions 1, 2 and 4, as* $(L,N,T)\to(\infty,\infty,\infty)$*,*

$$\sup_{\mathbf{w}}\left|\Pr^*(\sqrt{LNT}\,(\widehat{\mathbf{b}}_{mg}^*-\widehat{\mathbf{b}}_{bcmg})\le\mathbf{w})-\Pr(\sqrt{LNT}\,(\widehat{\mathbf{b}}_{bcmg}-\overline{\boldsymbol{\beta}})\le\mathbf{w})\right|=o_P(1).$$

Up to this point, the asymptotic properties and inference method for the mean group estimator have been well established. One can also use similar techniques to those for the homogeneous model to study the factor number selection issue for the heterogeneity setting, which is omitted in this paper due to the limitation of page space.

4 Simulation Studies

In this section, we first provide some details of numerical implementation in Section 4.1, and then conduct simulations for homogeneous balanced and unbalanced panels in Sections 4.2 and 4.3, respectively. Finally, the finite sample performance of the heterogeneous model is assessed in Section 4.4.

4.1 Numerical Implementation

On Estimation: For both our simulations and empirical study, the estimation procedure is as follows. For each dataset, we estimate $\boldsymbol{\beta}$ and the global and sector factors in a sequential manner. Specifically, in the first step, we estimate $\boldsymbol{\beta}$ using a pre-specified $d_{\max}$ (say, 20). We then estimate l^G and $\mathbf{l}^S$, together with $\mathbf{F}^G$ and $\mathbf{F}_i^S$'s. Finally, we update the estimate of $\boldsymbol{\beta}$ after obtaining the estimates of l^G and $\mathbf{l}^S$.

On Bandwidth ℓ: For the DWB procedure, we adopt the Bartlett kernel, and the bandwidth ℓ is set as $\lceil 1.75T^{1/3}\rceil$ following Palm et al. (2011). It is worth mentioning that Shao (2010), and Gao et al. (2023) discuss the optimal bandwidth ℓ under different settings in length. In summary, when the Bartlett kernel involved, the optimal bandwidth ℓ is at the order $T^{1/3}$ up to an unknown constant for our 3-dimensional panel data model. Moreover, Gao et al. (2023) conduct extensive simulations to show the DWB procedure is not sensitive to the choice of ℓ and the function form of $a(\cdot)$, so we will not further extend the following simulation studies along this line of research.

4.2 Simulation Results for Balanced Data

We now perform Monte Carlo simulations to investigate the finite sample properties of the theoretical findings of Section 2.

We first look at a balanced data set. Specifically, the data generating process is as follows:

$$y_{ijt} = \mathbf{x}_{ijt}^\top \boldsymbol{\beta} + \boldsymbol{\gamma}_{ij}^{G\top} \mathbf{f}_t^G + \boldsymbol{\gamma}_{ij}^{S\top} \mathbf{f}_{it}^S + \varepsilon_{ijt},$$

where $i \in [L]$, $j \in [N]$, and $t \in [T]$. Before introducing weak cross-sectional dependence, we first define two covariance matrices: $\boldsymbol{\Sigma}_v = \{0.2^{\|(i_1,j_1)-(i_2,j_2)\|}\}$ and $\boldsymbol{\Sigma}_\varepsilon = \{0.2^{\|(i_1,j_1)-(i_2,j_2)\|}\}$. Accordingly, we generate $\mathbf{V}_{\cdot\cdot t} = 0.1\mathbf{V}_{\cdot\cdot t-1} + \boldsymbol{\eta}_{v,t}$ and $\mathbf{E}_{\cdot\cdot t} = 0.2\mathbf{E}_{\cdot\cdot t-1} + \boldsymbol{\eta}_{\varepsilon,t}$, where $\mathbf{V}_{\cdot\cdot t} = (\mathbf{v}_{11t}, \ldots, \mathbf{v}_{1Nt}, \ldots, \mathbf{v}_{L1t}, \ldots, \mathbf{v}_{LNt})^\top$ with $\mathbf{v}_{ijt}$ being a $d \times 1$ vector, $\mathbf{E}_{\cdot\cdot t}$ is stacked by ε_{ijt}'s in the same way as $\mathbf{V}_{\cdot\cdot t}$, $\boldsymbol{\eta}_{v,t,\ell}$ is the ℓ^{th} column of $\boldsymbol{\eta}_{v,t}$, $\boldsymbol{\eta}_{v,t,\ell} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_v)$ for $\ell \in [d]$, and $\boldsymbol{\eta}_{\varepsilon,t} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$.

The factor loadings are generated as $\boldsymbol{\gamma}_{ij}^G \sim N(\mathbf{1}_{l^G\times 1}, \mathbf{I}_{l^G})$ and $\boldsymbol{\gamma}_{ij}^S \sim N(\mathbf{1}_{l_i^S\times 1}, \mathbf{I}_{l_i^S})$. The global and sector factors are generated respectively as $\mathbf{f}_t^G \sim N(\mathbf{0}_{l^G\times 1}, \mathbf{I}_{l^G})$ and $\mathbf{f}_{it}^S \sim N(\mathbf{0}_{l_i^S\times 1}, \mathbf{I}_{l_i^S})$. To introduce correlation between the regressors and the hierarchical factor structure, let $x_{ijt,1} = 1 + v_{ijt,1} + ||(\boldsymbol{\gamma}_{ij}^{G\top}, \boldsymbol{\gamma}_{ij}^{S\top})||^2 + ||(\mathbf{f}_t^{G\top}, \mathbf{f}_{it}^{S\top})||^2$ and $x_{ijt,2} = 1 + v_{ijt,2} + (\boldsymbol{\gamma}_{ij}^{G\top}, \boldsymbol{\gamma}_{ij}^{S\top})\mathbf{1}_{(l^G+l_i^S)\times 1} + (\mathbf{f}_t^{G\top}, \mathbf{f}_{it}^{S\top})\mathbf{1}_{(l^G+l_i^S)\times 1}$, where $x_{ijt,l}$ and $v_{ijt,l}$ stand for the l-th elements of $\mathbf{x}_{ijt}$ and $\mathbf{v}_{ijt}$, respectively. For each generated dataset, we run our hierarchical panel data regression, construct the confidence interval based on 500 bootstrap draws, and estimate the numbers of global and sector factors. We repeat the procedure M times.

For simplicity, we let $l^G = 1$ and let l_i^S (for $\forall i \in [L]$) be randomly chosen from $\{0, 1, 2, 3\}$ with equal probabilities. It is worth noting that $l_i^S = 0$ corresponds to the case where unobserved factors do not exist for the i^{th} sector. We let $L \in \{10, 20\}$, $N \in \{30, 60\}$, and $T \in \{60, 120\}$. Further let $\boldsymbol{\beta} = (1, 1)^\top$ and $M = 1000$.

We use the following criteria when evaluating the finite performance of our approach:

1. On coefficients:

   $\text{RMSE}_{\boldsymbol{\beta}} = \left\{\frac{1}{M}\sum_{m=1}^M \|\widehat{\mathbf{b}}_m - \boldsymbol{\beta}\|^2\right\}^{1/2}$,

   $\text{RMSE}_{\boldsymbol{\beta},bc} = \left\{\frac{1}{M}\sum_{m=1}^M \|\widehat{\mathbf{b}}_{bc,m} - \boldsymbol{\beta}\|^2\right\}^{1/2}$,

   $\text{Bias}_{\boldsymbol{\beta}} = \|\frac{1}{M}\sum_{m=1}^M \widehat{\mathbf{b}}_m - \boldsymbol{\beta}\|$,

   $\text{Bias}_{\boldsymbol{\beta},bc} = \|\frac{1}{M}\sum_{m=1}^M \widehat{\mathbf{b}}_{bc,m} - \boldsymbol{\beta}\|$,

   $\text{CR}_{\boldsymbol{\beta}} = \frac{1}{M}\sum_{m=1}^M \frac{1}{d}\sum_{\ell=1}^d \mathbb{I}(\widehat{b}_{m,\ell} - \beta_\ell \in \text{CI}_m)$.

2. On factor structure:

$\text{Acc}_{l^G} = \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}(\widehat{\ell}_m^G = l^G)$,

$\text{Acc}_{\mathbf{l}^S} = \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}(\widehat{\boldsymbol{\ell}}_m^S = \mathbf{l}^S)$,

$\text{RMSE}_{\mathbf{F}^G} = \left\{\frac{1}{M}\sum_{m=1}^{M}\|\mathbf{P}_{\widehat{\mathbf{C}}_m^G} - \mathbf{P}_{\mathbf{F}^G}\|^2\right\}^{1/2}$,

$\text{RMSE}_{\mathbf{F}^S} = \left\{\frac{1}{M}\sum_{m=1}^{M}\frac{1}{L}\sum_{i=1}^{L}\|\mathbf{P}_{\widehat{\mathbf{C}}_{i,m}^S} - \mathbf{P}_{\mathbf{F}_i^S}\|^2\right\}^{1/2}$.

Here the sub-index m indicates the quantity is obtained at the m^{th} replication, and $\widehat{b}_{m,\ell}$ and β_ℓ stand for the ℓ^{th} elements of $\widehat{\mathbf{b}}_m$ and $\boldsymbol{\beta}$ respectively. $\text{CR}_{\boldsymbol{\beta}}$ is used to evaluate the result of Section 2.3, while Acc_{l^G} and $\text{Acc}_{\mathbf{l}^S}$ are designed to evaluate the results of Section 2.4. Three RMSEs are designed to evaluate overall performances of the estimation procedure. Notably, for the bootstrap procedure, when calculating $\text{CR}_{\boldsymbol{\beta}}$, we use the true values of l^G and l_i^S instead of their estimates, and the nominal rate is set as 95%. As the sample size grows, we expect that $\text{CR}_{\boldsymbol{\beta}}$ moves towards 95%, the values of Acc_{l^G} and $\text{Acc}_{\mathbf{l}^S}$ converge to 1, and $\text{RMSE}_{\boldsymbol{\beta}}$, $\text{RMSE}_{\mathbf{F}^G}$ and $\text{RMSE}_{\mathbf{F}^S}$ converge to 0. As the estimation of unobservable factors is carried on in a sequential manner, we should see the measures of the global factors are better than those of sector factors.

Table 1: Estimation Results of Balanced Panel

L	N	T	$\text{RMSE}_{\boldsymbol{\beta}}$	$\text{RMSE}_{\boldsymbol{\beta},bc}$	$\text{Bias}_{\boldsymbol{\beta}}$	$\text{Bias}_{\boldsymbol{\beta},bc}$	$\text{CR}_{\boldsymbol{\beta}}$	$\text{RMSE}_{\mathbf{F}^G}$	$\text{RMSE}_{\mathbf{F}^S}$	Acc_{l^G}	$\text{Acc}_{\mathbf{l}^S}$
10	30	30	0.0138	0.0143	0.0019	0.0015	0.902	0.6519	0.9588	0.990	0.459
		60	0.0080	0.0085	0.0018	0.0014	0.901	0.6276	0.9202	1.000	0.520
		120	0.0054	0.0056	0.0016	0.0011	0.912	0.6119	0.8965	1.000	0.560
	60	30	0.0100	0.0100	0.0012	0.0008	0.914	0.6026	0.8952	0.994	0.562
		60	0.0064	0.0064	0.0006	0.0004	0.918	0.5622	0.8174	1.000	0.707
		120	0.0047	0.0045	0.0010	0.0007	0.920	0.5418	0.7525	1.000	0.828
20	30	30	0.0099	0.0098	0.0018	0.0016	0.908	0.5363	0.9474	1.000	0.449
		60	0.0062	0.0063	0.0018	0.0015	0.912	0.5223	0.9133	1.000	0.507
		120	0.0044	0.0044	0.0015	0.0012	0.924	0.5184	0.8861	1.000	0.562
	60	30	0.0076	0.0075	0.0004	0.0007	0.901	0.5053	0.8798	1.000	0.562
		60	0.0053	0.0051	0.0004	0.0008	0.922	0.4795	0.8030	1.000	0.704
		120	0.0039	0.0036	0.0011	0.0007	0.930	0.4536	0.7405	1.000	0.812

As shown in Table 1, $\text{RMSE}_{\boldsymbol{\beta}}$ is relatively small, and decreases as the sample size increases. $\text{CR}_{\boldsymbol{\beta}}$ converges to 95%, but is lower than the nominal rate overall, which is consistent with some existing results (such as the simulations presented in Shao, 2010). The numbers of factors are always correctly identified, which is not surprising in view of the fact that even the smallest sample size gives $10 \times 25 \times 60 = 15000$ observations. Also, when the sample size goes up, the values of $\text{RMSE}_{\mathbf{F}^G}$ and $\text{RMSE}_{\mathbf{F}^S}$ also converge to 0, but slow. It should be expected as the dimensions of $\mathbf{P}_{\mathbf{F}^G}$ and $\mathbf{P}_{\mathbf{F}_i^S}$ are large.

4.3 Simulation Results for Unbalanced Data

We next examine an unbalanced panel data set as mentioned in Section 2.5. Specifically, we consider two cases: Case 1: $N_i \in \{25, 26, \ldots, 35\}$, and Case 2: $N_i \in \{40, 41, \ldots, 60\}$, in which N_i is drawn from the given set with equal probability. The regressors are generated as $x_{ijt,1} = v_{ijt,1} + |\boldsymbol{\gamma}_{ij}^{G\top} \mathbf{1}_{l^G \times 1}|$ and $x_{ijt,2} = v_{ijt,2}$. The rest settings are identical to the balanced case. Since the estimators are asymptotically unbiased for this set of regressors, the results for bias corrected estimators are not reported. As shown in Table 2, the patten is very much similar to those presented in Table 1, so we do not repeat the discussions.

Table 2: Estimation Results of Unbalanced Panel

L	N	T	$\text{RMSE}_{\boldsymbol{\beta}}$	$\text{CR}_{\boldsymbol{\beta}}$	$\text{RMSE}_{\mathbf{F}^G}$	$\text{RMSE}_{\mathbf{F}^S}$	Acc_{l^G}	$\text{Acc}_{\mathbf{l}^S}$
10	Case 1	60	0.006	0.903	0.345	0.431	1.000	0.999
		120	0.004	0.921	0.327	0.392	1.000	0.999
	Case 2	60	0.004	0.910	0.315	0.413	1.000	1.000
		120	0.003	0.926	0.296	0.368	1.000	1.000
20	Case 1	60	0.004	0.903	0.283	0.424	1.000	0.998
		120	0.003	0.914	0.270	0.388	1.000	0.999
	Case 2	60	0.003	0.906	0.264	0.405	1.000	1.000
		120	0.002	0.929	0.247	0.363	1.000	1.000

4.4 Simulation Results for Heterogeneous Panel

To close our simulation investigation, we finally assess the proposed estimators of heterogeneous coefficients. Specifically, we consider the following data generating process:

$$y_{ijt} = \mathbf{x}_{ijt}^\top \boldsymbol{\beta}_i + \boldsymbol{\gamma}_{ij}^{G\top} \mathbf{f}_t^G + \boldsymbol{\gamma}_{ij}^{S\top} \mathbf{f}_{it}^S + \varepsilon_{ijt},$$

where the heterogeneous coefficients are $\boldsymbol{\beta}_i = (\sin(0.25\pi i), \sin(0.25\pi i))^\top$, $x_{ijt,1} = 1 + v_{ijt,1} + \mathbf{f}_t^{G\top} \mathbf{1}_{l_G}$ and $x_{ijt,2} = 1 + v_{ijt,2} + \mathbf{f}_{it}^{S\top} \mathbf{1}_{l_{S,i}}$. All other variables are generated in the same manner with those in Section 4.1. After replicating the experiments for 1000 times, we compute the root mean squared errors, biases and coverage rates for the mean group estimators:

$$\text{RMSE}_{mg} = \left\{ \tfrac{1}{M} \textstyle\sum_{m=1}^{M} \|\widehat{\mathbf{b}}_{mg,m} - \overline{\boldsymbol{\beta}}\|^2 \right\}^{1/2},$$

$$\text{RMSE}_{bcmg} = \left\{ \tfrac{1}{M} \textstyle\sum_{m=1}^{M} \|\widehat{\mathbf{b}}_{bcmg,m} - \overline{\boldsymbol{\beta}}\|^2 \right\}^{1/2},$$

$$\text{Bias}_{mg} = \|\tfrac{1}{M} \textstyle\sum_{m=1}^{M} \widehat{\mathbf{b}}_{mg,m} - \overline{\boldsymbol{\beta}}\|,$$

$$\text{Bias}_{bcmg} = \|\tfrac{1}{M} \textstyle\sum_{m=1}^{M} \widehat{\mathbf{b}}_{bcmg,m} - \overline{\boldsymbol{\beta}}\|,$$

$$\text{CR}_{mg} = \tfrac{1}{M} \textstyle\sum_{m=1}^{M} \tfrac{1}{d} \sum_{\ell=1}^{d} \mathbb{I}(\widehat{b}_{mg,\ell,m} - \overline{\beta}_\ell \in \text{CI}_m),$$

where $\widehat{b}_{mg,\ell,m}$ and $\overline{\beta}_\ell$ are the ℓ^{th} elements of $\widehat{\mathbf{b}}_{mg,m}$ and $\overline{\boldsymbol{\beta}}$, respectively. For factor estimates, we use the same criteria as in Section 4.2. As can be seen from Table 3, the proposed estimation and inference methods perform reasonably well with the heterogeneous coefficients.

Table 3: Estimation Results of Heterogeneous Panel

L	N	T	RMSE_{mg}	RMSE_{bcmg}	Bias_{mg}	Bias_{bcmg}	CR_{mg}	$\text{RMSE}_{\mathbf{F}^G}$	$\text{RMSE}_{\mathbf{F}^S}$	Acc_{l^G}	$\text{Acc}_{\mathbf{l}^S}$
10	30	30	0.0128	0.0164	0.0041	0.0038	0.887	0.6372	0.9593	1.000	0.447
		60	0.0087	0.0133	0.0026	0.0019	0.902	0.6293	0.9189	1.000	0.526
		120	0.0059	0.0129	0.0021	0.0016	0.908	0.6255	0.8956	1.000	0.565
	60	30	0.0099	0.0102	0.0021	0.0023	0.911	0.6388	0.8907	0.995	0.576
		60	0.0068	0.0068	0.0015	0.0010	0.914	0.6118	0.8267	1.000	0.706
		120	0.0047	0.0045	0.0013	0.0012	0.916	0.6075	0.7525	1.000	0.828
20	30	30	0.0151	0.0162	0.0019	0.0017	0.904	0.5334	0.9370	1.000	0.450
		60	0.0080	0.0081	0.0019	0.0010	0.909	0.5230	0.9023	1.000	0.514
		120	0.0068	0.0078	0.0010	0.0008	0.921	0.5196	0.8873	1.000	0.558
	60	30	0.0083	0.0084	0.0009	0.0009	0.901	0.5200	0.8641	1.000	0.575
		60	0.0060	0.0057	0.0009	0.0006	0.919	0.5076	0.8127	1.000	0.706
		120	0.0036	0.0043	0.0007	0.0005	0.931	0.5057	0.7610	1.000	0.812

5 An Empirical Study

Starting with the seminal studies by Baumol (1986), Barro (1991), and Barro and Sala-i Martin (1992), numerous studies have been devoted to testing whether income or productivity of poorer economies are converging to those of richer economies. As Durlauf (2003) puts it, "*Few issues in empirical growth economics have received as much attention as the question of whether countries exhibit convergence*". A main technique employed by these studies is "cross-country growth regressions", where aggregate- or industry-level cross-country data are used to regress the average growth rates of per capita income (or labour productivity) over a long period on the initial level of income per capita (or labour productivity) and some additional control variables[3]. A negative and significant coefficient on the initial conditions is taken to be evident in favour of β-convergence[4]. For excellent surveys of cross-country convergence studies, see Durlauf (2003) and Islam (2003) for example.

Despite the increasing availability of disaggregated data at industry level, the hierarchical structure of these data, to the best of our knowledge, has rarely been explored. Hierarchical panel data models deserve special attention in the convergence literature as industry level data with multi-level structure have become increasingly available for convergence analysis. Yet, little attempt has been made in this regard. Most existing studies focus exclusively on the effects of one level while ignoring effects from other level(s). Specifically, a substantial body of the literature (e.g., Mankiw et al., 1992; De la Fuente, 1999) uses aggregate-level data and focus exclusively on aggregate-level cross-country variations and attributes. As a result, they ignore industry-level attributes in influencing the convergence of aggregate-level income or productivity. Conversely, another strand of the literature (e.g., Bernard and Jones, 1996) uses industry-level data and focus exclusively on industry-level cross-country variations and attributes by running a separate regression for each industry. Consequently, these latter studies ignore aggregate-level attributes in influencing the convergence of industry-level income or productivity.

5.1 Data

We begin our empirical analysis by introducing the data. For labor productivity (or real value added per employee), we follow Rodrik (2013) and use the dataset of UNIDO's INDSTAT2, which provides data on value added (in nominal U.S dollars) and employment for 23 manufacturing industries at the ISIS two-digit level for a

[3] In this study we follow Rodrik (2013) and define labour productivity of an industry as the industry's real value added divided by its number of employees. As is well known, the value added of an industry, also referred to as gross domestic product (GDP)-by-industry, is the contribution of a private industry or government sector to overall GDP (The U.S. Bureau of Economic Analysis, 2006, available at `https://www.bea.gov/help/faq/184`). The definitions of labour productivity and value added, therefore, imply that an industry's labour productivity can be considered as the industry's "GDP per capita", which in turn implies that neoclassical growth models predict not only conditional convergence in GDP per capita among economies but also industry-level convergence in labour productivity among economies.

[4] As pointed out by Durlauf (2003), "*While β-convergence is not the only statistical measure of convergence that has been developed, it is the dominant measure*".

large number of countries[5]. With this dataset, real value added can be computed by deflating the nominal value added by the US producer price index, and labor productivity can then be obtained by further dividing real value added by employment (i.e., number of employees). Growth in labor productivity is then measured as percentage change in labor productivity.

Our control variables include a wide range of factors that have been found to be important for assessing convergence. These include human capital (as measured by school enrollment) (Barro, 1991; Sala-I-Martin et al., 2004), investment price (De Long and Summers, 1991; Jones, 1994), trade openness and terms of trade (Frankel and Romer, 1999; Dollar and Kraay, 2003), institutions (measured by civil liberties) (Rodrik et al., 2004; Acemoglu et al., 2019), natural resources (measured by oil reserves) (Easterly and Levine, 2003; Sachs and Warner, 2001), government consumption share (Sala-I-Martin et al., 2004; Salimans, 2012), and real exchange rate distortions (Sala-I-Martin et al., 2004).

Due to data availability, our sample starts at 1963 and ends at 2018. For the same reason, the number of countries varies across industries. The variable names, their means, and standard deviations are presented in Table 4. Table 5 summarizes the number of countries for each industry. It should be noted that since geographical factors are usually time-invariant, they will be captured by the factor structures and thus are not included in the control variables.

Table 4: Summary Statistics of the Dataset

	Abbreviation	Mean	Std
Labor productivity	LP	7.732	2.9858
Investment price (%)	IP	28.1833	21.1484
Government consumption share (%)	GCS	20.6441	11.8342
Openness measure	Open	-3.2222	12.2461
Secondary school enrolment (%)	SSE	52.722	31.7198
Civil liberties	CL	4.2661	1.5295
Terms of trade	TT	117.7786	42.4048
Real exchange rate distortions	DIS	124.1865	35.3769
Proved reserves (bbl/10^9)	Oil	5.5693	21.5011

When measuring dependent and independent variables, we follow Feng et al. (2022) and Salimans (2012) and treat them differently. Specifically, the dependent

[5]The International Standard Industrial Classification of All Economic Activities (ISIC) is the international reference classification of productive activities. Its main purpose is to provide a set of activity categories that can be utilized for the collection and reporting of statistics according to such activities. See `https://unstats.un.org/unsd/classifications/Econ/ISIC.cshtml` for more details.

Table 5: 23 Manufacturing Industries at the ISIC Two-Digit Level, and the Numbers of Countries of Each Industry

Industry Name	Abbreviation	NO. of Countries
Food and beverages	FB	78
Tobacco products	TP	73
Textiles	TE	78
Wearing apparel, fur	WAF	73
Leather, leather products and footwear	LLF	57
Wood products (excl. furniture)	WP	77
Paper and paper products	PPP	76
Printing and publishing	PP	77
Coke, refined petroleum products, nuclear fuel	CRN	73
Chemicals and chemical products	CCP	77
Rubber and plastics products	RPP	74
Non-metallic mineral products	NMP	78
Basic metals	BM	75
Fabricated metal products	FMP	78
Machinery and equipment n.e.c.	ME	74
Office, accounting and computing machinery	OACM	49
Electrical machinery and apparatus	EMA	72
Radio, television and communication equipment	RTCE	38
Medical, precision and optical instruments	MPOI	68
Motor vehicles, trailers, semi-trailers	MVTS	73
Other transport equipment	OTE	51
Furniture, manufacturing n.e.c.	FM	78
Recycling	RC	33

variable is measured as a five-year moving average of economic growth, while all explanatory variables are measured at the beginning of each five year period. This treatment has three advantages: (1) it can reduces the potential effects of short-term fluctuations; (2) it can maintains a high number of time series observations; and (3) perhaps more importantly, it can alleviate reverse causality or simultaneity between regressors and growth in labor productivity. Another commonly-used practice in the literature is to take a five-year simple moving average of both dependent and independent variables[6]. While this latter technique is capable of reducing

[6]A third popular method of looking at annual data in empirical growth literature is to use averaged five-year period data. But, as is stressed by Soto (2003) and Attanasio et al. (2000), "the use of n-year averages is not suitable because of the lost of information that it implies". In addition, as Soto (2003) and Attanasio et al. (2000) pointed out, attempting to use data on averaged five-year periods severely limited the number of observations to draw from in the data.

the potential effects of short-term fluctuations and maintaining a high number of time series observations, it may still suffer from reverse causality or simultaneity, because causality between regressors and growth could go the other way or some regressors and growth in productivity may be simultaneously determined (e.g., Bils and Klenow, 2000).

5.2 Results from the Conditional Convergence Regression

We start by investigating conditional convergence for the manufacturing industry as a whole. This can be done using equation (2.1) where all of the three components (i.e., initial productivity, control variables, and hierarchical factors) are included. The second column of Table 5 reports the number of the unobserved global and sector factors estimated using equations (2.7) and (2.9) respectively. As can be seen, we have identified one global factor that affects the growth in labour productivity of every individual manufacturing industry. As for the number of sector factors, it differs across industries, ranging from 1 to 10 with an average of 4.

To have a better idea of the importance of the global and industry-specific factors in explaining the total variance of the error terms (i.e., $\widehat{\varepsilon}_{ijt} = y_{ijt} - \mathbf{x}_{ijt}^{\top}\widehat{\mathbf{b}}$), we calculate the proportion of the total variance attributed to these factors and report them in the third column of Table 6. As this table shows, the global factor accounts for 28.59% of the total variance, while the sector factors together account for 56.67% of the total variance with the contribution of each sector factor ranging from 0.08% to 8.01%. The significant proportion explained by the sector factors shows that the use of global factors alone is not enough when multi-level data are employed for investigating convergence, thus justifying the use of a hierarchical model.

Table 7.A shows all the estimated coefficients for the manufacturing industry as a whole. We find that all the control variables have the expected signs. For example, the estimated coefficient of price for investment goods is significantly negative, suggesting that a relative low price of investment goods in the first year of each five-year period is strongly and positively related to subsequent growth in labour productivity. This finding is not surprising because a low investment price stimulates investment in machinery and equipment, which further spurs growth in labour productivity (De Long and Summers, 1991, 1992). To give another example, the estimated coefficient of secondary schooling enrollment is significantly positive. This latter finding is consistent with previous studies, which have documented that a large pool of workers with secondary education is indispensable for knowledge spillover to take place and for attracting imports of technologically advanced goods and foreign direct investment (Borensztein et al., 1998; Caselli and Coleman II,

2001). To give a third example, the estimated coefficient of trade openness is not statistically different from zero. This is consistent with Sala-I-Martin et al. (2004) who argue that the insignificance of the trade openness presumably reflects the crudity of this measure and perhaps the distinction between opening to international trade generating a one-time step increase in productivity as factors are reallocated according to comparative advantage versus an ongoing growth impact associated with greater openness.

Having discussed the hierarchical factors and control variables, in what follows we concentrate on the estimated coefficient on initial productivity for the manufacturing industry as a whole. As can be seen from Table 7.A, it is negative (-0.391%) and highly significant with a 95% confidence interval of (-0.845%, -0.080%). This suggests that when country characteristics are controlled, initial labour productivity is negatively related to the subsequent rate of growth in labour productivity. In other words, conditional divergence in labour productivity exists for the manufacturing industry as a whole. This finding is in line with that of Rodrik (2013) who, by applying a fixed effects panel data model to the UNIDO's INDSTAT dataset, also finds conditional convergence in labour productivity for the total manufacturing industry. It is also consistent with the income convergence literature (Islam, 1995; Sala-I-Martin et al., 2004) that finds that once country characteristics are controlled for, the coefficient on initial income becomes negative and statistically significant.

5.3 Results from the Unconditional Convergence Regression

Tests for unconditional convergence use a similar regression specification as tests for conditional convergence, but without controlling for country characteristics (i.e., equation (2.1) without the control variables). We start with results for the manufacturing industry as a whole. Table 6.B reports the number of the global factors, the numbers of sector factors, and their associated contributions in explaining the total variance of the error terms. As can be seen, the results presented in this table are very similar to those reported in Table 6.A. Specifically, we have identified one global factor that affects all industries. The number of sector factors vary across industries, ranging from 1 to 10 with an average of 5. In addition, the global factor accounts for 28.66% of the total variance of $\mathrm{Var}(\widehat{e}_{ijt})$, while all the industry-specific factors together account for 56.60% of the total variance with the contribution of each sector factor ranging from 0.08% to 8.00%.

Table 7.B shows the estimated coefficient on initial productivity for the manufacturing industry as a whole. As can be seen, It is positive (0.282%) and statistically

significant with a 95% confidence interval of (0.141%, 0.407%). This suggests that when country characteristics are not controlled for, initial labour productivity is positively related to the subsequent rate of growth in labour productivity. In other words, unconditional divergence in labour productivity exists for the manufacturing industry as a whole.

In order to confirm our results regarding unconditional convergence in labour productivity, we conduct two robustness checks. First, we follow Rodrik (2013) and exclude OCED countries from our sample of countries. The results, presented in Table 7.E show that our findings of unconditional divergence for both the total manufacturing and individual industries are very robust to the exclusion of OECD countries. Specifically, in Table 7.E, the estimated coefficient on initial productivity is positive and statistically significant for the total manufacturing industry, with a point estimate of 0.269% and a 95% confidence interval of (0.129% 0.414%).

Second, we conduct another robustness check by re-estimating equation (2.1) without the control variables for the following two subperiods: 1973-2018 and 1983-2018. The results are shown in Table 7.C-D respectively. As can be seen from Table 7.C-D, the estimated coefficients on initial productivity are positive and statistically significant for the total manufacturing industry regardless of the sub-period, confirming unconditional divergence for the total manufacturing industry.

In addition, we also examine the sector-specific unconditional convergence in labour productivity based on the heterogeneous panel data model (3.1). As can be seen from Table 8, most sector-specific estimates and their confidence intervals are positive, which suggests the growth rate of labour productivity is positively associated with its initial value for most sectors when the difference in country characteristics are not considered. This finding is in line with that from homogeneous model.

To summarize this section, we have examined the twin hypotheses of conditional and unconditional-convergence for manufacturing industries across countries. The empirical results presented in this section suggest that unconditional-convergence does not obtain. This finding is quite robust to the exclusion of OECD countries and to the use of different sample periods. On the other hand, there is strong and consistent evidence of convergence once factors that affect steady-state levels of labour productivity are controlled for.

Table 6: Estimation of the Global and Industry-Specific Factors

	Panel A (with controls)		Panel B (without controls)	
Global	No. of Factors	% of Var($\widehat{e}_{ijt}$)	No. of Factors	% of Var($\widehat{e}_{ijt}$)
	1	28.59%	1	28.66%
Industry	No. of Factors	% of Var($\widehat{e}_{ijt}$)	No. of Factors	% of Var($\widehat{e}_{ijt}$)
FB	1	0.08	1	0.08
TP	5	3.67	5	3.67
TE	1	1.26	1	1.26
WAF	1	0.45	1	0.45
LLF	1	0.34	1	0.34
WP	1	0.64	1	0.64
PPP	6	2.55	6	2.55
PP	1	0.07	1	0.25
CRN	7	5.03	7	5.03
CCP	1	0.06	1	0.30
RPP	3	1.65	3	1.65
NMP	1	1.41	1	1.40
BM	10	8.01	10	8.00
FMP	1	0.26	1	0.26
ME	10	5.36	10	5.36
OACM	8	6.00	8	6.00
EMA	5	2.31	5	2.31
RTCE	6	3.09	6	3.08
MPOI	2	3.59	2	3.59
MVTS	8	5.56	8	5.55
OTE	1	1.40	1	1.39
FM	3	3.25	3	3.25
RC	7	0.63	7	0.63
Sum of All Industries		56.67		56.60

Table 7: Coefficient Estimates Using Homogeneous Panel

		$\widehat{\beta}$	95% CI
Panel A (with controls for 1963-2018)	IniP	-0.391	(-0.845 -0.080)
	IP	-0.041	(-0.060 -0.019)
	GCS	-0.038	(-0.109 0.026)
	Open	-0.017	(-0.053 0.013)
	SSE	0.041	(0.003, 0.091)
	CL	0.157	(-0.125, 0.420)
	TT	0.028	(0.020, 0.041)
	DIS	0.012	(-0.005, 0.034)
	Oil	0.008	(-0.010, 0.026)
Panel B (without controls for 1963-2018)	IniP	0.282	(0.141, 0.407)
Panel C (without controls for 1973-2018)	IniP	0.240	(0.103, 0.362)
Panel D (without controls for 1983-2018)	IniP	0.175	(0.099, 0.267)
Panel E (without controls & excluding OECD for 1963-2018)	IniP	0.269	(0.129, 0.414)

2. IniP stands for the initial productivity.

Table 8: Coefficient Estimates Using Heterogeneous Panel

	Including OECD								Excluding OECD	
	Panel A (1963-2018)		Panel B (1973-2018)		Panel C (1983-2018)		Panel D (1993-2018)		Panel E (1963-2018)	
	$\widehat{\beta}_i$	CI	$\widehat{\beta}_i$	CI	$\widehat{\beta}_i$	CI	$\widehat{\beta}_i$	CI	$\widehat{\beta}_i$	CI
FB	0.294	(0.245, 0.395)	0.282	(0.220, 0.377)	0.225	(0.076, 0.406)	0.365	(-1.796, 3.883)	0.284	(0.211, 0.396)
TP	0.374	(0.221, 0.442)	0.315	(0.199, 0.419)	0.188	(-0.104, 0.402)	0.204	(-0.418, 0.389)	0.346	(0.160, 0.417)
TE	0.290	(0.137, 0.490)	0.225	(0.035, 0.422)	0.155	(-0.166, 0.274)	0.043	(-3.293, 1.780)	0.276	(0.120, 0.461)
WAF	0.238	(0.091, 0.376)	0.207	(-0.026, 0.329)	0.067	(-0.152, 0.283)	-0.007	(-0.216, 0.456)	0.221	(0.073, 0.370)
LLF	0.034	(-0.005, 0.128)	0.038	(-0.010, 0.210)	0.063	(-0.005, 0.285)	0.038	(-0.194, 0.389)	0.027	(-0.013, 0.162)
WP	0.283	(0.132, 0.475)	0.244	(0.052, 0.425)	-0.001	(-0.096, 0.231)	-0.077	(-0.110, 0.303)	0.267	(0.115, 0.464)
PPP	0.310	(0.178, 0.441)	0.238	(0.155, 0.339)	0.251	(0.078, 0.337)	0.168	(-0.041, 0.347)	0.300	(0.171, 0.403)
PP	0.166	(0.086, 0.287)	0.155	(0.037, 0.286)	0.086	(0.039, 0.259)	0.029	(-0.155, 0.299)	0.267	(0.134, 0.400)
CRN	0.268	(0.146, 0.454)	0.182	(0.083, 0.361)	0.179	(0.020, 0.356)	0.080	(-0.039, 0.407)	0.254	(0.103, 0.479)
CCP	0.199	(0.018, 0.416)	0.305	(0.146, 0.392)	0.166	(0.011, 0.390)	0.069	(-0.053, 0.400)	0.339	(0.205, 0.444)
RPP	0.224	(0.086, 0.417)	0.177	(0.025, 0.355)	0.111	(-0.238, 0.291)	-0.054	(-2.301, 4.546)	0.205	(0.070, 0.399)
NMP	0.351	(0.281, 0.504)	0.335	(0.235, 0.489)	0.177	(0.038, 0.373)	0.238	(-0.121, 0.329)	0.341	(0.263, 0.492)
BM	0.273	(0.089, 0.437)	0.197	(0.054, 0.384)	0.275	(-0.038, 0.455)	0.004	(-0.400, 1.165)	0.254	(0.065, 0.427)
FMP	0.314	(0.235, 0.449)	0.219	(0.125, 0.356)	0.278	(-0.014, 0.401)	-0.002	(-0.079, 0.437)	0.308	(0.222, 0.438)
ME	0.294	(-0.290, 0.289)	0.252	(-0.363, 0.324)	0.190	(-0.509, 0.352)	0.119	(-0.125, 0.667)	0.264	(-0.334, 0.253)
OACM	0.047	(-0.016, 0.094)	0.065	(-0.026, 0.139)	0.048	(-0.162, 0.213)	0.129	(-0.654, 1.426)	0.037	(-0.022, 0.089)
EMA	0.351	(0.223, 0.499)	0.335	(0.126, 0.475)	0.100	(-0.339, 0.481)	-0.030	(-0.170, 0.413)	0.336	(0.189, 0.491)
RTCE	0.022	(-0.171, 0.082)	-0.010	(-0.259, 0.105)	0.058	(-0.368, 0.207)	0.130	(-0.587, 0.364)	-0.190	(-0.305, 0.009)
MPOI	0.034	(-0.187, 0.223)	0.020	(-0.301, 0.222)	-0.127	(-0.890, 0.298)	-0.230	(-0.749, 0.484)	-0.033	(-0.292, 0.198)
MVTS	0.341	(0.064, 0.495)	0.299	(0.148, 0.419)	0.290	(-0.022, 0.433)	0.158	(-0.298, 0.514)	0.336	(0.012, 0.472)
OTE	0.143	(0.005, 0.356)	0.177	(0.025, 0.429)	0.183	(-0.046, 0.327)	0.305	(-0.165, 0.761)	0.132	(-0.002, 0.358)
FM	0.291	(0.196, 0.439)	0.277	(0.123, 0.402)	0.102	(-0.201, 0.312)	0.065	(-0.075, 0.421)	0.279	(0.177, 0.428)
RC	-0.001	(-0.140, 0.196)	-0.004	(-0.189, 0.269)	0.011	(-0.337, 0.408)	-0.461	(-7.559, 9.265)	-0.044	(-0.228, 0.158)

6 Conclusion

In this paper, we contribute to the hierarchical panel data literature in three ways. First, our 3D panel data model not only contains regressors but also specifies the nature of common shocks through the use of a two component factor structure (i.e., global factors and sector factors). Accordingly, we propose an approach to estimating the hierarchy, thus enabling us to have a better understanding of the relative importance of these two types of unobservable shocks in explaining the response variable. Second, we establish the asymptotic theories, and further provide a bootstrap procedure to construct the confidence interval for the parameters of interest while allowing for both types of CSD and time series autocorrelation.

Notably, these results are established assuming all three dimensions to diverge. In contrast, many previous studies only allow two dimensions to diverge, rendering their asymptotic derivations to a classical two dimensional panel data case. Third, we justify the theoretical findings using extensive simulated and real data examples. In the empirical study, we examine the twin hypotheses of conditional and unconditional-convergence for manufacturing industries across countries. The empirical results presented in this paper suggest that unconditional-convergence does not obtain. This finding is quite robust to the exclusion of OECD countries and to the use of different sample periods. On the other hand, there is strong and consistent evidence of convergence once factors that affect steady-state levels of labour productivity are controlled for.

References

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999), 'High wage workers and high wage firms', *Econometrica* **67**(2), 251–333.

Acemoglu, D., Naidu, S., Restrepo, P. and Robinson, J. A. (2019), 'Democracy does cause growth', *Journal of Political Economy* **127**(1), 47–100.

Ahn, S. C. and Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica* **81**(3), 1203–1227.

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**(3), 817–858.

Arellano, M. and Bond, S. (1991), 'Some tests of specification for panel data: Monte carlo evidence and an application to employment equations', *Review of Economic Studies* **58**(2), 277–297.

Attanasio, O. P., Picci, L. and Scorcu, A. E. (2000), 'Saving, growth, and investment: A macroeconomic analysis using a panel of countries', *Review of Economics and Statistics* **82**(2), 182–211.

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* **77**(4), 1229–1279.

Bai, J., Choi, S. H. and Liao, Y. (2020), ‘Standard errors for panel data models with unknown clusters’, *Journal of Econometrics* p. forthcoming.

Barro, R. J. (1991), ‘Economic growth in a cross section of countries’, *Quarterly Journal of Economics* **106**(2), 407–443.

Barro, R. J. and Sala-i Martin, X. (1992), ‘Convergence’, *Journal of Political Economy* **100**(2), 223–251.

Baumol, W. J. (1986), ‘Productivity growth, convergence, and welfare: What the long-run data show’, *American Economic Review* **76**(5), 1072–1085.

Bernard, A. and Jones, C. (1996), ‘Productivity across industries and countries: Time series theory and evidence’, *Review of Economics and Statistics* **78**(1), 135–46.

Bils, M. and Klenow, P. J. (2000), ‘Does schooling cause growth?’, *American Economic Review* **90**(5), 1160–1183.

Blundell, R. and Bond, S. (1998), ‘Initial conditions and moment restrictions in dynamic panel data models’, *Journal of Econometrics* **87**(1), 115–143.

Borensztein, E., De Gregorio, J. and Lee, J.-W. (1998), ‘How does foreign direct investment affect economic growth?1’, *Journal of International Economics* **45**(1), 115–135.

Breitung, J. and Eickmeier, S. (2016), ‘Analyzing international business and financial cycles using Multi-Level factor models: A comparison of alternative approaches’, *Advances in Econometrics* **35**, 177–214.

Caselli, F. and Coleman II, W. J. (2001), ‘The U.S. structural transformation and regional convergence: A reinterpretation’, *Journal of Political Economy* **109**(3), 584–616.

Chen, J., Gao, J. and Li, D. (2012), ‘A new diagnostic test for cross-section uncorrelatedness in nonparametric panel data models’, *Econometric Theory* **28**(5), 1144–1163.

Chen, M., Fernández-Val, I. and Weidner, M. (2021), ‘Nonlinear factor models for network and panel data’, *Journal of Econometrics* **220**(2), 296–324.

Choi, I., Lin, R. and Shin, Y. (2023), ‘Canonical correlation-based model selection for the multilevel factors’, *Journal of Econometrics* **233**(1), 22–44.

De la Fuente, A. (1999), ‘Convergence across countries and regions: Theory and empirics’, *European Investment Bank (EIB)* **5**(2), 25–45.

De Long, J. B. and Summers, L. (1991), ‘Equipment investment and economic growth’, *Quarterly Journal of Economics* **106**(2), 445–502.

De Long, J. B. and Summers, L. (1992), ‘Equipment investment and economic growth: How strong is the nexus?’, *Brookings Papers on Economic Activity* **2**, 157–199.

Dollar, D. and Kraay, A. (2003), ‘Institutions, trade, and growth’, *Journal of Monetary Economics* **50**(1), 133–162.

Durlauf, S. N. (2003), 'The convergence hypothesis after 10 years', *Revista Economica de Castilla-La Mancha* **2**, 55–74.

Easterly, W. and Levine, R. (2003), 'Tropics, germs, and crops: How endowments influence economic development', *Journal of Monetary Economics* **50**(1), 3–39.

Feng, G., Gao, J. and Peng, B. (2022), 'An integrated panel data approach to modelling economic growth', *Journal of Econometrics* **228**(2), 379–397.

Frankel, J. A. and Romer, D. H. (1999), 'Does trade cause growth?', *American Economic Review* **89**(3), 379–399.

Gao, J., Peng, B. and Yan, Y. (2023), Higher-order expansions and inference for panel data models. https://ssrn.com/abstract=4098007.

Golub, G. H. and Van Loan, C. F. (2013), *Matrix Computations (4th Edition)*, The Johns Hopkins University Press.

Gonçalves, S. (2011), 'The moving blocks bootstrap for panel linear regression models with individual fixed effects', *Econometric Theory* **27**(5), 1048–1082.

Hansen, B. E. (1991), 'Strong laws for dependent heterogeneous processes', *Econometric Theory* **7**(2), 213–221.

Hsiao, C. (2003), *Analysis of Panel Aata*, second edn, New York: Cambridge University Press.

Islam, N. (1995), 'Growth empirics: A panel data approach', *Quarterly Journal of Economics* **110**(4), 1127–1170.

Islam, N. (2003), 'What have we learnt from the convergence debate?', *Journal of Economic Surveys* **17**(3), 309–362.

Jin, S., Lu, X. and Su, L. (2023), Three-dimensional factor models with global and local factors. https://teacher.econ.cuhk.edu.hk/~xunlu/.

Jones, C. I. (1994), 'Economic growth and the relative price of capital', *Journal of Monetary Economics* **34**(3), 359–382.

Kapetanios, G., Serlenga, L. and Shin, Y. (2021), 'Estimation and inference for multi-dimensional heterogeneous panel datasets with hierarchical multi-factor error structure', *Journal of Econometrics* **220**(2), 504–531.

Lam, C. and Yao, Q. (2012), 'Factor modeling for high-dimensional time series: Inference for the number of factors', *Annals of Statistics* **40**(2), 694–726.

Lin, R. (2023), Estimation and inference for a multi-dimensional panel data model with multilevel factors. https://ssrn.com/abstract=4518900.

Lu, X., Miao, K. and Su, L. (2021), 'Determination of different types of fixed effects in three-dimensional panels', *Econometrics Reviews* **40**(9), 867–898.

Mankiw, N. G., Romer, D. and Weil, D. N. (1992), 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics* **107**(2), 407–437.

Matyas et al. (2017), *The Econometrics of Multi-dimensional Panels: Theory and Applications*, Springer, Chamg.

Palm, F., Smeekes, S. and Urbain, J.-P. (2011), 'Cross-sectional dependence robust block bootstrap panel unit root tests', *Journal of Econometrics* **163**(1), 85–104.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

Rodrik, D. (2013), 'Unconditional convergence in manufacturing', *Quarterly Journal of Economics* **128**(1), 165–204.

Rodrik, D., Subramanian, A. and Trebbi, F. (2004), 'Institutions rule: The primacy of institutions over geography and integration in economic development', *Journal of Economic Growth* **9**, 131–165.

Sachs, J. D. and Warner, A. (2001), 'The curse of natural resources', *European Economic Review* **45**(4-6), 827–838.

Sala-I-Martin, X., Doppelhofer, G. and Miller, R. I. (2004), 'Determinants of long-term growth: A bayesian averaging of classical estimates (bace) approach', *American Economic Review* **94**(4), 813–835.

Salimans, T. (2012), 'Variable selection and functional form uncertainty in cross-country growth regressions', *Journal of Econometrics* **171**(2), 267–280.

Shao, X. (2010), 'The dependent wild bootstrap', *Journal of the American Statistical Association* **105**(489), 218–235.

Soto, M. (2003), 'Taxing capital flows: An empirical comparative analysis', *Journal of Development Economics* **72**(1), 203–221.

Su, L., Chen, Y. and Ullah, A. (2009), 'Functional coefficient estimation with both categorical and continuous data', *Nonparametric Econometric Methods (Advances in Econometrics)* **25**, 131–167.

Online Supplementary Appendix to "Estimation and Inference for Three-Dimensional Panel Data Models"

Guohua Feng*, Jiti Gao†, Fei Liu♯ and Bin Peng†

*University of North Texas, †Monash University and ♯Nankai University

The appendix is organized as follows. In Appendix A.1, we outline the structure of the proofs. Appendix A.2 provides the preliminary lemmas. The proofs are provided in Appendix A.3.

A.1 Outline of the Derivation

The outline of the theoretical development is as follows:

1. We first work with the homogeneous case and establish some preliminary results in Lemma A.2, which are then used to establish Lemma 2.1. After that, we derive the results in Lemma 2.2 and Theorem 2.1 assuming that l^G and l_i^S's are given. It then completes the proofs for Sections 2.1 and 2.2.

2. Having established the asymptotic distribution, we work on the bias corrected estimator (i.e., the proof of Theorem 2.2), and we then complete the proof of Section 2.3. Next, we relax the assumption about l^G and l_i^S's, and show how to estimate them in practice. In order to estimate l^G and l_i^S's, we first derive two additional lemmas (Lemma A.3 and Lemma A.4), and then provide the proof for Theorem 2.3, which concludes our study on the homogeneous case.

3. Finally, we work with the heterogeneous case.

A.2 Preliminary Lemmas

Lemma A.1. *Suppose that* $\mathbf{A}$ *and* $\mathbf{A}+\mathbf{E}$ *are* $n\times n$ *symmetric matrices and that* $\mathbf{Q}=(\mathbf{Q}_1,\mathbf{Q}_2)$*, where* $\mathbf{Q}_1$ *is* $n\times r$ *and* $\mathbf{Q}_2$ *is* $n\times(n-r)$*, is an orthogonal matrix such that* $\mathrm{span}(\mathbf{Q}_1)$ *is an invariant subspace for* $\mathbf{A}$*. Decompose* $\mathbf{Q}^\top\mathbf{A}\mathbf{Q}$ *and* $\mathbf{Q}^\top\mathbf{E}\mathbf{Q}$ *as* $\mathbf{Q}^\top\mathbf{A}\mathbf{Q}=\mathrm{diag}(\mathbf{D}_1,\mathbf{D}_2)$ *and* $\mathbf{Q}^\top\mathbf{E}\mathbf{Q}=\{\mathbf{E}_{ij}\}_{2\times 2}$*. Let* $\mathrm{sep}(\mathbf{D}_1,\mathbf{D}_2)=\min_{\lambda_1\in\lambda(\mathbf{D}_1),\ \lambda_2\in\lambda(\mathbf{D}_2)}|\lambda_1-\lambda_2|$*. If* $\mathrm{sep}(\mathbf{D}_1,\mathbf{D}_2)>0$ *and* $\|\mathbf{E}\|_2\le\mathrm{sep}(\mathbf{D}_1,\mathbf{D}_2)/5$*, then there exists a* $(n-r)\times r$ *matrix* $\mathbf{P}$ *with* $\|\mathbf{P}\|_2\le 4\|\mathbf{E}_{21}\|_2/\mathrm{sep}(\mathbf{D}_1,\mathbf{D}_2)$*, such that the columns of* $\mathbf{Q}_1^0=(\mathbf{Q}_1+\mathbf{Q}_2\mathbf{P})(\mathbf{I}_r+\mathbf{P}^\top\mathbf{P})^{-1/2}$ *define an orthonormal basis for a subspace that is invariant for* $\mathbf{A}+\mathbf{E}$*.*

Lemma A.2. *Under Assumption 1, as* $(L,N,T)\to(\infty,\infty,\infty)$*, the following results hold:*

1. $\sup_{\mathbf{C}}\frac{1}{LNT}\sum_{i=1}^L\sum_{j=1}^N\mathbf{E}_{ij\cdot}^\top\mathbf{P}_{\mathbf{C}_i}\mathbf{E}_{ij\cdot}=O_P\left(\frac{1}{T}\vee\frac{1}{N}\right)$,

2. $\sup_{\mathbf{C}} \left| \frac{1}{LNT} \sum_{i=1}^{L} \sum_{j=1}^{N} \mathbf{X}_{ij.}^{\top} \mathbf{P}_{\mathbf{C}_i} \mathbf{E}_{ij.} \right| = O_P\left(\frac{1}{\sqrt{T}} \vee \frac{1}{\sqrt{N}} \right)$,

3. $\sup_{\mathbf{C}} \left| \frac{1}{LNT} \sum_{i=1}^{L} \sum_{j=1}^{N} \boldsymbol{\gamma}_{ij}^{G\top} \mathbf{F}^{G\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{E}_{ij.} \right| = O_P\left(\frac{1}{\sqrt{T}} \vee \frac{1}{\sqrt{N}} \right)$,

4. $\sup_{\mathbf{C}} \left| \frac{1}{LNT} \sum_{i=1}^{L} \sum_{j=1}^{N} \boldsymbol{\gamma}_{ij}^{S\top} \mathbf{F}_i^{S\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{E}_{ij.} \right| = O_P\left(\frac{1}{\sqrt{T}} \vee \frac{1}{\sqrt{N}} \right)$.

Lemma A.3. *Let $\widehat{\lambda}_1^G, \ldots, \widehat{\lambda}_{d_{\max}}^G$ be the $d_{\max}$ largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}^G$, $\widehat{\mathbf{C}}^G$ include the eigenvectors corresponding to $\widetilde{\mathbf{V}}^G = \operatorname{diag}\{\widehat{\lambda}_1^G, \ldots, \widehat{\lambda}_{l^G}^G\}$, and $\lambda_\ell^G = \frac{1}{T} \mathbf{H}_\ell^{G\top} \mathbf{F}^{G\top} \boldsymbol{\Sigma}^G \mathbf{F}^G \mathbf{H}_\ell^G$ with $\mathbf{H}^G = (\mathbf{H}_1^G, \ldots, \mathbf{H}_{l^G}^G) = \frac{1}{LNT} \boldsymbol{\Gamma}^{G\top} \boldsymbol{\Gamma}^G \cdot \mathbf{F}^{G\top} \widehat{\mathbf{C}}^G (\widetilde{\mathbf{V}}^G)^{-1}$. Then under Assumptions 1, 2.1 and 3.1, as $(L, N, T) \to (\infty, \infty, \infty)$, we have the following results:*

1. $|\widehat{\lambda}_\ell^G - \lambda_\ell^G| = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T \wedge L}} \right)$ *for $\ell = 1, \ldots, l^G$,*

2. $\widehat{\lambda}_\ell^G = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|^2 + \frac{\log(LN)}{T \wedge L} \right)$ *for $\ell = l^G + 1, \ldots, d_{\max}$.*

Lemma A.4. *Let $\widehat{\lambda}_{i,1}^S, \ldots, \widehat{\lambda}_{i,d_{\max}}^S$ be the $d_{\max}$ largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}_i^S$, $\widehat{\mathbf{C}}_i^S$ include the eigenvectors corresponding to $\widetilde{\mathbf{V}}_i^S = \operatorname{diag}\{\widehat{\lambda}_{i,1}^S, \ldots, \widehat{\lambda}_{i,l_i^S}^S\}$, and $\lambda_{i,\ell}^S = \frac{1}{T} \mathbf{H}_{i,\ell}^{S\top} \mathbf{F}_i^{S\top} \boldsymbol{\Sigma}_i^S \mathbf{F}_i^S \mathbf{H}_{i,\ell}^S$ with $\mathbf{H}_i^S = (\mathbf{H}_{i,1}^S, \ldots, \mathbf{H}_{i,l_i^S}^S) = \frac{1}{NT} \boldsymbol{\Gamma}_{i.}^{S\top} \boldsymbol{\Gamma}_{i.}^S \cdot \mathbf{F}_i^{S\top} \widehat{\mathbf{C}}_i^S (\widetilde{\mathbf{V}}_i^S)^{-1}$. Then under Assumptions 1, 2.1, and 3, as $(L, N, T) \to (\infty, \infty, \infty)$, we have the following results for $\forall i \in [L]$,*

1. $|\widehat{\lambda}_{i,\ell}^S - \lambda_{i,\ell}^S| = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T \wedge L}} + T^{\nu} \right)$ *for $\ell = 1, \ldots, l_i^S$,*

2. $\widehat{\lambda}_{i,\ell}^S = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|^2 + \frac{\log(LN)}{T \wedge L} + T^{2\nu} \right)$ *for $\ell = l_i^S + 1, \ldots, d_{\max}$.*

A.3 Proofs

Proof of Lemma A.1:

The proof is given in Theorem 8.1.10 of Golub and Van Loan (2013), and is therefore omitted. ■

Proof of Lemma A.2:

(1). Write

$$
\begin{aligned}
&\sup_{\mathbf{C}} \frac{1}{LNT} \sum_{i=1}^{L} \sum_{j=1}^{N} \mathbf{E}_{ij.}^{\top} \mathbf{P}_{\mathbf{C}_i} \mathbf{E}_{ij.} = \sup_{\mathbf{C}} \operatorname{Tr} \left\{ \frac{1}{LNT} \sum_{i=1}^{L} \sum_{j=1}^{N} \mathbf{P}_{\mathbf{C}_i} \mathbf{E}_{ij.} \mathbf{E}_{ij.}^{\top} \right\} \\
&= \sup_{\mathbf{C}} \frac{1}{LNT} \sum_{i=1}^{L} \operatorname{Tr} \left\{ \mathbf{P}_{\mathbf{C}_i} \mathbf{E}_{i..}^{\top} \mathbf{E}_{i..} \right\} \leq O(1) \sup_{\mathbf{C}} \frac{1}{LNT} \sum_{i=1}^{L} \|\mathbf{P}_{\mathbf{C}_i}\|_2 \cdot \|\mathbf{E}_{i..}\|_2^2 \\
&= O_P\left(\frac{1}{T} \vee \frac{1}{N} \right),
\end{aligned}
$$

where $\mathbf{E}_{i..}$ has been defined in Assumption 1.1, the first inequality follows from $|\operatorname{Tr}\{\mathbf{A}\}| \leq \operatorname{rank}(\mathbf{A}) \cdot \|\mathbf{A}\|_2$, and the last equality follows from Assumption 1.1 and the fact that $\|\mathbf{P}_{\mathbf{C}_i}\|_2 = 1$.

(2). As d is a fixed positive integer, without loss of generality suppose that $d = 1$ (an assumption that is used only for this result).

$$\sup_{\mathbf{C}} \left| \frac{1}{LNT} \sum_{i=1}^{L}\sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^{\top} \mathbf{P}_{\mathbf{C}_i} \mathbf{E}_{ij\cdot} \right| = \sup_{\mathbf{C}} \left| \mathrm{Tr}\left\{ \frac{1}{LNT} \sum_{i=1}^{L}\sum_{j=1}^{N} \mathbf{P}_{\mathbf{C}_i} \mathbf{E}_{ij\cdot} \mathbf{X}_{ij\cdot}^{\top} \right\} \right|$$
$$= \sup_{\mathbf{C}} \left| \mathrm{Tr}\left\{ \frac{1}{LNT} \sum_{i=1}^{L} \mathbf{P}_{\mathbf{C}_i} \mathbf{E}_{i\cdot\cdot}^{\top} \mathbf{X}_{i\cdot\cdot} \right\} \right| \le O(1) \sup_{\mathbf{C}} \frac{1}{LNT} \sum_{i=1}^{L} \|\mathbf{P}_{\mathbf{C}_i}\|_2 \cdot \|\mathbf{E}_{i\cdot\cdot}\|_2 \cdot \|\mathbf{X}_{i\cdot\cdot}\|_2$$
$$\le O_P\left(\frac{1}{\sqrt{T}} \vee \frac{1}{\sqrt{N}} \right),$$

where $\mathbf{X}_{i\cdot\cdot} = (\mathbf{X}_{i1\cdot}, \ldots, \mathbf{X}_{iN\cdot})^{\top}$, the first inequality follows from $|\mathrm{Tr}\{\mathbf{A}\}| \le \mathrm{rank}(\mathbf{A}) \cdot \|\mathbf{A}\|_2$, and the second inequality follows from $\|\mathbf{P}_{\mathbf{C}_i}\|_2 = 1$ and Assumption 1.1.

(3). Write

$$\sup_{\mathbf{C}} \left| \frac{1}{LNT} \sum_{i=1}^{L}\sum_{j=1}^{N} \boldsymbol{\gamma}_{ij}^{G\top} \mathbf{F}^{G\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{E}_{ij\cdot} \right| = \sup_{\mathbf{C}} \left| \mathrm{Tr}\left\{ \frac{1}{LNT} \sum_{i=1}^{L} \mathbf{F}^{G\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{E}_{i\cdot\cdot}^{\top} \boldsymbol{\Gamma}_{i\cdot}^{G} \right\} \right|$$
$$\le O(1) \sup_{\mathbf{C}} \frac{1}{LNT} \sum_{i=1}^{L} \|\mathbf{F}^{G}\|_2 \cdot \|\mathbf{M}_{\mathbf{C}_i}\|_2 \cdot \|\mathbf{E}_{i\cdot\cdot}\|_2 \cdot \|\boldsymbol{\Gamma}_{i\cdot}^{G}\|_2 = O_P\left(\frac{1}{\sqrt{T}} \vee \frac{1}{\sqrt{N}} \right),$$

where $\mathbf{F}^G$ has been defined in (2.3), $\boldsymbol{\Gamma}_{i\cdot}^{G}$ has been defined in Assumption 1, the first inequality follows from $|\mathrm{Tr}\{\mathbf{A}\}| \le \mathrm{rank}(\mathbf{A}) \cdot \|\mathbf{A}\|_2$, and the last equality follows from Assumptions 1.1-1.3. Based on the above development, the result follows.

(4). The fourth result can be proved in a similar way as for the third result. ■

Proof of Lemma 2.1:

In what follows, we let $\boldsymbol{\gamma}_{ij} = (\boldsymbol{\gamma}_{ij}^{G\top}, \boldsymbol{\gamma}_{ij}^{S\top})^{\top}$, and $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_L)$. The rest of notation has been defined in Section 2.1, so we do not repeat here.

We expand $Q(\mathbf{b}, \mathbf{C})$ as follows:

$$\begin{aligned} Q(\mathbf{b}, \mathbf{C}) = & \sum_{i=1}^{L}\sum_{j=1}^{N} (\boldsymbol{\beta} - \mathbf{b})^{\top} \mathbf{X}_{ij\cdot}^{\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{X}_{ij\cdot} (\boldsymbol{\beta} - \mathbf{b}) + \sum_{i=1}^{L}\sum_{j=1}^{N} \boldsymbol{\gamma}_{ij}^{\top} \mathbf{F}_i^{\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{F}_i \boldsymbol{\gamma}_{ij} \\ & + \sum_{i=1}^{L}\sum_{j=1}^{N} \mathbf{E}_{ij\cdot}^{\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{E}_{ij\cdot} + 2 \sum_{i=1}^{L}\sum_{j=1}^{N} (\boldsymbol{\beta} - \mathbf{b})^{\top} \mathbf{X}_{ij\cdot}^{\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{F}_i \boldsymbol{\gamma}_{ij} \\ & + 2 \sum_{i=1}^{L}\sum_{j=1}^{N} (\boldsymbol{\beta} - \mathbf{b})^{\top} \mathbf{X}_{ij\cdot}^{\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{E}_{ij\cdot} + 2 \sum_{i=1}^{L}\sum_{j=1}^{N} \boldsymbol{\gamma}_{ij}^{\top} \mathbf{F}_i^{\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{E}_{ij\cdot}. \end{aligned}$$

Using Lemma A.2, we obtain

$$\frac{1}{LNT} Q(\mathbf{b}, \mathbf{C}) - \frac{1}{LNT} Q(\boldsymbol{\beta}, \mathbf{F})$$
$$= \frac{1}{LNT} \sum_{i=1}^{L}\sum_{j=1}^{N} (\boldsymbol{\beta} - \mathbf{b})^{\top} \mathbf{X}_{ij\cdot}^{\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{X}_{ij\cdot} (\boldsymbol{\beta} - \mathbf{b}) + \frac{1}{LNT} \sum_{i=1}^{L}\sum_{j=1}^{N} \boldsymbol{\gamma}_{ij}^{\top} \mathbf{F}_i^{\top} \mathbf{M}_{\mathbf{C}_i} \mathbf{F}_i \boldsymbol{\gamma}_{ij}$$

$$+\frac{2}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\boldsymbol{\beta}-\mathbf{b})^\top\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{C}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij}+\frac{2}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\boldsymbol{\beta}-\mathbf{b})^\top\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{C}_i}\mathbf{E}_{ij\cdot}$$
$$+O_P\left(\frac{1}{\sqrt{T}}\vee\frac{1}{\sqrt{N}}\right). \tag{A.1}$$

We now focus on the right hand side of (A.1). Since $\boldsymbol{\beta}$ belongs to $\mathbb{R}^d$, we consider two cases below:

$$\text{Case 1: } \|\boldsymbol{\beta}-\mathbf{b}\|\le c \quad \text{and} \quad \text{Case 2: } \|\boldsymbol{\beta}-\mathbf{b}\|> c,$$

where c is a large positive constant. Note that for Case 1, using Lemma A.2 and Assumption 1.1, equation (A.1) can further be simplified as follows:

$$\begin{aligned}
&\frac{1}{LNT}Q(\mathbf{b},\mathbf{C})-\frac{1}{LNT}Q(\boldsymbol{\beta},\mathbf{F})\\
&=\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\boldsymbol{\beta}-\mathbf{b})^\top\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{C}_i}\mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\mathbf{b})+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\boldsymbol{\gamma}_{ij}^\top\mathbf{F}_i^\top\mathbf{M}_{\mathbf{C}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij}\\
&\quad+\frac{2}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\boldsymbol{\beta}-\mathbf{b})^\top\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{C}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij}+O_P\left(\frac{1}{\sqrt{T}}\vee\frac{1}{\sqrt{N}}\right)\\
&=(\boldsymbol{\beta}-\mathbf{b})^\top\frac{1}{LNT}\sum_{i=1}^{L}\mathbf{D}_i(\boldsymbol{\beta}-\mathbf{b})+\frac{1}{L}\sum_{i=1}^{L}\boldsymbol{\theta}_i^\top\mathbf{B}_i\boldsymbol{\theta}_i+O_P\left(\frac{1}{\sqrt{T}}\vee\frac{1}{\sqrt{N}}\right),
\end{aligned}\tag{A.2}$$

where $\mathbf{D}_i$ is defined in Assumption 1, and $\mathbf{B}_i$ and $\boldsymbol{\theta}_i$ are defined in the same fashion as those on page 1265 of Bai (2009). By (A.2) and Assumption 1.4, it is easy to see that $\widehat{\mathbf{b}}-\boldsymbol{\beta}=o_P(1)$ if we show that $\widehat{\mathbf{b}}$ cannot belong to Case 2, which is exactly what we are about to do.

For Case 2, we write (A.1) as follows:

$$\begin{aligned}
&\frac{1}{LNT}Q(\mathbf{b},\mathbf{C})-\frac{1}{LNT}Q(\boldsymbol{\beta},\mathbf{F})\\
&=(\boldsymbol{\beta}-\mathbf{b})^\top\frac{1}{LNT}\sum_{i=1}^{L}\mathbf{D}_i(\boldsymbol{\beta}-\mathbf{b})+\frac{1}{L}\sum_{i=1}^{L}\boldsymbol{\theta}_i^\top\mathbf{B}_i\boldsymbol{\theta}_i\\
&\quad+\frac{2}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\boldsymbol{\beta}-\mathbf{b})^\top\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{C}_i}\mathbf{E}_{ij\cdot}+O_P\left(\frac{1}{\sqrt{T}}\vee\frac{1}{\sqrt{N}}\right)\\
&\ge a_0c^2+\frac{1}{L}\sum_{i=1}^{L}\boldsymbol{\theta}_i^\top\mathbf{B}_i\boldsymbol{\theta}_i+\frac{2}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\boldsymbol{\beta}-\mathbf{b})^\top\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{C}_i}\mathbf{E}_{ij\cdot}\\
&\quad+O_P\left(\frac{1}{\sqrt{T}}\vee\frac{1}{\sqrt{N}}\right),
\end{aligned}\tag{A.3}$$

where a_0 is a positive constant by Assumption 1.4. Apparently, $\widehat{\mathbf{b}}$ cannot belong to the Case 2 by comparing the right hand sides of (A.1) and (A.3).

Noting that in the above development, we can tight the value of c to $c_1\left(\frac{1}{\sqrt{T}}\vee\frac{1}{\sqrt{N}}\right)$, where c_1 is a large positive constant. It then yields the rate that we aim to achieve. The proof is now completed.

(2). By (2.5), we write

$$
\begin{aligned}
\widehat{\mathbf{C}}_i\widehat{\mathbf{V}}_i &= \frac{1}{NT}\sum_{j=1}^N \mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}^\top\widehat{\mathbf{C}}_i \\
&+\frac{1}{NT}\sum_{j=1}^N \mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\boldsymbol{\gamma}_{ij}^\top\mathbf{F}_i^\top\widehat{\mathbf{C}}_i+\frac{1}{NT}\sum_{j=1}^N \mathbf{F}_i\boldsymbol{\gamma}_{ij}(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}^\top\widehat{\mathbf{C}}_i \\
&+\frac{1}{NT}\sum_{j=1}^N \mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\mathbf{E}_{ij\cdot}^\top\widehat{\mathbf{C}}_i+\frac{1}{NT}\sum_{j=1}^N \mathbf{E}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}^\top\widehat{\mathbf{C}}_i \\
&+\frac{1}{NT}\sum_{j=1}^N \mathbf{F}_i\boldsymbol{\gamma}_{ij}\boldsymbol{\gamma}_{ij}^\top\mathbf{F}_i^\top\widehat{\mathbf{C}}_i+\frac{1}{NT}\sum_{j=1}^N \mathbf{F}_i\boldsymbol{\gamma}_{ij}\mathbf{E}_{ij\cdot}^\top\widehat{\mathbf{C}}_i \\
&+\frac{1}{NT}\sum_{j=1}^N \mathbf{E}_{ij\cdot}\boldsymbol{\gamma}_{ij}^\top\mathbf{F}_i^\top\widehat{\mathbf{C}}_i+\frac{1}{NT}\sum_{j=1}^N \mathbf{E}_{ij\cdot}\mathbf{E}_{ij\cdot}^\top\widehat{\mathbf{C}}_i \\
&:= \mathbf{J}_{i,1}+\cdots+\mathbf{J}_{i,9},
\end{aligned}\tag{A.4}
$$

where the definitions of $\mathbf{J}_{i,1}$ to $\mathbf{J}_{i,9}$ are self evident.

For $\mathbf{J}_{i,1}$, write

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\|\mathbf{J}_{i,1}\| &= \frac{1}{\sqrt{T}}\left\|\frac{1}{NT}\sum_{j=1}^N \mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}^\top\widehat{\mathbf{C}}_i\right\| \\
&\le O(1)\frac{1}{NT}\sum_{j=1}^N\|\mathbf{X}_{ij\cdot}\|^2\cdot\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|^2 = O_P(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|^2),
\end{aligned}\tag{A.5}
$$

where the first inequality follows from the triangle inequality and the fact that $\frac{1}{T}\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i = \mathbf{I}_{l^G+l_i^S}$, and the second equality follows from the fact that $\frac{1}{NT}\sum_{j=1}^N\|\mathbf{X}_{ij\cdot}\|^2 = O_P(1)$ by Assumption 1.1.

Similarly, we can obtain that

$$
\frac{1}{\sqrt{T}}\sum_{m=2}^5\|\mathbf{J}_{i,m}\| = O_P(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|).
$$

For $\mathbf{J}_{i,7}$, write

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\|\mathbf{J}_{i,7}\|_2 &= \frac{1}{\sqrt{T}}\left\|\frac{1}{NT}\sum_{j=1}^N \mathbf{F}_i\boldsymbol{\gamma}_{ij}\mathbf{E}_{ij\cdot}^\top\widehat{\mathbf{C}}_i\right\|_2 \le O(1)\frac{1}{NT}\|\mathbf{F}_i\boldsymbol{\Gamma}_{i\cdot}^\top\mathbf{E}_{i\cdot\cdot}\|_2 \\
&= O_P\left(\frac{\|\mathbf{E}_{i\cdot\cdot}\|_2}{\sqrt{NT}}\right),
\end{aligned}\tag{A.6}
$$

where the first inequality follows from the fact that $\frac{1}{T}\widehat{\mathbf{C}}_i^\top \widehat{\mathbf{C}}_i = \mathbf{I}_{l^G + l_i^S}$, and the second equality follows from the fact that $\|\mathbf{F}_i\| = O_P(\sqrt{T})$ and $\|\mathbf{\Gamma}_{i\cdot}\| = O_P(\sqrt{N})$ by Assumptions 1.2-1.3 respectively. Similarly, we have

$$\frac{1}{\sqrt{T}}\|\mathbf{J}_{i,8}\|_2 = O_P\left(\frac{\|\mathbf{E}_{i\cdot\cdot}\|_2}{\sqrt{NT}}\right) \quad \text{and} \quad \frac{1}{\sqrt{T}}\|\mathbf{J}_{i,9}\|_2 = O_P\left(\frac{\|\mathbf{E}_{i\cdot\cdot}\|_2^2}{NT}\right). \tag{A.7}$$

Thus, we can conclude for $\forall i \in [L]$

$$\begin{aligned}\frac{1}{\sqrt{T}}\|\widehat{\mathbf{C}}_i - \mathbf{F}_i\mathbf{H}_i\|_2 &= O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\|\mathbf{E}_{i\cdot\cdot}\|_2}{\sqrt{NT}}\right)\\ &= O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{1}{\sqrt{N \wedge T}}\right),\end{aligned} \tag{A.8}$$

where $\mathbf{H}_i = \frac{1}{NT}\mathbf{\Gamma}_{i\cdot}^\top \mathbf{\Gamma}_{i\cdot}\mathbf{F}_i^\top \widehat{\mathbf{C}}_i \widehat{\mathbf{V}}_i^{-1}$, and the second equality follows from Assumption 1.1. In connection with the fact that

$$\begin{aligned}\mathbf{P}_{\mathbf{F}_i} = &(\mathbf{F}_i - \widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} + \widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})\\ &\cdot[(\mathbf{F}_i - \widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} + \widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})^\top(\mathbf{F}_i - \widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} + \widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})]^{-1}(\mathbf{F}_i - \widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} + \widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})^\top,\end{aligned}$$

Lemma 2.1.(2) follows immediately. ■

Proof of Lemma 2.2:

Note that $\widehat{\mathbf{b}}$ can be expanded as follows:

$$\begin{aligned}\widehat{\mathbf{b}} = \boldsymbol{\beta} &+ \left(\sum_{i=1}^L\sum_{j=1}^N \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{X}_{ij\cdot}\right)^{-1}\sum_{i=1}^L\sum_{j=1}^N \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{E}_{ij\cdot}\\ &+ \left(\sum_{i=1}^L\sum_{j=1}^N \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{X}_{ij\cdot}\right)^{-1}\sum_{i=1}^L\sum_{j=1}^N \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij}.\end{aligned}$$

We then start with the term $\frac{1}{LNT}\sum_{i=1}^L\sum_{j=1}^N \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij}$, which can be expanded as follows:

$$\begin{aligned}\frac{1}{LNT}\sum_{i=1}^L\sum_{j=1}^N \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij} &= -\frac{1}{LNT}\sum_{i=1}^L\sum_{j=1}^N \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}(\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} - \mathbf{F}_i)\boldsymbol{\gamma}_{ij}\\ &= -\frac{1}{LNT}\sum_{i=1}^L\sum_{j=1}^N \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\sum_{\ell=1,\ell\neq 6}^{9}\mathbf{J}_{i,\ell}\mathbf{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij}\\ &:= -(\mathbf{A}_1 + \cdots + \mathbf{A}_8),\end{aligned} \tag{A.9}$$

where $\mathbf{J}_{i,\ell}$'s have been defined in (A.4), $\mathbf{\Pi}_i^{-1} = (\frac{1}{T}\mathbf{F}_i^\top\widehat{\mathbf{C}}_i)^{-1}(\frac{1}{N}\mathbf{\Gamma}_{i\cdot}^\top\mathbf{\Gamma}_{i\cdot})^{-1}$, and the definitions of $\mathbf{A}_\ell$'s should be obvious. Also, simple algebra shows that

$$\left\|\frac{1}{LNT}\sum_{i=1}^L\sum_{j=1}^N \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij}\right\|_2 = \left\|\frac{1}{LNT}\sum_{i=1}^L\sum_{j=1}^N \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}(\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} - \mathbf{F}_i)\boldsymbol{\gamma}_{ij}\right\|_2$$

$$= O_P(1)\frac{1}{L}\sum_{i=1}^{L}\frac{1}{\sqrt{T}}\|\mathbf{M}_{\widehat{\mathbf{C}}_i}(\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} - \mathbf{F}_i)\|_2.$$

Then any term on the right-hand side of (A.9) will be negligible, if we can show that it is

$$o_P(1)\frac{1}{L}\sum_{i=1}^{L}\frac{1}{\sqrt{T}}\|\mathbf{M}_{\widehat{\mathbf{C}}_i}(\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} - \mathbf{F}_i)\|_2. \tag{A.10}$$

The reason is that we can keep expanding the term $\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} - \mathbf{F}_i$ recursively.

It is easy to see that $\|\mathbf{A}_1\|_2 = o_P(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|)$ using Assumptions 1.1-1.3, and Lemma 2.1. Thus, we start from $\mathbf{A}_2$, and write

$$\begin{aligned}\mathbf{A}_2 &= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^{\top}\mathbf{M}_{\widehat{\mathbf{C}}_i}\frac{1}{NT}\sum_{j_2=1}^{N}\mathbf{X}_{ij_2\cdot}(\boldsymbol{\beta} - \widehat{\mathbf{b}})\boldsymbol{\gamma}_{ij_2}^{\top}\mathbf{F}_i^{\top}\widehat{\mathbf{C}}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1}\\ &= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\mathbf{X}_{ij_1\cdot}^{\top}\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{X}_{ij_2\cdot}\boldsymbol{\gamma}_{ij_2}^{\top}(\boldsymbol{\Gamma}_{i\cdot}^{\top}\boldsymbol{\Gamma}_{i\cdot})^{-1}\boldsymbol{\gamma}_{ij_1}(\boldsymbol{\beta} - \widehat{\mathbf{b}}).\end{aligned}$$

We will come back to $\mathbf{A}_2$ later on.

For $\mathbf{A}_3$, write

$$\begin{aligned}\|\mathbf{A}_3\|_2 &= \left\|\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^{\top}\mathbf{M}_{\widehat{\mathbf{C}}_i}\frac{1}{NT}\sum_{j_2=1}^{N}\mathbf{F}_i\boldsymbol{\gamma}_{ij_2}(\boldsymbol{\beta} - \widehat{\mathbf{b}})^{\top}\mathbf{X}_{ij_2\cdot}^{\top}\widehat{\mathbf{C}}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1}\right\|_2\\ &\le \frac{\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|}{LNT^2}\sum_{i=1}^{L}\frac{1}{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\|\mathbf{M}_{\widehat{\mathbf{C}}_i}(\mathbf{F}_i - \widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})\|_2\|\mathbf{X}_{ij_1\cdot}\|_2\|\boldsymbol{\gamma}_{ij_2}\|_2\|\mathbf{X}_{ij_2\cdot}^{\top}\widehat{\mathbf{C}}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1}\|_2\\ &= O_P(1)\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|\frac{1}{L}\sum_{i=1}^{L}\frac{1}{\sqrt{T}}\|\mathbf{M}_{\widehat{\mathbf{C}}_i}(\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} - \mathbf{F}_i)\|_2,\end{aligned} \tag{A.11}$$

where the second equality can be easily shown using Assumption 1. Using Lemma 2.1 and the arguments made for (A.10), it is straightforward to show that the term $\mathbf{A}_3$ is negligible.

For $\mathbf{A}_4$, write

$$\begin{aligned}\|\mathbf{A}_4\|_2 &= \left\|\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^{\top}\mathbf{M}_{\widehat{\mathbf{C}}_i}\frac{1}{NT}\sum_{j_2=1}^{N}\mathbf{X}_{ij_2\cdot}(\boldsymbol{\beta} - \widehat{\mathbf{b}})\mathbf{E}_{ij_2\cdot}^{\top}\widehat{\mathbf{C}}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1}\right\|_2\\ &= \left\|\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^{\top}\mathbf{M}_{\widehat{\mathbf{C}}_i}\frac{1}{NT}\mathbf{X}_{i,\boldsymbol{\beta}-\widehat{\mathbf{b}}}^{\top}\mathbf{E}_{i\cdot\cdot}\widehat{\mathbf{C}}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij}\right\|_2\\ &\le \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\|\mathbf{X}_{ij\cdot}^{\top}\mathbf{M}_{\widehat{\mathbf{C}}_i}\|_2\cdot\frac{1}{NT}\|\mathbf{X}_{i,\boldsymbol{\beta}-\widehat{\mathbf{b}}}^{\top}\mathbf{E}_{i\cdot\cdot}\mathbf{F}_i\mathbf{H}_i\|_2\cdot\|\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij}\|_2\\ &\quad+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\|\mathbf{X}_{ij\cdot}^{\top}\mathbf{M}_{\widehat{\mathbf{C}}_i}\|_2\cdot\frac{1}{NT}\|\mathbf{X}_{i,\boldsymbol{\beta}-\widehat{\mathbf{b}}}^{\top}\mathbf{E}_{i\cdot\cdot}(\widehat{\mathbf{C}}_i - \mathbf{F}_i\mathbf{H}_i)\|_2\cdot\|\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij}\|_2\end{aligned}$$

$$
\begin{aligned}
&= o_P(1)\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\| \\
&\quad + O_P(1)\frac{\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\|\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\|_2\cdot\frac{1}{N\sqrt{T}}\sqrt{NT}\|\mathbf{E}_{i\cdot\cdot}\|_2\cdot\frac{1}{\sqrt{T}}\|\widehat{\mathbf{C}}_i-\mathbf{F}_i\mathbf{H}_i\|_2 \\
&= o_P(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|), \qquad (A.12)
\end{aligned}
$$

where $\mathbf{X}_{i,\boldsymbol{\beta}-\widehat{\mathbf{b}}} = (\mathbf{X}_{i1\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}}),\ldots,\mathbf{X}_{iN\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}}))^\top$, the third equality follows from Assumption 2.1, and the last equality follows from Assumption 1.1 and the first result of the theorem. Similarly, we can show that

$$
\|\mathbf{A}_5\|_2 = o_P(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|). \qquad (A.13)
$$

We now consider $\mathbf{A}_6$. Write

$$
\begin{aligned}
\mathbf{A}_6 &= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\frac{1}{NT}\sum_{j_2=1}^{N}\mathbf{F}_i\boldsymbol{\gamma}_{ij_2}\mathbf{E}_{ij_2\cdot}^\top\widehat{\mathbf{C}}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1} \\
&= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}(\mathbf{F}_i-\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})\frac{1}{NT}\sum_{j_2=1}^{N}\boldsymbol{\gamma}_{ij_2}\mathbf{E}_{ij_2\cdot}^\top\mathbf{F}_i\mathbf{H}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1} \\
&\quad + \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}(\mathbf{F}_i-\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})\frac{1}{NT}\sum_{j_2=1}^{N}\boldsymbol{\gamma}_{ij_2}\mathbf{E}_{ij_2\cdot}^\top(\widehat{\mathbf{C}}_i-\mathbf{F}_i\mathbf{H}_i)\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1} \\
&:= \mathbf{A}_{6,1}+\mathbf{A}_{6,2}.
\end{aligned}
$$

First, note that

$$
\begin{aligned}
\sum_{i=1}^{L}\mathbb{E}\|\boldsymbol{\Gamma}_{i\cdot}^\top\mathbf{E}_{i\cdot\cdot}\mathbf{F}_i\|^2 &= \sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\sum_{t_1,t_2=1}^{T}E[\boldsymbol{\gamma}_{ij_1}^{G\top}\varepsilon_{ij_1t_1}\mathbf{f}_{it_1}^G\boldsymbol{\gamma}_{ij_2}^{G\top}\varepsilon_{ij_2t_2}\mathbf{f}_{it_2}^G] \\
&= O(LNT), \qquad (A.14)
\end{aligned}
$$

where the last equality follows from Assumption 2.1. Additionally, using the same expansion (A.4) and arguments that are analogous to those in the proof of (A.8), we can readily show

$$
\begin{aligned}
\frac{1}{T}\left\|\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}(\mathbf{F}_i-\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})\right\|_2 &\le \frac{1}{T}\left\|\mathbf{X}_{ij_1\cdot}^\top(\mathbf{P}_{\widehat{\mathbf{C}}_i}-\mathbf{P}_{\mathbf{F}_i})(\mathbf{F}_i-\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})\right\|_2 \\
&\quad + \frac{1}{T}\left\|\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\mathbf{F}_i}(\mathbf{F}_i-\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})\right\|_2 \\
&= O_P\left(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|+\frac{1}{N\wedge T}\right). \qquad (A.15)
\end{aligned}
$$

Thus, by (A.14) and (A.15), we are able to write

$$
\|\mathbf{A}_{6,1}\|_2 = \left\|\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}(\mathbf{F}_i-\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})\frac{1}{NT}\sum_{j_2=1}^{N}\boldsymbol{\gamma}_{ij_2}\mathbf{E}_{ij_2\cdot}^\top\mathbf{F}_i\mathbf{H}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1}\right\|_2
$$

$$\leq \frac{1}{LNT}\sum_{i=1}^{L}\frac{1}{NT}\sum_{j=1}^{N}\left\|\mathbf{X}_{ij\cdot}\mathbf{M}_{\widehat{\mathbf{C}}_i}(\mathbf{F}_i-\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})\right\|_2\cdot\|\boldsymbol{\Gamma}_{i\cdot}^\top\mathbf{E}_{i\cdot\cdot}\mathbf{F}_i\|_2\cdot\|\mathbf{H}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij}\|_2$$
$$= o_P\left(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|+\frac{1}{\sqrt{LNT}}\right),$$

where the last equality holds under the conditions $\frac{\sqrt{L}}{N}\to 0$ and $\frac{\sqrt{L}}{T}\to 0$. Analogously, we can obtain

$$\|\mathbf{A}_{6,2}\|_2 = o_P\left(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|+\frac{1}{\sqrt{LNT}}\right).$$

Therefore, it is easy to see that $\mathbf{A}_6$ is negligible.

For $\mathbf{A}_7$, we write

$$\mathbf{A}_7 = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\frac{1}{NT}\sum_{j_2=1}^{N}\mathbf{E}_{ij_2\cdot}\boldsymbol{\gamma}_{ij_2}^\top\mathbf{F}_i^\top\widehat{\mathbf{C}}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1}$$
$$= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{E}_{ij_2\cdot}\boldsymbol{\gamma}_{ij_2}^\top(\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot})^{-1}\boldsymbol{\gamma}_{ij_1}.$$

We will come back to this term later.

Finally, we consider $\mathbf{A}_8$.

$$\begin{aligned}\mathbf{A}_8 &= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\frac{1}{NT}\sum_{j_2=1}^{N}\mathbf{E}_{ij_2\cdot}\mathbf{E}_{ij_2\cdot}^\top\widehat{\mathbf{C}}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1}\\ &= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\frac{1}{NT}\sum_{j_2=1}^{N}\boldsymbol{\Sigma}_{\varepsilon,ij_2}\widehat{\mathbf{C}}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1}\\ &\quad+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\frac{1}{NT}\sum_{j_2=1}^{N}(\mathbf{E}_{ij_2\cdot}\mathbf{E}_{ij_2\cdot}^\top-\boldsymbol{\Sigma}_{\varepsilon,ij_2})\widehat{\mathbf{C}}_i\boldsymbol{\Pi}_i^{-1}\boldsymbol{\gamma}_{ij_1},\\ &:= \mathbf{A}_{8,1}+\mathbf{A}_{8,2},\end{aligned} \tag{A.16}$$

where $\boldsymbol{\Sigma}_{\varepsilon,ij} = E[\mathbf{E}_{ij\cdot}\mathbf{E}_{ij\cdot}^\top]$. We further note that

$$\mathbf{A}_{8,1} = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\boldsymbol{\Sigma}_{\varepsilon,ij_2}\widehat{\mathbf{C}}_i(\mathbf{F}_i^\top\widehat{\mathbf{C}}_i)^{-1}(\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot})^{-1}\boldsymbol{\gamma}_{ij_1}$$
$$= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\mathbf{X}_{ij_1\cdot}^\top\mathbf{M}_{\mathbf{F}_i}\boldsymbol{\Sigma}_{\varepsilon,ij_2}\mathbf{F}_i(\mathbf{F}_i^\top\mathbf{F}_i)^{-1}(\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot})^{-1}\boldsymbol{\gamma}_{ij_1}+o_P\left(\frac{1}{\sqrt{LNT}}\right),$$

where the second equality can be proved easily using (A.8) and the condition $\frac{L}{T}, \frac{LN}{T^2}\to 0$. Furthermore, applying the same procedure as used by Lemma A.5 of Bai (2009), we can see that $\mathbf{A}_{8,2}$ is negligible. Therefore,

$$\mathbf{A}_8 = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\mathbf{X}_{ij_1\bullet}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,ij_2}\mathbf{F}_i(\mathbf{F}_i^\top\mathbf{F}_i)^{-1}(\mathbf{\Gamma}_{i\bullet}^\top\mathbf{\Gamma}_{i\bullet})^{-1}\boldsymbol{\gamma}_{ij_1} + o_P\left(\frac{1}{\sqrt{LNT}}\right). \tag{A.17}$$

We now put everything together and obtain that

$$\begin{aligned}\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\bullet}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij} &= \widehat{\mathbf{\Sigma}}_2\cdot(\widehat{\mathbf{b}}-\boldsymbol{\beta})\\ &\quad -\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\mathbf{X}_{ij_1\bullet}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{E}_{ij_2\bullet}\boldsymbol{\gamma}_{ij_2}^\top(\mathbf{\Gamma}_{i\bullet}^\top\mathbf{\Gamma}_{i\bullet})^{-1}\boldsymbol{\gamma}_{ij_1}\\ &\quad -\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\mathbf{X}_{ij_1\bullet}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,ij_2}\mathbf{F}_i(\mathbf{F}_i^\top\mathbf{F}_i)^{-1}(\mathbf{\Gamma}_{i\bullet}^\top\mathbf{\Gamma}_{i\bullet})^{-1}\boldsymbol{\gamma}_{ij_1}\\ &\quad + \text{negligible terms},\end{aligned}$$

where $\widehat{\mathbf{\Sigma}}_2 = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\mathbf{X}_{ij_1\bullet}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{X}_{ij_2\bullet}\boldsymbol{\gamma}_{ij_2}^\top(\mathbf{\Gamma}_{i\bullet}^\top\mathbf{\Gamma}_{i\bullet})^{-1}\boldsymbol{\gamma}_{ij_1}$. Thus, we have

$$\widehat{\mathbf{b}}-\boldsymbol{\beta} = \widehat{\mathbf{\Sigma}}_1^{-1}\cdot\widehat{\mathbf{\Sigma}}_2\cdot(\widehat{\mathbf{b}}-\boldsymbol{\beta}) + \widehat{\mathbf{\Sigma}}_1^{-1}\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij\bullet}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{E}_{ij\bullet} - \widehat{\mathbf{\Sigma}}_1^{-1}\widetilde{\mathbf{W}} + \text{negligible terms},$$

where

$$\widehat{\mathbf{\Sigma}}_1 = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\bullet}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{X}_{ij\bullet},\quad \widetilde{\mathbf{W}} = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\mathbf{X}_{ij_1\bullet}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,ij_2}\mathbf{F}_i(\mathbf{F}_i^\top\mathbf{F}_i)^{-1}(\mathbf{\Gamma}_{i\bullet}^\top\mathbf{\Gamma}_{i\bullet})^{-1}\boldsymbol{\gamma}_{ij_1}.$$

Thus,

$$\widehat{\mathbf{b}}-\boldsymbol{\beta} = (\widehat{\mathbf{\Sigma}}_1-\widehat{\mathbf{\Sigma}}_2)^{-1}\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij\bullet}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{E}_{ij\bullet} - (\widehat{\mathbf{\Sigma}}_1-\widehat{\mathbf{\Sigma}}_2)^{-1}\widetilde{\mathbf{W}} + \text{negligible terms}.$$

We now concentrate on $\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij\bullet}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{E}_{ij\bullet}$. It is sufficient to focus on

$$\begin{aligned}&\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij\bullet}^\top(\mathbf{P}_{\widehat{\mathbf{C}}_i}-\mathbf{P}_{\mathbf{F}_i})\mathbf{E}_{ij\bullet}\\ =\ &\frac{1}{LNT^2}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij\bullet}^\top(\widehat{\mathbf{C}}_i-\mathbf{F}_i\mathbf{H}_i)\mathbf{H}_i^\top\mathbf{F}_i^\top\mathbf{E}_{ij\bullet}\\ &+\frac{1}{LNT^2}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij\bullet}^\top(\widehat{\mathbf{C}}_i-\mathbf{F}_i\mathbf{H}_i)(\widehat{\mathbf{C}}_i-\mathbf{F}_i\mathbf{H}_i)^\top\mathbf{E}_{ij\bullet}\\ &+\frac{1}{LNT^2}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij\bullet}^\top\mathbf{F}_i\mathbf{H}_i(\widehat{\mathbf{C}}_i-\mathbf{F}_i\mathbf{H}_i)^\top\mathbf{E}_{ij\bullet}\end{aligned}$$

$$\frac{1}{LNT^2}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^\top\mathbf{F}_i(\mathbf{H}_i\mathbf{H}_i^\top - T(\mathbf{F}_i^\top\mathbf{F}_i)^{-1})\mathbf{F}_i^\top\mathbf{E}_{ij.}$$
$$:= \mathbf{W}_1 + \mathbf{W}_2 + \mathbf{W}_3 + \mathbf{W}_4,$$

where the definitions of $\mathbf{W}_1$ to $\mathbf{W}_4$ should be obvious. In addition, let $\mathbf{W}_{\ell,k}$ be the k^{th} row of $\mathbf{W}_\ell$ for $\ell = 1,2,3,4$ below.

By expanding $\widehat{\mathbf{C}}_i - \mathbf{F}_i\mathbf{H}_i$ as for $\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij}$ above, one can show that $\|\mathbf{W}_{1,k}\| = o_P\left(\frac{1}{\sqrt{LNT}}\right)$. Similarly, we can show that $\|\mathbf{W}_{2,k}\| = o_P\left(\frac{1}{\sqrt{LNT}}\right)$ and $\|\mathbf{W}_{4,k}\| = o_P\left(\frac{1}{\sqrt{LNT}}\right)$. It remains to consider $\mathbf{W}_3$.

$$\mathbf{W}_3 = \frac{1}{LNT^2}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^\top\mathbf{F}_i\mathbf{H}_i(\widehat{\mathbf{C}}_i - \mathbf{F}_i\mathbf{H}_i)^\top\mathbf{E}_{ij.}$$
$$= \frac{1}{LNT^2}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^\top\mathbf{F}_i\mathbf{H}_i\mathbf{H}_i^\top(\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1} - \mathbf{F}_i)^\top\mathbf{E}_{ij.}.$$

This term can be expanded in the same way as for $\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij.}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij}$. Applying the same procedure as above, one can conclude that

$$\sqrt{LNT}\mathbf{W}_3 = \frac{\sqrt{T}L}{\sqrt{LN}}\cdot\frac{1}{L}\sum_{i=1}^{L}\frac{1}{N}\sum_{j_1,j_2=1}^{N}\frac{\mathbf{Z}_{ij_1.}^\top\mathbf{F}_i}{T}\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i.}^\top\boldsymbol{\Gamma}_{i.}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij_2}\frac{\mathbf{E}_{ij_2.}^\top\mathbf{E}_{ij_1.}}{T} + o_P(1)$$
$$= \frac{\sqrt{TL}}{\sqrt{N}}\cdot\frac{1}{L}\sum_{i=1}^{L}\frac{1}{N}\sum_{j_1,j_2=1}^{N}\frac{\mathbf{Z}_{ij_1.}^\top\mathbf{F}_i}{T}\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i.}^\top\boldsymbol{\Gamma}_{i.}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij_2}\sigma_{i,j_1j_2} + o_P(1),$$

where $\sigma_{i,j_1j_2} = E[\varepsilon_{ij_1t}\varepsilon_{ij_2t}]$.

Collecting the above results, the proof is completed. ■

Proof of Theorem 2.1:

We first study the bias correction for $\mathbf{a}_1$. Recall that

$$\widehat{\mathbf{a}}_1 = -\widehat{\boldsymbol{\Omega}}_2^{-1}\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{\widehat{\mathbf{Z}}_{ij_1.}^\top\widehat{\mathbf{C}}_i}{T}\left(\frac{\widehat{\boldsymbol{\Gamma}}_{i.}^\top\widehat{\boldsymbol{\Gamma}}_{i.}}{N}\right)^{-1}\widehat{\boldsymbol{\gamma}}_{ij_2}\widehat{\sigma}_{i,j_1j_2},$$

where $\widehat{\boldsymbol{\Omega}}_2 = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij.}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\widehat{\mathbf{Z}}_{ij.}$, $\widehat{\mathbf{Z}}_{ij.} = \mathbf{X}_{ij.} - \sum_{k=1}^{N}\mathbf{X}_{ik.}\widehat{\boldsymbol{\gamma}}_{ij}^\top(\widehat{\boldsymbol{\Gamma}}_{i.}^\top\widehat{\boldsymbol{\Gamma}}_{i.})^{-1}\widehat{\boldsymbol{\gamma}}_{ik}$ and $\widehat{\sigma}_{i,j_1j_2} = T^{-1}(\mathbf{Y}_{ij_1.} - \mathbf{X}_{ij_1.}\widehat{\mathbf{b}} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_1})^\top(\mathbf{Y}_{ij_2.} - \mathbf{X}_{ij_2.}\widehat{\mathbf{b}} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_2})$. For notational simplicity, we further define $a_{\gamma,ijk} = \boldsymbol{\gamma}_{ij}^\top(\boldsymbol{\Gamma}_{i.}^\top\boldsymbol{\Gamma}_{i.})^{-1}\boldsymbol{\gamma}_{ik}$ and $\widehat{a}_{\gamma,ijk} = \widehat{\boldsymbol{\gamma}}_{ij}^\top(\widehat{\boldsymbol{\Gamma}}_{i.}^\top\widehat{\boldsymbol{\Gamma}}_{i.})^{-1}\widehat{\boldsymbol{\gamma}}_{ik}$. We now derive the convergence of $\widehat{\boldsymbol{\Omega}}_2$. Write

$$\widehat{\boldsymbol{\Omega}}_2 = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{Z}_{ij.} + \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^\top(\mathbf{P}_{\mathbf{F}_i} - \mathbf{P}_{\widehat{\mathbf{C}}_i})\mathbf{Z}_{ij.}$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\widehat{\mathbf{Z}}_{ij.} - \mathbf{Z}_{ij.})^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{Z}_{ij.} + \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij.}^\top\mathbf{M}_{\mathbf{F}_i}(\widehat{\mathbf{Z}}_{ij.} - \mathbf{Z}_{ij.})$$

$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\widehat{\mathbf{Z}}_{ij\cdot}-\mathbf{Z}_{ij\cdot})^\top(\mathbf{P}_{\mathbf{F}_i}-\mathbf{P}_{\widehat{\mathbf{C}}_i})\mathbf{Z}_{ij\cdot}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}_{ij\cdot}^\top(\mathbf{P}_{\mathbf{F}_i}-\mathbf{P}_{\widehat{\mathbf{C}}_i})(\widehat{\mathbf{Z}}_{ij\cdot}-\mathbf{Z}_{ij\cdot})$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\widehat{\mathbf{Z}}_{ij\cdot}-\mathbf{Z}_{ij\cdot})^\top\mathbf{M}_{\mathbf{F}_i}(\widehat{\mathbf{Z}}_{ij\cdot}-\mathbf{Z}_{ij\cdot})$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}(\widehat{\mathbf{Z}}_{ij\cdot}-\mathbf{Z}_{ij\cdot})^\top(\mathbf{P}_{\mathbf{F}_i}-\mathbf{P}_{\widehat{\mathbf{C}}_i})(\widehat{\mathbf{Z}}_{ij\cdot}-\mathbf{Z}_{ij\cdot}):=\mathcal{L}_1+\cdots+\mathcal{L}_8,$$

where the definitions of $\mathcal{L}_1$, …, $\mathcal{L}_8$ are obvious.

By the definition of $\mathbf{\Omega}_2$ and Assumption 2.2, we can first obtain $\mathcal{L}_1=\mathbf{\Omega}_2+o_P(\sqrt{\frac{N}{LT}})$ under the condition $\frac{LT}{N^2}\to 0$. For $\mathcal{L}_2$, directly using Lemma 2.2.(1) gives $\mathcal{L}_2=o_P(\sqrt{\frac{N}{LT}})$, if $\frac{LT}{N^2}\to 0$ and $\frac{L}{N}\to 0$. For $\mathcal{L}_3$, we write

$$\mathcal{L}_3=-\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{k=1}^{N}(\widehat{a}_{\gamma,ijk}-a_{\gamma,ijk})\mathbf{X}_{ik\cdot}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{Z}_{ij\cdot}.$$

Then, we need to study $\widehat{\mathbf{\Gamma}}_{i\cdot}$ and establish the convergence of $\mathcal{L}_3$. Write

$$\begin{aligned}\widehat{\mathbf{\Gamma}}_{i\cdot}&=\frac{1}{T}(\mathbf{Y}_{i\cdot\cdot}-\mathbf{X}_{i,\widehat{\mathbf{b}}})\widehat{\mathbf{C}}_i\\&=\frac{1}{T}\mathbf{\Gamma}_{i\cdot}\mathbf{F}_i^\top\widehat{\mathbf{C}}_i+\frac{1}{T}\mathbf{E}_{i\cdot\cdot}^\top\widehat{\mathbf{C}}_i+\frac{1}{T}\mathbf{X}_{i,(\boldsymbol{\beta}-\widehat{\mathbf{b}})}^\top\widehat{\mathbf{C}}_i\\&=\mathbf{\Gamma}_{i\cdot}\mathbf{H}_i^{-1\top}+\frac{1}{T}\mathbf{\Gamma}_{i\cdot}(\mathbf{F}_i-\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})^\top\widehat{\mathbf{C}}_i+\frac{1}{T}\mathbf{E}_{i\cdot\cdot}^\top\widehat{\mathbf{C}}_i+\frac{1}{T}\mathbf{X}_{i,(\boldsymbol{\beta}-\widehat{\mathbf{b}})}^\top\widehat{\mathbf{C}}_i,\end{aligned}\tag{A.18}$$

where $\mathbf{Y}_{i\cdot\cdot}=(\mathbf{Y}_{i1\cdot},\ldots,\mathbf{Y}_{iN\cdot})^\top$ and $\mathbf{X}_{i,\widehat{\mathbf{b}}}=(\mathbf{X}_{i1\cdot}\widehat{\mathbf{b}},\ldots,\mathbf{X}_{iN\cdot}\widehat{\mathbf{b}})^\top$. Using Lemma 2.2, we obtain the following results for the second term on the right-hand side of (A.18),

$$\begin{aligned}\frac{1}{\sqrt{N}}\left\|\frac{1}{T}\mathbf{\Gamma}_{i\cdot}(\mathbf{F}_i-\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1})^\top\widehat{\mathbf{C}}_i\right\|_2&\le\frac{1}{\sqrt{N}}\|\mathbf{\Gamma}_{i\cdot}\|_2\cdot\frac{1}{\sqrt{T}}\left\|\mathbf{F}_i-\widehat{\mathbf{C}}_i\mathbf{H}_i^{-1}\right\|_2\cdot\frac{1}{\sqrt{T}}\|\widehat{\mathbf{C}}_i\|_2\\&=O_P\left(\frac{1}{\sqrt{N\wedge T}}\right).\end{aligned}$$

By Assumption 1.1, we have $\frac{1}{\sqrt{N}}\|\frac{1}{T}\mathbf{E}_{i\cdot\cdot}^\top\widehat{\mathbf{C}}_i\|=O_P\left(\frac{1}{\sqrt{N\wedge T}}\right)$. In light of Lemma 2.2.(2), the fourth term on the right-hand side of (A.18) is negligible. Combining these results, we have

$$\frac{1}{\sqrt{N}}\|\widehat{\mathbf{\Gamma}}_{i\cdot}-\mathbf{\Gamma}_{i\cdot}\mathbf{H}_i^{-1\top}\|_2=O_P\left(\frac{1}{\sqrt{N\wedge T}}\right).\tag{A.19}$$

It immediately yields that $\mathcal{L}_3=o_P\left(\sqrt{\frac{N}{LT}}\right)$. Since $\mathcal{L}_4=\mathcal{L}_3^\top$, we have $\mathcal{L}_4=o_P(\sqrt{\frac{N}{LT}})$. Using analogous arguments, we can show that $\mathcal{L}_5$, $\mathcal{L}_6$, $\mathcal{L}_7$ and $\mathcal{L}_8$ are also negligible. Therefore, we obtain

$$\widehat{\mathbf{\Omega}}_2=\mathbf{\Omega}_2+o_P\left(\sqrt{\frac{N}{LT}}\right).\tag{A.20}$$

Then, we shall prove

$$\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\left(\frac{\widehat{\mathbf{Z}}_{ij_1\boldsymbol{.}}^\top\widehat{\mathbf{C}}_i}{T}\left(\frac{\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}^\top\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}}{N}\right)^{-1}\widehat{\gamma}_{ij_2}\widehat{\sigma}_{i,j_1j_2}-\frac{\mathbf{Z}_{ij_1\boldsymbol{.}}^\top\mathbf{F}_i}{T}\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\boldsymbol{.}}^\top\mathbf{\Gamma}_{i\boldsymbol{.}}}{N}\right)^{-1}\gamma_{ij_2}\sigma_{i,j_1j_2}\right)$$
$$=o_P\left(\sqrt{\frac{N}{LT}}\right). \tag{A.21}$$

For notational simplicity, we use $\widetilde{\mathbf{a}}_1$ to denote the term on the left-hand side of (A.21). It is now convenient to add $\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}$ back and consider the following expansion:

$$\begin{aligned}
\widetilde{\mathbf{a}}_1 =& \frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\left(\frac{\widehat{\mathbf{Z}}_{ij_1\boldsymbol{.}}^\top\widehat{\mathbf{C}}_i}{T}\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}\left(\frac{\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}^\top\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}}{N}\right)^{-1}\widehat{\gamma}_{ij_2}\widehat{\sigma}_{i,j_1j_2}\right.\\
&\left.-\frac{\mathbf{Z}_{ij_1\boldsymbol{.}}^\top\mathbf{F}_i}{T}\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\boldsymbol{.}}^\top\mathbf{\Gamma}_{i\boldsymbol{.}}}{N}\right)^{-1}\gamma_{ij_2}\sigma_{i,j_1j_2}\right)\\
=& \frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\boldsymbol{.}}^\top\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\boldsymbol{.}}^\top\mathbf{\Gamma}_{i\boldsymbol{.}}}{N}\right)^{-1}\gamma_{ij_2}\left(\widehat{\sigma}_{i,j_1j_2}-\sigma_{i,j_1j_2}\right)\\
&+\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\boldsymbol{.}}^\top\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\boldsymbol{.}}^\top\mathbf{\Gamma}_{i\boldsymbol{.}}}{N}\right)^{-1}\mathbf{H}_i\left(\widehat{\gamma}_{ij_2}-\mathbf{H}_i^{-1}\gamma_{ij_2}\right)\widehat{\sigma}_{i,j_1j_2}\\
&+\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\boldsymbol{.}}^\top\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\mathbf{H}_i^{-1\top}\left(\left(\frac{\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}^\top\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}}{N}\right)^{-1}-\mathbf{H}_i^\top\left(\frac{\mathbf{\Gamma}_{i\boldsymbol{.}}^\top\mathbf{\Gamma}_{i\boldsymbol{.}}}{N}\right)^{-1}\mathbf{H}_i\right)\widehat{\gamma}_{ij_2}\widehat{\sigma}_{i,j_1j_2}\\
&+\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\boldsymbol{.}}^\top\mathbf{F}_i\mathbf{H}_i\left(\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}-\mathbf{H}_i^{-1}\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\mathbf{H}_i^{-1\top}\right)\left(\frac{\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}^\top\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}}{N}\right)^{-1}\widehat{\gamma}_{ij_2}\widehat{\sigma}_{i,j_1j_2}\\
&+\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\boldsymbol{.}}^\top(\widehat{\mathbf{C}}_i-\mathbf{F}_i\mathbf{H}_i)\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}\left(\frac{\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}^\top\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}}{N}\right)^{-1}\widehat{\gamma}_{ij_2}\widehat{\sigma}_{i,j_1j_2}\\
&+\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}(\widehat{\mathbf{Z}}_{ij_1\boldsymbol{.}}-\mathbf{Z}_{ij_1\boldsymbol{.}})^\top\widehat{\mathbf{C}}_i\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}\left(\frac{\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}^\top\widehat{\mathbf{\Gamma}}_{i\boldsymbol{.}}}{N}\right)^{-1}\widehat{\gamma}_{ij_2}\widehat{\sigma}_{i,j_1j_2}\\
:=& \mathcal{M}_1+\cdots+\mathcal{M}_6,
\end{aligned}$$

where the second equality holds by the following identity:

$$\widehat{a}\widehat{b}\widehat{c}\widehat{d}\widehat{e}\widehat{f}-abcdef=abcde(\widehat{f}-f)+abcd(\widehat{e}-e)\widehat{f}+abc(\widehat{d}-d)\widehat{e}\widehat{f}+ab(\widehat{c}-c)\widehat{d}\widehat{e}\widehat{f}+a(\widehat{b}-b)\widehat{c}\widehat{d}\widehat{e}\widehat{f}$$
$$+(\widehat{a}-a)\widehat{b}\widehat{c}\widehat{d}\widehat{e}\widehat{f}. \tag{A.22}$$

We then study these terms one by one. For $\mathcal{M}_1$, we continue to expand $\widehat{\sigma}_{i,j_1j_2}-\sigma_{i,j_1j_2}$.

Note that

$$
\begin{aligned}
\widehat{\sigma}_{i,j_1j_2} - \sigma_{i,j_1j_2} &= \frac{1}{T}(\mathbf{Y}_{ij_1\cdot} - \mathbf{X}_{ij_1\cdot}\widehat{\mathbf{b}} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_1})^\top(\mathbf{Y}_{ij_2\cdot} - \mathbf{X}_{ij_2\cdot}\widehat{\mathbf{b}} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_2}) - \sigma_{i,j_1j_2} \\
&= \frac{1}{T}(\boldsymbol{\beta} - \widehat{\mathbf{b}})^\top\mathbf{X}_{ij_1\cdot}^\top\mathbf{X}_{ij_2\cdot}(\boldsymbol{\beta} - \widehat{\mathbf{b}}) + \frac{1}{T}(\boldsymbol{\beta} - \widehat{\mathbf{b}})^\top\mathbf{X}_{ij_1\cdot}^\top(\mathbf{F}_i\boldsymbol{\gamma}_{ij_2} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_2}) \\
&\quad + \frac{1}{T}(\mathbf{F}_i\boldsymbol{\gamma}_{ij_1} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_1})^\top\mathbf{X}_{ij_2\cdot}(\boldsymbol{\beta} - \widehat{\mathbf{b}}) + \frac{1}{T}(\boldsymbol{\beta} - \widehat{\mathbf{b}})^\top\mathbf{X}_{ij_1\cdot}^\top\mathbf{E}_{ij_2\cdot} + \frac{1}{T}\mathbf{E}_{ij_1\cdot}^\top\mathbf{X}_{ij_2\cdot}(\boldsymbol{\beta} - \widehat{\mathbf{b}}) \\
&\quad + \frac{1}{T}(\mathbf{F}_i\boldsymbol{\gamma}_{ij_1} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_1})^\top(\mathbf{F}_i\boldsymbol{\gamma}_{ij_2} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_2}) + \frac{1}{T}(\mathbf{F}_i\boldsymbol{\gamma}_{ij_1} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_1})^\top\mathbf{E}_{ij_2\cdot} \\
&\quad + \frac{1}{T}\mathbf{E}_{ij_1\cdot}^\top(\mathbf{F}_i\boldsymbol{\gamma}_{ij_2} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_2}) + \left(\frac{1}{T}\mathbf{E}_{ij_1\cdot}^\top\mathbf{E}_{ij_2\cdot} - \sigma_{i,j_1j_2}\right) \\
&:= c_{1,i,j_1j_2} + \cdots + c_{9,i,j_1j_2}.
\end{aligned} \tag{A.23}
$$

With this expansion, we have

$$
\mathcal{M}_1 = \frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\cdot}^\top\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij_2}\sum_{l=1}^{9}c_{l,i,j_1j_2} := \mathcal{M}_{1,1}\cdots+\mathcal{M}_{1,9}.
$$

Directly using Lemma 2.2 and (A.19), we can readily obtain

$$
\mathcal{M}_{1,1} = O_P(N\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|^2),\ \mathcal{M}_{1,2} = O_P\left(\frac{N\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|}{\sqrt{N \wedge T}}\right),\ \mathcal{M}_{1,3} = O_P\left(\frac{N\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|}{\sqrt{N \wedge T}}\right),
$$

and

$$
\mathcal{M}_{1,4} = O_P\left(\frac{N}{\sqrt{T}}\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|\right),\ \mathcal{M}_{1,5} = O_P\left(\frac{N}{\sqrt{T}}\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|\right).
$$

Since Lemma 2.2.(2) implies that $\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| = O_P\left(\frac{1}{\sqrt{LNT \wedge N \wedge T}}\right)$, these terms are all $o_P\left(\sqrt{\frac{N}{LT}}\right)$ under the conditions $\frac{LT}{N^2} \to 0$, $\frac{LN}{T^2} \to 0$ and $\frac{L}{N} \to 0$. By substituting the expansions (A.4) and (A.18) into $\mathcal{M}_{1,6}$, we can show that this term is also negligible. For $\mathcal{M}_{1,7}$, note that

$$
\begin{aligned}
\mathcal{M}_{1,7} &= \frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\cdot}^\top\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij_2}\frac{1}{T}(\mathbf{F}_i\boldsymbol{\gamma}_{ij_1} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij_1})^\top\mathbf{E}_{ij_2\cdot} \\
&= -\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\cdot}^\top\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij_2}\widehat{\boldsymbol{\gamma}}_{ij_1}^\top\frac{1}{T}(\widehat{\mathbf{C}}_i - \mathbf{F}_i\mathbf{H}_i)^\top\mathbf{E}_{ij_2\cdot} \\
&\quad + \frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\cdot}^\top\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij_2}(\boldsymbol{\gamma}_{ij_1} - \mathbf{H}_i\widehat{\boldsymbol{\gamma}}_{ij_1})^\top\frac{1}{T}\mathbf{F}_i^\top\mathbf{E}_{ij_2\cdot}.
\end{aligned}
$$

Using expansion (A.4) and the similar arguments to those in the proof of (A.8), we can readily obtain $\|T^{-1}(\widehat{\mathbf{C}}_i - \mathbf{F}_i\mathbf{H}_i)^\top\mathbf{E}_{ij_2\cdot}\|_2 = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{1}{\sqrt{(N \wedge T)T}}\right)$, which yields that the first term in $\mathcal{M}_{1,7}$ is $o_P(\sqrt{\frac{N}{LT}})$. For the second term, since $\sum_{i=1}^{L}\|\sum_{j_2=1}^{N}\boldsymbol{\gamma}_{ij_2}\mathbf{F}_i^\top\mathbf{E}_{ij_2\cdot}\|^2 =$

$O(LNT)$ under Assumption 2.1, using Cauchy–Schwarz inequality gives

$$\frac{1}{LN}\left\|\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\cdot}^{\top}\mathbf{F}_i\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\cdot}^{\top}\mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}\gamma_{ij_2}(\boldsymbol{\gamma}_{ij_1}-\mathbf{H}_i\widehat{\boldsymbol{\gamma}}_{ij_1})^{\top}\frac{1}{T}\mathbf{F}_i^{\top}\mathbf{E}_{ij_2\cdot}\right\|_2$$
$$\le\frac{1}{LN}\left(\sum_{i=1}^{L}\left\|\sum_{j_1=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\cdot}^{\top}\mathbf{F}_i\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\cdot}^{\top}\mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}(\boldsymbol{\gamma}_{ij_1}-\mathbf{H}_i\widehat{\boldsymbol{\gamma}}_{ij_1})^{\top}\right\|_2^2\right)^{\frac{1}{2}}$$
$$\times\left(\sum_{i=1}^{L}\left\|\sum_{j_2=1}^{N}\frac{1}{T}\gamma_{ij_2}\mathbf{F}_i^{\top}\mathbf{E}_{ij_2\cdot}\right\|^2\right)^{\frac{1}{2}}=O_P\left(\sqrt{\frac{N}{(N\wedge T)T}}\right)=o_P\left(\sqrt{\frac{N}{LT}}\right),\qquad\text{(A.24)}$$

where the last equality holds when $\frac{L}{N}\to 0$ and $\frac{L}{T}\to 0$. Thus, it follows that $\mathcal{M}_{1,7}=o_P\left(\sqrt{\frac{N}{LT}}\right)$. Analogously, we also have $\mathcal{M}_{1,8}=o_P\left(\sqrt{\frac{N}{LT}}\right)$. For $\mathcal{M}_{1,9}$, we first note that by Assumption 2.1,

$$\sum_{i=1}^{L}\mathbb{E}\left[\left\|\sum_{j_1,j_2=1}^{N}\mathbf{Z}_{ij_1t}\gamma_{ij_2}^{\top}\left(\frac{1}{T}\mathbf{E}_{ij_1\cdot}^{\top}\mathbf{E}_{ij_2\cdot}-\sigma_{i,j_1j_2}\right)\right\|^2\right]$$
$$=O(1)\frac{1}{T^2}\sum_{i=1}^{L}\sum_{j_1,j_2,j_3,j_4=1}^{N}\sum_{t_1,t_2=1}^{T}|cov(\varepsilon_{ij_1t_1}\varepsilon_{ij_2t_1},\varepsilon_{ij_3t_2}\varepsilon_{ij_4t_2})|$$
$$=O\left(\frac{LN^2}{T}\right).$$

Together with Cauchy–Schwarz inequality, it yields

$$\|\mathcal{M}_{1,9}\|_2\le\frac{1}{LNT}\left(\sum_{i=1}^{L}\sum_{t=1}^{T}\left\|\mathbf{F}_{it}^{\top}\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\cdot}^{\top}\mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}\right\|^2\right)^{\frac{1}{2}}$$
$$\times\left(\sum_{i=1}^{L}\sum_{t=1}^{T}\left\|\sum_{j_1,j_2=1}^{N}\mathbf{Z}_{ij_1t}\gamma_{ij_2}^{\top}\left(\frac{1}{T}\mathbf{E}_{ij_1\cdot}^{\top}\mathbf{E}_{ij_2\cdot}-\sigma_{i,j_1j_2}\right)\right\|^2\right)^{\frac{1}{2}}$$
$$=O_P\left(\frac{1}{\sqrt{T}}\right)=o_P\left(\sqrt{\frac{N}{LT}}\right),$$

where the last equality holds under the condition $\frac{L}{N}\to 0$. Combining these results gives $\mathcal{M}_1=o_P\left(\sqrt{\frac{N}{LT}}\right)$.

For $\mathcal{M}_2$, write

$$\mathcal{M}_2=\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\cdot}^{\top}\mathbf{F}_i\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\cdot}^{\top}\mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}\mathbf{H}_i\left(\widehat{\boldsymbol{\gamma}}_{ij_2}-\mathbf{H}_i^{-1}\boldsymbol{\gamma}_{ij_2}\right)\sigma_{i,j_1j_2}$$

$$+\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}\mathbf{Z}_{ij_1\cdot}^{\top}\mathbf{F}_i\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^{\top}\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\mathbf{H}_i\left(\widehat{\boldsymbol{\gamma}}_{ij_2}-\mathbf{H}_i^{-1}\boldsymbol{\gamma}_{ij_2}\right)\left(\widehat{\sigma}_{i,j_1j_2}-\sigma_{i,j_1j_2}\right)$$

$$:=\mathcal{M}_{2,1}+\mathcal{M}_{2,2}.$$

Using the expansion in (A.23) and analogous arguments in the proof of $\mathcal{M}_1$'s convergence, we can show that $\mathcal{M}_{2,2}$ is negligible. For the first term, note that Assumption 2.1 implies that $\max_{i,j_2}\sum_{j_1=1}^{N}|\sigma_{i,j_1j_2}|=O(1)$. Together with (A.19) and Cauchy–Schwarz inequality,

$$\begin{aligned}\|\mathcal{M}_{2,1}\|_2\leq&\frac{1}{LN}\left(\sum_{i=1}^{L}\left\|\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^{\top}\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\mathbf{H}_i\right\|_2^2\sum_{j_2=1}^{N}\|\widehat{\boldsymbol{\gamma}}_{ij_2}-\mathbf{H}_i^{-1}\boldsymbol{\gamma}_{ij_2}\|_2^2\right)^{\frac{1}{2}}\\&\times\left(\sum_{i=1}^{L}\sum_{j_2=1}^{N}\left(\sum_{j_1=1}^{N}\left\|\frac{1}{T}\mathbf{Z}_{ij_1\cdot}^{\top}\mathbf{F}_i\right\|_2|\sigma_{i,j_1j_2}|\right)^2\right)^{\frac{1}{2}}\\=&O_P\left(\frac{1}{\sqrt{N\wedge T}}\right).\end{aligned}$$

Therefore, we have proved $\mathcal{M}_2=o_P\left(\sqrt{\frac{N}{LT}}\right)$. Analogously, we can show that $\mathcal{M}_3$, $\mathcal{M}_4$ and $\mathcal{M}_5$ are also negligible. For $\mathcal{M}_6$, we can use the expansion in (A.23) and analogous arguments in the proof of $\mathcal{M}_1$'s convergence to show that

$$\mathcal{M}_6=\frac{1}{LN}\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}(\widehat{\mathbf{Z}}_{ij_1\cdot}-\mathbf{Z}_{ij_1\cdot})^{\top}\mathbf{F}_i\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^{\top}\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij_2}\sigma_{i,j_1j_2}+o_P(1).$$

For the leading term in $\mathcal{M}_6$, by Assumption 2.1, (A.19) and Cauchy–Schwarz inequality,

$$\begin{aligned}&\frac{1}{LN}\left\|\sum_{i=1}^{L}\sum_{j_1,j_2=1}^{N}\frac{1}{T}(\widehat{\mathbf{Z}}_{ij_1\cdot}-\mathbf{Z}_{ij_1\cdot})^{\top}\mathbf{F}_i\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^{\top}\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij_2}\sigma_{i,j_1j_2}\right\|_2\\\leq&\frac{1}{LN}\left(\sum_{i=1}^{L}\sum_{j_1=1}^{N}\left\|\frac{1}{T}(\widehat{\mathbf{Z}}_{ij_1\cdot}-\mathbf{Z}_{ij_1\cdot})^{\top}\mathbf{F}_i\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^{\top}\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\right\|_2^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{L}\sum_{j_1=1}^{N}\left(\sum_{j_2=1}^{N}\|\boldsymbol{\gamma}_{ij_2}\||\sigma_{i,j_1j_2}|\right)^2\right)^{\frac{1}{2}}\\=&O_P\left(\frac{1}{\sqrt{N\wedge T}}\right).\end{aligned}$$

Combining these results, we can readily obtain $\widetilde{\mathbf{a}}_1=o_P\left(\sqrt{\frac{N}{LT}}\right)$, which immediately leads to the desired result in (A.21). Together with the result $\widehat{\boldsymbol{\Omega}}_2=\boldsymbol{\Omega}_2+o_P\left(\sqrt{\frac{N}{LT}}\right)$

which has already been established at an earlier stage, it yields

$$\widehat{\mathbf{a}}_1 = \mathbf{a}_1 + o_P\left(\sqrt{\frac{N}{LT}}\right). \tag{A.25}$$

We then proceed with the bias term $\mathbf{a}_2$. Recall that we construct the following estimator

$$\widehat{\mathbf{a}}_2 = -\widehat{\mathbf{\Omega}}_2^{-1}\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\widehat{\mathbf{\Sigma}}_{\varepsilon,i}\widehat{\mathbf{C}}_i\left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top\widehat{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widehat{\gamma}_{ij},$$

where the (t,s)-th element of $\widehat{\mathbf{\Sigma}}_{\varepsilon,i}$is given by $\widehat{\sigma}_{\varepsilon,i,ts} = N^{-1}(\mathbf{Y}_{i\cdot t} - \mathbf{X}_{i\cdot t}\,\widehat{\mathbf{b}} - \widehat{\mathbf{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{it})^\top(\mathbf{Y}_{i\cdot s} - \mathbf{X}_{i\cdot s}\,\widehat{\mathbf{b}} - \widehat{\mathbf{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{is})$. Since we have established the convergence of $\widehat{\mathbf{\Omega}}_2$, it suffices only to show

$$\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\left(\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\widehat{\mathbf{C}}_i}\widehat{\mathbf{\Sigma}}_{\varepsilon,i}\widehat{\mathbf{C}}_i\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}\left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top\widehat{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widehat{\gamma}_{ij} - \mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,i}\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\cdot}^\top\mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}\gamma_{ij}\right)$$
$$= o_P\left(\sqrt{\frac{T}{LN}}\right). \tag{A.26}$$

Let $\widetilde{\mathbf{a}}_2$ be the term on the left-hand side of (A.26). It follows the expansion identity (A.22) that

$$\begin{aligned}
\widetilde{\mathbf{a}}_2 =\ & \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,i}\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\cdot}^\top\mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}\mathbf{H}_i(\widehat{\gamma}_{ij} - \mathbf{H}_i^{-1}\gamma_{ij}) \\
& + \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,i}\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\mathbf{H}_i^{-1\top}\left(\left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top\widehat{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1} - \mathbf{H}_i^\top\left(\frac{\mathbf{\Gamma}_{i\cdot}^\top\mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}\mathbf{H}_i\right)\widehat{\gamma}_{ij} \\
& + \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,i}\mathbf{F}_i\mathbf{H}_i\left(\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1} - \mathbf{H}_i^{-1}\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\mathbf{H}_i^{-1\top}\right)\left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top\widehat{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widehat{\gamma}_{ij} \\
& + \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,i}(\widehat{\mathbf{C}}_i - \mathbf{F}_i\mathbf{H}_i)\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}\left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top\widehat{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widehat{\gamma}_{ij} \\
& + \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{F}_i}(\widehat{\mathbf{\Sigma}}_{\varepsilon,i} - \mathbf{\Sigma}_{\varepsilon,i})\widehat{\mathbf{C}}_i\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}\left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top\widehat{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widehat{\gamma}_{ij} \\
& - \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^\top(\mathbf{P}_{\widehat{\mathbf{C}}_i} - \mathbf{P}_{\mathbf{F}_i})\widehat{\mathbf{\Sigma}}_{\varepsilon,i}\widehat{\mathbf{C}}_i\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}\left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top\widehat{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widehat{\gamma}_{ij} \\
:=\ & \mathcal{K}_1 + \cdots + \mathcal{K}_6,
\end{aligned}$$

where the definitions of $\mathcal{K}_1,\cdots,\mathcal{K}_6$ are obvious. We study these terms one by one. For notational simplicity let $\sigma_{\varepsilon,i,ts} = N^{-1}\sum_{j=1}^N E[\varepsilon_{ijt}\varepsilon_{ijs}]$ and $\mathbf{X}_{ijt}^* = \mathbf{X}_{ijt} - \mathbf{X}_{ij\cdot}^\top\mathbf{F}_i(\mathbf{F}_i^\top\mathbf{F}_i)^{-1}\mathbf{f}_{it}$, where $\mathbf{f}_{it}$ is the t-th column of $\mathbf{F}_i^\top$. Then, for $\mathcal{K}_1$, we write

$$\mathcal{K}_1 = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sigma_{\varepsilon,i,ts}\mathbf{X}_{ijt}^{*}\mathbf{f}_{is}^\top\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\mathbf{H}_i(\widehat{\boldsymbol{\gamma}}_{ij}-\mathbf{H}_i^{-1}\boldsymbol{\gamma}_{ij}).$$

By Assumption 2.1, we have $\max_i\sum_{t=1}^{T}\sum_{s=1}^{T}|\sigma_{\varepsilon,i,ts}| = O(T)$. Together with (A.19) and Cauchy–Schwarz inequality, it yields

$$\begin{aligned}\|\mathcal{K}_1\|_2 &\leq \frac{1}{LNT}\left(\sum_{i=1}^{L}\left\|\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\right\|^2\left(\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\sigma_{\varepsilon,i,ts}|\|\mathbf{X}_{ijt}^{*}\mathbf{f}_{it}^\top\|_2\right)^2\right)^{\frac{1}{2}}\\ &\times\left(\sum_{i=1}^{L}\sum_{j=1}^{N}\left\|\mathbf{H}_i(\widehat{\boldsymbol{\gamma}}_{ij}-\mathbf{H}_i^{-1}\boldsymbol{\gamma}_{ij})\right\|_2^2\right)^{\frac{1}{2}} = O_P\left(\frac{1}{\sqrt{N\wedge T}}\right) = o_P\left(\sqrt{\frac{T}{LN}}\right), \qquad \text{(A.27)}\end{aligned}$$

where the last equality holds under the conditions $\frac{L}{T}\to 0$ and $\frac{LN}{T^2}\to 0$.

Analogously, we can show that both $\mathcal{K}_2$ and $\mathcal{K}_3$ are also negligible, and obtain $\mathcal{K}_4$'s leading term

$$\begin{aligned}\mathcal{K}_4 &= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\mathbf{F}_i}\boldsymbol{\Sigma}_{\varepsilon,i}(\widehat{\mathbf{C}}_i-\mathbf{F}_i\mathbf{H}_i)\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij}+o_P\left(\sqrt{\frac{T}{LN}}\right)\\ &= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sigma_{\varepsilon,i,ts}\mathbf{X}_{ijt}^{*}(\widehat{\mathbf{c}}_{is}-\mathbf{H}_i^\top\mathbf{f}_{is})^\top\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Gamma}_{i\cdot}^\top\boldsymbol{\Gamma}_{i\cdot}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij}+o_P\left(\sqrt{\frac{T}{LN}}\right),\end{aligned}$$

where $\widehat{\mathbf{c}}_{is}$ is the s-th column of $\widehat{\mathbf{C}}_i^\top$.

Since Assumption 2.1 implies that $\max_{i,s}\sum_{t=1}^{T}|\sigma_{\varepsilon,i,ts}| = O(1)$, using Lemma 2.2.(1) and Cauchy–Schwarz inequality, we can readily show that $\mathcal{K}_4 = o_P\left(\sqrt{\frac{T}{LN}}\right)$. Analogously, we also have $\mathcal{K}_6 = o_P\left(\sqrt{\frac{T}{LN}}\right)$.

Then, we shall prove the convergence of $\mathcal{K}_5$. For $\mathcal{K}_5$, we make the following expansion for $\widehat{\sigma}_{\varepsilon,i,ts}-\sigma_{\varepsilon,i,ts}$:

$$\begin{aligned}\widehat{\sigma}_{\varepsilon,i,ts}-\sigma_{\varepsilon,i,ts} &= \frac{1}{N}(\mathbf{Y}_{i\cdot t}-\mathbf{X}_{i\cdot t}\widehat{\mathbf{b}}-\widehat{\boldsymbol{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{it})^\top(\mathbf{Y}_{i\cdot s}-\mathbf{X}_{i\cdot s}\widehat{\mathbf{b}}-\widehat{\boldsymbol{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{is})\\ &= \frac{1}{N}(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{i\cdot t}^\top\mathbf{X}_{i\cdot s}(\boldsymbol{\beta}-\widehat{\mathbf{b}})+\frac{1}{N}(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{i\cdot t}^\top(\boldsymbol{\Gamma}_{i\cdot}\mathbf{f}_{is}-\widehat{\boldsymbol{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{is})\\ &\quad+\frac{1}{N}(\boldsymbol{\Gamma}_{i\cdot}\mathbf{f}_{it}-\widehat{\boldsymbol{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{it})^\top\mathbf{X}_{i\cdot s}(\boldsymbol{\beta}-\widehat{\mathbf{b}})+\frac{1}{N}(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{i\cdot t}^\top\mathbf{E}_{i\cdot s}+\frac{1}{N}\mathbf{E}_{i\cdot t}^\top\mathbf{X}_{i\cdot s}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\\ &\quad+\frac{1}{N}(\boldsymbol{\Gamma}_{i\cdot}\mathbf{f}_{it}-\widehat{\boldsymbol{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{it})^\top(\boldsymbol{\Gamma}_{i\cdot}\mathbf{f}_{is}-\widehat{\boldsymbol{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{is})+\frac{1}{N}(\boldsymbol{\Gamma}_{i\cdot}\mathbf{f}_{it}-\widehat{\boldsymbol{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{it})^\top\mathbf{E}_{i\cdot s}\\ &\quad+\frac{1}{N}\mathbf{E}_{i\cdot t}^\top(\boldsymbol{\Gamma}_{i\cdot}\mathbf{f}_{is}-\widehat{\boldsymbol{\Gamma}}_{i\cdot}\widehat{\mathbf{c}}_{is})+\left(\frac{1}{N}\mathbf{E}_{i\cdot t}^\top\mathbf{E}_{i\cdot s}-\sigma_{\varepsilon,i,st}\right)\\ &:= d_{i,ts,1}+\cdots+d_{i,ts,9}.\end{aligned}$$

Thus, we can write

$$\mathcal{K}_5 = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{t,s=1}^{T}\mathbf{X}_{ijt}^*\widehat{\mathbf{c}}_{is}^\top\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}\left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top\widehat{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widehat{\boldsymbol{\gamma}}_{ij}(d_{i,ts,1}+\cdots+d_{i,ts,9}) := \mathcal{K}_{5,1}\cdots+\mathcal{K}_{5,9}.$$

Since Lemma 2.2.(2) implies that $\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\| = O_P\left(\frac{1}{\sqrt{LNT\wedge N\wedge T}}\right)$, we can show that $\mathcal{K}_{5,1},\ldots,\mathcal{K}_{5,5}$ are all $o_P(\sqrt{\frac{T}{LN}})$ under the conditions $\frac{LN}{T^2}\to 0$, $\frac{LT}{N^2}\to 0$ and $\frac{L}{T}\to 0$. Therefore, these terms are all negligible.

By substituting the expansions (A.4) and (A.18) into $\mathcal{K}_{5,6}$, we can show that this term is also negligible. For $\mathcal{K}_{5,7}$ and $\mathcal{K}_{5,8}$, we can use the arguments that are analogous to those in the proof of (A.24) to show that $\mathcal{K}_{5,7} = o_P(\sqrt{\frac{T}{LN}})$ and $\mathcal{K}_{5,8} = o_P(\sqrt{\frac{T}{LN}})$. For the last term, we write

$$\begin{aligned}\mathcal{K}_{5,9} &= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{t,s=1}^{T}\mathbf{X}_{ijt}^*\widehat{\mathbf{c}}_{is}^\top\left(\frac{\widehat{\mathbf{C}}_i^\top\widehat{\mathbf{C}}_i}{T}\right)^{-1}\left(\frac{\widehat{\mathbf{\Gamma}}_{i\cdot}^\top\widehat{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widehat{\boldsymbol{\gamma}}_{ij}\left(\frac{1}{N}\mathbf{E}_{i\cdot t}^\top\mathbf{E}_{i\cdot s}-\sigma_{\varepsilon,i,st}\right)\\
&= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{t,s=1}^{T}\mathbf{X}_{ijt}^*\mathbf{f}_{is}^\top\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\cdot}^\top\mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}\boldsymbol{\gamma}_{ij}\left(\frac{1}{N}\mathbf{E}_{i\cdot t}^\top\mathbf{E}_{i\cdot s}-\sigma_{\varepsilon,i,st}\right)+o_P\left(\sqrt{\frac{T}{LN}}\right)\\
&= \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{t,s=1}^{T}\sum_{l_1,l_2=1}^{l^G+l_i^S}\mathbf{X}_{ijt}^*\mathcal{R}_{i,l_1l_2}f_{is,l_1}\gamma_{ij,l_2}\left(\frac{1}{N}\mathbf{E}_{i\cdot t}^\top\mathbf{E}_{i\cdot s}-\sigma_{\varepsilon,i,st}\right)+o_P\left(\sqrt{\frac{T}{LN}}\right),\end{aligned}$$

where $\mathcal{R}_{i,l_1l_2}$ is the (l_1,l_2)-th element of $\left(\frac{\mathbf{F}_i^\top\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\cdot}^\top\mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}$, and $f_{is,l}$ and $\gamma_{ij,l}$ are the l-th elements of $\mathbf{f}_{is}$ and $\boldsymbol{\gamma}_{ij}$, respectively. For the leading term in $\mathcal{K}_{5,9}$, by Assumption 2.1 and Cauchy–Schwarz inequality, we have

$$\begin{aligned}&\frac{1}{LNT}\left\|\sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{t,s=1}^{T}\sum_{l_1,l_2=1}^{l^G+l_i^S}\mathbf{X}_{ijt}^*\mathcal{R}_{i,l_1l_2}f_{is,l_1}\gamma_{ij,l_2}\left(\frac{1}{N}\mathbf{E}_{i\cdot t}^\top\mathbf{E}_{i\cdot s}-\sigma_{\varepsilon,i,st}\right)\right\|\\
&\le \frac{1}{LNT}\left(\sum_{i=1}^{L}\sum_{l_1,l_2=1}^{l^G+l_i^S}\left\|\sum_{j=1}^{N}\sum_{t,s=1}^{T}\mathbf{X}_{ijt}^*f_{is,l_1}\gamma_{ij,l_2}\left(\frac{1}{N}\mathbf{E}_{i\cdot t}^\top\mathbf{E}_{i\cdot s}-\sigma_{\varepsilon,i,st}\right)\right\|^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{L}\sum_{l_1,l_2=1}^{l^G+l_i^S}\|\mathcal{R}_{i,l_1l_2}\|^2\right)^{\frac{1}{2}}\\
&= O_P\left(\frac{1}{\sqrt{N}}\right) = o_P\left(\sqrt{\frac{T}{LN}}\right),\end{aligned}$$

when $\frac{L}{T}\to 0$. In summary of these results, we have $\mathcal{K}_5 = o_P(\sqrt{\frac{T}{LN}})$. Therefore, it follows that

$$\widehat{\mathbf{a}}_2 = \mathbf{a}_2 + o_P\left(\sqrt{\frac{T}{LN}}\right). \tag{A.28}$$

Combining the results that are established in Lemma 2.2.(2), (A.25) and (A.28) leads to the desired property of $\widehat{\mathbf{b}}_{bc}$ in Corollary 2.1. ■

Proof of Theorem 2.2:

Write

$$
\begin{aligned}
\widehat{\mathbf{b}}^* &= \widehat{\mathbf{b}}_{bc} + \left(\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\widehat{\mathbf{Z}}_{ij\cdot}\right)^{-1}\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}[\mathbf{E}_{ij\cdot}\circ\boldsymbol{\xi}] \\
&\quad + \left(\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\widehat{\mathbf{Z}}_{ij\cdot}\right)^{-1}\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}[(\mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}}_{bc}))\circ\boldsymbol{\xi}] \\
&\quad + \left(\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\widehat{\mathbf{Z}}_{ij\cdot}\right)^{-1}\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}[(\mathbf{F}_i\boldsymbol{\gamma}_{ij}-\widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij})\circ\boldsymbol{\xi}] \\
&:= \mathbf{B}_1+\mathbf{B}_2+\mathbf{B}_3, \qquad (A.29)
\end{aligned}
$$

where the definitions of $\mathbf{B}_1$-$\mathbf{B}_3$ are obvious.

Next, we focus on the terms on the right hand side of (A.29). $\mathbf{B}_2$ is negligible in an obvious manner, so we focus on $\mathbf{B}_1$ and $\mathbf{B}_3$ below.

We start with $\mathbf{B}_1$, and write

$$
\begin{aligned}
&\frac{1}{LNT}\mathbb{E}^*\left[\left(\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}[\mathbf{E}_{ij\cdot}\circ\boldsymbol{\xi}]\right)\left(\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}[\mathbf{E}_{ij\cdot}\circ\boldsymbol{\xi}]\right)^\top\right] \\
&= \frac{1}{LNT}\sum_{i_1=1}^{L}\sum_{j_1=1}^{N}\sum_{i_2=1}^{L}\sum_{j_2=1}^{N}\widehat{\mathbf{Z}}_{i_1j_1\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_{i_1}}\mathbf{E}_{i_1j_1\cdot}\mathbf{E}_{i_2j_2\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_{i_2}}\widehat{\mathbf{Z}}_{i_2j_2\cdot}^\top \\
&\quad + \frac{1}{LNT}\sum_{i_1=1}^{L}\sum_{j_1=1}^{N}\sum_{i_2=1}^{L}\sum_{j_2=1}^{N}\widehat{\mathbf{Z}}_{i_1j_1\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_{i_1}}[\mathbf{E}_{i_1j_1\cdot}\mathbf{E}_{i_2j_2\cdot}^\top\circ(\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^\top]-\mathbf{1}_T\mathbf{1}_T^\top)]\mathbf{M}_{\widehat{\mathbf{C}}_{i_2}}\widehat{\mathbf{Z}}_{i_2j_2\cdot} \\
&:= \mathbf{B}_{11}+\mathbf{B}_{12},
\end{aligned}
$$

where the definitions of $\mathbf{B}_{11}$ and $\mathbf{B}_{12}$ are obvious, and the first term of the right hand is the leading term if we can show $\|\mathbf{B}_{12}\| = o_P(1)$. Note that

$$
\mathbf{B}_{12} = \frac{1}{LNT}\sum_{i_1=1}^{L}\sum_{j_1=1}^{N}\sum_{i_2=1}^{L}\sum_{j_2=1}^{N}\mathbf{Z}_{i_1j_1\cdot}^{*\top}[(\mathbf{E}_{i_1j_1\cdot}\mathbf{E}_{i_2j_2\cdot}^\top)\circ(\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^\top]-\mathbf{1}_T\mathbf{1}_T^\top)]\mathbf{Z}_{i_2j_2\cdot}^*(1+o_P(1)),
$$

where $\mathbf{Z}_{ij\cdot}^* = (\mathbf{I}-T^{-1}\mathbf{F}_i\mathbf{F}_i^\top)\mathbf{Z}_{ij\cdot}$ and the equality follows from the facts that Assumption 2.2 and Lemma 2.2.1. Also, note that

$$
\begin{aligned}
&\frac{1}{LNT}\sum_{i_1=1}^{L}\sum_{j_1=1}^{N}\sum_{i_2=1}^{L}\sum_{j_2=1}^{N}\mathbf{Z}_{i_1j_1\cdot}^{*\top}[(\mathbf{E}_{i_1j_1\cdot}\mathbf{E}_{i_2j_2\cdot}^\top)\circ(\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^\top]-\mathbf{1}_T\mathbf{1}_T^\top)]\mathbf{Z}_{i_2j_2\cdot}^* \\
&= \frac{1}{LNT}\sum_{i_1=1}^{L}\sum_{j_1=1}^{N}\sum_{i_2=1}^{L}\sum_{j_2=1}^{N}\sum_{t=1}^{d_T}\sum_{s=1}^{T-t}\mathbf{z}_{i_1j_1t}^*\mathbf{z}_{i_2j_2,s+t}^{*\top}\varepsilon_{i_1j_1s}\varepsilon_{i_2j_2,s+t}\left[a\left(\frac{t}{\ell}\right)-1\right] \\
&\quad + \frac{1}{LNT}\sum_{i_1=1}^{L}\sum_{j_1=1}^{N}\sum_{i_2=1}^{L}\sum_{j_2=1}^{N}\sum_{t=d_T+1}^{T}\sum_{s=1}^{T-t}\mathbf{z}_{ijt}^*\mathbf{z}_{ij,s+t}^{*\top}\varepsilon_{i_1j_1s}\varepsilon_{i_2j_2,s+t}\left[a\left(\frac{t}{\ell}\right)-1\right]
\end{aligned}
$$

$$:= \mathbf{B}_{121} + \mathbf{B}_{122},$$

where $\mathbf{z}^*_{ijt}$ is the t-th column of $\mathbf{Z}^{*\top}_{ij\cdot}$.

By the fact that $a(w)$ is Lipschitz continuous on $[-1,1]$, and by letting $d_T^2/\ell \to 0$ and $d_T \to \infty$, we have

$$\|\mathbb{E}[\mathbf{B}_{121}]\| \leq O(1)\frac{d_T^2}{\ell} = o(1).$$

For $\mathbf{B}_{122}$, applying the mixing condition on $\varepsilon_{i_1j_11}\varepsilon_{i_2j_2,t}$ such as (A.7) of Su et al. (2009), we can show that $\|\mathbb{E}[\mathbf{B}_{122}]\| = o_P(1)$. Therefore, we have proved $\|\mathbb{E}[\mathbf{B}_{12}]\| = o(1)$.

Similarly, we can show that

$$\begin{aligned}
&\mathbb{E}^*\Bigg[\left(\frac{1}{\sqrt{LNT}}\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}^\top_{ij\cdot}\mathbf{M}_{\widehat{\mathbf{C}}_i}(\mathbf{F}_i\boldsymbol{\gamma}_{ij} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij})\circ\boldsymbol{\xi})\right) \\
&\quad\cdot\left(\frac{1}{\sqrt{LNT}}\sum_{i=1}^{L}\sum_{j=1}^{N}\widehat{\mathbf{Z}}^\top_{ij\cdot}\mathbf{M}_{\widehat{\mathbf{C}}_i}(\mathbf{F}_i\boldsymbol{\gamma}_{ij} - \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\gamma}}_{ij})\circ\boldsymbol{\xi})\right)^\top\Bigg] \\
&= \frac{1}{LNT}\sum_{i_1=1}^{L}\sum_{j_1=1}^{N}\sum_{i_2=1}^{L}\sum_{j_2=1}^{N}\widehat{\mathbf{Z}}^\top_{i_1j_1\cdot}\mathbf{M}_{\widehat{\mathbf{C}}_{i_1}}(\mathbf{F}_{i_1}\boldsymbol{\gamma}_{i_1j_1} - \widehat{\mathbf{C}}_{i_1}\widehat{\boldsymbol{\gamma}}_{i_1j_1})(\mathbf{F}_{i_2}\boldsymbol{\gamma}_{i_2j_2} - \widehat{\mathbf{C}}_{i_2}\widehat{\boldsymbol{\gamma}}_{i_2j_2})^\top\mathbf{M}_{\widehat{\mathbf{C}}_{i_2}}\widehat{\mathbf{Z}}_{i_2j_2\cdot} \\
&\quad + o_P(1).
\end{aligned}$$

Using the above development and expanding the terms $\widehat{\mathbf{C}}_i$ and $\widehat{\boldsymbol{\gamma}}_{ij}$ as in the proof of Lemma 2.2, we can show $\mathbf{B}_3$ is negligible. Then, all we need is to show that

$$\mathbf{B}^* \equiv \frac{1}{\sqrt{LNT}}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{Z}^{*\top}_{ij\cdot}[\mathbf{E}_{ij\cdot}\circ\boldsymbol{\xi}] \to_{D^*} N(\mathbf{0},\boldsymbol{\Omega}_1),$$

which can be done by verifying the Lindeberg condition using the large- and small-block technique. Applying the Cramér-Wold device, and letting $\boldsymbol{\eta}$ be a $d\times 1$ vector satisfying $\|\boldsymbol{\eta}\| = 1$, we consider

$$\widetilde{B}^* \equiv \boldsymbol{\eta}^\top\mathbf{B}^* = \frac{1}{\sqrt{LNT}}\sum_{i=1}^{L}\sum_{j=1}^{N}\boldsymbol{\eta}^\top\mathbf{Z}^{*\top}_{ij\cdot}[\mathbf{E}_{ij\cdot}\circ\boldsymbol{\xi}].$$

The goal is to show that

$$\widetilde{B}^* \to_{D^*} N(0, \boldsymbol{\eta}^\top\boldsymbol{\Omega}_1\boldsymbol{\eta}), \tag{A.30}$$

which in connection with Assumption 2.2 immediately yields the result.

We now rewrite $\widetilde{B}^*$ as follows.

$$\widetilde{B}^* = \sum_{j=1}^{K}\nu_j^* + \sum_{j=1}^{K}\varpi_j^*, \tag{A.31}$$

where

$$\nu_j^* = \sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{t=B_j+1}^{B_j+r_1} \frac{1}{\sqrt{LNT}} z_{\eta,ijt}\varepsilon_{ijt}\xi_t, \quad \varpi_j^* = \sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{t=B_j+r_1+1}^{B_j+r_1+r_2} \frac{1}{\sqrt{LNT}} z_{\eta,ijt}\varepsilon_{ijt}\xi_t,$$

and $z_{\eta,ijt} \equiv \boldsymbol{\eta}^\top \mathbf{z}_{ijt}^*$. Moreover, $B_j = (j-1)(r_1+r_2)$, and without loss of generality we suppose that $K = T/(r_1+r_2)$ is an integer for simplicity. Otherwise, one needs to include the remaining terms in (A.31) which are negligible for an obvious reason. In addition, we let

$$(r_1, r_2) \to (\infty, \infty), \quad \left(\frac{r_2}{r_1}, \frac{r_1}{T}\right) \to (0,0), \quad r_1 \geq \ell, \tag{A.32}$$

so the blocks ϖ_j^*'s are mutually independent by the construction of ξ_t's. Note that by $\frac{r_2}{r_1} \to 0$ of (A.32),

$$\frac{Kr_2}{T} \to 0 \quad \text{and} \quad \frac{Kr_1}{T} \to 1.$$

By construction, a direct calculation on the small blocks shows that

$$\mathbb{E}\mathbb{E}^*\left[\left(\sum_{j=1}^{K}\varpi_j^*\right)^2\right] = \sum_{j=1}^{K}\mathbb{E}\mathbb{E}^*[(\varpi_j^*)^2] = O(1)\frac{Kr_2}{T} = o(1).$$

Therefore, the term $\sum_{j=1}^{K}\varpi_j^*$ of (A.31) is negligible.

Next, we employ the Lindeberg CLT to establish the asymptotic normality of $\sum_{j=1}^{K}\nu_j^*$. We have shown that $(\widetilde{B}^*)^2 = \boldsymbol{\eta}^\top\boldsymbol{\Omega}_1\boldsymbol{\eta} + o_P(1)$ and $\sum_{j=1}^{K}\varpi_j^*$ of (A.31) is negligible, so it is easy to know that

$$\mathbb{E}^*(\sum_{j=1}^{K}\nu_j^*)^2 = \boldsymbol{\eta}^\top\boldsymbol{\Omega}_1\boldsymbol{\eta} + o_P(1).$$

Similar arguments can be seen in (A.8)-(A.9) of Chen et al. (2012). That said, we just need to verify that for $\forall\epsilon > 0$

$$\sum_{j=1}^{K}\mathbb{E}^*\left[(\nu_j^*)^2 \cdot I\left(|\nu_j^*| > \epsilon\right)\right] = o_P(1). \tag{A.33}$$

Before proceeding further, we point out that the series $\frac{1}{\sqrt{LNT}}z_{\eta,ijt}\varepsilon_{ijt}\xi_t$ is in fact a mixingale sequence by Definition 1 of Hansen (1991), where the term $|\frac{1}{\sqrt{LNT}}z_{\eta,ijt}\varepsilon_{ijt}\xi_t|$ is equivalent to c_i in the notation of Hansen (1991). This is not hard to justify given $\{\xi_t\}$ is an ℓ-dependent series. When m in the notation of Hansen (1991) is greater than ℓ, all the requirements of Definition 1 of Hansen (1991) are fulfilled. Thus, it allows us to invoke the asymptotic properties associated to the mixingale sequence in the following development.

Write

$$
\begin{aligned}
&\sum_{j=1}^{K}\mathbb{E}^*[(\nu_j^*)^2\cdot I(|\nu_j^*|>\epsilon)]\\
&\leq \sum_{j=1}^{K}\{\mathbb{E}^*|(\nu_j^*)^2|^{\delta/2}\}^{2/\delta}\cdot\{\mathbb{E}^*[I(|\nu_j^*|>\epsilon)]\}^{(\delta-2)/\delta}\leq \sum_{j=1}^{K}\{\mathbb{E}^*|(\nu_j^*)^2|^{\delta/2}\}^{2/\delta}\left\{\frac{\mathbb{E}^*|\nu_j^*|^\delta}{\epsilon^\delta}\right\}^{(\delta-2)/\delta}\\
&=\epsilon^{\delta-2}\sum_{j=1}^{K}\mathbb{E}^*|\nu_j^*|^\delta=\epsilon^{\delta-2}\sum_{j=1}^{K}\left\{\mathbb{E}^*\left[\left(\sum_{t=B_j+1}^{B_j+r_1}\frac{1}{\sqrt{LNT}}\sum_{i=1}^{L}\sum_{j=1}^{N}z_{\eta,ijt}\varepsilon_{ijt}\xi_t\right)^\delta\right]\right\}^{\frac{1}{\delta}\cdot\delta}\\
&\leq O(1)\epsilon^{\delta-2}\sum_{j=1}^{K}\left\{\sum_{t=B_j+1}^{B_j+r_1}\left(\frac{1}{\sqrt{LNT}}\sum_{i=1}^{L}\sum_{j=1}^{N}z_{\eta,ijt}\varepsilon_{ijt}\right)^2\right\}^{\frac{1}{2}\cdot\delta}\\
&\leq O(1)\epsilon^{\delta-2}\sum_{j=1}^{K}r_1^{\delta/2-1}\sum_{t=B_j+1}^{B_j+r_1}\left(\frac{1}{\sqrt{LNT}}\sum_{i=1}^{L}\sum_{j=1}^{N}z_{\eta,ijt}\varepsilon_{ijt}\right)^\delta\\
&\leq O(1)\epsilon^{\delta-2}\frac{r_1^{\delta/2-1}}{T^{\delta/2-1}}\cdot\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{LN}}\sum_{i=1}^{L}\sum_{j=1}^{N}z_{\eta,ijt}\varepsilon_{ijt}\right)^\delta\\
&=O_P(1)\frac{r_1^{\delta/2-1}}{T^{\delta/2-1}}=o_P(1),
\end{aligned}
$$

where the first inequality follows from the Hölder inequality, the second inequality follows from the Chebyshev's inequality, the third inequality follows from Lemma 2 of Hansen (1991), and the last equality follows from $r_1/(Th)\to 0$ and $\delta>2$ (say, letting $\delta=4$). Thus, we can conclude the validity of (A.33).

Based on the above development, we are ready to conclude that (A.30) holds. The proof is now completed. ■

Proof of Lemma A.3:

Before proving the two results of this lemma, we first derive some preliminary results. For (2.6), we conduct the PCA analysis as follows:

$$
\widehat{\mathbf{C}}\widehat{\mathbf{V}}^G=\widehat{\mathbf{\Sigma}}^G\widehat{\mathbf{C}}, \tag{A.34}
$$

where $\widehat{\mathbf{V}}^G=\text{diag}\{\widehat{\lambda}_1^G,\ldots,\widehat{\lambda}_{d_{\max}}^G\}$, and $\frac{1}{T}\widehat{\mathbf{C}}^\top\widehat{\mathbf{C}}=\mathbf{I}_{d_{\max}}$. We expand the right hand side of (A.34) as follows:

$$
\begin{aligned}
\widehat{\mathbf{C}}\widehat{\mathbf{V}}^G=&\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}\widehat{\mathbf{C}}\\
&+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}\widehat{\mathbf{C}}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}^G\boldsymbol{\gamma}_{ij}^G(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}^\top\widehat{\mathbf{C}}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij.}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}\widehat{\mathbf{C}}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}_i^{S}\boldsymbol{\gamma}_{ij}^{S}(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij.}^\top\widehat{\mathbf{C}}\\
&+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij.}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\mathbf{E}_{ij.}^\top\widehat{\mathbf{C}}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{E}_{ij.}(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij.}^\top\widehat{\mathbf{C}}\\
&+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}^{G}\boldsymbol{\gamma}_{ij}^{G}\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}\widehat{\mathbf{C}}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}^{G}\boldsymbol{\gamma}_{ij}^{G}\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}\widehat{\mathbf{C}}\\
&+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}_i^{S}\boldsymbol{\gamma}_{ij}^{S}\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}\widehat{\mathbf{C}}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}^{G}\boldsymbol{\gamma}_{ij}^{G}\mathbf{E}_{ij.}^\top\widehat{\mathbf{C}}\\
&+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{E}_{ij.}\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}\widehat{\mathbf{C}}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}_i^{S}\boldsymbol{\gamma}_{ij}^{S}\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}\widehat{\mathbf{C}}\\
&+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}_i^{S}\boldsymbol{\gamma}_{ij}^{S}\mathbf{E}_{ij.}^\top\widehat{\mathbf{C}}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{E}_{ij.}\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}\widehat{\mathbf{C}}\\
&+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{E}_{ij.}\mathbf{E}_{ij.}^\top\widehat{\mathbf{C}}:=\mathbf{J}_1+\cdots+\mathbf{J}_{16},
\end{aligned}
\tag{A.35}
$$

where the definitions of $\mathbf{J}_1$ to $\mathbf{J}_{16}$ should be obvious. In what follows, we examine the terms on the right hand side of (A.35) one by one.

For $\mathbf{J}_1$, write

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\|\mathbf{J}_1\|_2&=\frac{1}{\sqrt{T}}\left\|\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij.}(\boldsymbol{\beta}-\widehat{\mathbf{b}})(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij.}\widehat{\mathbf{C}}\right\|_2\\
&\le O(1)\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\|\mathbf{X}_{ij.}\|_2^2\cdot\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|^2=O_P(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|^2),
\end{aligned}
$$

where the last step follows from $\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\|\mathbf{X}_{ij.}\|_2^2=O_P(1)$ by Assumption 1.1.

For $\mathbf{J}_2$, write

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\|\mathbf{J}_2\|_2&=\frac{1}{\sqrt{T}}\left\|\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij.}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}\widehat{\mathbf{C}}\right\|_2\\
&\le O(1)\left\{\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\|\mathbf{X}_{ij.}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\|_2^2\right\}^{1/2}\left\{\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\|\mathbf{F}^{G}\boldsymbol{\gamma}_{ij}^{G}\|_2^2\right\}^{1/2}\\
&=O_P(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|),
\end{aligned}
$$

where the first inequality follows from Cauchy-Schwarz inequality, and the last step follows from Assumptions 1.1-1.3. Similarly, we can obtain that

$$
\sum_{\ell=3}^{7}\frac{1}{\sqrt{T}}\|\mathbf{J}_\ell\|_2=O_P(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|).
$$

To analyse $\mathbf{J}_9$, let

$$\boldsymbol{\Gamma}^S = \mathrm{diag}\{\boldsymbol{\Gamma}^S_{1\cdot}, \dots, \boldsymbol{\Gamma}^S_{L\cdot}\} \quad \text{with} \quad \boldsymbol{\Gamma}^S_{i\cdot} = (\boldsymbol{\gamma}^S_{i1}, \dots, \boldsymbol{\gamma}^S_{iN})^\top,$$

and write

$$\begin{aligned}
\|\boldsymbol{\Gamma}^S\|_2 &= \sqrt{\lambda_{\max}(\boldsymbol{\Gamma}^{S\top}\boldsymbol{\Gamma}^S)} = \sqrt{\lambda_{\max}(\mathrm{diag}\{\boldsymbol{\Gamma}^{S\top}_{1\cdot}\boldsymbol{\Gamma}^S_{1\cdot}, \dots, \boldsymbol{\Gamma}^{S\top}_{L\cdot}\boldsymbol{\Gamma}^S_{L\cdot}\})} \\
&\le \sqrt{\max_i \lambda_{\max}(\boldsymbol{\Gamma}^{S\top}_{i\cdot}\boldsymbol{\Gamma}^S_{i\cdot})} = \sqrt{\max_i \lambda_{\max}\left(\sum_{j=1}^N \boldsymbol{\gamma}^S_{ij}\boldsymbol{\gamma}^{S\top}_{ij}\right)} \\
&\le \sqrt{\max_i \sum_{j=1}^N \|\boldsymbol{\gamma}^S_{ij}\|^2} = O_P(\sqrt{N\log(LN)}), \qquad \text{(A.36)}
\end{aligned}$$

where the last equality follows from Assumption 3.1. Then we can write

$$\begin{aligned}
\frac{1}{\sqrt{T}}\|\mathbf{J}_9\|_2 &= \frac{1}{\sqrt{T}}\left\|\frac{1}{LNT}\sum_{i=1}^L\sum_{j=1}^N \mathbf{F}^G\boldsymbol{\gamma}^G_{ij}\boldsymbol{\gamma}^{S\top}_{ij}\mathbf{F}^{S\top}_i\widehat{\mathbf{C}}\right\|_2 \\
&\le O(1)\frac{1}{LNT}\left\|\mathbf{F}^G\boldsymbol{\Gamma}^{G\top}\boldsymbol{\Gamma}^S\mathbf{F}^{S\top}\right\|_2 \\
&\le O_P(1)\frac{1}{LNT}\cdot\sqrt{LNT}\cdot\sqrt{N\log(LN)}\cdot(\sqrt{T}\vee\sqrt{L}) \\
&= O_P\left(\frac{\sqrt{\log(LN)}}{\sqrt{N\wedge T}}\right),
\end{aligned}$$

where $\mathbf{F}^S$ has been defined in Assumption 3.1, and the second inequality follows from (A.36) and Assumptions 1.2, 1.3, and 3.1. Similarly, we can show that

$$\frac{1}{\sqrt{T}}\|\mathbf{J}_{10}\|_2 = O_P\left(\frac{\sqrt{\log(LN)}}{\sqrt{N\wedge T}}\right).$$

We now consider $\mathbf{J}_{11}$, and let

$$\mathbf{E} = (\mathbf{E}_{11\cdot}, \dots, \mathbf{E}_{1N\cdot}, \dots, \mathbf{E}_{L1\cdot}, \dots, \mathbf{E}_{LN\cdot})^\top.$$

Write

$$\frac{1}{\sqrt{T}}\|\mathbf{J}_{11}\|_2 = \frac{1}{\sqrt{T}}\left\|\frac{1}{LNT}\sum_{i=1}^L\sum_{j=1}^N \mathbf{F}^G\boldsymbol{\gamma}^G_{ij}\mathbf{E}^\top_{ij\cdot}\widehat{\mathbf{C}}\right\|_2 \le O(1)\frac{1}{LNT}\left\|\mathbf{F}^G\boldsymbol{\Gamma}^{G\top}\mathbf{E}\right\|_2 = O_P\left(\frac{1}{\sqrt{LN}}\right),$$

where the last step follows from $\|\mathbf{F}^G\| = O_P(\sqrt{T})$ by Assumption 1.2 and the fact that

$$\mathbb{E}\|\boldsymbol{\Gamma}^{G\top}\mathbf{E}\|^2 = \sum_{t=1}^T \mathbb{E}\left\|\sum_{i=1}^L\sum_{j=1}^N \boldsymbol{\gamma}^G_{ij}\varepsilon_{ijt}\right\|^2 = \sum_{i_1=1}^L\sum_{j_1=1}^N\sum_{i_2=1}^L\sum_{j_2=1}^N\sum_{t=1}^T \mathbb{E}[\boldsymbol{\gamma}^{G\top}_{i_1j_1}\boldsymbol{\gamma}^G_{i_2j_2}\varepsilon_{i_1j_1t}\varepsilon_{i_2j_2t}]$$

$$\leq O(1)\sum_{i_1=1}^{L}\sum_{j_1=1}^{N}\sum_{i_2=1}^{L}\sum_{j_2=1}^{N}\sum_{t=1}^{T}|\mathbb{E}[\varepsilon_{i_1j_1t}\varepsilon_{i_2j_2t}]| = O(LNT)$$

using Assumption 2.1. Similarly, we obtain that $\frac{1}{\sqrt{T}}\|\mathbf{J}_{12}\|_2 = O_P\left(\frac{1}{\sqrt{LN}}\right)$.

For $\mathbf{J}_{13}$, write

$$\frac{1}{\sqrt{T}}\|\mathbf{J}_{13}\|_2 = \frac{1}{\sqrt{T}}\left\|\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}_i^S\boldsymbol{\gamma}_{ij}^S\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}\widehat{\mathbf{C}}\right\|_2 \leq O(1)\frac{1}{LNT}\|\mathbf{F}^S\boldsymbol{\Gamma}^{S\top}\boldsymbol{\Gamma}^S\mathbf{F}^{S\top}\|_2$$
$$= \frac{1}{LNT}\|\mathbf{F}^S\|_2^2\cdot\|\boldsymbol{\Gamma}^S\|_2^2 = O_P\left(\frac{\log(LN)}{T\wedge L}\right),$$

where we have used (A.36).

For $\mathbf{J}_{14}$, write

$$\frac{1}{\sqrt{T}}\|\mathbf{J}_{14}\|_2 = \frac{1}{\sqrt{T}}\left\|\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}_i^S\boldsymbol{\gamma}_{ij}^S\mathbf{E}_{ij\cdot}^\top\widehat{\mathbf{C}}\right\|_2 \leq O(1)\frac{1}{LNT}\|\mathbf{F}^S\boldsymbol{\Gamma}^{S\top}\mathbf{E}\|_2 = O_P\left(\frac{1}{\sqrt{LN}}\right),$$

where the last step follows from the fact that

$$\mathbb{E}\|\mathbf{F}^S\boldsymbol{\Gamma}^{S\top}\mathbf{E}\|^2 = \sum_{i_1=1}^{L}\sum_{j_1=1}^{N}\sum_{i_2=1}^{L}\sum_{j_2=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\boldsymbol{\gamma}_{i_1j_1}^{S\top}\mathbf{F}_{i_1}^{S\top}\mathbf{F}_{i_2}^S\boldsymbol{\gamma}_{i_2j_2}^S\varepsilon_{i_1j_1t}\varepsilon_{i_2j_2t}]$$
$$\leq O(1)T^2\sum_{i_1=1}^{L}\sum_{j_1=1}^{N}\sum_{i_2=1}^{L}\sum_{j_2=1}^{N}|\mathbb{E}[\varepsilon_{i_1j_11}\varepsilon_{i_2j_21}]| = O(LNT^2),$$

using Assumption 2.1. In the same fashion, we can show that

$$\frac{1}{\sqrt{T}}\|\mathbf{J}_{15}\|_2 = O_P\left(\frac{1}{\sqrt{LN}}\right).$$

For $\mathbf{J}_{16}$, write

$$\frac{1}{\sqrt{T}}\|\mathbf{J}_{16}\|_2 = \frac{1}{\sqrt{T}}\left\|\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{E}_{ij\cdot}\mathbf{E}_{ij\cdot}^\top\widehat{\mathbf{C}}\right\|_2 \leq O(1)\frac{1}{LNT}\|\mathbf{E}^\top\mathbf{E}\|_2 = O_P\left(\frac{1}{LN}\vee\frac{1}{T}\right),$$

where the last step follows from the fact that

$$\frac{1}{L^2N^2T^2}\mathbb{E}\|\mathcal{E}^\top\mathcal{E}\|^2 = \frac{1}{L^2N^2T^2}\sum_{t_1,t_2}\sum_{i_1,j_1}\sum_{i_2,j_2}\mathbb{E}[\varepsilon_{i_1j_1t_1}\varepsilon_{i_1j_1t_2}\varepsilon_{i_2j_2t_1}\varepsilon_{i_2j_2t_2}]$$
$$= \frac{1}{L^2N^2T^2}\sum_{t_1,t_2}\sum_{i,j}\mathbb{E}[\varepsilon_{ijt_1}^2\varepsilon_{ijt_2}^2] + \frac{1}{L^2N^2}\sum_{(i_1,j_1)\neq(i_2,j_2)}\sigma_{i_1j_1,i_2j_2}^2$$
$$+\frac{1}{L^2N^2T^2}\sum_{t_1,t_2}\sum_{(i_1,j_1)\neq(i_2,j_2)}E[(\varepsilon_{i_1j_1t_1}\varepsilon_{i_2j_2t_1}-\sigma_{i_1j_1,i_2j_2})(\varepsilon_{i_1j_1t_2}\varepsilon_{i_2j_2t_2}-\sigma_{i_1j_1,i_2j_2})]$$

$$= \frac{1}{L^2N^2T^2}\sum_t \left(\sum_{i,j} E[\varepsilon_{ijt}^4] + \sum_{(i_1,j_1)\neq(i_2,j_2)} E[(\varepsilon_{i_1j_1t}\varepsilon_{i_2j_2t} - \sigma_{i_1j_1,i_2j_2})^2] \right)$$
$$+\frac{1}{L^2N^2T^2}\sum_{t_1\neq t_2}\sum_{i,j} E[\varepsilon_{ijt_1}^2\varepsilon_{ijt_2}^2]$$
$$+\frac{1}{L^2N^2T^2}\sum_{t_1\neq t_2}\sum_{(i_1,j_1)\neq(i_2,j_2)} E[(\varepsilon_{i_1j_1t_1}\varepsilon_{i_2j_2t_1} - \sigma_{i_1j_1,i_2j_2})(\varepsilon_{i_1j_1t_2}\varepsilon_{i_2j_2t_2} - \sigma_{i_1j_1,i_2j_2})]$$
$$+\frac{1}{L^2N^2}\sum_{(i_1,j_1)\neq(i_2,j_2)} \sigma_{i_1j_1,i_2j_2}^2 = O\left(\frac{1}{T}+\frac{1}{LN}\right),$$

in which $\mathbb{E}[\varepsilon_{i_1j_1t}\varepsilon_{i_2j_2t}] = \sigma_{i_1j_1,i_2j_2}$, and we have used the α-mixing condition of Assumption 2.1 in dealing with the term $\varepsilon_{i_1j_1t}\varepsilon_{i_2j_2t} - \sigma_{i_1j_1,i_2j_2}$.

Based on the above development, we can conclude that

$$\frac{1}{\sqrt{T}}\|\widehat{\mathbf{C}}\widehat{\mathbf{V}}^G - \mathbf{J}_8\|_2 = O_P\left(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}}\right). \quad \text{(A.37)}$$

Equation (A.37) immediately yields that

$$\left\|\widehat{\mathbf{V}}^G - \frac{1}{T}\widehat{\mathbf{C}}^\top\mathbf{F}^G\cdot\frac{1}{LN}\boldsymbol{\Gamma}^{G\top}\boldsymbol{\Gamma}^G\cdot\frac{1}{T}\mathbf{F}^{G\top}\widehat{\mathbf{C}}\right\| = O_P\left(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}}\right), \text{(A.38)}$$

which implies that $\widehat{V}^G$ is at most of rank l^G. By left multiplying (A.37) by $\frac{1}{T}F^{G\top}$, we obtain that

$$\left\|\frac{1}{T}\mathbf{F}^{G\top}\widehat{\mathbf{C}}\widehat{\mathbf{V}}^G - \frac{1}{T}\mathbf{F}^{G\top}\mathbf{F}^G\cdot\frac{1}{LN}\boldsymbol{\Gamma}^{G\top}\boldsymbol{\Gamma}^G\cdot\frac{1}{T}\mathbf{F}^{G\top}\widehat{\mathbf{C}}\right\| = O_P\left(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}}\right).$$

Note that $\frac{1}{T}\mathbf{F}^{G\top}\widehat{\mathbf{C}}$ is of rank l^G, which further indicates that $\widehat{\mathbf{V}}^G$ has at least l^G non-zero elements on the main diagonal which converge to the eigenvalues of $\frac{1}{T}\mathbf{F}^{G\top}\mathbf{F}^G\cdot\frac{1}{LN}\boldsymbol{\Gamma}^{G\top}\boldsymbol{\Gamma}^G$. We can then conclude that $\widehat{\mathbf{V}}^G$ is of rank l^G in limit.

We are now ready to investigate the following two parts of the proof of this lemma.

(1). We focus on the first l^G columns of $\widehat{\mathbf{C}}$, and denote them as $\widehat{\mathbf{C}}^G$ as in the description of this lemma. Correspondingly, we let $\widetilde{\mathbf{V}}^G$ be the leading $l^G\times l^G$ principal submatrix of $\widehat{\mathbf{V}}^G$. Further let $\widehat{\mathbf{C}}_\ell^G$ be the ℓ^{th} column of $\widehat{\mathbf{C}}^G$. By (A.37), we can further write

$$\frac{1}{\sqrt{T}}\|\widehat{\mathbf{C}}^G - \mathbf{F}^G\mathbf{H}^G\|_2 = O_P\left(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}}\right), \quad \text{(A.39)}$$

where $\mathbf{H}^G = \frac{1}{LN}\boldsymbol{\Gamma}^{G\top}\boldsymbol{\Gamma}^G\cdot\frac{1}{T}\mathbf{F}^{G\top}\widehat{\mathbf{C}}^G\cdot(\widetilde{\mathbf{V}}^G)^{-1}$. Also, in what follows, let

$$\boldsymbol{\Sigma}^G = \frac{1}{LNT}\mathbf{F}^G\boldsymbol{\Gamma}^{G\top}\boldsymbol{\Gamma}^G\mathbf{F}^{G\top}.$$

Note that $\widehat{\Sigma}^G$ admits the next expansion:

$$\widehat{\boldsymbol{\Sigma}}^G = \frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}^\top$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}^G\boldsymbol{\gamma}_{ij}^G(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}^\top$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}_i^S\boldsymbol{\gamma}_{ij}^S(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}^\top$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{X}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})\mathbf{E}_{ij\cdot}^\top+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{E}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}^\top$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}^G\boldsymbol{\gamma}_{ij}^G\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}^G\boldsymbol{\gamma}_{ij}^G\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}_i^S\boldsymbol{\gamma}_{ij}^S\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}^G\boldsymbol{\gamma}_{ij}^G\mathbf{E}_{ij\cdot}^\top$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{E}_{ij\cdot}\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}_i^S\boldsymbol{\gamma}_{ij}^S\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{F}_i^S\boldsymbol{\gamma}_{ij}^S\mathbf{E}_{ij\cdot}^\top+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{E}_{ij\cdot}\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}+\frac{1}{LNT}\sum_{i=1}^{L}\sum_{j=1}^{N}\mathbf{E}_{ij\cdot}\mathbf{E}_{ij\cdot}^\top$$

By the development of the terms $\mathbf{J}_1-\mathbf{J}_{16}$ above, it is easy to show that

$$\|\widehat{\boldsymbol{\Sigma}}^G-\boldsymbol{\Sigma}^G\|_2 = O_P\left(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|+\frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}}\right). \quad \text{(A.40)}$$

In the context of this lemma, we have also defined $\widehat{\lambda}_\ell^G$ and λ_ℓ^G. These notations and results will be repeatedly used below.

Let's now consider $\widehat{\lambda}_\ell^G-\lambda_\ell^G$, and write

$$\begin{aligned}
&\widehat{\lambda}_\ell^G-\lambda_\ell^G\\
&=\frac{1}{\sqrt{T}}(\widehat{\mathbf{C}}_\ell^G-\mathbf{F}^G\mathbf{H}_\ell^G+\mathbf{F}^G\mathbf{H}_\ell^G)^\top(\widehat{\boldsymbol{\Sigma}}^G-\boldsymbol{\Sigma}^G+\boldsymbol{\Sigma}^G)\frac{1}{\sqrt{T}}(\widehat{\mathbf{C}}_\ell^G-\mathbf{F}^G\mathbf{H}_\ell^G+\mathbf{F}^G\mathbf{H}_\ell^G)\\
&\quad-\frac{1}{T}\mathbf{H}_\ell^{G\top}\mathbf{F}^{G\top}\boldsymbol{\Sigma}^G\mathbf{F}^G\mathbf{H}_\ell^G=\frac{1}{\sqrt{T}}(\widehat{\mathbf{C}}_\ell^G-\mathbf{F}^G\mathbf{H}_\ell^G)^\top(\widehat{\boldsymbol{\Sigma}}^G-\boldsymbol{\Sigma}^G)\frac{1}{\sqrt{T}}(\widehat{\mathbf{C}}_\ell^G-\mathbf{F}^G\mathbf{H}_\ell^G)\\
&\quad+\frac{2}{\sqrt{T}}(\widehat{\mathbf{C}}_\ell^G-\mathbf{F}^G\mathbf{H}_\ell^G)^\top(\widehat{\boldsymbol{\Sigma}}^G-\boldsymbol{\Sigma}^G)\frac{1}{\sqrt{T}}\mathbf{F}^G\mathbf{H}_\ell^G\\
&\quad+\frac{1}{\sqrt{T}}(\widehat{\mathbf{C}}_\ell^G-\mathbf{F}^G\mathbf{H}_\ell^G)^\top\boldsymbol{\Sigma}^G\frac{1}{\sqrt{T}}(\widehat{\mathbf{C}}_\ell^G-\mathbf{F}^G\mathbf{H}_\ell^G)\\
&\quad+\frac{2}{\sqrt{T}}(\widehat{\mathbf{C}}_\ell^G-\mathbf{F}^G\mathbf{H}_\ell^G)^\top\boldsymbol{\Sigma}^G\frac{1}{\sqrt{T}}\mathbf{F}^G\mathbf{H}_\ell^G+\frac{1}{\sqrt{T}}(\mathbf{F}^G\mathbf{H}_\ell^G)^\top(\widehat{\boldsymbol{\Sigma}}^G-\boldsymbol{\Sigma}^G)\frac{1}{\sqrt{T}}\mathbf{F}^G\mathbf{H}_\ell^G\\
&:=\mathbf{A}_1+2\mathbf{A}_2+\mathbf{A}_3+2\mathbf{A}_4+\mathbf{A}_5,
\end{aligned}$$

where the definitions of $\mathbf{A}_1$ to $\mathbf{A}_5$ are obvious.

By (A.39) and (A.40), we can immediately conclude that $|\mathbf{A}_1| = o_P(|\mathbf{A}_5|)$, $|\mathbf{A}_2| = o_P(|\mathbf{A}_5|)$, and $|\mathbf{A}_3| = o_P(|\mathbf{A}_4|)$. Thus, we focus on $\mathbf{A}_4$ and $\mathbf{A}_5$ below. For $\mathbf{A}_4$, we write

$$|\mathbf{A}_4| \le \frac{1}{\sqrt{T}}\|\widehat{\mathbf{C}}_\ell^G - \mathbf{F}^G\mathbf{H}_\ell^G\|_2 \cdot \|\mathbf{\Sigma}^G\|_2 \cdot \frac{1}{\sqrt{T}}\|\mathbf{F}^G\mathbf{H}_\ell^G\|_2 = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}}\right),$$

where the last equality follows from (A.39) and the fact that $\|\mathbf{\Sigma}^G\|_2 = O_P(1)$ and $\frac{1}{\sqrt{T}}\|\mathbf{F}^G\mathbf{H}_\ell^G\|_2 = O_P(1)$ by the construction. For $\mathbf{A}_5$, write

$$\begin{aligned}|\mathbf{A}_5| &= \left|\frac{1}{\sqrt{T}}(\mathbf{F}^G\mathbf{H}_\ell^G)^\top(\widehat{\mathbf{\Sigma}}^G - \mathbf{\Sigma}^G)\frac{1}{\sqrt{T}}\mathbf{F}^G\mathbf{H}_\ell^G\right| \le \|\widehat{\mathbf{\Sigma}}^G - \mathbf{\Sigma}^G\|_2 \cdot \frac{1}{T}\|\mathbf{F}^G\mathbf{H}_\ell^G\|_2^2\\ &= O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}}\right),\end{aligned}$$

where the last step follows from (A.40).

This concludes the proof of the first result of this lemma.

(2). To investigate the second result, we start the proof by introducing some notations. We denote $\mathbf{F}^{G\perp}$ as a $T\times(T-l^G)$ matrix such that $\frac{1}{T}(\mathbf{F}^{G\perp}, \mathbf{F}^G\mathbf{R})^\top(\mathbf{F}^{G\perp}, \mathbf{F}^G\mathbf{R}) = \mathbf{I}_T$, where $\mathbf{R}$ is a $l^G\times l^G$ rotation matrix. The matrices $\frac{1}{\sqrt{T}}\mathbf{F}^{G\perp}$, $\frac{1}{\sqrt{T}}\mathbf{F}^G\mathbf{R}$, $\mathbf{\Sigma}^G$, and $\widehat{\mathbf{\Sigma}}^G - \mathbf{\Sigma}^G$ correspond to $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{A}$, and $\mathbf{E}$ of Lemma A.1. Thus, the counterpart of the matrix $\mathbf{Q}_1^0$ becomes

$$\widehat{\mathbf{C}}^{G\perp} = \frac{1}{\sqrt{T}}(\mathbf{F}^{G\perp} + \mathbf{F}^G\mathbf{R}\mathbf{P})(\mathbf{I}_{T-l^G} + \mathbf{P}^\top\mathbf{P})^{-1/2},$$

in which

$$\|\mathbf{P}\|_2 \le \frac{4}{\mathrm{sep}(0, \frac{1}{T}\mathbf{R}^\top\mathbf{F}^{G\top}\mathbf{\Sigma}^G\mathbf{F}^G\mathbf{R})}\|\widehat{\mathbf{\Sigma}}^G - \mathbf{\Sigma}^G\|_2 = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}}\right). \quad (A.41)$$

Moreover, $\widehat{\mathbf{C}}^{G\perp}$ is an orthonormal basis for a subspace that is invariant for $\widehat{\mathbf{\Sigma}}^G$. In addition, note that

$$\begin{aligned}&\left\|\widehat{\mathbf{C}}^{G\perp} - \frac{1}{\sqrt{T}}\mathbf{F}^{G\perp}\right\|_2\\ &= \frac{1}{\sqrt{T}}\left\|\left[(\mathbf{F}^{G\perp} + \mathbf{F}^G\mathbf{R}\mathbf{P}) - \mathbf{F}^{G\perp}(\mathbf{I}_{T-l^G} + \mathbf{P}^\top\mathbf{P})^{1/2}\right](\mathbf{I}_{T-l^G} + \mathbf{P}^\top\mathbf{P})^{-1/2}\right\|_2\\ &\le \frac{1}{\sqrt{T}}\left\|\mathbf{F}^{G\perp}\left[\mathbf{I}_{T-l^G} - (\mathbf{I}_{T-l^G} + \mathbf{P}^\top\mathbf{P})^{1/2}\right](\mathbf{I}_{T-l^G} + \mathbf{P}^\top\mathbf{P})^{-1/2}\right\|_2\\ &\quad + \frac{1}{\sqrt{T}}\left\|\mathbf{F}^G\mathbf{R}\mathbf{P}(\mathbf{I}_{T-l^G} + \mathbf{P}^\top\mathbf{P})^{-1/2}\right\|_2\\ &\le O_P(1)\left\|\left[\mathbf{I}_{T-l^G} - (\mathbf{I}_{T-l^G} + \mathbf{P}^\top\mathbf{P})^{1/2}\right](\mathbf{I}_{T-l^G} + \mathbf{P}^\top\mathbf{P})^{-1/2}\right\|_2\\ &\quad + \left\|\mathbf{P}(\mathbf{I}_{T-l^G} + \mathbf{P}^\top\mathbf{P})^{-1/2}\right\|_2\\ &\le O_P(1)\left\|\mathbf{I}_{T-l^G} - (\mathbf{I}_{T-l^G} + \mathbf{P}^\top\mathbf{P})^{1/2}\right\|_2 + O_P(1)\,\|\mathbf{P}\|_2\end{aligned}$$

$$= O_P(\|\mathbf{P}\|_2) = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T \wedge L}}\right),$$

where the last equality follows from (A.41).

Now, let $\widehat{\mathbf{C}}_\ell^{G\perp}$ be the ℓ^{th} column of $\widehat{\mathbf{C}}^{G\perp}$. Since $\widehat{\mathbf{C}}^{G\perp}$ is an orthonormal basis for a subspace that is invariant for $\widehat{\boldsymbol{\Sigma}}^G$, for $\ell = 1, \dots, T - l^G$ we write

$$\begin{aligned}
\widehat{\lambda}_{l^G+\ell}^G &= \left(\widehat{\mathbf{C}}_\ell^{G\perp} - \frac{1}{\sqrt{T}}\mathbf{F}_\ell^{G\perp} + \frac{1}{\sqrt{T}}\mathbf{F}_\ell^{G\perp}\right)^\top (\widehat{\boldsymbol{\Sigma}}^G - \boldsymbol{\Sigma}^G + \boldsymbol{\Sigma}^G) \\
&\quad \cdot \left(\widehat{\mathbf{C}}_\ell^{G\perp} - \frac{1}{\sqrt{T}}\mathbf{F}_\ell^{G\perp} + \frac{1}{\sqrt{T}}\mathbf{F}_\ell^{G\perp}\right) \\
&\leq \left\|\widehat{\mathbf{C}}^{G\perp} - \frac{1}{\sqrt{T}}\mathbf{F}^{G\perp}\right\|_2^2 \cdot \|\widehat{\boldsymbol{\Sigma}}^G - \boldsymbol{\Sigma}^G\|_2 \\
&\quad + 2\left\|\widehat{\mathbf{C}}^{G\perp} - \frac{1}{\sqrt{T}}\mathbf{F}^{G\perp}\right\|_2 \cdot \|\widehat{\boldsymbol{\Sigma}}^G - \boldsymbol{\Sigma}^G\|_2 \cdot \frac{1}{\sqrt{T}}\|\mathbf{F}^{G\perp}\|_2 \\
&\quad + \left\|\widehat{\mathbf{C}}^{G\perp} - \frac{1}{\sqrt{T}}\mathbf{F}^{G\perp}\right\|_2^2 \cdot \|\boldsymbol{\Sigma}^G\|_2 \\
&= O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|^2 + \frac{\log(LN)}{T \wedge L}\right).
\end{aligned}$$

The proof of the second result of this lemma is now completed. ■

Proof of Lemma A.4:

Note that PCA yields the following equation:

$$\widehat{\mathbf{C}}_i \widehat{\mathbf{V}}_i^S = \widehat{\boldsymbol{\Sigma}}_i^S \widehat{\mathbf{C}}_i, \tag{A.42}$$

where $\widehat{\mathbf{V}}_i^S = \operatorname{diag}\{\widehat{\lambda}_{i,1}^S, \dots, \widehat{\lambda}_{i,d_{\max}}^S\}$. Below we expand the right hand side of (A.42) and examine the terms one by one.

$$\begin{aligned}
\widehat{\mathbf{C}}_i \widehat{\mathbf{V}}_i &= \mathbf{M}_{\widehat{\mathbf{F}}^G} \frac{1}{NT} \sum_{j=1}^N \mathbf{X}_{ij\cdot}(\boldsymbol{\beta} - \widehat{\mathbf{b}})(\boldsymbol{\beta} - \widehat{\mathbf{b}})^\top \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{F}}^G} \widehat{\mathbf{C}}_i \\
&\quad + \mathbf{M}_{\widehat{\mathbf{F}}^G} \frac{1}{NT} \sum_{j=1}^N \mathbf{X}_{ij\cdot}(\boldsymbol{\beta} - \widehat{\mathbf{b}})\boldsymbol{\gamma}_{ij}^{G\top} \mathbf{F}^{G\top} \mathbf{M}_{\widehat{\mathbf{F}}^G} \widehat{\mathbf{C}}_i \\
&\quad + \mathbf{M}_{\widehat{\mathbf{F}}^G} \frac{1}{NT} \sum_{j=1}^N \mathbf{F}^G \boldsymbol{\gamma}_{ij}^G (\boldsymbol{\beta} - \widehat{\mathbf{b}})^\top \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{F}}^G} \widehat{\mathbf{C}}_i \\
&\quad + \mathbf{M}_{\widehat{\mathbf{F}}^G} \frac{1}{NT} \sum_{j=1}^N \mathbf{X}_{ij\cdot}(\boldsymbol{\beta} - \widehat{\mathbf{b}})\boldsymbol{\gamma}_{ij}^{S\top} \mathbf{F}_i^{S\top} \mathbf{M}_{\widehat{\mathbf{F}}^G} \widehat{\mathbf{C}}_i \\
&\quad + \mathbf{M}_{\widehat{\mathbf{F}}^G} \frac{1}{NT} \sum_{j=1}^N \mathbf{F}_i^S \boldsymbol{\gamma}_{ij}^S (\boldsymbol{\beta} - \widehat{\mathbf{b}})^\top \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{F}}^G} \widehat{\mathbf{C}}_i \\
&\quad + \mathbf{M}_{\widehat{\mathbf{F}}^G} \frac{1}{NT} \sum_{j=1}^N \mathbf{X}_{ij\cdot}(\boldsymbol{\beta} - \widehat{\mathbf{b}}) \mathbf{E}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{F}}^G} \widehat{\mathbf{C}}_i
\end{aligned}$$

$$
\begin{aligned}
&+\mathbf{M}_{\widehat{\mathbf{F}}^G}\frac{1}{NT}\sum_{j=1}^{N}\mathbf{E}_{ij\cdot}(\boldsymbol{\beta}-\widehat{\mathbf{b}})^\top\mathbf{X}_{ij\cdot}^\top\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i \\
&+\mathbf{M}_{\widehat{\mathbf{F}}^G}\frac{1}{NT}\sum_{j=1}^{N}\mathbf{F}^G\boldsymbol{\gamma}_{ij}^G\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i+\mathbf{M}_{\widehat{\mathbf{F}}^G}\frac{1}{NT}\sum_{j=1}^{N}\mathbf{F}^G\boldsymbol{\gamma}_{ij}^G\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i \\
&+\mathbf{M}_{\widehat{\mathbf{F}}^G}\frac{1}{NT}\sum_{j=1}^{N}\mathbf{F}_i^S\boldsymbol{\gamma}_{ij}^S\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i+\mathbf{M}_{\widehat{\mathbf{F}}^G}\frac{1}{NT}\sum_{j=1}^{N}\mathbf{F}^G\boldsymbol{\gamma}_{ij}^G\mathbf{E}_{ij\cdot}^\top\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i \\
&+\mathbf{M}_{\widehat{\mathbf{F}}^G}\frac{1}{NT}\sum_{j=1}^{N}\mathbf{E}_{ij\cdot}\boldsymbol{\gamma}_{ij}^{G\top}\mathbf{F}^{G\top}\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i+\mathbf{M}_{\widehat{\mathbf{F}}^G}\frac{1}{NT}\sum_{j=1}^{N}\mathbf{F}_i^S\boldsymbol{\gamma}_{ij}^S\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i \\
&+\mathbf{M}_{\widehat{\mathbf{F}}^G}\frac{1}{NT}\sum_{j=1}^{N}\mathbf{F}_i^S\boldsymbol{\gamma}_{ij}^S\mathbf{E}_{ij\cdot}^\top\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i+\mathbf{M}_{\widehat{\mathbf{F}}^G}\frac{1}{NT}\sum_{j=1}^{N}\mathbf{E}_{ij\cdot}\boldsymbol{\gamma}_{ij}^{S\top}\mathbf{F}_i^{S\top}\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i \\
&+\mathbf{M}_{\widehat{\mathbf{F}}^G}\frac{1}{NT}\sum_{j=1}^{N}\mathbf{E}_{ij\cdot}\mathbf{E}_{ij\cdot}^\top\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i:=\mathbf{J}_{i,1}^S+\cdots+\mathbf{J}_{i,16}^S,
\end{aligned}
$$

where the definitions of $\mathbf{J}_{i,1}^S-\mathbf{J}_{i,16}^S$ are obvious.

In view of the fact that $\|\mathbf{M}_{\widehat{\mathbf{F}}^G}\|_2=1$, applying the same arguments as those for $\mathbf{J}_1$ to $\mathbf{J}_7$ of Lemma A.3, we can show that

$$
\frac{1}{\sqrt{T}}\sum_{\ell=1}^{7}\|\mathbf{J}_{i,\ell}^S\|_2=O_P(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|).
$$

For $\mathbf{J}_{i,8}^S$, write

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\|\mathbf{J}_{i,8}^S\|_2&=\frac{1}{\sqrt{T}}\left\|\frac{1}{NT}\mathbf{M}_{\widehat{\mathbf{F}}^G}\mathbf{F}^G\boldsymbol{\Gamma}_{i\cdot}^{G\top}\boldsymbol{\Gamma}_{i\cdot}^G\mathbf{F}^{G\top}\mathbf{M}_{\widehat{\mathbf{F}}^G}\widehat{\mathbf{C}}_i\right\|_2 \\
&\le O(1)\frac{1}{NT}\|\mathbf{M}_{\widehat{\mathbf{F}}^G}\mathbf{F}^G\boldsymbol{\Gamma}_{i\cdot}^{G\top}\boldsymbol{\Gamma}_{i\cdot}^G\mathbf{F}^{G\top}\mathbf{M}_{\widehat{\mathbf{F}}^G}\|_2 \\
&\le O(1)\frac{1}{N}\|\boldsymbol{\Gamma}_{i\cdot}^G\|_2^2\cdot\frac{1}{T}\|\mathbf{M}_{\widehat{\mathbf{F}}^G}(\mathbf{F}^G-\widehat{\mathbf{C}}^G(\mathbf{H}^G)^{-1})\|_2^2 \\
&=O_P\left(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|^2+\frac{\log(LN)}{T\wedge L}\right),
\end{aligned}
$$

where $\boldsymbol{\Gamma}_{i\cdot}^G$ has been defined in Assumption 1, $\mathbf{H}^G$ has been defined in Lemma A.3, and the last equality follows from (A.39). Similarly, we obtain that

$$
\frac{1}{\sqrt{T}}\sum_{\ell=9}^{12}\|\mathbf{J}_{i,\ell}^S\|_2=O_P\left(\|\boldsymbol{\beta}-\widehat{\mathbf{b}}\|+\frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}}\right).
$$

Again, in view of the fact that $\|\mathbf{M}_{\widehat{\mathbf{F}}^G}\|_2=1$, applying the same procedure as that used for $\mathbf{J}_{i,7}$ to $\mathbf{J}_{i,9}$ of Lemma 2.2, we can show that

$$
\frac{1}{\sqrt{T}}\sum_{\ell=14}^{16}\|\mathbf{J}_{i,\ell}^S\|=O_P\left(\frac{1}{\sqrt{N\wedge T}}\right).
$$

Then, we just need to concentrate on $\mathbf{J}_{i,13}^S$.

$$\mathbf{J}_{i,13}^S = \frac{1}{NT}\mathbf{F}_i^S\boldsymbol{\Gamma}_{i\cdot}^{S\top}\boldsymbol{\Gamma}_{i\cdot}^S\mathbf{F}_i^{S\top} + \frac{1}{NT}\mathbf{P}_{\widehat{\mathbf{C}}^G}\mathbf{F}_i^S\boldsymbol{\Gamma}_{i\cdot}^{S\top}\boldsymbol{\Gamma}_{i\cdot}^S\mathbf{F}_i^{S\top}\mathbf{P}_{\widehat{\mathbf{C}}^G}$$
$$-\frac{1}{NT}\mathbf{P}_{\widehat{\mathbf{C}}^G}\mathbf{F}_i^S\boldsymbol{\Gamma}_{i\cdot}^{S\top}\boldsymbol{\Gamma}_{i\cdot}^S\mathbf{F}_i^{S\top} - \frac{1}{NT}\mathbf{F}_i^S\boldsymbol{\Gamma}_{i\cdot}^{S\top}\boldsymbol{\Gamma}_{i\cdot}^S\mathbf{F}_i^{S\top}\mathbf{P}_{\widehat{\mathbf{C}}^G}.$$

For the terms on the right hand side, we can obtain that

$$\begin{aligned}
\frac{1}{NT}\|\mathbf{F}_i^S\boldsymbol{\Gamma}_{i\cdot}^{S\top}\boldsymbol{\Gamma}_{i\cdot}^S\mathbf{F}_i^{S\top}\mathbf{P}_{\widehat{\mathbf{C}}^G}\|_2 &\le \frac{1}{NT}\cdot\frac{1}{T}\|\mathbf{F}_i^S\boldsymbol{\Gamma}_{i\cdot}^{S\top}\boldsymbol{\Gamma}_{i\cdot}^S\mathbf{F}_i^{S\top}(\widehat{\mathbf{C}}^G - \mathbf{F}^G\mathbf{H}^G)\widehat{\mathbf{C}}^{G\top}\|_2 \\
&\quad + \frac{1}{NT}\cdot\frac{1}{T}\|\mathbf{F}_i^S\boldsymbol{\Gamma}_{i\cdot}^{S\top}\boldsymbol{\Gamma}_{i\cdot}^S\mathbf{F}_i^{S\top}\mathbf{F}^G\mathbf{H}^G\widehat{\mathbf{C}}^{G\top}\|_2 \\
&= \frac{1}{NT}\cdot\|\mathbf{F}_i^S\|_2^2\cdot\|\boldsymbol{\Gamma}_{i\cdot}^S\|_2^2\cdot\frac{1}{\sqrt{T}}\|\widehat{\mathbf{C}}^G - \mathbf{F}^G\mathbf{H}^G\|_2\cdot\frac{1}{\sqrt{T}}\|\widehat{\mathbf{C}}^G\|_2 \\
&\quad + \frac{1}{NT}\cdot\|\mathbf{F}_i^S\|_2\cdot\|\boldsymbol{\Gamma}_{i\cdot}^S\|_2^2\cdot\frac{1}{T}\|\mathbf{F}_i^{S\top}\mathbf{F}^G\|_2\cdot\|\widehat{\mathbf{C}}^G\|_2 \\
&= O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}} + T^\nu\right),
\end{aligned}$$

where the second equality follows from (A.39) and Assumption 3.2. Similarly,

$$\begin{aligned}
\frac{1}{NT}\|\mathbf{P}_{\widehat{\mathbf{C}}^G}\mathbf{F}_i^S\boldsymbol{\Gamma}_{i\cdot}^{S\top}\boldsymbol{\Gamma}_{i\cdot}^S\mathbf{F}_i^{S\top}\mathbf{P}_{\widehat{\mathbf{C}}^G}\|_2 &\le O_P(1)\Big\{\frac{1}{T}\|\widehat{\mathbf{C}}^G - \mathbf{F}^G\mathbf{H}^G\|_2^2 + \frac{1}{T^2}\|\mathbf{F}_i^{S\top}\mathbf{F}^G\|_2^2\Big\} \\
&= O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|^2 + \frac{\log(LN)}{T\wedge L} + T^{2\nu}\right).
\end{aligned}$$

Thus, we obtain

$$\frac{1}{\sqrt{T}}\left\|\mathbf{J}_{i,13}^S - \frac{1}{NT}\mathbf{F}_i^S\boldsymbol{\Gamma}_{i\cdot}^{S\top}\boldsymbol{\Gamma}_{i\cdot}^S\mathbf{F}_i^{S\top}\widehat{\mathbf{C}}_i\right\|_2 = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}} + T^\nu\right).$$

Applying the same argument as that used for Lemma A.3, we can conclude that $\widehat{\mathbf{V}}_i^S$ is of rank l_i^S in limit, and

$$\frac{1}{\sqrt{T}}\|\widehat{\mathbf{C}}_i^S - \mathbf{F}_i^S\mathbf{H}_i^S\|_2 = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T\wedge L}} + T^\nu\right),$$

where $\widehat{\mathbf{C}}_i^S$ and $\mathbf{H}_i^S$ have been defined in the context of this lemma. The rest of the proof is identical to those in Lemma A.3, and therefore is omitted. The proof is now completed.■

Proof of Theorem 2.3:

Note that $\Pr(\widehat{\ell}^G = l^G, \widehat{\boldsymbol{\ell}}^S = \mathbf{l}^S) = \Pr(\widehat{\boldsymbol{\ell}}^S = \mathbf{l}^S | \widehat{\ell}^G = l^G)\Pr(\widehat{\ell}^G = l^G)$. Therefore, in what follows, we first show that $\Pr(\widehat{\ell}^G = l^G) \to 1$, and then prove that $\Pr(\widehat{\boldsymbol{\ell}}^S = \mathbf{l}^S | \widehat{\ell}^G = l^G) \to 1$ in the second step.

Step 1. First, consider the case when $l^G = 0$. By Lemma A.3, we have

$$\widehat{\lambda}^G_{l^G+\ell} = \widehat{\lambda}^G_\ell = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\| + \frac{\sqrt{\log(LN)}}{\sqrt{T \wedge L}}\right),$$

for $\ell = 1, \ldots, d_{\max}$, which is less than ω with a probability approaching one. By the construction of the mock eigenvalue, we immediately obtain that $\Pr(\widehat{l}^G = 0) \to 1$.

Next, we consider the case with $l^G > 0$. Note that for $\ell = 1, \ldots, l^G$,

$$\begin{aligned}
\lambda^G_\ell &= \frac{1}{T}\mathbf{H}^{G\top}_\ell \mathbf{F}^{G\top}\boldsymbol{\Sigma}^G\mathbf{F}^G\mathbf{H}^G_\ell \\
&= \frac{1}{T}\mathbf{H}^{G\top}_\ell \mathbf{F}^{G\top}\mathbf{F}^G \cdot \frac{1}{LN}\boldsymbol{\Gamma}^{G\top}\boldsymbol{\Gamma}^G \cdot \frac{1}{T}\mathbf{F}^{G\top}\mathbf{F}^G\mathbf{H}^G_\ell \\
&\asymp \frac{1}{T}\widehat{\mathbf{C}}^{G\top}_\ell \mathbf{F}^G \cdot \frac{1}{LN}\boldsymbol{\Gamma}^{G\top}\boldsymbol{\Gamma}^G \cdot \frac{1}{T}\mathbf{F}^{G\top}\widehat{\mathbf{C}}^G_\ell \asymp 1,
\end{aligned}$$

where the first $\asymp$ follows from (A.39), and the second $\asymp$ follows from (A.38).

By Lemma A.3, $\widehat{\lambda}^G_\ell \asymp \lambda^G_\ell$ for $\ell = 1, \ldots, l^G$, which are larger than ω with a probability approaching one. Thus, for $\ell = 1, \ldots, l^G - 1$ we can conclude that

$$\frac{\widehat{\lambda}^G_{\ell+1}}{\widehat{\lambda}^G_\ell}\mathbb{I}(\widehat{\lambda}^G_\ell \geq \omega) + \mathbb{I}(\widehat{\lambda}^G_\ell < \omega) \asymp 1.$$

For $\ell = l^G + 1, \ldots, d_{\max}$, by Lemma A.3, $\widehat{\lambda}^G_\ell = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|^2 + \frac{\log(LN)}{T \wedge L} + T^{2\nu}\right)$, which is less than ω with a probability approaching one. Thus,

$$\frac{\widehat{\lambda}^G_{\ell+1}}{\widehat{\lambda}^G_\ell}\mathbb{I}(\widehat{\lambda}^G_\ell \geq \omega) + \mathbb{I}(\widehat{\lambda}^G_\ell < \omega) = 1$$

for $\ell = l^G + 1, \ldots, d_{\max}$ by construction. In addition, for $\ell = l^G$, it is straightforward to obtain that

$$\frac{\widehat{\lambda}_{l^G+1}}{\widehat{\lambda}_{l^G}} = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|^2 + \frac{\log(LN)}{T \wedge L} + T^{2\nu}\right)$$

using the facts that $\widehat{\lambda}^G_{l^G} \asymp 1$ and $\widehat{\lambda}^G_{l^G+1} = O_P\left(\|\boldsymbol{\beta} - \widehat{\mathbf{b}}\|^2 + \frac{\log(LN)}{T \wedge L} + T^{2\nu}\right)$. Thus, we are ready to conclude that $P(\widehat{l}^G = l^G) \to 1$.

Step 2. Below, we consider two cases: (i) there is at least one $\widehat{\ell}^S_i < l^S_i$ in $\widehat{\ell}^S$, and (ii) there is at least one $\widehat{\ell}^S_i > l^S_i$ in $\widehat{\ell}^S$. Note that case (i) does not rule out the possibility that other estimated numbers of factors may be larger than the true value. Similarly, case (ii) does not rule out the possibility that other estimated numbers of factors may be less than the true value. If we can rule out both cases with a probability approaching one, then $\Pr(\widehat{\boldsymbol{\ell}}^S = \mathbf{l}^S | \widehat{\ell}^G = l^G) \to 1$.

We now consider case (i), and suppose that $\widehat{\ell}^S_i < l^S_i$. By Lemma A.4, we can show that $\frac{\widehat{\lambda}^S_{i,\widehat{\ell}^S_i+1}}{\widehat{\lambda}^S_{i,\widehat{\ell}^S_i}}\mathbb{I}(\widehat{\lambda}^S_{i,\widehat{\ell}^S_i} \geq \omega) + \mathbb{I}(\widehat{\lambda}^S_{i,\widehat{\ell}^S_i} < \omega) \asymp 1$.

By replacing $\widehat{\ell}_i^S$ of $\widehat{\boldsymbol{\ell}}^S$ with l_i^S, we find another $\widetilde{\boldsymbol{\ell}}^S = (\widehat{\ell}_1^S, \ldots, \widehat{\ell}_{i-1}^S, l_i^S, \widehat{\ell}_{i+1}^S, \ldots \widehat{\ell}_L^S)$, which yields a smaller value for the objective function considered in (2.9) with a probability approaching one. However, this is contradictory to the definition of $\widehat{\ell}^S$.

Next, we consider case (ii), and suppose that $\widehat{\ell}_i^S > l_i^S$. Again, Lemma A.4 yields that $\frac{\widehat{\lambda}^S_{i,\widehat{\ell}_i^S+1}}{\widehat{\lambda}^S_{i,\widehat{\ell}_i^S}}\mathbb{I}(\widehat{\lambda}^S_{i,\widehat{\ell}_i^S} \geq \omega) + \mathbb{I}(\widehat{\lambda}^S_{i,\widehat{\ell}_i^S} < \omega) = 1$.

By replacing $\widehat{\ell}_i^S$ of $\widehat{\boldsymbol{\ell}}^S$ with l_i^S, we find another $\widetilde{\boldsymbol{\ell}}^S = (\widehat{\ell}_1^S, \ldots, \widehat{\ell}_{i-1}^S, l_i^S, \widehat{\ell}_{i+1}^S, \ldots \widehat{\ell}_L^S)$, which yields a smaller value for the objective function considered in (2.9) with a probability approaching one. However, it is contradictory to the definition of $\widehat{\ell}^S$. Based on the above development, we conclude that $\Pr(\widehat{\boldsymbol{\ell}}^S = \mathbf{l}^S | \widehat{\ell}^G = l^G) \to 1$.

In view of Steps 1 and 2, the proof is now completed. ■

Proof of Lemma 3.1:

In this proof, we use the following steps to establish Lemma 3.1:

(a) a uniform rate of convergence is derived for $\widehat{\mathbf{b}}_i$: $\max_i \|\widehat{\mathbf{b}}_i - \boldsymbol{\beta}_i\| = O_P\left(\frac{\log(LN)^3}{N \wedge T}\right)$;

(b) the asymptotic distribution of $\widehat{\mathbf{b}}_{mg}$ is proved.

Step (a): We first extend the consistency results in Lemma 2.1 for the heterogeneous model. Using the arguments that are analogous to those in the proof of Lemma 2.1.(1), we can expand $\frac{1}{NT}Q_i(\mathbf{b}_i, \mathbf{C}_i) - \frac{1}{NT}Q_i(\boldsymbol{\beta}_i, \mathbf{F}_i)$ and obtain a preliminary rate of convergence for $\widehat{\mathbf{b}}_i$:

$$\widehat{\mathbf{b}}_i - \boldsymbol{\beta}_i = O_P\left(\frac{1}{\sqrt{N \wedge T}}\right). \tag{A.43}$$

We then proceed with the expansion of factor estimators. Write

$$\begin{aligned}
\widetilde{\mathbf{C}}_i \widetilde{\mathbf{V}}_i &= \frac{1}{NT}\sum_{j=1}^N \mathbf{X}_{ij.}(\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i)(\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i)^\top \mathbf{X}_{ij.}^\top \widetilde{\mathbf{C}}_i \\
&+ \frac{1}{NT}\sum_{j=1}^N \mathbf{X}_{ij.}(\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i)\boldsymbol{\gamma}_{ij}^\top \mathbf{F}_i^\top \widetilde{\mathbf{C}}_i + \frac{1}{NT}\sum_{j=1}^N \mathbf{F}_i \gamma_{ij}(\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i)^\top \mathbf{X}_{ij.}^\top \widetilde{\mathbf{C}}_i \\
&+ \frac{1}{NT}\sum_{j=1}^N \mathbf{X}_{ij.}(\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i)\mathbf{E}_{ij.}^\top \widetilde{\mathbf{C}}_i + \frac{1}{NT}\sum_{j=1}^N \mathbf{E}_{ij.}(\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i)^\top \mathbf{X}_{ij.}^\top \widetilde{\mathbf{C}}_i \\
&+ \frac{1}{NT}\sum_{j=1}^N \mathbf{F}_i \gamma_{ij}\gamma_{ij}^\top \mathbf{F}_i^\top \widetilde{\mathbf{C}}_i + \frac{1}{NT}\sum_{j=1}^N \mathbf{F}_i \gamma_{ij} \mathbf{E}_{ij.}^\top \widetilde{\mathbf{C}}_i \\
&+ \frac{1}{NT}\sum_{j=1}^N \mathbf{E}_{ij.}\boldsymbol{\gamma}_{ij}^\top \mathbf{F}_i^\top \widetilde{\mathbf{C}}_i + \frac{1}{NT}\sum_{j=1}^N \mathbf{E}_{ij.}\mathbf{E}_{ij.}^\top \widetilde{\mathbf{C}}_i \\
&:= \widetilde{\mathbf{J}}_{i,1} + \cdots + \widetilde{\mathbf{J}}_{i,9}.
\end{aligned} \tag{A.44}$$

By (A.43), Assumption 4, and using analogous arguments to those in (A.5), (A.6) and (A.7), we can readily show

$$\max_i \left\| \frac{1}{\sqrt{T}} \widetilde{\mathbf{J}}_{i,1} \right\| = O_P(\max_i \|\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i\|^2 \log(LN)^2 \log L),$$

$$\max_i \sum_{m=2}^{3} \left\| \frac{1}{\sqrt{T}} \widetilde{\mathbf{J}}_{i,m} \right\| = O_P(\max_i \|\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i\| \log(LN)^2 (\log L)^2),$$

$$\max_i \sum_{m=4}^{5} \left\| \frac{1}{\sqrt{T}} \widetilde{\mathbf{J}}_{i,m} \right\|_2 = O_P(\max_i \|\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i\| \log(LN) \log L),$$

$$\max_i \sum_{m=7}^{9} \left\| \frac{1}{\sqrt{T}} \widetilde{\mathbf{J}}_{i,m} \right\|_2 = O_P\left(\frac{\log(LN)(\log L)^2}{\sqrt{N \wedge T}} \right). \tag{A.45}$$

Therefore, we have the following uniform convergence rate:

$$\max_i \frac{1}{\sqrt{T}} \|\widetilde{\mathbf{C}}_i - \mathbf{F}_i \widetilde{\mathbf{H}}_i\|_2 = O_P\left(\left(\max_i \|\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i\| + \frac{1}{\sqrt{N \wedge T} \log(LN)} \right) \log(LN)^2 (\log L)^2 \right), \tag{A.46}$$

where $\widetilde{\mathbf{H}}_i = \frac{1}{NT} \boldsymbol{\Gamma}_{i\cdot}^\top \boldsymbol{\Gamma}_{i\cdot} \mathbf{F}_i^\top \widetilde{\mathbf{C}}_i \widetilde{\mathbf{V}}_i^{-1}$. It immediately yields

$$\max_i \|\mathbf{P}_{\widetilde{\mathbf{C}}_i} - \mathbf{P}_{\mathbf{F}_i}\|_2 = O_P\left(\left(\max_i \|\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i\| + \frac{1}{\sqrt{N \wedge T} \log(LN)} \right) \log(LN)^2 (\log L)^2 \right).$$

We then consider the following expansion for $\widehat{\mathbf{b}}_i - \boldsymbol{\beta}_i$:

$$\begin{aligned} \widehat{\mathbf{b}}_i - \boldsymbol{\beta}_i &= \left(\sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widetilde{\mathbf{C}}_i} \mathbf{X}_{ij\cdot} \right)^{-1} \sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widetilde{\mathbf{C}}_i} \mathbf{F}_i \boldsymbol{\gamma}_{ij} \\ &\quad + \left(\sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widetilde{\mathbf{C}}_i} \mathbf{X}_{ij\cdot} \right)^{-1} \sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widetilde{\mathbf{C}}_i} \mathbf{E}_{ij\cdot} = \boldsymbol{\pi}_{i,1} + \boldsymbol{\pi}_{i,2}. \end{aligned}$$

For the first term, we have

$$\begin{aligned} \frac{1}{NT} \sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widetilde{\mathbf{C}}_i} \mathbf{F}_i \boldsymbol{\gamma}_{ij} &= -\frac{1}{NT} \sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widetilde{\mathbf{C}}_i} (\widetilde{\mathbf{C}}_i \widetilde{\mathbf{H}}_i^{-1} - \mathbf{F}_i) \boldsymbol{\gamma}_{ij} \\ &= -\frac{1}{NT} \sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widetilde{\mathbf{C}}_i} \sum_{\ell=1,\ell\neq 6}^{9} \widetilde{\mathbf{J}}_{i,\ell} \widetilde{\boldsymbol{\Pi}}_i^{-1} \boldsymbol{\gamma}_{ij} := -(\widetilde{\mathbf{A}}_{i,1} + \cdots + \widetilde{\mathbf{A}}_{i,8}), \end{aligned} \tag{A.47}$$

where $\widetilde{\boldsymbol{\Pi}}_i^{-1} = (\frac{1}{T} \mathbf{F}_i^\top \widetilde{\mathbf{C}}_i)^{-1} (\frac{1}{N} \boldsymbol{\Gamma}_{i\cdot}^\top \boldsymbol{\Gamma}_{i\cdot})^{-1}$.

For $\widetilde{\mathbf{A}}_{i,1}$, $\widetilde{\mathbf{A}}_{i,3}$, $\widetilde{\mathbf{A}}_{i,4}$ and $\widetilde{\mathbf{A}}_{i,5}$, we can use arguments that are closely related to those in the proof of (A.11), (A.12) and (A.13) to show that these terms can uniformly have negligible orders ($o_P(\max_i \|\widehat{\mathbf{b}}_i - \boldsymbol{\beta}_i\|)$). For $\widetilde{\mathbf{A}}_{i,2}$,

$$\widetilde{\mathbf{A}}_{i,2} = \frac{1}{NT} \sum_{j_1, j_2 = 1}^{N} \mathbf{X}_{ij_1 \cdot}^{\top} \mathbf{M}_{\widetilde{\mathbf{C}}_i} \mathbf{X}_{ij_2 \cdot} \boldsymbol{\gamma}_{ij_2}^{\top} (\boldsymbol{\Gamma}_{i \cdot}^{\top} \boldsymbol{\Gamma}_{i \cdot})^{-1} \boldsymbol{\gamma}_{ij_1} (\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i). \tag{A.48}$$

We leave it for discussion later on.

For $\widetilde{\mathbf{A}}_{i,6}$, we write

$$\begin{aligned}
\widetilde{\mathbf{A}}_{i,6} &= \frac{1}{NT} \sum_{j_1=1}^{N} \mathbf{X}_{ij_1 \cdot}^{\top} \mathbf{M}_{\widetilde{\mathbf{C}}_i} \frac{1}{NT} \sum_{j_2=1}^{N} \mathbf{F}_i \gamma_{ij_2} \mathbf{E}_{ij_2 \cdot}^{\top} \widetilde{\mathbf{C}}_i \widetilde{\boldsymbol{\Pi}}_i^{-1} \gamma_{ij_1} \\
&= \frac{1}{NT} \sum_{j_1=1}^{N} \mathbf{X}_{ij_1 \cdot}^{\top} \mathbf{M}_{\widetilde{\mathbf{C}}_i} (\mathbf{F}_i - \widetilde{\mathbf{C}}_i \widetilde{\mathbf{H}}_i^{-1}) \frac{1}{NT} \sum_{j_2=1}^{N} \gamma_{ij_2} \mathbf{E}_{ij_2 \cdot}^{\top} \mathbf{F}_i \widetilde{\mathbf{H}}_i \widetilde{\boldsymbol{\Pi}}_i^{-1} \gamma_{ij_1} \\
&\quad + \frac{1}{NT} \sum_{j_1=1}^{N} \mathbf{X}_{ij_1 \cdot}^{\top} \mathbf{M}_{\widetilde{\mathbf{C}}_i} (\mathbf{F}_i - \widetilde{\mathbf{C}}_i \widetilde{\mathbf{H}}_i^{-1}) \frac{1}{NT} \sum_{j_2=1}^{N} \gamma_{ij_2} \mathbf{E}_{ij_2 \cdot}^{\top} (\widetilde{\mathbf{C}}_i - \mathbf{F}_i \widetilde{\mathbf{H}}_i) \widetilde{\boldsymbol{\Pi}}_i^{-1} \gamma_{ij_1} \\
&:= \widetilde{\mathbf{A}}_{i,6,1} + \widetilde{\mathbf{A}}_{i,6,2}.
\end{aligned} \tag{A.49}$$

By (A.46) and Assumption 4, we can show that $\max_i \|\widetilde{\mathbf{A}}_{i,6,2}\| = o_P(\max_i \|\widehat{\mathbf{b}}_i - \boldsymbol{\beta}_i\| + \frac{\log(LN)^3}{N \wedge T})$. For the first term $\widetilde{\mathbf{A}}_{i,6,1}$, simple algebra yields $\max_i \|NT^{-1} \sum_{j_2=1}^{N} \mathbf{E}_{ij_2 \cdot}^{\top} \mathbf{F}_i\| = O_P((NT)^{-1/2} \log(LN))$. Hence, we have

$$\max_i \|\widetilde{\mathbf{A}}_{i,6,1}\| = O_P\left(\max_i \|\mathbf{F}_i - \widetilde{\mathbf{C}}_i \widetilde{\mathbf{H}}_i^{-1}\|_2 \frac{\log(LN)^3 \log L}{\sqrt{NT}} \right), \tag{A.50}$$

which is also negligible.

Using analogous arguments, we can derive the probability order of $\max_i \|\widetilde{\mathbf{A}}_{i,7}\|$. Specifically, write

$$\begin{aligned}
\widetilde{\mathbf{A}}_{i,7} = &\frac{1}{NT} \sum_{j_1, j_2 = 1}^{N} \mathbf{X}_{ij_1 \cdot}^{\top} \mathbf{M}_{\mathbf{F}_i} \mathbf{E}_{ij_2 \cdot} \boldsymbol{\gamma}_{ij_2}^{\top} (\boldsymbol{\Gamma}_{i \cdot}^{\top} \boldsymbol{\Gamma}_{i \cdot})^{-1} \boldsymbol{\gamma}_{ij_1} \\
&+ \frac{1}{NT} \sum_{j_1, j_2 = 1}^{N} \mathbf{X}_{ij_1 \cdot}^{\top} (\mathbf{P}_{\mathbf{F}_i} - \mathbf{P}_{\widetilde{\mathbf{C}}_i}) \mathbf{E}_{ij_2 \cdot} \boldsymbol{\gamma}_{ij_2}^{\top} (\boldsymbol{\Gamma}_{i \cdot}^{\top} \boldsymbol{\Gamma}_{i \cdot})^{-1} \boldsymbol{\gamma}_{ij_1} := \widetilde{\mathbf{A}}_{i,7,1} + \widetilde{\mathbf{A}}_{i,7,2}.
\end{aligned}$$

For $\widetilde{\mathbf{A}}_{i,7,1}$, it is straightforward to see $\max_i \|(NT)^{-1} \sum_{j_2=1}^{N} \mathbf{F}_i^{\top} \mathbf{E}_{ij_2 \cdot}\| = O_P((NT)^{-1/2} \log L)$ and $\max_{i, j_1} \|(NT)^{-1} \sum_{j_2=1}^{N} \mathbf{X}_{ij_1}^{\top} \mathbf{E}_{ij_2 \cdot}\| = O_P((NT)^{-1/2} \log(LN))$. Hence, $\max_i \|\widetilde{\mathbf{A}}_{i,7,1}\| = O_P((NT)^{-1/2} \log(LN)^3)$. In addition, by expanding $\mathbf{P}_{\mathbf{F}_i} - \mathbf{P}_{\widetilde{\mathbf{C}}_i}$, we can show $\max_i \|\widetilde{\mathbf{A}}_{i,7,2}\| = o_P(\frac{\log(LN)^3}{N \wedge T})$. These results jointly yield

$$\max_i \|\widetilde{\mathbf{A}}_{i,7}\| = O_P\left(\frac{\log(LN)^3}{N \wedge T}) \right). \tag{A.51}$$

We then proceed with $\widetilde{\mathbf{A}}_{i,8}$. Write

$$\widetilde{\mathbf{A}}_{i,8} = \frac{1}{NT}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\boldsymbol{\cdot}}^\top\mathbf{M}_{\widetilde{\mathbf{C}}_i}\frac{1}{NT}\sum_{j_2=1}^{N}\boldsymbol{\Sigma}_{\varepsilon,ij_2}\widetilde{\mathbf{C}}_i\widetilde{\boldsymbol{\Pi}}_i^{-1}\boldsymbol{\gamma}_{ij_1}$$
$$+\frac{1}{NT}\sum_{j_1=1}^{N}\mathbf{X}_{ij_1\boldsymbol{\cdot}}^\top\mathbf{M}_{\widetilde{\mathbf{C}}_i}\frac{1}{NT}\sum_{j_2=1}^{N}(\mathbf{E}_{ij_2\boldsymbol{\cdot}}\mathbf{E}_{ij_2\boldsymbol{\cdot}}^\top-\boldsymbol{\Sigma}_{\varepsilon,ij_2})\widetilde{\mathbf{C}}_i\widetilde{\boldsymbol{\Pi}}_i^{-1}\boldsymbol{\gamma}_{ij_1}.$$

Analogously to (A.17), using (A.46) and Lemma A.5 of Bai (2009), we can readily obtain $\max_i\|\widetilde{\mathbf{A}}_{i,8}\| = O_P\left(\frac{\log(LN)^3}{N\wedge T}\right)$.

For $\boldsymbol{\pi}_{i,2}$, we define $\widetilde{\mathbf{A}}_{i,9} = \frac{1}{NT}\sum_{j=1}^{N}\mathbf{X}_{ij\boldsymbol{\cdot}}^\top\mathbf{M}_{\widetilde{\mathbf{C}}_i}\mathbf{E}_{ij\boldsymbol{\cdot}}$ for the simplicity of notation. Then, analogous arguments in the proof of (A.51) can be used here to show $\max_i\|\widetilde{\mathbf{A}}_{i,9}\| = O_P\left(\frac{\log(LN)}{N\wedge T}\right)$. In summary of the established results for $\widehat{\mathbf{b}}_i-\boldsymbol{\beta}_i$, we have

$$\max_i\|\widehat{\mathbf{b}}_i-\boldsymbol{\beta}_i\| = \max_i\left\|(\widetilde{\boldsymbol{\Sigma}}_{i,1}-\widetilde{\boldsymbol{\Sigma}}_{i,2})^{-1}\sum_{\ell=1,\neq 2}^{9}\widetilde{\mathbf{A}}_{i,\ell}\right\| = O_P\left(\frac{\log(LN)^3}{N\wedge T}\right), \quad \text{(A.52)}$$

where

$$\widetilde{\boldsymbol{\Sigma}}_{i,1} = \frac{1}{NT}\sum_{j=1}^{N}\mathbf{X}_{ij\boldsymbol{\cdot}}^\top\mathbf{M}_{\widetilde{\mathbf{C}}_i}\mathbf{X}_{ij\boldsymbol{\cdot}}, \quad \widetilde{\boldsymbol{\Sigma}}_{i,2} = \frac{1}{NT}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\mathbf{X}_{ij_1\boldsymbol{\cdot}}^\top\mathbf{M}_{\widetilde{\mathbf{C}}_i}\mathbf{X}_{ij_2\boldsymbol{\cdot}}\boldsymbol{\gamma}_{ij_2}^\top(\boldsymbol{\Gamma}_{i\boldsymbol{\cdot}}^\top\boldsymbol{\Gamma}_{i\boldsymbol{\cdot}})^{-1}\boldsymbol{\gamma}_{ij_1}.$$

Step (b): To study the asymptotic property of the mean group estimator, we start with the following weighted average of the individual estimators:

$$\frac{1}{L}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\widetilde{\boldsymbol{\Sigma}}_{i,1}(\widehat{\mathbf{b}}_i-\boldsymbol{\beta}_i) = \frac{1}{LNT}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{X}_{ij\boldsymbol{\cdot}}^\top\mathbf{M}_{\widetilde{\mathbf{C}}_i}\mathbf{F}_i\boldsymbol{\gamma}_{ij}$$
$$+\frac{1}{LNT}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{X}_{ij\boldsymbol{\cdot}}^\top\mathbf{M}_{\widetilde{\mathbf{C}}_i}\mathbf{E}_{ij\boldsymbol{\cdot}} := \boldsymbol{\Pi}_{mg,1}+\boldsymbol{\Pi}_{mg,2}, \quad \text{(A.53)}$$

where $\boldsymbol{\Omega}_{2,i}$ is defined in Assumption 4.

We first proceed with $\boldsymbol{\Pi}_{mg,1}$. Write

$$\boldsymbol{\Pi}_{mg,1} = -\frac{1}{LNT}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{X}_{ij\boldsymbol{\cdot}}^\top\mathbf{M}_{\widetilde{\mathbf{C}}_i}(\widetilde{\mathbf{C}}_i\widetilde{\mathbf{H}}_i^{-1}-\mathbf{F}_i)\boldsymbol{\gamma}_{ij}$$
$$= -\frac{1}{LNT}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{X}_{ij\boldsymbol{\cdot}}^\top\mathbf{M}_{\widetilde{\mathbf{C}}_i}\sum_{\ell=1,\ell\neq 6}^{9}\widetilde{\mathbf{J}}_{i,\ell}\widetilde{\boldsymbol{\Pi}}_i^{-1}\boldsymbol{\gamma}_{ij}$$
$$= -\frac{1}{L}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}(\widetilde{\mathbf{A}}_{i,1}+\cdots+\widetilde{\mathbf{A}}_{i,8}) := \boldsymbol{\Pi}_{mg,1,1}+\cdots+\boldsymbol{\Pi}_{mg,1,8},$$

where $\widetilde{\mathbf{A}}_{i,1},\ldots,\widetilde{\mathbf{A}}_{i,8}$ are defined in (A.47). We study these terms one by one.

By (A.45) and Assumption 4, it is easy to see

$$\frac{1}{L}\left\|\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\widetilde{\mathbf{A}}_{i,1}\right\| = O_P\left(\max_i \|\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i\|^2 \log(LN)^2(\log L)^2\right) = o_P\left(\frac{1}{\sqrt{LNT}}\right).$$

Hence, $\boldsymbol{\Pi}_{mg,1,1}$ is negligible. By (A.48), we have the following expression for $\boldsymbol{\Pi}_{mg,1,2}$:

$$\begin{aligned}\boldsymbol{\Pi}_{mg,1,2} &= -\frac{1}{L}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\widetilde{\mathbf{A}}_{i,2} \\ &= \frac{1}{L}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\widetilde{\boldsymbol{\Sigma}}_{i,2}(\widehat{\mathbf{b}}_i - \boldsymbol{\beta}_i),\end{aligned}$$

which will be utilized at a later stage.

For $\boldsymbol{\Pi}_{mg,1,3}$, we have

$$\begin{aligned}\left\|\frac{1}{L}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\widetilde{\mathbf{A}}_{i,3}\right\| &= \left\|\frac{1}{LN^2T^2}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\sum_{j_1,j_2=1}^{N}\mathbf{X}_{ij_1\boldsymbol{\cdot}}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}(\mathbf{F}_i - \widetilde{\mathbf{C}}_i\widetilde{\mathbf{H}}_i^{-1})\gamma_{ij_2}(\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_i)^{\top}\mathbf{X}_{ij_2\boldsymbol{\cdot}}^{\top}\widetilde{\mathbf{C}}_i\widetilde{\boldsymbol{\Pi}}_i^{-1}\gamma_{ij_1}\right\| \\ &= o_P\left(\frac{1}{\sqrt{LNT}}\right),\end{aligned}$$

which immediately yields $\|\boldsymbol{\Pi}_{mg,1,3}\| = o_P\left(\frac{1}{\sqrt{LNT}}\right)$.

Using the arguments that are closely related to (A.12), we can readily obtain the following results

$$\|\boldsymbol{\Pi}_{mg,1,4}\| = o_P\left(\frac{1}{\sqrt{LNT}}\right), \quad \|\boldsymbol{\Pi}_{mg,1,5}\| = o_P\left(\frac{1}{\sqrt{LNT}}\right).$$

For $\boldsymbol{\Pi}_{mg,1,6}$, write

$$\begin{aligned}\boldsymbol{\Pi}_{mg,1,6} &= -\frac{1}{L}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}(\widetilde{\mathbf{A}}_{i,6,1} + \widetilde{\mathbf{A}}_{i,6,2}) = -\frac{1}{L}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\widetilde{\mathbf{A}}_{i,6,1} + o_P\left(\frac{1}{\sqrt{LNT}}\right) \\ &= O_P\left(\max_i \|\mathbf{F}_i - \widetilde{\mathbf{C}}_i\widetilde{\mathbf{H}}_i^{-1}\|_2 \frac{\log(LN)^3\log L}{\sqrt{NT}}\right) + o_P\left(\frac{1}{\sqrt{LNT}}\right),\end{aligned}$$

where $\widetilde{\mathbf{A}}_{i,6,1}$ and $\widetilde{\mathbf{A}}_{i,6,2}$ are defined in (A.52) and the third equality holds by (A.50).

Together with (A.46) and (A.52), it implies that $\boldsymbol{\Pi}_{mg,1,6}$ is negligible. Using analogous arguments in the proof of (A.51), we can formulate the leading-order term of $\boldsymbol{\Pi}_{mg,1,7}$:

$$\boldsymbol{\Pi}_{mg,1,7} = -\frac{1}{LNT}\sum_{i=1}^{L}\boldsymbol{\Omega}_{2,i}^{-1}\sum_{j_1,j_2=1}^{N}\mathbf{X}_{ij_1\boldsymbol{\cdot}}^{\top}\mathbf{M}_{\mathbf{F}_i}\mathbf{E}_{ij_2\boldsymbol{\cdot}}\boldsymbol{\gamma}_{ij_2}^{\top}(\boldsymbol{\Gamma}_{i\boldsymbol{\cdot}}^{\top}\boldsymbol{\Gamma}_{i\boldsymbol{\cdot}})^{-1}\boldsymbol{\gamma}_{ij_1} + o_P\left(\frac{1}{\sqrt{LNT}}\right).$$

Analogously to (A.17), it follows (A.46) and (A.52) that

$$\mathbf{\Pi}_{mg,1,8} = -\frac{1}{LNT}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}\sum_{j_1,j_2=1}^{N}\mathbf{X}_{ij_1\cdot}^{\top}\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,ij_2}\mathbf{F}_i(\mathbf{F}_i^{\top}\mathbf{F}_i)^{-1}(\mathbf{\Gamma}_{i\cdot}^{\top}\mathbf{\Gamma}_{i\cdot})^{-1}\gamma_{ij_1} + o_P\left(\frac{1}{\sqrt{LNT}}\right).$$

Linking the results that we have established for $\mathbf{\Pi}_{mg,1,1}$, …, $\mathbf{\Pi}_{mg,1,8}$ with (A.53), we can readily obtain

$$\frac{1}{L}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}(\widetilde{\mathbf{\Sigma}}_{i,1}-\widetilde{\mathbf{\Sigma}}_{i,2})(\widehat{\mathbf{b}}_i-\boldsymbol{\beta}_i) = \frac{1}{LNT}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{Z}_{ij\cdot}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\mathbf{E}_{ij\cdot}$$
$$-\frac{1}{L}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}\widetilde{\mathbf{W}}_i + \text{negligible terms},$$

where $\widetilde{\mathbf{W}}_i = \frac{1}{NT}\sum_{j_1,j_2=1}^{N}\mathbf{X}_{ij_1\cdot}^{\top}\mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,ij_2}\mathbf{F}_i(\mathbf{F}_i^{\top}\mathbf{F}_i)^{-1}(\mathbf{\Gamma}_{i\cdot}^{\top}\mathbf{\Gamma}_{i\cdot})^{-1}\gamma_{ij_1}$.

In addition, by Assumption 4 and (A.52), we have

$$\begin{aligned}\widehat{\mathbf{b}}_{mg}-\overline{\boldsymbol{\beta}} &= \frac{1}{L}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}(\widetilde{\mathbf{\Sigma}}_{i,1}-\widetilde{\mathbf{\Sigma}}_{i,2})(\widehat{\mathbf{b}}_i-\boldsymbol{\beta}_i) + \frac{1}{L}\sum_{i=1}^{L}(\widehat{\mathbf{\Omega}}_{2,i}^{-1}-\mathbf{\Omega}_{2,i}^{-1})(\widetilde{\mathbf{\Sigma}}_{i,1}-\widetilde{\mathbf{\Sigma}}_{i,2})(\widehat{\mathbf{b}}_i-\boldsymbol{\beta}_i)\\ &= \frac{1}{LNT}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{Z}_{ij\cdot}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\mathbf{E}_{ij\cdot} - \frac{1}{L}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}\widetilde{\mathbf{W}}_i + \text{negligible terms},\end{aligned} \quad (A.54)$$

where $\widehat{\mathbf{\Omega}}_{2,i} = \frac{1}{NT}\sum_{j=1}^{N}\mathbf{Z}_{ij\cdot}^{\top}\mathbf{M}_{\mathbf{F}_i}\mathbf{Z}_{ij\cdot}$.

We now shift our fucus to $\frac{1}{LNT}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{Z}_{ij\cdot}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\mathbf{E}_{ij\cdot}$. As discussed in the proof of Lemma 2.2, we can expand $\mathbf{P}_{\widetilde{\mathbf{C}}_i}-\mathbf{P}_{\mathbf{F}_i}$ and $\widetilde{\mathbf{C}}_i-\mathbf{F}_i\widetilde{\mathbf{H}}_i$ sequentially to obtain that

$$\begin{aligned}\frac{1}{LNT}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{Z}_{ij\cdot}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\mathbf{E}_{ij\cdot} &= -\frac{1}{L}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}\sum_{j_1,j_2=1}^{N}\mathbf{Z}_{ij_1\cdot}^{\top}\mathbf{F}_i\left(\mathbf{F}_i^{\top}\mathbf{F}_i\right)^{-1}\left(\mathbf{\Gamma}_{i\cdot}^{\top}\mathbf{\Gamma}_{i\cdot}\right)^{-1}\gamma_{ij_2}\sigma_{i,j_1j_2}\\ &\quad + \frac{1}{LNT}\sum_{i=1}^{L}\mathbf{\Omega}_{2,i}^{-1}\sum_{j=1}^{N}\mathbf{Z}_{ij\cdot}^{\top}\mathbf{M}_{\mathbf{F}_i}\mathbf{E}_{ij\cdot} + o_P\left(\frac{1}{\sqrt{LNT}}\right).\end{aligned}$$

Together with (A.54), it leads to the desired result in Lemma 3.1. ■

Proof of Theorem 3.1:

To establish Theorem 3.1, it suffices to show

$$\widehat{\mathbf{a}}_{1,mg}-\mathbf{a}_{1,mg} = o_P\left(\sqrt{\frac{LT}{N}}\right) \text{ and } \widehat{\mathbf{a}}_{2,mg}-\mathbf{a}_{2,mg} = o_P\left(\sqrt{\frac{LN}{T}}\right). \quad (A.55)$$

We first study $\widehat{\mathbf{a}}_{1,mg}-\mathbf{a}_{1,mg}$. For simplicity of notation, define

$$\mathbf{a}_{1,i}^{*} = -\frac{1}{N}\sum_{j_1,j_2=1}^{N}\frac{\mathbf{Z}_{ij_1\cdot}^{\top}\mathbf{F}_i}{T}\left(\frac{\mathbf{F}_i^{\top}\mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\cdot}^{\top}\mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}\gamma_{ij_2}\sigma_{i,j_1j_2},$$

$$\widehat{\mathbf{a}}_{i,1}^* = -\frac{1}{N}\sum_{j_1,j_2=1}^{N} \frac{\widetilde{\mathbf{Z}}_{ij_1\cdot}^\top \widetilde{\mathbf{C}}_i}{T}\left(\frac{\widetilde{\mathbf{\Gamma}}_{i\cdot}^\top \widetilde{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1} \widetilde{\gamma}_{ij_2}\widetilde{\sigma}_{i,j_1j_2}.$$

With these notations, we write

$$\widehat{\mathbf{a}}_{1,mg} - \mathbf{a}_{1,mg} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{\Omega}_{2,i}^{-1}(\widehat{\mathbf{a}}_{1,i}^* - \mathbf{a}_{1,i}^*) + \frac{1}{L}\sum_{i=1}^{L}(\widetilde{\mathbf{\Omega}}_{2,i}^{-1} - \mathbf{\Omega}_{2,i}^{-1})\mathbf{a}_{1,i}^* + \frac{1}{L}\sum_{i=1}^{L}(\widetilde{\mathbf{\Omega}}_{2,i}^{-1} - \mathbf{\Omega}_{2,i}^{-1})(\widehat{\mathbf{a}}_{1,i}^* - \mathbf{a}_{1,i}^*), \tag{A.56}$$

where $\widetilde{\mathbf{\Omega}}_{2,i}$ is defined in (3.4). Using arguments that are closely related to those in the proof of (A.25), we obtain $\frac{1}{L}\sum_{i=1}^{L} \mathbf{\Omega}_{2,i}^{-1}(\widehat{\mathbf{a}}_{1,i}^* - \mathbf{a}_{1,i}^*) = o_P(\sqrt{\frac{N}{LT}})$. In addition, by expanding $\widetilde{\mathbf{\Omega}}_{2,i} - \mathbf{\Omega}_{2,i}$, we can show that $\max_i \|\widetilde{\mathbf{\Omega}}_{2,i} - \mathbf{\Omega}_{2,i}\| = o_P(\frac{1}{\log(LN)^3}\sqrt{\frac{N}{LT}})$ whose proofs are similar to those for (A.20) and therefore omitted here. In summary of these results, the first condition in (A.55) holds.

To study $\widehat{\mathbf{a}}_{2,mg} - \mathbf{a}_{2,mg}$, we define the following notation:

$$\mathbf{a}_{2,i}^* = -\frac{1}{NT}\sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\mathbf{F}_i}\mathbf{\Sigma}_{\varepsilon,i}\mathbf{F}_i\left(\frac{\mathbf{F}_i^\top \mathbf{F}_i}{T}\right)^{-1}\left(\frac{\mathbf{\Gamma}_{i\cdot}^\top \mathbf{\Gamma}_{i\cdot}}{N}\right)^{-1}\gamma_{ij},$$

$$\widehat{\mathbf{a}}_{i,2}^* = -\frac{1}{NT}\sum_{j=1}^{N} \mathbf{X}_{ij\cdot}^\top \mathbf{M}_{\widehat{\mathbf{C}}_i}\widetilde{\mathbf{\Sigma}}_{\varepsilon,i}\widetilde{\mathbf{C}}_i\left(\frac{\widetilde{\mathbf{\Gamma}}_{i\cdot}^\top \widetilde{\mathbf{\Gamma}}_{i\cdot}}{N}\right)^{-1}\widetilde{\gamma}_{ij}.$$

Then, we write

$$\widehat{\mathbf{a}}_{2,mg} - \mathbf{a}_{2,mg} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{\Omega}_{2,i}^{-1}(\widehat{\mathbf{a}}_{2,i}^* - \mathbf{a}_{2,i}^*) + \frac{1}{L}\sum_{i=1}^{L}(\widetilde{\mathbf{\Omega}}_{2,i}^{-1} - \mathbf{\Omega}_{2,i}^{-1})\mathbf{a}_{2,i}^* + \frac{1}{L}\sum_{i=1}^{L}(\widetilde{\mathbf{\Omega}}_{2,i}^{-1} - \mathbf{\Omega}_{2,i}^{-1})(\widehat{\mathbf{a}}_{2,i}^* - \mathbf{a}_{2,i}^*).$$

Analogously to (A.28), we can easily show that each term in $\widehat{\mathbf{a}}_{2,mg} - \mathbf{a}_{2,mg}$ is $o_P(\sqrt{\frac{T}{LN}})$ by expanding $\widehat{\mathbf{a}}_{2,i}^* - \mathbf{a}_{2,i}^*$. The proofs are omitted here to avoid unnecessary repetitions. Therefore, the second condition in (A.55) also holds, which completes the proof of Theorem 3.1. ■

Proof of Theorem 3.2:

Recall that we generate the bootstrap sample by $\mathbf{Y}_{ij\cdot}^* = \widetilde{\mathbf{Z}}_{ij\cdot}\widehat{\mathbf{b}}_{i,bc} + \widetilde{\mathbf{U}}_{ij\cdot} \circ \boldsymbol{\xi}$, where $\widetilde{\mathbf{U}}_{ij\cdot} = (\mathbf{Y}_{ij\cdot} - \mathbf{X}_{ij\cdot}\widehat{\mathbf{b}}_{i,bc} - \widetilde{\mathbf{C}}_i\widetilde{\gamma}_{ij})$. By the definition of the bootstrap mean group estimator, we can write

$$
\begin{aligned}
\widehat{\mathbf{b}}^{*}_{mg} - \widehat{\mathbf{b}}_{bcmg} &= \frac{1}{L}\sum_{i=1}^{L}\left(\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}\right)^{-1}\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\widetilde{\mathbf{U}}_{ij\boldsymbol{\cdot}}\circ\boldsymbol{\xi} \\
&= \frac{1}{L}\sum_{i=1}^{L}\left(\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}\right)^{-1}\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}[\mathbf{E}_{ij\boldsymbol{\cdot}}\circ\boldsymbol{\xi}] \\
&\quad+\frac{1}{L}\sum_{i=1}^{L}\left(\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}\right)^{-1}\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}[(\mathbf{X}_{ij\boldsymbol{\cdot}}(\boldsymbol{\beta}_i - \widehat{\mathbf{b}}_{i,bc}))\circ\boldsymbol{\xi}] \\
&\quad+\frac{1}{L}\sum_{i=1}^{L}\left(\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}\right)^{-1}\sum_{j=1}^{N}\widetilde{\mathbf{Z}}_{ij\boldsymbol{\cdot}}^{\top}\mathbf{M}_{\widetilde{\mathbf{C}}_i}[(\mathbf{F}_i\boldsymbol{\gamma}_{ij} - \widetilde{\mathbf{C}}_i\widetilde{\boldsymbol{\gamma}}_{ij})\circ\boldsymbol{\xi}],
\end{aligned}
$$

which gives a structure that is almost identical to $\widehat{\mathbf{b}}^{*} - \widehat{\mathbf{b}}_{bc}$, as derived in (A.29). Therefore, similar arguments can be used to complete the rest of the proof of Theorem 3.2. The details are omitted here. ■