Panel Data Models with Time-Varying Latent Group Structures*

Yiren Wang[a], Peter C.B. Phillips[b] and Liangjun Su[c]
[a]School of Economics, Singapore Management University, Singapore
[b]Yale University, University of Auckland,
University of Southampton, Singapore Management University
[c]School of Economics and Management, Tsinghua University, China

November 27, 2022

Abstract

This paper considers a linear panel model with interactive fixed effects such that individual heterogeneity is captured by latent group structure and time heterogeneity is captured by an unknown structural break. We allow the model to have different numbers of groups and/or different group memberships before and after the break. With the preliminary estimates by nuclear norm regularization followed by row- and column-wise linear regressions, we estimate the break point based on the idea of binary segmentation and the latent group structures together with the number of groups before and after the break by sequential testing K-means algorithm simultaneously. We show that the break point, the number of groups and the group membership can be estimated correctly with probability approaching one. Finite sample performance of the methodology is illustrated via Monte Carlo simulations and a real dataset application.

Key words: Interactive fixed effects, latent group structure, structural break, nuclear norm regularization, sequential testing K-means algorithm.

JEL Classification: C23, C33, C38, C51

*Phillips acknowledges support from a Lee Kong Chian Fellowship at SMU, the Kelly Fund at the University of Auckland, and the NSF under Grant No. SES 18-50860. Su gratefully acknowledges the support from the National Natural Science Foundation of China under Grant No. 72133002. Address correspondence to: Liangjun Su, School of Economics and Management, Tsinghua University, Beijing, 100084, China. Email: sulj@sem.tsinghua.edu.cn.

1 Introduction

Heterogeneous panel data regressions have seen increasing use in economic analyses because they can capture rich heterogeneity across both cross section and time. But models with complete heterogeneity along either the cross section or time dimension tend to possess too many parameters to be identified, which results in slow convergence or inefficient estimates. For this reason, more and more researchers advocate the use of panel data models with certain structures imposed along either the cross section or time dimension. On the one hand, the recent burgeoning on panels with latent group structures can be motivated from the observation that different groups of individuals may respond differently to an exogenous shock. For instance, Durlauf and Johnson (1995), Berthelemy and Varoudakis (1996), and Ben-David (1998) show economies in different groups of income per capita and/or education level may converge to different steady state equilibria. Chu (2012), Zhang and Cheng (2019) and Klapper and Love (2011) show an exogenous shock like policy implementation has different impact on different individuals. Long et al. (2012) discuss the influence of 2008 financial crisis on the economic growth is different for emerging and developed economies. On the other hand, the recent popularity of panels with structural changes can be motivated from the occurrence of financial crises, technological progress, and economic transitions, etc, during the time periods covered by the data. See Qian and Su (2016) for a survey on panel data models that consider the estimation and tests of structural changes.

Even though there exists a large literature on the study of individual or time heterogeneity alone in the slope coefficients of a panel data model, few of them consider both types of heterogeneity simultaneously. Exceptions include Keane and Neal (2020) and Lu and Su (2022) who consider linear panel data models with two-dimensional unobserved heterogeneity in the slope coefficients that are modelled via the usual additive structures, and Chernozhukov et al. (2020) and Wang et al. (2022) who model the slope coefficients via the use of low-rank matrices for conditional mean and quantile regressions, respectively. In addition, Okui and Wang (2021) and Lumsdaine et al. (2022) consider both individual heterogeneity and time heterogeneity by modeling them as a grouped pattern and structural breaks, respectively. Specifically, Okui and Wang (2021) develop a new panel data model with latent groups where the number of groups and the group memberships do not change over time but the coefficients within each group can change over time and they may have different breaking dates; Lumsdaine et al. (2022) consider the panels with a grouped pattern of heterogeneity when the latent group membership structure and/or the values of slope coefficients change at a break point. Both papers provide algorithms to recover the latent group structure based on linear panel models with or without individual fixed effects, but cannot allow for the presence of more complicated fixed effects such as the interactive fixed effects (IFEs) to capture the strong cross-sectional dependence in the data.

In this paper, we proposes a linear panel data model with IFEs such that the slope coefficients exhibit two-way heterogeneity. Following the lead of Okui and Wang (2021) and Lumsdaine et al. (2022) and to encourage the parameter parsimony, we use a latent group structure to capture the individual heterogeneity and an unknown structural break to capture the time heterogeneity. As for the latent group structure, we allow the model to have different group numbers and different group membership before and after the break. Given the complicated structure of the model, we propose to estimate the break point, the number of groups before and after the break, the group membership before and after the break, and the group-specific parameters in multiple steps. Our key insight is that for each of the p regressors, their slope coefficients, when allowed to vary across both cross section and time dimension, can be written as a factor structure with a fixed number of factors so that they can be stacked into a low-rank matrix.

In the first step, we explore the low-rank nature of the slope matrices and propose to obtain their initial estimates by the nuclear norm regularization (NNR), a popular machine learning technique in computer sciences. Such initial matrix estimates are consistent in terms of Frobenius norm but do not have the pointwise or uniform convergence for their elements. Despite this, by applying singular value decomposition (SVD) to these estimates, we can obtain estimates of the associated factors and factor loadings that are also consistent in terms of Frobenius norm. In the second step, we use the first-step initial estimates of the factors and factor loadings to run the row- and column-wise linear regressions to update the estimates of the factors and factor loadings which now possess pointwise and uniform consistency and can be used for subsequent analyses. In the third step, we estimate the break point by using the celebrated idea of binary segmentation as commonly used for break point estimation in the time series literature. Once the break point is estimated, the full-sample is naturally split into two sub-samples. In the fourth step, we follow the lead of Lin and Ng (2012) and Jin et al. (2022) to focus on each sub-sample before and after the estimated break point and propose a sequential testing K-means algorithm to recover the latent group structure and obtain the number of groups simultaneously. In the last step, we use the estimated group structure to estimate the group-specific parameters. Asymptotic analyses show that the break point, the number of groups and the group membership can be consistently estimated in Steps 3-4, so that the final step estimates for the group-specific coefficients can enjoy the oracle property. This means they have the same asymptotic distributions as the ones obtained by knowing the break point and the latent group structures before and after the break points.

This paper relates to two branches of literature. First, our paper contributes to the panel data literature on one-way heterogeneity, especially with either latent group structures or structural breaks. As for the latent group structures, there are several popular ways to recover the latent groups. The first approach is K-means algorithm. Lin and Ng (2012) apply the K-means algorithm to linear panel data models with grouped slope coefficients and propose an information criterion and a sequential testing approach to estimate the true number of groups. Sarafidis and Weber (2015) analyze the unknown grouped slopes in the large N and fixed T framework, and Zhang et al. (2019) provide an iterative algorithm based on K-means clustering for panel quantile regression model. Bonhomme and Manresa (2015) and Ando and Bai (2016) consider panels with grouped fixed effects. The second approach is the Classifier-Lasso (C-Lasso) that has become a popular clustering method since Su et al. (2016). This method is extended by Lu and Su (2017), Su and Ju (2018), Su et al. (2019), Wang et al. (2019), and Huang et al. (2020) to various contexts. In addition, clustering algorithm in regression via data-driven segmentation (CARDS) approach and binary segmentation are also considered in Ke et al. (2015), Wang et al. (2018), Ke et al. (2016) and Wang and Su (2021), among others. As for the panel data models with structural breaks, binary segmentation has become a common approach to estimate the break point. See Bai (2010), Lin and Hsu (2011), Kim (2011), Kim (2014) and Baltagi et al. (2017), among others. These papers focus on the case of one break point in the model. In contrast, Qian and Su (2016) and Li et al. (2016) allow for multiple breaks in linear panel data models with either the classical fixed effects or the IFEs, and propose the adaptive grouped fused lasso (AGFL) approach to estimate the break points. Compared to existing panel literature on one-way heterogeneity, we allow for two-way heterogeneity in our model. In particular, we allow not only different membership structures in different time blocks but also the change of number of groups over time. As a result, our model is more flexible than the vast existing models that allow for only latent group structures or structural breaks, but not both.

Second, this paper contributes to the recent burgeoning literature that models two-way heterogeneity in the slope coefficients of a panel data model. As mentioned above, there are two approaches to model the two-way heterogeneity in the slope coefficients. One is to model them as an additive structure so that both the individual and time effects enter the slope coefficients additively, as in Keane and Neal (2020) and Lu and Su (2022). The other is to impose certain low-rank structures on the slope coefficient matrices in which case one models each slope coefficient via the use of IFEs as used to model the strong cross sectional dependence in the panels. In view of the low-rank structures, we can resort to the NNR that has attracted increasing attention recently in panel data analyses. NNR has been used in recent researches in econometrics, see Bai and Ng (2019), Moon and Weidner (2018), Feng (2019), Chernozhukov et al. (2020), Belloni et al. (2019), Miao et al. (2022), and Hong et al. (2022), among others. But none of these papers impose any latent group structures in the slope coefficients. With latent group structures and structural breaks imposed, Okui and Wang (2021) allow the slope coefficients within each group to have common breaks and the break points to vary across different groups, and they propose to estimate the latent group structures, the structural breaks, and the group-specific regression parameters by the grouped adaptive group fused lasso (GAGFL). Note that neither the number of groups nor the group memberships is allowed to change over time in Okui and Wang (2021). In a companion paper, Lumsdaine et al. (2022) allow the latent group membership structure and/or the values of slope coefficients to change at a break point, and propose an estimation algorithm similar to the K-means of Bonhomme and Manresa (2015). Note that both Okui and Wang (2021) and Lumsdaine et al. (2022) allow for at most one-way heterogeneity (individual FEs) in the intercept and neither allows for IFEs to capture strong cross section dependence. In contrast, this paper proposes the algorithm to detect the unknown break point and to recover the group structure based on linear panel model with IFEs, which leads to a more general model. In addition, Lumsdaine et al. (2022) first assume the number of groups is known in the estimation algorithm and then estimate the number of groups via an information criterion but they do not establish the consistency result for such an estimate. Instead, we estimate the number of groups and group membership simultaneously by the sequential testing K-means algorithm and establish the consistency of the number of groups estimator.

The rest of the paper is organized as follows. We first introduce the linear panel model with time-varying latent group structures in Section 2 and provide the estimation algorithm in Section 3. The asymptotic properties are given in Section 4. In Section 5, we propose an alternative approach to detect the break point, provide the test statistics for the null that the slope coefficient has no structure change against the alternative with one break point, and discuss the estimation for the model with multiple breaks. In Sections 6 and 7, we show the finite sample performance of our method by Monte Carlo simulations and an empirical application, respectively. Section 8 concludes. All proofs are related to the online appendix.

Notation. $\|\cdot\|_1$, $\|\cdot\|_{op}$, $\|\cdot\|_\infty$, $\|\cdot\|_{\max}$ $\|\cdot\|_2$, $\|\cdot\|_F$, $\|\cdot\|_*$ denote the norm induced by 1-norms, the norm induced by 2-norms the norm induced by ∞-norms, the maximum norm, the Euclidean norm, the Frobenius norm and the nuclear norm. $\odot$ is the element-wise product. $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor and ceiling functions, respectively. $a \vee b$ and $a \wedge b$ return the maximum and the minimum of a and b. The symbol $\lesssim$ means "left side bounded by a positive constant times the right side", and $a_n \gg b_n$ means $b_n a_n^{-1} = o(1)$. Let $A = \{A_{it}\}$ as a matrix with its (i,t)-th entry denoted as A_{it}, and we denote $\{A_j\}_{j\in[p]\cup\{0\}}$ to be the collection of matrix A_j for all $j \in \{0, 1, \cdots, p\}$. For a specific $A \in \mathbb{R}^{m\times n}$, $P_A = A(A'A)^{-1}A'$ and $M_A = I_m - P_A$. When A is symmetric, $\lambda_{\max}(A)$, $\lambda_{\min}(A)$ and $\lambda_n(A)$ denote its largest, smallest and n-th largest eigenvalues, respectively. The operators $\rightsquigarrow$ and $\xrightarrow{p}$ denote convergence in distribution and in probability, respectively. We also use $[n]$ to denote the set $\{1, \cdots, n\}$ for any positive integer n. Besides, w.p.a.1 is "with probability approaching 1" and a.s. denotes "almost surely" for short.

2 Model Setup

In this paper, we consider the following linear panel model with IFEs:

$$Y_{it} = \Theta^0_{0,it} + X'_{it}\Theta^0_{it} + e_{it}, \tag{2.1}$$

where $i \in [N]$, $t \in [T]$, Y_{it} is the dependent variable, $X_{it} = (X_{1,it}, \cdots, X_{p,it})'$ is a $p \times 1$ vector of regressors, $\Theta^0_{it} = (\Theta^0_{1,it}, \cdots, \Theta^0_{p,it})'$ is a $p \times 1$ vector of slope coefficients, $\Theta^0_{0,it} = \lambda_i^{0\prime} f_t^0$ is an intercept term that exhibits a factor structure with r_0 factors, and e_{it} is the error term. Here, we assume r_0 is a fixed integer that does not change as $(N,T) \to \infty$. Let $\Lambda^0 = \left(\lambda^0_1, \cdots, \lambda^0_N\right)'$ and $F^0 = \left(f^0_1, \cdots, f^0_T\right)'$. Let $Y = \{Y_{it}\}$, $X_j = \{X_{j,it}\}$, $\Theta^0_j = \left\{\Theta^0_{j,it}\right\}$ and $E = \{e_{it}\}$, all of which are

$N \times T$ matrices. Then we can rewrite (2.1) in terms of matrices as follows:

$$Y = \Theta_0^0 + \sum_{j=1}^{p} X_j \odot \Theta_j^0 + E. \tag{2.2}$$

We assume that the slope coefficients follow time-varying latent group structures, i.e.,

$$\Theta_{it}^0 = \sum_{k \in [K_t]} \alpha_{kt} \mathbf{1}\{i \in G_{kt}\},$$

where $\{G_{kt}\}_{k \in K_t}$ forms a partition of $[N]$ for each specific time t with K_t being the number of groups at time t. Moreover, we assume that the group-specific slope coefficients α_{kt} or the memberships change at an unknown time point T_1, i.e.,

$$\alpha_{kt} = \begin{cases} \alpha_k^{(1)}, & \text{for} \quad t = 1, \ldots, T_1, \\ \alpha_k^{(2)}, & \text{for} \quad t = T_1 + 1, \ldots, T, \end{cases}$$

$$G_{kt} = \begin{cases} G_k^{(1)}, & \text{for} \quad t = 1, \ldots, T_1, \ k = 1, \ldots, K^{(1)}, \\ G_k^{(2)}, & \text{for} \quad t = T_1 + 1, \ldots, T, \ k = 1, \ldots, K^{(2)}, \end{cases}$$

with $K^{(1)}$ and $K^{(2)}$ being the number of latent groups before and after the break point, respectively. Let $g_i^{(1)}$ and $g_i^{(2)}$ respectively denote the individual group indices before and after the break:

$$g_i^{(1)} = \sum_{k \in K^{(1)}} k \mathbf{1}\left\{i \in G_k^{(1)}\right\} \quad \text{and} \quad g_i^{(2)} = \sum_{k \in K^{(2)}} k \mathbf{1}\left\{i \in G_k^{(2)}\right\}.$$

Let r_j be the rank of Θ_j^0 for $j \in [p] \cup \{0\}$. It is easy to see that Θ_j^0 exhibits a low-rank for all j. By the SVD, we have

$$\Theta_j^0 = \sqrt{NT} \mathcal{U}_j^0 \Sigma_j^0 \mathcal{V}_j^{0\prime} := U_j^0 V_j^{0\prime}, \quad j \in [p] \cup \{0\},$$

where $\mathcal{U}_j^0 \in \mathbb{R}^{N \times r_j}$, $\mathcal{V}_j^0 \in \mathbb{R}^{T \times r_j}$, $\Sigma_j^0 = \text{diag}(\sigma_{1,j}, \cdots, \sigma_{r_j,j})$, $U_j^0 = \sqrt{N} \mathcal{U}_j^0 \Sigma_j^0$ with each row being $u_{i,j}^{0\prime}$, and $V_j^0 = \sqrt{T} \mathcal{V}_j^0$ with each row being $v_{t,j}^{0\prime}$.

Note that we allow $\{\Theta_{it}^0\}_{i=1}^N$ to exhibit latent group structures before and after the break. For a particular $j \in [p]$, the $N \times T$ matrix Θ_j^0 may have no group structure before or after the break, or no break, or more or fewer groups after the break. Let $K_j^{(1)}$ and $K_j^{(2)}$ denote the number of groups before and after the break, respectively, for $\left\{\Theta_{j,it}^0\right\}_{i=1}^N$. Let $\mathcal{G}_j^{(\ell)} = \left\{G_{1,j}^{(\ell)}, \cdots, G_{K_j^{(\ell)},j}^{(\ell)}\right\}$, $\ell = 1, 2$, denote the associated latent group structures. Define $N_{k,j}^{(\ell)} = \left|G_{k,j}^{(\ell)}\right|$ and $\pi_{k,j}^{(\ell)} = \frac{N_{k,j}^{(\ell)}}{N}$ for $\ell = 1, 2$. Further define $\tau_T := \frac{T_1}{T}$. Below we show that Θ_j^0 has low-rank structure in all cases.

Case 1. Θ_j^0 exhibits neither structural break nor group structure.

In this case, $K_j^{(1)} = K_j^{(2)} = 1$, and $\Theta_{j,it}^0 = \alpha_j \ \forall (i, t) \in [N] \times [T]$. By the SVD, we have

$$\mathcal{U}_j = \frac{1}{\sqrt{N}} \iota_N \in \mathbb{R}^{N \times 1}, \quad \Sigma_j = \alpha_j, \quad \mathcal{V}_j = \frac{1}{\sqrt{T}} \iota_T \in \mathbb{R}^{T \times 1},$$

$$U_j = \alpha_j \iota_N \in \mathbb{R}^{N\times 1}, \quad V_j = \iota_N \in \mathbb{R}^{T\times 1},$$

where $\iota_d = (1, \cdots, 1)' \in \mathbb{R}^{d\times 1}$ for any natural number d. Obviously, $r_j = 1$ under Case 1.

Case 2. Θ_j^0 exhibits no structural break but a group structure.
In this case, $K_j^{(1)} = K_j^{(2)} = K_j$, $G_{k,j}^{(1)} = G_{k,j}^{(2)} = G_{k,j}$, $N_{k,j}^{(1)} = N_{k,j}^{(2)} = N_{k,j}$, $\pi_{k,j}^{(1)} = \pi_{k,j}^{(2)} = \pi_{k,j}$ $\forall k \in [K_j]$, and $\Theta_{j,it}^0 = \sum\limits_{k\in[K_j]} \alpha_{k,j} \mathbf{1}\{i \in G_{k,j}\}$ for $t \in [T]$. Therefore, we have

$$\mathcal{U}_{j,i} = \frac{\sum_{k\in[K_j]} \alpha_{k,j} \mathbf{1}\{i \in G_{k,j}\}}{\sqrt{\sum_{k\in[K_j]} N_{k,j} (\alpha_{k,j})^2}}, \quad \Sigma_j = \sqrt{\sum_{k\in[K_j]} \pi_{k,j} (\alpha_{k,j})^2}, \quad \mathcal{V}_j = \frac{1}{\sqrt{T}} \iota_T,$$
$$u_{i,j} = \sum_{k\in[K_j]} \alpha_{k,j} \mathbf{1}\{i \in G_{k,j}\}, \quad V_j = \iota_T,$$

where $\mathcal{U}_{j,i}$ is the i-th element in $\mathcal{U}_j$. Obviously, $r_j = 1$ under this case.

Case 3. Θ_j^0 exhibits both structural break and group structure.

- **(i)** $K_j^{(1)} \neq K_j^{(2)}$.
  Under this scenario, we have different number of groups before and after the break.
- **(ii)** $K_j^{(1)} = K_j^{(2)} = K_j$ and $G_{k,j}^{(1)} \neq G_{k,j}^{(2)}$.
  Under this scenario, we have the same number of groups before and after the break, but the group membership changes after the break point.
- **(iii)** $K_j^{(1)} = K_j^{(2)} = K_j$, $G_{k,j}^{(1)} = G_{k,j}^{(2)} = G_{k,j}$ for $\forall k \in [K_j]$, and $\alpha_{k,j}^{(1)} \neq \alpha_{k,j}^{(2)}$ for at least one $k \in [K_j]$.
  Under this scenario, even though neither the number of groups nor group membership changes after the break, there exists at least one group whose slope coefficients change.

For any positive integer d, we use $\mathbf{0}_d$ to denote a $d \times 1$ vector of zeros. The following lemma lays down the foundation for the detection of break point in our model.

Lemma 2.1 *For any $j \in [p]$ such that Θ_j^0 lies in Case 3 above, we have $rank(\Theta_j^0) \leq 2$. When $rank(\Theta_j^0) = 2$, we have*

(i) $\Theta_j^0 = \mathcal{U}_j \Sigma_j \mathcal{V}_j' = U_j V_j'$ *where* $U_j = \mathcal{U}_j \Sigma_j / \sqrt{T}$, $V_j = \sqrt{T}\mathcal{V}_j = D_j R_j$, $D_j = \begin{bmatrix} \frac{1}{\sqrt{\tau_T}} \iota_{T_1} & \mathbf{0}_{T_1} \\ \mathbf{0}_{T-T_1} & \frac{1}{\sqrt{1-\tau_T}} \iota_{T-T_1} \end{bmatrix}$ *and* $R_j' R_j = I_2$;

(ii) $\left\| \frac{v_{t,j}^0}{\left\| v_{t,j}^0 \right\|_2} - \frac{v_{t^*,j}^0}{\left\| v_{t^*,j}^0 \right\|_2} \right\|_2 = \sqrt{2}$ *for any $t \leq T_1$ and $t^* > T_1$.*

By Lemma 2.1 for Case 3 and the above analyses for Cases 1 and 2, we conclude that Θ_j^0 is a low-rank matrix with rank equal to or less than 2. In view of the low-rank structure of the slope matrices, we propose to adopt the NNR to obtain the preliminary estimates below. Moreover, under Case 3, Lemma 2.1(ii) indicates that singular vectors of the slope matrix with rank equals to 2 contain the structural break information.

3 Estimation

In this section we provide the estimation algorithm. We first assume that the ranks r_j for $j \in [p] \cup \{0\}$ are known, and then propose a singular value thresholding (SVT) procedure to estimate them. After we recover the break point and the latent group structures, we propose consistent estimates of the group-specific parameters.

3.1 Estimation Algorithm

Given r_j, $\forall j \in [p] \cup \{0\}$, we propose the following four-step procedure to estimate the break point and to recover the latent group structures before and after the break.

Step 1: **Nuclear Norm Regularization.** We run the nuclear norm regularized regression and obtain the preliminary estimates $\left\{\tilde{\Theta}_j\right\}_{j\in\{0\}\cup[p]}$, i.e.,

$$\{\tilde{\Theta}_j\}_{j\in[p]\cup\{0\}} = \underset{\{\Theta_j\}_{j=0}^{p}}{\arg\min} \frac{1}{NT} \left\| Y - \sum_{j=1}^{p} X_j \odot \Theta_j - \Theta_0 \right\|_F^2 + \sum_{j=0}^{p} \nu_j \left\|\Theta_j\right\|_* , \tag{3.1}$$

where ν_j is the tuning parameter for $j \in [p] \cup \{0\}$. For each j, conduct the SVD: $\frac{1}{\sqrt{NT}}\tilde{\Theta}_j = \hat{\tilde{\mathcal{U}}}_j \hat{\tilde{\Sigma}}_j \hat{\tilde{\mathcal{V}}}_j'$, where $\hat{\tilde{\Sigma}}_j$ is a diagonal matrix with diagonal elements being the descending singular values of $\tilde{\Theta}_j$. Moreover, let $\tilde{\mathcal{V}}_j$ consist the first r_j columns of $\hat{\tilde{\mathcal{V}}}_j$ and $\tilde{V}_j = \sqrt{T}\tilde{\mathcal{V}}_j$. Let $\tilde{v}_{t,j}'$ denote the t-th row of $\tilde{V}_j$ for $t \in [T]$.

Step 2: **Row- and Column-Wise Regressions.** First run the row-wise regressions of Y_{it} on $\left(\tilde{v}_{t,0}, \{\tilde{v}_{t,j}X_{j,it}\}_{j\in[p]}\right)$ to obtain $\{\dot{u}_{i,j}\}_{j\in[p]\cup\{0\}}$ for $i \in [N]$. Then run the column-wise regressions of Y_{it} on $\left(\dot{u}_{i,0}, \{\dot{u}_{i,j}X_{j,it}\}_{j\in[p]}\right)$ to obtain $\{\dot{v}_{t,j}\}_{j\in[p]\cup\{0\}}$ for $t \in [T]$. Let $\dot{\Theta}_{j,it} = \dot{u}_{i,j}'\dot{v}_{t,j}$ for $(i,t) \in [N] \times [T]$ and $j \in [p] \cup \{0\}$. Specifically, the row- and column-wise regressions are given by

$$\{\dot{u}_{i,j}\}_{j\in[p]\cup\{0\}} = \underset{\{u_{i,j}\}_{j\in[p]\cup\{0\}}}{\arg\min} \frac{1}{T} \sum_{t\in[T]} \left(Y_{it} - u_{i,0}'\tilde{v}_{t,0} - \sum_{j=1}^{p} u_{i,j}'\tilde{v}_{t,j}X_{j,it} \right)^2 \quad i \in [N], \tag{3.2}$$

$$\{\dot{v}_{t,j}\}_{j\in[p]\cup\{0\}} = \underset{\{v_{t,j}\}_{j\in[p]\cup\{0\}}}{\arg\min} \frac{1}{N} \sum_{i\in[N]} \left(Y_{it} - v_{t,0}'\dot{u}_{i,0} - \sum_{j=1}^{p} v_{t,j}'\dot{u}_{i,j}X_{j,it} \right)^2 \quad t \in [T]. \tag{3.3}$$

Step 3: **Break Point Estimation.** We estimate the break point as follows

$$\hat{T}_1 = \underset{s\in\{2,\cdots,T-1\}}{\arg\min} \frac{1}{pNT} \sum_{j\in[p]} \sum_{i\in[N]} \left\{ \sum_{t=1}^{s} \left(\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}_{j,i}^{(1s)} \right)^2 + \sum_{t=s+1}^{T} \left(\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}_{j,i}^{(2s)} \right)^2 \right\}, \quad (3.4)$$

where $\bar{\dot{\Theta}}_{j,i}^{(1s)} = \frac{1}{s}\sum_{t=1}^{s} \dot{\Theta}_{j,it}$ and $\bar{\dot{\Theta}}_{j,i}^{(2s)} = \frac{1}{T-s}\sum_{t=s+1}^{T} \dot{\Theta}_{j,it}$.

Step 4: **Sequential Testing K-means (STK).** In this step, we estimate the number of groups and the group membership before and after the break by using the STK algorithm. For each $j \in [p]$, define $\dot{\Theta}_{j,i}^{(1)} = (\dot{\Theta}_{j,i1}, \cdots, \dot{\Theta}_{j,i\hat{T}_1})'$, $\dot{\Theta}_{j,i}^{(2)} = (\dot{\Theta}_{j,i,\hat{T}_1+1}, \cdots, \dot{\Theta}_{j,iT})'$, $\dot{\beta}_i^{(1)} = \frac{1}{\sqrt{\hat{T}_1}}(\dot{\Theta}_{1,i}^{(1)\prime}, \cdots, \dot{\Theta}_{p,i}^{(1)\prime})'$, and $\dot{\beta}_i^{(2)} = \frac{1}{\sqrt{\hat{T}_2}}(\dot{\Theta}_{1,i}^{(2)\prime}, \cdots, \dot{\Theta}_{p,i}^{(2)\prime})'$. Let z_α be some predetermined number which will be specified in the next subsection. Given the sub-sample before and after the estimated break point, initialize $m = 1$ and classify each sub-sample into m groups by K-means algorithm with group membership obtained as $\hat{\mathcal{G}}_m^{(\ell)} := \{\hat{G}_{k,m}^{(\ell)}\}_{k\in[m]}$. Next, we construct test statistic $\hat{\Gamma}_m^{(\ell)}$, compare it to z_α, set $m = m + 1$ and go to the next iteration if $\hat{\Gamma}_m^{(\ell)} > z_\alpha$ and stop the STK algorithm otherwise. At last, define $\hat{K}^{(\ell)} = m$ and $\hat{\mathcal{G}}^{(\ell)} = \hat{\mathcal{G}}_m^{(\ell)}$. In the next subsection, we will present each step of the STK algorithm in detail.

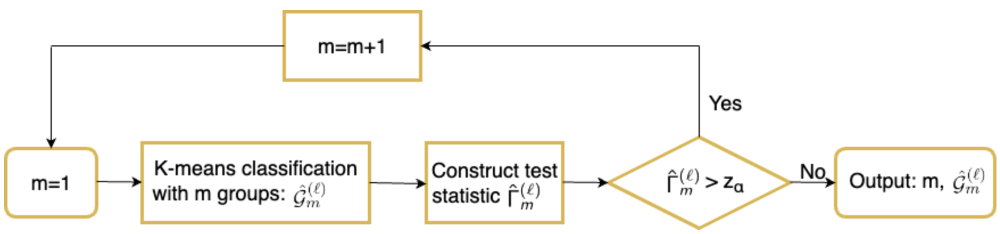

Figure 1: The flow chart of STK algorithm

Several remarks are in order. First, we assume the ranks of the intercept and slope matrices are known in Step 1 but will propose consistent estimates for them by the SVT below. Second, we obtain the preliminary estimates by nuclear norm regularization based on the low-rank structure of the intercept and slope matrices in the model. These estimates are consistent in terms of Frobenius norm but we cannot establish the pointwise or uniform convergence for their elements. Despite this, we can conduct SVD to obtain preliminary estimates of the factors and factor loadings to be used subsequently. Third, we conduct the row- and column-wise linear regressions to obtain updated estimates of the factors and factor loadings where we can establish their pointwise and uniform convergence rates. Fourth, with the consistent estimates obtained in the second step, we can estimate the break point in Step 3 consistently by following the idea of binary segmentation. Fifth, the STK algorithm in Step 4 will yield the estimated number of groups and group memberships at the same time.

In the latent group literature, it is standard and popular to assume the number of groups in the K-means algorithm is known and then estimate the number of groups by using certain information criteria. In this case, one needs to consider not only under- and just-fitting cases, but also over-fitting cases. It is well known that the major difficulty with this approach is to show that the over-fitting case occurs with probability approaching zero. As for the STK algorithm, it ensures us to focus on under-and just-fitting cases, which helps to avoid the theoretical difficulty caused by K-means classification with a larger than the true number of groups. Besides, although we adopt this sequential algorithm, the error from the previous iteration will not accumulate in the following iterations owing to fact that the classification in each iteration is new and not based on the K-means result in previous iterations.

3.2 The STK algorithm

In this subsection, we describe the K-means algorithm and the construction of test statistics $\hat{\Gamma}_m^{(\ell)}$ in the STK algorithm for $\ell \in \{1, 2\}$.

First, we define the objective function for the K-means algorithm with m clusters at each iteration. Let $a_{k,m}^{(\ell)}$ be a $p\hat{T}_1 \times 1$ and $p(T - \hat{T}_1) \times 1$ vector for $\ell = 1, 2$, respectively. We obtain the group membership with m groups by solving the following minimization problem:

$$\left\{ \dot{a}_{k,m}^{(\ell)} \right\}_{k\in[m]} = \underset{\left\{ a_{k,m}^{(\ell)} \right\}_{k\in[m]}}{\arg\min} \ \frac{1}{N} \sum_{i\in[N]} \min_{k\in[m]} \left\| \dot{\beta}_i^{(\ell)} - a_{k,m}^{(\ell)} \right\|_2^2, \tag{3.5}$$

which yields the membership estimates for each individual at the m-th iteration as

$$\hat{g}_{i,m}^{(\ell)} = \underset{k\in[m]}{\arg\min} \left\| \dot{\beta}_i^{(\ell)} - \dot{a}_{k,m}^{(\ell)} \right\|_2 \quad \forall i \in [N]. \tag{3.6}$$

Let $\hat{G}_{k,m}^{(\ell)} := \left\{ i \in [N] : \hat{g}_{i,m}^{(\ell)} = k \right\}$.

Second, we discuss the construction of the test statistic based on the idea of homogeneity test for several sub-samples. At iteration m, we have m potential subgroups $(\hat{G}_{1,m}^{(\ell)}, \cdots, \hat{G}_{m,m}^{(\ell)})$ after the K-means classification for $\ell = 1$ and 2. Let $\hat{\mathcal{T}}_1 = [\hat{T}_1]$, $\hat{\mathcal{T}}_2 = [T]\backslash[\hat{T}_1]$, $\hat{\mathcal{T}}_{1,-1} = \hat{\mathcal{T}}_1\backslash\{\hat{T}_1\}$, $\hat{\mathcal{T}}_{2,-1} = \hat{\mathcal{T}}_2\backslash\{T\}$, $\hat{\mathcal{T}}_{1,j} = \left\{1 + j, \cdots, \hat{T}_1\right\}$, and $\hat{\mathcal{T}}_{2,j} = \left\{\hat{T}_1 + 1 + j, \cdots, T\right\}$ for some specific $j \in \hat{\mathcal{T}}_{\ell,-1}$. Based on these estimated subgroups, we can obtain the estimates of the coefficients, factors and factor loadings for each subgroup in regime ℓ as follows:

$$\left(\left\{ \hat{\theta}_{i,k,m}^{(\ell)} \right\}_{i\in\hat{G}_{k,m}^{(\ell)}}, \hat{\Lambda}_{k,m}^{(\ell)}, \hat{F}_{k,m}^{(\ell)} \right) = \underset{\{\theta_i, \lambda_i, f_t\}_{i\in\hat{G}_{k,m}^{(\ell)}, t\in\hat{\mathcal{T}}_\ell}}{\arg\min} \sum_{i\in\hat{G}_{k,m}^{(\ell)}} \sum_{t\in\hat{\mathcal{T}}_\ell} \left(Y_{it} - X_{it}'\theta_i - \lambda_i' f_t \right)^2,$$

where $\hat{\Lambda}_{k,m}^{(\ell)} = \{\hat{\lambda}_{i,k,m}^{(\ell)}\}_{i\in\hat{G}_{k,m}^{(\ell)}}$ and $\hat{F}_{k,m}^{(\ell)} = \{\hat{f}_{t,k,m}^{(\ell)}\}_{t\in\hat{\mathcal{T}}_\ell}$. For all $i \in [N]$ and $t \in [T]$, define the residuals

$$\hat{e}_{it} = \sum_{\ell=1}^{2} \left(Y_{it} - \hat{f}_{t,k,m}^{(\ell)\prime} \hat{\lambda}_{i,k,m}^{(\ell)} - X_{it}'\hat{\theta}_{i,k,m}^{(\ell)} \right) \mathbf{1} \left\{ t \in \hat{\mathcal{T}}_\ell \right\}.$$

Let $\hat{X}_i^{(1)} = \left(X_{i1}, \cdots, X_{i\hat{T}_1}\right)'$, $\hat{X}_i^{(2)} = \left(X_{i,\hat{T}_1+1}, \cdots, X_{iT}\right)'$, and $\hat{T}_2 = T - \hat{T}_1$. Define

$$\hat{\bar{\theta}}_{k,m}^{(\ell)} = \frac{1}{\left|\hat{G}_{k,m}^{(\ell)}\right|} \sum_{i \in \hat{G}_{k,m}^{(\ell)}} \hat{\theta}_{i,k,m}^{(\ell)}, \quad M_{\hat{F}_{k,m}^{(\ell)}} = I_{\hat{T}_\ell} - \frac{1}{\hat{T}_\ell} \hat{F}_{k,m}^{(\ell)} \hat{F}_{k,m}^{(\ell)\prime},$$

$$\hat{S}_{ii,k,m}^{(\ell)} = \frac{1}{\hat{T}_\ell} \left(\hat{X}_i^{(\ell)}\right)' M_{\hat{F}_{k,m}^{(\ell)}} \hat{X}_i^{(\ell)}, \quad \hat{a}_{ii,k}^{(\ell)} = \hat{\lambda}_{i,k,m}^{(\ell)\prime} \left(\frac{\hat{\Lambda}_{k,m}^{(\ell)\prime} \hat{\Lambda}_{k,m}^{(\ell)}}{\left|\hat{G}_{k,m}^{(\ell)}\right|}\right)^{-1} \hat{\lambda}_{i,k,m}^{(\ell)}.$$

Let $\hat{z}_{it}^{(\ell)\prime}$ being the t-th row of $M_{\hat{F}_{k,m}^{(\ell)}} \hat{X}_i^{(\ell)}$. For each subgroup $\hat{G}_{k,m}^{(\ell)}$ with $k \in [m]$, we follow the lead of Pesaran and Yamagata (2008) and Ando and Bai (2015) and define $\hat{\Gamma}_{k,m}^{(\ell)}$ as follows:

$$\hat{\Gamma}_{k,m}^{(\ell)} = \sqrt{\left|\hat{G}_{k,m}^{(\ell)}\right|} \left(\frac{\frac{1}{\left|\hat{G}_{k,m}^{(\ell)}\right|} \sum_{i \in \hat{G}_{k,m}^{(\ell)}} \hat{\mathbb{S}}_{i,k,m}^{(\ell)} - p}{\sqrt{2p}}\right)$$

where

$$\hat{\mathbb{S}}_{i,k,m}^{(\ell)} = \hat{T}_\ell \left(\hat{\theta}_{i,k,m}^{(\ell)} - \hat{\bar{\theta}}_{k,m}^{(\ell)}\right)' \hat{S}_{ii,k,m}^{(\ell)} \left(\hat{\Omega}_{i,k,m}^{(\ell)}\right)^{-1} \hat{S}_{ii,k,m}^{(\ell)} \left(\hat{\theta}_{i,k}^{(\ell)} - \hat{\bar{\theta}}_{k}^{(\ell)}\right) \left(1 - \frac{\hat{a}_{ii,k}^{(\ell)}}{\left|\hat{G}_{k,m}^{(\ell)}\right|}\right)^2,$$

$$\hat{\Omega}_{i,k,m}^{(\ell)} = \frac{1}{\hat{T}_\ell} \sum_{t \in \hat{\mathcal{T}}_\ell} \hat{z}_{it}^{(\ell)} \hat{z}_{it}^{(\ell)\prime} \hat{e}_{it}^2 + \frac{1}{\hat{T}_\ell} \sum_{j \in \hat{\mathcal{T}}_{\ell,-1}} k(j/S_T) \sum_{t \in \hat{\mathcal{T}}_{\ell,j}} \left(\hat{z}_{it}^{(\ell)} \hat{z}_{i,t+j}^{(\ell)\prime} \hat{e}_{it} \hat{e}_{i,t+j} + \hat{z}_{i,t-j}^{(\ell)} \hat{z}_{it}^{(\ell)\prime} \hat{e}_{i,t-j} \hat{e}_{i,t}\right),$$

and $k(\cdot)$ is a kernel function with S_T being the bandwidth. Noted that the above expression for $\hat{\Omega}_{i,k,m}^{(\ell)}$ is the traditional HAC estimator. Let $\hat{\Gamma}_m^{(\ell)} = \max_{k \in [m]} \left(\hat{\Gamma}_{k,m}^{(\ell)}\right)^2$.

We will show below that $\hat{\Gamma}_m^{(\ell)}$ is asymptotically distributed as the maximum of m independent $\chi^2(1)$ random variables under the null hypothesis that the slope coefficients in each of the m sub-samples is homogeneous, while it diverges to infinity under the alternative. Let z_α denote be the critical value at significance level α, which is calculated from the maximum of m independent $\chi^2(1)$ random variables. We reject the null of m subgroups in favor of more groups at level α if $\hat{\Gamma}_m^{(\ell)} > z_\alpha$.

3.3 Rank Estimation

To obtain the rank estimator, we use the low-rank estimators from (3.1) and estimate r_j by the singular value thresholding (SVT):

$$\hat{r}_j = \sum_{i=1}^{N \wedge T} \mathbf{1}\left\{\sigma_i\left(\tilde{\Theta}_j\right) \geq 0.2\left(\nu_j \left\|\tilde{\Theta}_j\right\|_{op}\right)^{1/2}\right\} \quad \forall j \in \{0\} \cup [p],$$

where $\sigma_i(A)$ denotes the i-th largest singular value of A and $N \wedge T = \min(N, T)$. By arguments as used in the proof of Proposition D.1 in Chernozhukov et al. (2020) and that of Theorem 3.2 in Hong et al. (2022), we can show that $\mathbb{P}(\hat{r}_j = r_j) \to 1$ as $(N, T) \to \infty$.

3.4 Parameters Estimation

Once we obtain the estimated break point, the number of groups and the group membership before and after the estimated break point, we can estimate the group-specific slope coefficients $\left\{\alpha_k^{(\ell)}\right\}_{k\in[\hat{K}^{(\ell)}]}$ along with the factors and factor loadings as follows:

$$\left(\hat{\Lambda}^{(\ell)}, \hat{F}^{(\ell)}, \left\{\hat{\alpha}_k^{(\ell)}\right\}_{k\in[\hat{K}^{(\ell)}]}\right) = \arg\min \mathbb{L}\left(\Lambda, F, \left\{a_k^{(\ell)}\right\}_{k\in[\hat{K}^{(\ell)}]}\right) \tag{3.7}$$

where $\mathbb{L}\left(\Lambda, F, \left\{a_k^{(\ell)}\right\}_{k\in[\hat{K}^{(\ell)}]}\right) = \frac{1}{N\hat{T}_\ell}\sum_{k=1}^{\hat{K}^{(\ell)}}\sum_{i\in\hat{G}_k^{(\ell)}}\sum_{t\in\hat{\mathcal{T}}_\ell}\left(Y_{it} - \lambda_i' f_t - X_{it}' a_k^{(\ell)}\right)^2$. Here, we ignore the fact that the prior- and post-break regimes share the same set of factor loadings and estimate group-specific parameters separately for the two regimes at the cost of sacrificing some efficiency for the factor loading estimates. Alternatively, we can pool the observations before and after the break to estimate the parameters as follows:

$$\left(\hat{\Lambda}, \hat{F}, \left\{\hat{\alpha}_k^{(1)}\right\}_{k\in[\hat{K}^{(1)}]}, \left\{\hat{\alpha}_k^{(2)}\right\}_{k\in[\hat{K}^{(2)}]}\right) = \arg\min \mathbb{L}\left(\Lambda, F, \left\{a_k^{(1)}\right\}_{k\in[\hat{K}^{(1)}]}, \left\{a_k^{(2)}\right\}_{k\in[\hat{K}^{(2)}]}\right)$$

where

$$\mathbb{L}\left(\Lambda, F, \left\{a_k^{(1)}\right\}_{k\in[\hat{K}^{(1)}]}, \left\{a_k^{(2)}\right\}_{k\in[\hat{K}^{(2)}]}\right) = \mathbb{L}\left(\Lambda, F, \left\{a_k^{(1)}\right\}_{k\in[\hat{K}^{(1)}]}\right) + \mathbb{L}\left(\Lambda, F, \left\{a_k^{(2)}\right\}_{k\in[\hat{K}^{(2)}]}\right). \tag{3.8}$$

In either case, as one can imagine, due to the presence of group structures, the establishment of the asymptotic properties of the post-classification estimators of the group-specific slope coefficients becomes much more involved than that in Bai (2009) and Moon and Weidner (2017). For this reason, we will focus on the estimates defined in (3.7).

4 Asymptotic Theory

In this section, we study the asymptotic properties of the estimators introduced in the last section.

4.1 Basic Assumptions

Define $e_i = (e_{i1}, \cdots, e_{iT})'$ and $X_{j,i} = (X_{j,i1}, \cdots, X_{j,iT})'$. Let V_j^0 be a $T\times r_j$ matrix with its t-th row being $v_{t,j}^{0\prime}$, and U_j^0 be the $N\times r_j$ matrix with its i-th row being $u_{i,j}^{0\prime}$. Throughout the paper, we treat the factors $\left\{V_j^0\right\}_{j\in[p]\cup\{0\}}$ as random and their loadings $\left\{U_j^0\right\}_{j\in[p]\cup\{0\}}$ as deterministic. Let $\mathscr{D} := \sigma\left(\left\{V_j^0\right\}_{j\in[p]\cup\{0\}}\right)$ denote the minimum σ-field generated by $\left\{V_j^0\right\}_{j\in[p]\cup\{0\}}$. Similarly, let $\mathscr{G}_t := \sigma\left(\mathscr{D}, \{X_{is}\}_{i\in[N], s\le t+1}, \{e_{is}\}_{i\in[N], s\le t}\right)$. Let M and C be generic bounded positive constants which may vary across lines.

Assumption 1 *(i) $\{e_{it}, X_{it}\}_{t\in[T]}$ are conditionally independent across i given $\mathscr{D}$.*

(ii) $\mathbb{E}(e_{it}|X_{it},\mathscr{D})=0$.

(iii) *For each* i, $\{(e_{it},X_{it}),t\geq 1\}$ *is strong mixing conditional on* $\mathscr{D}$ *with the mixing coefficient* $\alpha_i(\cdot)$ *satisfying* $\max_{i\in[N]}\alpha_i(z)\leq M\alpha^z$ *for some constant* $\alpha\in(0,1)$.

(iv) *There exists a constant* $C>0$ *such that*

$$\max_{i\in[N],j\in[p]}\frac{1}{T}\sum_{t\in[T]}\xi_{it}^2\leq C\ a.s.\quad and\quad \max_{t\in[T],j\in[p]}\frac{1}{N}\sum_{i\in[N]}\xi_{it}^2\leq C\ a.s.$$

for $\xi_{it}=e_{it}$, $X_{j,it}$ *and* $X_{j,it}e_{it}\ \forall j\in[p]$.

(v) $\max_{i\in[N],t\in[T]}\mathbb{E}\left[|\xi_{it}|^q\,|\mathscr{D}\right]\leq M$ *a.s. for some* $q>8$ *where* $\xi_{it}=e_{it}$, $X_{j,it}$ *and* $X_{j,it}e_{it}\ \forall j\in[p]$.

(vi) *As* $(N,T)\to\infty$, $\sqrt{N}(\log N)^2T^{-1}\to 0$ *and* $T(\log N)^2N^{-3/2}\to 0$.

Assumption 1* *(i), (iv) and (v) are same as Assumption 1(i), (iv) and (v). Besides,*

(ii) $\mathbb{E}(e_{it}|\mathscr{G}_{t-1})=0\ \forall(i,t)\in[N]\times[T]$, *and* $\max_{i\in[N],t\in[T]}\mathbb{E}\left(e_{it}^2|\mathscr{G}_{t-1}\right)\leq M$ *a.s..*

(iii) $\{e_{it}\}_{i\in[N]}$ *is conditionally independent across* t *given* $\mathscr{D}$.

Assumption 1(i) imposes conditional independence on $\{e_{it},X_{it}\}_{t\in[T]}$ across the cross sectional units. Assumption 1(ii) imposes the moment condition. Assumption 1(iii) impose conditional strong mixing conditions along the time dimension. See Prakasa Rao (2009) for the definition of conditional strong mixing and Su and Chen (2013) for an application in the panel setup. Assumption 1(iv)-(v) imposes some conditions which restricts the tail behavior of ξ_{it}. Note that we don't restrict either the regressors or error terms to be bounded. Assumption 1(vi) imposes some restrictions on N and T but does not restrict N and T diverge to infinity at the the same rate. It is possible to allow N to diverge to infinity faster but not too faster than T, and vice versa.

Assumption 1* is for the study of dynamic panel data models. To be specific, Assumption 1*(ii) requires that the error sequence $\{e_{it},t\geq 1\}$ is a martingale difference sequence (m.d.s.) with respect to the filter $\mathscr{G}_t$, which allows for lagged dependent variables in X_{it}. Assumption 1*(iii) imposes the conditional independence of error term across t. In the panel for the least-squares based PCA estimation, there will be the endogeneity issue if we allow for both dynamics and serially correlated errors.

Assumption 2 *rank*$\left(\Theta_j^0\right)=r_j\leq\bar{r}$ *for* $j\in[p]\cup\{0\}$ *and some fixed* $\bar{r}$, *and* $\max_{j\in[p]\cup\{0\}}\left\|\Theta_j^0\right\|_{\max}\leq M$.

Assumption 2 imposes the low-rank conditions on the coefficient matrices, which facilitates the use of NNR to obtain the preliminary estimates in the first step. As discussed in the previous section, we see that the low-rank assumption for the slope matrices is satisfied for the model in Section 2. Moreover, we follow Ma et al. (2020) and assume the elements of the coefficient matrices are uniformly bounded to simplify the proofs. The boundedness of the slope coefficients is reasonable given that their cardinality does not grow with the sample size. The boundedness assumption for the intercept coefficient can be relaxed at the cost of more lengthy arguments.

Assumption 3 *Let $\sigma_{l,j}$ denote the l-th largest singular values of Θ_j^0 for $j \in [p] \cup \{0\}$. There exist some constants C_σ and c_σ such that*

$$\infty > C_\sigma \geq \limsup_{N,T} \max_{j \in [p]} \sigma_{1,j} \geq \liminf_{N,T} \min_{j \in [p]} \sigma_{r_j,j} \geq c_\sigma > 0.$$

Assumption 3 imposes some conditions on the singular values of the coefficient matrices. It implies that we only allow pervasive factors when these matrices are written as a factor structure. This condition can be easily verified given the latent group structures of the slope coefficients.

Let $\Theta_j^0 = R_j \Sigma_j S_j'$ be the SVD for Θ_j^0, $\forall j \in [p] \cup \{0\}$. Further decompose $R_j = (R_{j,r}, R_{j,0})$, $S_j = (S_{j,r}, S_{j,0})$ with $(R_{j,r}, S_{j,r})$ being the singular vectors corresponding to nonzero singular values and $(R_{j,0}, S_{j,0})$ being the singular vectors corresponding to zero singular values. Hence, for any matrix $W \in \mathbb{R}^{N \times T}$, we define

$$\mathcal{P}_j^{\perp}(W) = R_{j,0} R_{j,0}' W S_{j,0} S_{j,0}', \quad \mathcal{P}_j(W) = W - \mathcal{P}_j^{\perp}(W),$$

where $\mathcal{P}_j(W)$ can be seen as the linear projection of matrix W into the low-rank space with $\mathcal{P}_j^{\perp}(W)$ being its orthogonal space.

Let $\Delta_{\Theta_j} = \Theta_j - \Theta_j^0$ for any Θ_j. Based on the spaces constructed above, with some positive constants C_1 and C_2, we define the restricted set for full-sample parameters as follows:

$$\begin{aligned}\mathcal{R}(C_1, C_2) := \Bigg\{ (\{\Delta_{\Theta_j}\}_{j \in [p] \cup \{0\}}) : & \sum_{j \in [p] \cup \{0\}} \left\| \mathcal{P}_j^{\perp}(\Delta_{\Theta_j}) \right\|_* \leq C_1 \sum_{j \in [p] \cup \{0\}} \left\| \mathcal{P}_j(\Delta_{\Theta_j}) \right\|_*, \\ & \sum_{j \in [p] \cup \{0\}} \|\Theta_j\|_F^2 \geq C_2 \sqrt{NT} \Bigg\}. \end{aligned} \tag{4.1}$$

Lemma B.4 in the Appendix shows that our nuclear norm estimators are in a restricted set larger than (4.1), which eliminates the restriction for the Frobenius norm in the $\mathcal{R}(C_1, C_2)$. The restrictive set means the projection to the orthogonal low-rank space of the estimator error can not be larger than its projection to the low-rank space. Theorem 4.1 below will greatly rely on this property.

Assumption 4 *For any $C_2 > 0$, there are constants C_3 and C_4 such that uniformly for $(\{\Delta_{\Theta_j}\}_{j \in [p] \cup \{0\}}) \in \mathcal{R}(3, C_2)$, we have*

$$\left\| \Delta_{\Theta_0} + \sum_{j=1}^{p} \Delta_{\Theta_j} \odot X_j \right\|_F^2 \geq C_3 \sum_{j \in [p] \cup \{0\}} \left\| \Delta_{\Theta_j} \right\|_F^2 - C_4(N+T) \quad \textit{w.p.a.1.}$$

Assumption 4 imposes the restricted strong convexity condition, which is similar to Assumption 3.1 in Chernozhukov et al. (2020). The latter authors also provide some sufficient conditions to verify such an assumption.

Let $r = \sum_{j\in[p]\cup\{0\}} r_j$. Define the following $r \times r$ matrices:

$$\Phi_i = \frac{1}{T}\sum_{t=1}^{T} \phi_{it}^0 \phi_{it}^{0\prime} \;\; \forall i \in [N] \text{ and } \Psi_t = \frac{1}{N}\sum_{i\in[N]} \psi_{it}^0 \psi_{it}^{0\prime} \;\; \forall t \in [T],$$

where $\phi_{it}^0 = (v_{t,0}^{0\prime}, v_{t,1}^{0\prime} X_{1,it}, \cdots, v_{t,p}^{0\prime} X_{p,it})'$, $\psi_{it}^0 = (u_{i,0}^{0\prime}, u_{i,1}^{0\prime} X_{1,it}, \cdots, u_{i,p}^{0\prime} X_{p,it})'$.

Assumption 5 *There exist constants C_ϕ and c_ϕ such that*

$$\infty > C_\phi \geq \limsup_{T} \max_{t\in[T]} \lambda_{\max}(\Psi_t) \geq \liminf_{T} \min_{t\in[T]} \lambda_{\min}(\Psi_t) \geq c_\phi > 0,$$
$$\infty > C_\phi \geq \limsup_{N} \max_{i\in[N]} \lambda_{\max}(\Phi_i) \geq \liminf_{N} \min_{i\in[N]} \lambda_{\min}(\Phi_i) \geq c_\phi > 0.$$

Assumption 5 is similar to Assumption 8 in Ma et al. (2020).

4.2 Asymptotic Properties of the NNR Estimators and Singular Vector Estimators

Let $\eta_{N,1} = \frac{\sqrt{\log T}}{\sqrt{N\wedge T}}$ and $\eta_{N,2} = \frac{\sqrt{\log N\vee T}}{\sqrt{N\wedge T}}(NT)^{1/q}$. Let $\tilde{\sigma}_{k,j}$ denotes the k-th largest singular value of $\tilde{\Theta}_j$ for $j \in [p]\cup\{0\}$. Our first main result is about the consistency of the first-stage NNR Estimators and the second-stage singular vector estimators.

Theorem 4.1 *Suppose that Assumptions 1-4 hold. Then $\forall j \in [p]\cup\{0\}$, we have*

(i) $\frac{1}{\sqrt{NT}}\left\|\tilde{\Theta}_j - \Theta_j^0\right\|_F = O_p(\eta_{N,1})$, $\max_{k\in[r_j]} |\tilde{\sigma}_{k,j} - \sigma_{k,j}| = O_p(\eta_{N,1})$, *and* $\left\|V_j^0 - \tilde{V}_j O_j\right\|_F = O_p(\sqrt{T}\eta_{N,1})$ *where O_j is an orthogonal matrix defined in the proof.*

If in addition Assumption 5 is also satisfied, then we have

(ii) $\max_{i\in[N]} \left\|\dot{u}_{i,j} - O_j u_{i,j}^0\right\|_2 = O_p(\eta_{N,2})$, $\max_{t\in[T]} \left\|\dot{v}_{t,j} - O_j v_{t,j}^0\right\|_2 = O_p(\eta_{N,2})$,

(iii) $\max_{i\in[N],t\in[T]} \left|\dot{\Theta}_{j,it} - \Theta_{j,it}^0\right| = O_p(\eta_{N,2})$.

Remark 1. Theorem 4.1(i) reports the error bounds for $\tilde{\Theta}_j$, $\tilde{\sigma}_{k,j}$, and $\tilde{V}_j$. The $\log T$ term in the numerator of $\eta_{N,1}$ is due to the use of some exponential inequality for strong mixing process. Theorem 4.1(ii) and (iii) reports the uniform convergence rate of the factor and factor loading estimators. The extra $(NT)^{1/q}$ term in the $\eta_{N,2}$ is by the nonboundedness of $X_{j,it}$ in Assumption 1(v), and it disappears when $X_{j,it}$ is assumed to be uniformly bounded.

4.3 Consistency of the Break Point Estimate

Recall that $g_i^{(1)}$ and $g_i^{(2)}$ denote the true group individual i belongs to before and after the break, respectively. To estimate the break point consistently, we add the following condition.

Assumption 6 *(i)* $\sqrt{\frac{1}{N}\sum_{i\in[N]}\left\|\alpha_{g_i^{(1)}}-\alpha_{g_i^{(2)})}\right\|_2^2}=C_5\zeta_{NT}$, *where* C_5 *is a positive constant and* $\zeta_{NT}\gg\eta_{N,2}$.

(ii) $\tau_T:=\frac{T_1}{T}\to\tau\in(0,1)$ *as* $T\to\infty$.

Remark 2. Assumption 6(i) imposes conditions on the break size in order to identify the break point. Note that we allow the average break size to shrink to zero at the rate slower than $\sqrt{\frac{\log(N\vee T}{N\wedge T}}(NT)^{1/q}$. This rate is of much bigger magnitude than the optimal $(NT)^{-1/2}$-rate that can be detected in the panel threshold regressions (PTRs) for several reasons. First, in PTRs, the slope coefficients are usually assumed to be homogeneous so that each individual is subject to the same change in the slope coefficients and one can use the cross-sectional information effectively. In contrast, we allow for heterogeneous slope coefficients here and the change can occur only for a subset of cross section units but not all. In addition, in the presence of latent group structure, we not only allow the slope coefficients of some specific groups to change with group membership fixed, but also allow the slope coefficient to remain the same for some groups while the group memberships change after the break. Second, our break point estimation relies on the binary segmentation idea borrowed from the time series literature where one can allow break sizes of bigger magnitude than $T^{-1/2}$ in order to identify the break ratio consistently but not the break point consistently. As we can see, even though we require bigger break sizes, we can estimate the break date consistently by using information from both the cross section and time dimensions. Third, as mentioned above, the additional term $\log(N\vee T)$ in the above rate is mainly due to the use of some exponential inequality and the term $(NT)^{1/q}$ is due to the fact that we only assume the existence of q-th order moments for some random variables.

The following theorem indicates that we can estimate the break date T_1 consistently.

Theorem 4.2 *Suppose Assumptions 1-6 hold, with the true break point being* T_1 *and the estimator defined in (3.4). Then* $\mathbb{P}\left(\hat{T}_1=T_1\right)\to 1$ *as* $(N,T)\to\infty$.

Theorem 4.2 shows that we can estimate the true break date consistently w.p.a.1 despite the fact that we allow the break size to shrink to zero at certain rate.

4.4 Consistency of the Estimates of the Number of Groups and the Latent Group Structures

To study the asymptotic properties of the estimates of the number of groups and the recovery of the latent group structures, we first add the following definition.

Definition 4.3 *Fix $K^{(\ell)} > 1$ and $m \leq K^{(\ell)}$. The estimated group structure $\hat{\mathcal{G}}_m^{(\ell)}$ satisfies the non-splitting property (NSP) if for any pair of individuals in the same true group, the estimated group labels are the same.*

Definition 4.3 describes the non-splitting property introduced by Jin et al. (2022). The latter authors show that the STK algorithm yields estimated group structure enjoying the NSP.

To proceed, we add following assumptions.

Assumption 7 *(i) Let k_s and k_{s^*} be different group indices. Assume that*

$$\min_{1 \leq k_s < k_{s^*} \leq K^{(\ell)}} \left\| \alpha_{k_s}^{(\ell)} - \alpha_{k_{s^*}}^{(\ell)} \right\|_2 \geq C_5, \ \ell \in \{1, 2\}.$$

(ii) Let $N_k^{(\ell)}$ be the number of individuals in group k, $\forall k \in [K^{(\ell)}]$. Define $\pi_k^{(\ell)} = \frac{N_k^{(\ell)}}{N}$ for $\ell = 1, 2$. Assume $K^{(\ell)}$ is fixed and

$$\infty > \bar{C} \geq \limsup_N \sup_{k \in [K^{(\ell)}]} \pi_k^{(\ell)} \geq \liminf_N \inf_{k \in [K^{(\ell)}]} \pi_k^{(\ell)} \geq \underline{c} > 0, \ \ell = 1, 2.$$

(iii) For any combination of the collection of n true groups with $n \geq 2$, we have

$$\frac{T_\ell}{\sqrt{N}} \sum_{s=1}^{n} N_{k_s}^{(\ell)} \left\| \sum_{s^* \neq s} \left(\alpha_{k_{s^*}}^{(\ell)} - \alpha_{k_s}^{(\ell)} \right) \right\|_2^2 \to \infty, \ \ell = 1, 2.$$

Remark 3. Assumption 7(i) and 7(ii) are the standard assumption for the K-means algorithm, which are comparable to Assumption 4 in Su et al. (2020). Assumption 7(i) assumes that the minimum distance of two distinct groups is bounded away from 0. This greatly facilitates the subsequent analyses. For Assumption 7(iii), it can be shown to hold under mild conditions. Below we explain this assumption in detail. When $n = 2$, it's clear that

$$\begin{aligned} \frac{T_\ell}{\sqrt{N}} \sum_{s=1}^{n} N_{k_s}^{(\ell)} \left\| \sum_{s^* \neq s} \left(\alpha_{k_{s^*}}^{(\ell)} - \alpha_{k_s}^{(\ell)} \right) \right\|_2^2 &= \frac{T_\ell}{\sqrt{N}} \left(N_{k_1}^{(\ell)} \left\| \alpha_{k_2}^{(\ell)} - \alpha_{k_1}^{(\ell)} \right\|_2^2 + N_{k_2}^{(\ell)} \left\| \alpha_{k_1}^{(\ell)} - \alpha_{k_2}^{(\ell)} \right\|_2^2 \right) \\ &\geq \frac{C_5^2 T_\ell (N_{k_1}^{(\ell)} + N_{k_2}^{(\ell)})}{\sqrt{N}} = O(T\sqrt{N}) \end{aligned}$$

by combining Assumptions 6(ii), 7(i), and 7(ii). When $n > 2$, for a special case such that the term $\left\| \sum_{s^* \neq s} \left(\alpha_{k_{s^*}}^{(\ell)} - \alpha_{k_s}^{(\ell)} \right) \right\|_2$ equals 0 for a specific $s = s_0 \in [n]$, it's clear that it will be non-zero for $\forall s \in [n] \backslash \{s_0\}$. Hence, if we assume $\left\| \sum_{s^* \neq s} \left(\alpha_{k_{s^*}}^{(\ell)} - \alpha_{k_s}^{(\ell)} \right) \right\|_2$ is lower bounded by a constant for $\forall s \in [n] \backslash \{s_0\}$, Assumption 7(iii) will holds naturally. Except this special case, similar arguments follow for other cases.

Assumption 8 *Let* $\mathcal{T}_1 = [T_1]$, $\mathcal{T}_2 = [T] \backslash [T_1]$.

(i) $\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell} f_t^0 f_t^{0\prime} \xrightarrow{p} \Sigma_F^{(\ell)} > 0$, *as* $T \to \infty$. $\frac{1}{N_k^{(\ell)}}\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)} \xrightarrow{p} \Sigma_{\Lambda,k}^{(\ell)} > 0$ *as* $N \to \infty$, *where* $\Lambda_k^{0,(\ell)}$ *is a stack of* λ_i^0 *for all individuals in group* k *and* $k \in [K^{(\ell)}]$.

(ii) There exists a constant $C > 0$ *such that* $\max_{i\in[N],j\in[p]} \frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell} \xi_{it}^2 \leq C$ *a.s. for* $\xi_{it} = e_{it}$ *and* $X_{j,it}$, $\forall j \in [p]$.

Assumption 8(i) imposes some standard assumptions on the factors and factor loadings. Assumption 8(ii) is similar as Assumption 1(iv), which strengthen the Assumption 1(iv) to hold for two time regimes.

The next theorem reports the asymptotic properties of the STK estimators.

Theorem 4.4 *Fix* $\alpha = \alpha_N \in (0,1)$. *Suppose that Assumption 1* and Assumptions 2-8 hold. Then for* $\ell \in \{1,2\}$, *we have*

(i) if $m = K^{(\ell)}$,

(a) $\max_{i\in[N]} \mathbf{1}\left\{\hat{g}_{i,K^{(\ell)}}^{(\ell)} \neq g_i^{(\ell)}\right\} = 0$ *w.p.a.1,*

(b) $\hat{\Gamma}_{K^{(\ell)}}^{(\ell)}$ *is asymptotically distributed as the maximum of* $K^{(\ell)}$ *independent* $\chi^2(1)$ *random variables,*

(c) $\mathbb{P}\left(\hat{K}^{(\ell)} \leq K^{(\ell)}\right) \geq 1 - \alpha + o(1)$,

(ii) if $m < K^{(\ell)}$, $\hat{\Gamma}_m^{(\ell)} \to \infty$ *w.p.a.1. Thus* $\mathbb{P}\left(\hat{K}^{(\ell)} \neq K^{(\ell)}\right) \leq \alpha + o(1)$.

Remark 4. Theorem 4.4 studies the asymptotic properties of the STK algorithm. At iteration m such that $m < K^{(\ell)}$, w.p.a.1, the test statistics $\hat{\Gamma}_m^{(\ell)}$ diverges to infinity, which means the iteration will continue at $(m+1)$-th iteration. At iteration m such that $m = K^{(\ell)}$, the test statistics $\hat{\Gamma}_m^{(\ell)}$ is stochastically bounded and the iteration stops w.p.a.1 provided $\alpha = \alpha_N \to 0$. It follows that $\mathbb{P}\left(\hat{K}^{(\ell)} = K^{(\ell)}\right) \to 1$ as along as we set $\alpha_N \to 0$ as $N \to \infty$. As aforementioned, Theorem 4.4 ensures the application of K-means algorithm only for the under-fitting and just-fitting cases and it helps us to avoid the theoretical challenge in handling the over-fitting case in the classification.

Remark 5. To allow the dynamic panels, we focus on Assumption 1*, where the error term is martingale difference sequence. Under this assumption, the HAC estimator $\hat{\Omega}_{i,k,m}^{(\ell)}$ degenerate to $\frac{1}{\hat{T}_\ell}\sum_{t\in\hat{\mathcal{T}}_\ell} \hat{z}_{it}^{(\ell)}\hat{z}_{it}^{(\ell)\prime}\hat{e}_{it}^2$. If we allow the serial correlated error in the non-dynamic panels, same results in Theorem 4.4 hold. We skip the proof for the non-dynamic panels with serial correlated error for brevity. For the kernel function and bandwidth, we can follow Andrews (1991) and let $k(\cdot)$ belongs to the following class of kernels:

$$\mathcal{K} = \left\{k(\cdot) : \mathbb{R} \mapsto [-1,1] \mid k(0) = 1,\ k(u) = k(-u),\ \int |k(u)|\, du < \infty,\right.$$

$$k(\cdot) \text{ is continuous at 0 and at all but a finite number of other points}\Big\}.$$

See, e.g., Andrews (1991) and White (2014) for detail.

4.5 Distribution Theory for the Group-specific Slope Estimators

For $\ell \in \{1,2\}$, let $\left\{\hat{\alpha}_k^{*(\ell)}\right\}_{k\in K^{(\ell)}}$ be the oracle estimators of the group-specific slope coefficients before and after the break point by using the true break and membership information for all individuals. To study the asymptotic distribution theory for $\left\{\hat{\alpha}_k^{(\ell)}\right\}_{k\in K^{(\ell)}}$, $\ell \in \{1,2\}$, we only need to show that for the oracle estimators $\left\{\hat{\alpha}_k^{*(\ell)}\right\}_{k\in K^{(\ell)}}$ based on Theorems 4.2 and 4.4 by extending the result of Bai (2009) and Moon and Weidner (2017).

To proceed, we define some notation. For $\ell \in \{1,2\}$, we first define the matrix notation for each subgroup. For $j \in [p]$, let $X_{j,i}^{(1)} = (X_{j,i1},\cdots,X_{j,iT_1})'$, $X_{j,i}^{(2)} = \left(X_{j,i(T_1+1)},\cdots,X_{j,iT}\right)'$, $e_i^{(1)} = (e_{i1},\cdots,e_{iT_1})'$ and $e_i^{(2)} = \left(e_{i(T_1+1)},\cdots,e_{iT}\right)'$. Then we use $\mathbb{X}_{j,k}^{(\ell)} \in \mathbb{R}^{N_k^{(\ell)}\times T_\ell}$ and $E_k^{(\ell)} \in \mathbb{R}^{N_k^{(\ell)}\times T_\ell}$ to denote the regressor and error matrix for subgroup $k \in [K^{(\ell)}]$ with each row being $X_{j,i}^{(\ell)}$ and $e_i^{(\ell)}$, respectively. Let $\mathcal{X}_{j,k}^{(\ell)} = M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}M_{F^{0,(\ell)}} \in \mathbb{R}^{N_k^{(\ell)}\times T_\ell}$ with entries $\mathcal{X}_{j,k,it}^{(\ell)}$, which follows by $\mathcal{X}_{k,it}^{(\ell)} = \left(\mathcal{X}_{1,k,it}^{(\ell)},\cdots,\mathcal{X}_{p,k,it}^{(\ell)}\right)'$. Further define

$$\begin{aligned}
\mathbb{B}_{NT,1,j,k}^{(\ell)} &= \frac{1}{N_k^{(\ell)}}tr\left[P_{F^{0,(\ell)}}\mathbb{E}\left(E_k^{(\ell)\prime}\mathbb{X}_{j,k}^{(\ell)}|\mathscr{D}\right)\right],\\
\mathbb{B}_{NT,2,j,k}^{(\ell)} &= \frac{1}{T_\ell}tr\left[\mathbb{E}\left(E_k^{(\ell)}E_k^{(\ell)\prime}|\mathscr{D}\right)M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}F^{0,(\ell)}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\Lambda_k^{0,(\ell)\prime}\right],\\
\mathbb{B}_{NT,3,j,k}^{(\ell)} &= \frac{1}{N_k^{(\ell)}}tr\left[\mathbb{E}\left(E_k^{(\ell)}E_k^{(\ell)\prime}|\mathscr{D}\right)M_{F^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right],\\
\mathbb{B}_{NT,m,k}^{(\ell)} &= \left(\mathbb{B}_{NT,m,1,k}^{(\ell)},\cdots,\mathbb{B}_{NT,m,p,k}^{(\ell)}\right)', \quad \forall m \in \{1,2,3\},\\
\Omega_k^{(\ell)} &= \frac{1}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathcal{T}_\ell}e_{it}^2\mathcal{X}_{k,it}^{(\ell)}\mathcal{X}_{k,it}^{(\ell)\prime}.
\end{aligned}$$

Let $\mathbb{W}_{NT,k}^{(\ell)}$ be a $p\times p$ matrix with (j_1,j_2)-th entry being $\frac{1}{N_k^{(\ell)}T_\ell}tr\left(M_{F^{0,(\ell)}}\mathbb{X}_{j_1,k}^{(\ell)\prime}M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j_2,k}^{(\ell)}\right)$. Then we define the overall bias term for each subgroup as $\mathbb{B}_{NT,k}^{(\ell)} = -\rho_k^{(\ell)}\mathbb{B}_{NT,1,k}^{(\ell)} - \left(\rho_k^{(\ell)}\right)^{-1}\mathbb{B}_{NT,2,k}^{(\ell)} - \rho_k^{(\ell)}\mathbb{B}_{NT,3,k}^{(\ell)}$ with $\rho_k^{(\ell)} = \sqrt{\frac{N_k^{(\ell)}}{T_\ell}}$. To state the main result in this subsection, we add the following assumption.

Assumption 9 *(i) As $(N,T)\to\infty$, $T(\log T)N^{-4/3}\to 0$.*

(ii) $\text{plim}_{(N,T)\to\infty}\left[\frac{1}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathcal{T}_\ell}X_{it}X_{it}'\right] > 0,\ \ell\in\{1,2\}, k\in K^{(\ell)}.$

(iii) *For $\ell \in \{1,2\}$ and $k \in [K^{(\ell)}]$, separate the p regressors of each subgroups into p_1 "low-rank regressors" $\mathbb{X}_{j,k}^{(\ell)}$ such that rank$(\mathbb{X}_{j,k}^{(\ell)}) = 1,\ \forall j \in \{1,\cdots,p_1\}$, and "high-rank regressors" $\mathbb{X}_{j,k}^{(\ell)}$ such that rank$(\mathbb{X}_{j,k}^{(\ell)}) > 1,\ \forall j \in \{p_1+1,\cdots,p\}$. Let $p_2 := p-p_1$. These two types of regressors satisfy:*

(iii.a) *Consider the linear combinations $b \cdot \mathbb{X}_{high,k}^{(\ell)} := \sum_{j=p_1+1}^{p} b_j \mathbb{X}_{j,k}^{(\ell)}$ for high-rank regressors with p_2-vectors b such that $\|b\|_2 = 1$ and $b = (b_{p_1+1},\cdots,b_p)'$. With a positive constant C_b, we assume that*

$$\min_{\{\|b\|_2=1\}} \sum_{n=2r_0+p_1+1}^{N} \lambda_n \left[\frac{\left(b \cdot \mathbb{X}_{high,k}^{(\ell)}\right)\left(b \cdot \mathbb{X}_{high,k}^{(\ell)}\right)'}{NT_\ell} \right] \geq C_b \quad w.p.a.1.$$

(iii.b) *For $j \in [p_1]$, write $\mathbb{X}_{j,k}^{(\ell)} = w_{j,k}^{(\ell)} v_{j,k}^{(\ell)\prime}$ with $N_k^{(\ell)}$-vectors $w_j^{(\ell)}$ and T_ℓ-vectors $v_j^{(\ell)}$. Let $w_k^{(\ell)} = \left(w_{1,k}^{(\ell)},\cdots,w_{p_1,k}^{(\ell)}\right) \in \mathbb{R}^{N\times p_1}$, $v^{(\ell)} = \left(v_1^{(\ell)},\cdots,v_{p_1}^{(\ell)}\right) \in \mathbb{R}^{T_\ell \times p_1}$, $M_{w_k^{(\ell)}} = I_{N_k^{(\ell)}} - w_k^{(\ell)}\left(w_k^{(\ell)\prime} w_k^{(\ell)}\right)^{-1} w_k^{(\ell)\prime}$ and $M_{v^{(\ell)}} = I_{T_\ell} - v^{(\ell)}\left(v^{(\ell)\prime} v^{(\ell)}\right)^{-1} v^{(\ell)\prime}$. For a positive constant C_B, we assume that $\left(N_k^{(\ell)}\right)^{-1} \Lambda_k^{0,(\ell)\prime} M_{w_k^{(\ell)}} \Lambda_k^{0,(\ell)} > C_B I_{r_0}$ and $T_\ell^{-1} F^{0,(\ell)\prime} M_{w^{(\ell)}} F^{0,(\ell)} > C_B I_{r_0}$ w.p.a.1.*

(iv) *For $\forall j \in [p]$, $\ell \in \{1,2\}$, $k \in K^{(\ell)}$,*

$$\frac{1}{N_k^{(\ell)}\left(T_\ell\right)^2} \sum_{i\in G_k^{(\ell)}} \sum_{t_1\in\mathcal{T}_\ell} \sum_{t_2\in\mathcal{T}_\ell} \sum_{s_1\in\mathcal{T}_\ell} \sum_{s_2\in\mathcal{T}_\ell} \left| Cov\left(e_{it_1}\tilde{X}_{j,it_2}, e_{is_1}\tilde{X}_{j,is_2}\right)\right| = O_p(1),$$

where $\tilde{X}_{j,it} = X_{j,it} - \mathbb{E}\left(X_{j,it}|\mathscr{D}\right)$.

Assumption 9 imposes some conditions to derive the asymptotic distribution theory for the panel model with IFEs which allows for dynamics. Assumption 9(i) strengthens Assumption 1(vi) a bit. Assumption 9(ii) is the standard non-collinearity condition for regressors, which is analogous to Assumption 4(i) in Moon and Weidner (2017). Assumption 9(iii) is the identification assumption which is comparable to Assumption 4 in Moon and Weidner (2017). With strong mixing condition shown in Assumption 1(iii), we can verify Assumption 9(iv).

The following theorem establishes the asymptotic distribution of $\left\{\hat{\alpha}_k^{(\ell)}\right\}_{k\in K^{(\ell)}}$.

Theorem 4.5 *Suppose that Assumption 1 or 1* and Assumptions 2-9 hold. For $\ell \in \{1,2\}$, the estimators $\left\{\hat{\alpha}_k^{(\ell)}\right\}_{k\in K^{(\ell)}}$ are asymptotically equivalent to the oracle estimators $\left\{\hat{\alpha}_k^{*(\ell)}\right\}_{k\in K^{(\ell)}}$, and we have*

$$\mathbb{W}_{NT}^{(\ell)} \mathbb{D}_{NT}^{(\ell)} \begin{pmatrix} \hat{\alpha}_1^{(\ell)} - \alpha_1^{(\ell)} \\ \vdots \\ \hat{\alpha}_{K^{(\ell)}}^{(\ell)} - \alpha_{K^{(\ell)}}^{(\ell)} \end{pmatrix} - \mathbb{B}_{NT}^{(\ell)} \rightsquigarrow \mathcal{N}\left(0, \Omega^{(\ell)}\right),$$

such that $\mathbb{D}_{NT}^{(\ell)} = diag\left(\sqrt{N_1^{(\ell)} T_\ell}, \cdots, \sqrt{N_{K^{(\ell)}}^{(\ell)} T_\ell}\right)$, $\mathbb{W}_{NT}^{(\ell)} = diag\left(\mathbb{W}_{NT,1}^{(\ell)}, \cdots, \mathbb{W}_{NT,K^{(\ell)}}^{(\ell)}\right)$, $\mathbb{B}_{NT}^{(\ell)} = diag\left(\mathbb{B}_{NT,1}^{(\ell)}, \cdots, \mathbb{B}_{NT,K^{(\ell)}}^{(\ell)}\right)$ *and* $\Omega^{(\ell)} = diag\left(\Omega_1^{(\ell)}, \cdots, \Omega_{K^{(\ell)}}^{(\ell)}\right)$.

Remark 6. Theorem 4.5 establishes the asymptotic distribution for the estimators of the group-specific slope coefficients before and after the break. It shows that parameter estimates from our algorithm enjoy the oracle property given the results in Theorems 4.2 and 4.4. In the appendix, we sketch the proof by following Moon and Weidner (2017) and Lu and Su (2016).

5 Alternatives and Extensions

In this section we first consider an alternative method to estimate the break point and then discuss several possible extensions.

5.1 Alternative for Break Point Detection

The algorithm proposed in Section 3 uses low-rank estimates of Θ_j^0 to find the break point estimates. However, by Lemma 2.1(ii), we observe that the right singular vector matrix of Θ_j^0, i.e., V_j^0, contains the structural break information when $r_j = 2$. For this reason, we can propose an alternative way to estimate the break point under the case that the maximum rank of slope matrix in the model being 2. Let $\dot{v}_{t,j}^* := \frac{\dot{v}_{t,j}}{\|\dot{v}_{t,j}\|_2}$ and $\dot{v}_t^* := \left(\dot{v}_{t,1}^{*\prime}, \cdots, \dot{v}_{t,p}^{*\prime}\right)'$, with the true values being $v_{t,j}^* := \frac{O_j v_{t,j}^0}{\left\|O_j v_{t,j}^0\right\|_2}$ and $v_t^* := \left(v_{t,1}^{*\prime}, \cdots, v_{t,p}^{*\prime}\right)'$, respectively.

Step 3*: **Break Point Estimation by Singular Vectors.** We estimate the break point as follows:

$$\tilde{T}_1 = \underset{s \in \{2, \cdots, T-1\}}{\arg\min} \frac{1}{T} \left\{ \sum_{t=1}^{s} \left\| \dot{v}_t^* - \bar{\dot{v}}^{*(1)s} \right\|_2^2 + \sum_{t=s+1}^{T} \left\| \dot{v}_t^* - \bar{\dot{v}}^{*(2)s} \right\|_2^2 \right\}, \tag{5.1}$$

where $\bar{\dot{v}}^{*(1)s} = \frac{1}{s} \sum_{t=1}^{s} \dot{v}_t^*$ and $\bar{\dot{v}}^{*(2)s} = \frac{1}{T-s} \sum_{t=s+1}^{T} \dot{v}_t^*$.

The following two theorems state the consistency of $\dot{v}_t^*$ and $\tilde{T}_1$, respectively.

Theorem 5.1 *Suppose that Assumptions 1-5 hold. Then* $\forall j \in [p] \cup \{0\}$, *we have*

$$\max_{t \in [T]} \|\dot{v}_t^* - v_t^*\|_2 = O_p(\eta_{N,2}).$$

Theorem 5.2 *Suppose that Assumptions 1-6 hold. Then* $\mathbb{P}\left(\tilde{T}_1 = T_1\right) \to 1$ *as* $(N, T) \to \infty$.

Since the singular vectors of slope matrices contain the structural change information, Theorem 5.1 indicates that we can consistently estimate the break point by using the factor estimates instead of the slope matrix estimates in (3.4). Given Theorem 5.1 and Lemma 2.1(iii), we can prove Theorem 5.2 with arguments analogous to those used in the proof of Theorem 4.4.

5.2 Test for the Presence of a Structural Break

In Section 2, we consider time-varying latent group structures with one break point. In this subsection, we propose a test for the null that the slope coefficients are time-invariant against the alternative that there's one structural break as assumed in Section 2.

Since various scenarios can occur once we allow for the presence of a structural break in the latent group structures, and the number of group may and may not change under the alternative and so do some of the group-specific coefficients. As a first try, one may ignore the information on the latent group structures and test for the possible time-varying feature of the slope coefficients. In this case, we can rewrite Θ_{it}^0 as follows:

$$\Theta_{it}^0 = \Theta_i^0 + c_{it},$$

where $\Theta_i^0 := \frac{1}{T}\sum_{t\in[T]}\Theta_{it}^0$. Then we specify the null and alternative hypothesis respectively as

$$\begin{aligned} H_0 &: c_{it} = 0 \text{ for all } i \in [N], \quad \text{and} \\ H_1 &: c_{it} \neq 0 \text{ for some } i \in [N]. \end{aligned} \tag{5.2}$$

To construct the test statistics, we can follow the idea of Bai and Perron (1998) and consider a sup-F test. Let $\mathcal{T}_\epsilon := \{T_1 : \epsilon T \leq T_1 \leq (1-\epsilon)T\}$, where $\epsilon > 0$ is a tuning parameter that avoids breaks at the end of the sample. Define

$$F_{NT}(1|0) := \max_{i\in[N]} \sup_{T_1\in\mathcal{T}_\epsilon} F_i(T_1),$$

where

$$F_i(T_1) = \frac{T-2p}{p}\left[\tilde{\beta}_i^{(1)}(T_1) - \tilde{\beta}_i^{(2)}(T_1)\right]'\left[\hat{\Sigma}_i(T_1)\right]^{-1}\left[\tilde{\beta}_i^{(1)}(T_1) - \tilde{\beta}_i^{(2)}(T_1)\right],$$

$\tilde{\beta}_i^{(1)}(T_1)$ and $\tilde{\beta}_i^{(2)}(T_1)$ are the PCA slope estimators of Θ_i^0 in the linear panels with IFEs for each individual i with the prior-break observations $\{(i,t) : i \in [N], t \in [T_+]\}$ and post-break observations $\{(i,t) : i \in [N], t \in [T]\backslash[T_+]\}$, respectively,[1] and $\hat{\Sigma}_i(T_+)$ is the consistent estimator for the asymptotic variance of $\tilde{\beta}_i^{(1)}(T_1) - \tilde{\beta}_i^{(2)}(T_1)$. Following Bai and Perron (1998), we conjecture that the asymptotic distribution for $\sup_{T_1\in\mathcal{T}_\epsilon} F_i(T_1)$ is associated with the p-vector of Wiener processes on $[0,1]$, based on which one can derives the corresponding distribution of $F_{NT}(1|0)$.

Alternatively, we can estimate the model with latent group structures by assuming the presence of a break point at T_1. Then we obtain the estimates of the group-specific parameters $\left\{\alpha_j^{(1)}(T_1)\right\}_{j\in K^{(1)}}$ prior to the potential break point T_1 and those of the group-specific parameters $\left\{\alpha_j^{(2)}(T_1)\right\}_{j\in K^{(2)}}$ after the potential break point T_1. It is possible to construct a test statistic based on the contrast of these two sets of estimates or the corresponding residual sum of squares (RSS) and then take the supremum over $T_1 \in \mathcal{T}_\epsilon$. As one can imagine, this approach is also quite

[1] See Section C in the appendix for the detail of the PCA estimation in linear panels with IFEs.

involved as one has to determine the number of groups before and after the break, $K^{(1)}$ and $K^{(1)}$, at each T_1. It is not clear how the estimation errors from these estimates and those of the factors and factor loadings with slow convergence rates affect the asymptotic properties of the estimators of the group-specific parameters.

Last, it is also possible to estimate the model with latent group structures under the case of no structural change to obtain the restricted residuals. If there exists a structural change in the latent group structure, it should be reflected into the restricted residuals obtained under the null. Then we can consider the regression of the restricted residuals on the regressors and construct an LM-type test statistic to check the goodness of fit for such an auxiliary regression model as in Su and Chen (2013). See also Su and Ullah (2013) and Su and Wang (2020) for similar ideas for model specification testing. We leave this for future research.

5.3 The Case of Multiple Breaks

In Section 2, we only consider a one-time structural break in the latent group structures. In practice it is possible to have multiple breaks especially if T is large. Here we generalize the model in Section 2 to allow for multiple breaks. In this case, we have

$$\alpha_{kt} = \begin{cases} \alpha_k^{(1)}, & \text{for} \quad t = 1, \ldots, T_1, \\ \alpha_k^{(2)}, & \text{for} \quad t = T_1 + 1, \ldots, T_2, \\ \vdots \\ \alpha_k^{(b+1)}, & \text{for} \quad t = T_b + 1, \ldots, T, \end{cases}$$

where $b \geq 1$ denotes the number of breaks.

To estimate the number of breaks and the break points $T_1, \cdots, T_b$, in principle we can follow the sequential method proposed by Bai and Perron (1998). First, using the full-sample data, we can construct $F_{NT}(1|0)$ defined in the previous subsection and estimate the break point as in (3.4). Second, for each regime before and after the estimated break point, we test the hypothesis in (5.2) and estimate the break point for each regime separately. At last, we repeat this sequential method until we can not reject the null for all sub-samples. At the end, we can obtain the break point estimates $\left\{\hat{T}_a\right\}_{a\in[\hat{b}]}$ where $\hat{b}$ is the estimated number of breaks. We conjecture that we can establish the consistency of $\hat{b}$ and $\left\{\hat{T}_a\right\}$.

After we obtain the estimated number of breaks and break points, for each sub-sample

$$\left\{(i,t) : i \in [N], t \in \{\hat{T}_{a-1} + 1, \cdots, \hat{T}_a\}\right\},$$

$a \in [\hat{b} + 1]$ with $\hat{T}_0 := 0$ and $\hat{T}_{\hat{b}+1} := T$, we can continue Step 4 in the estimation algorithm in Section 3 to obtain the estimated group structure for each sub-sample.

6 Monte Carlo Simulations

In this section, we show the simulation results of low-rank estimates, break point estimates, group membership estimates and the number of groups estimates with 1000 replication, and we choose the tuning parameter ν_j by the similar procedure described in Chernozhukov et al. (2020). We will focus on the linear panel model $Y_{it} = \lambda_i' f_t + X_{it}'\Theta_{it} + e_{it}$, where $X_{it} = (X_{1,it}, X_{2,it})'$ and $\Theta_{it} = (\Theta_{1,it}, \Theta_{2,it})'$.

6.1 DGP

We focus on the following four main DGPs:

DGP 1: [**Static panel with homoskedasticity**] $X_{1,it} \sim i.i.d.\ U(-2,2)$, $X_{2,it} \sim i.i.d.\ U(-2,2)$, error term $e_{it} \sim i.i.d.\ \mathcal{N}(0,1)$. For Θ_1, we randomly choose the break point T_1 from $0.4T$ to $0.6T$.

DGP 2: [**Static panel with heteroscedasticity**] Compared to the DGP 1, error term $e_{it} \sim i.i.d.\ \mathcal{N}(0, \sigma_{it}^2)$ such that $\sigma_{it}^1 \sim i.i.d.\ U(0.5,1)$. The settings for the regressors and break point are the same as those in DGP 1.

DGP 3: [**Serially correlated error**] Compared to the DGP 2, error term $e_{it} = 0.2e_{i,t-1} + \eta_{it}$, where $\eta_{it} \sim i.i.d.\ \mathcal{N}(0,1)$ and all other settings are the same as in DGP 2.

DGP 4: [**Dynamic panel**] In this case, $X_{1,it} = Y_{i,t-1}$ with $Y_{i,0} \sim i.i.d. \mathcal{N}(0,1)$. $X_{2,it} \sim i.i.d. U(-2,2)$, and $e_{it} \sim i.i.d.\ \mathcal{N}(0, 0.5)$.

For each DGP above, λ_i and f_t are each $i.i.d.\ \mathcal{N}(0,1)$ and mutually independent. We define the slope coefficient based on three subcases below.

DGP X.1: In this case, the group membership and the number of groups don't change after the break point and only the value of the slope coefficient changes. We set the number of groups to be 2, the ratio of individuals among the two groups is $N_1 : N_2 = 0.5 : 0.5$, and the group membership G_1 is obtained by a random draw from $[N]$ without replacement. For DGP 1.1, 2.1 and 3.1,

$$\Theta_{1,it} = \Theta_{2,it} = \begin{cases} 0.1, & i \in G_1,\ t \in \{1, \cdots, T_1\}, \\ 0.9, & i \in G_2,\ t \in \{1, \cdots, T_1\}, \\ 0.1 + \dfrac{2\log(N \vee T)}{\sqrt{N \wedge T}}, & i \in G_1,\ t \in \{T_1 + 1, \cdots, T\}, \\ 0.9 + \dfrac{2\log(N \vee T)}{\sqrt{N \wedge T}}, & i \in G_2,\ t \in \{T_1 + 1, \cdots, T\}. \end{cases}$$

For DGP 4.1, $\Theta_{2,it}$ is same as other DGPs X.1 for $X \in \{1,2,3\}$, and

$$\Theta_{1,it} = \begin{cases} 0.1, & i \in G_1,\ t \in \{1,\cdots,T_1\}, \\ 0.7, & i \in G_2,\ t \in \{1,\cdots,T_1\}, \\ 0.1 + \dfrac{\log(N \vee T)}{\sqrt{N \wedge T}}, & i \in G_1,\ t \in \{T_1+1,\cdots,T\}, \\ 0.7 - \dfrac{\log(N \vee T)}{\sqrt{N \wedge T}}, & i \in G_2,\ t \in \{T_1+1,\cdots,T\}. \end{cases}$$

DGP X.2: Compared to DGP X.1, the values of the slope coefficients for different groups do not change after the break point, but the group membership changes. The number of groups is 2, the ratio of individuals among the group groups is still $N_1 : N_2 = 0.5 : 0.5$. Nevertheless, $\{G_1^{(1)}, G_2^{(1)}\}$ is different from $\{G_1^{(2)}, G_2^{(2)}\}$ so that elements in both $G_1^{(1)}$ and $G_1^{(2)}$ are independent draws from $[N]$ without replacement. In addition, for DGPs 1.2, 2.2, and 3.2,

$$\Theta_{1,it} = \Theta_{2,it} = \begin{cases} 0.1, & i \in G_1^{(1)},\ t \in \{1,\cdots,T_1\}, \\ 0.9, & i \in G_2^{(1)},\ t \in \{1,\cdots,T_1\}, \\ 0.1, & i \in G_1^{(2)},\ t \in \{T_1+1,\cdots,T\}, \\ 0.9, & i \in G_2^{(2)},\ t \in \{T_1+1,\cdots,T\}. \end{cases}$$

For DGP 4.2, $\Theta_{2,it}$ is defined same as other DGPs X.2 for $X \in \{1,2,3\}$, and

$$\Theta_{1,it} = \begin{cases} 0.1, & i \in G_1^{(1)},\ t \in \{1,\cdots,T_1\}, \\ 0.7, & i \in G_2^{(1)},\ t \in \{1,\cdots,T_1\}, \\ 0.1, & i \in G_1^{(2)},\ t \in \{T_1+1,\cdots,T\}, \\ 0.7, & i \in G_2^{(2)},\ t \in \{T_1+1,\cdots,T\}. \end{cases}$$

DGP X.3: Under this scenario, the number of groups changes after the breaking. We set $N_1^{(1)} : N_2^{(1)} = 0.5 : 0.5$ and $N_1^{(2)} : N_2^{(2)} : N_3^{(2)} = 0.4 : 0.3 : 0.3$ before and after the break, respectively. Specifically, for DGPs 1.3, 2.3, and 3.3, we have

$$\Theta_{1,it} = \Theta_{2,it} = \begin{cases} 0.1, & i \in G_1^{(1)},\ t \in \{1,\cdots,T_1\}, \\ 0.9, & i \in G_2^{(1)},\ t \in \{1,\cdots,T_1\}, \\ 0.1, & i \in G_1^{(2)},\ t \in \{T_1+1,\cdots,T\}, \\ 0.5, & i \in G_2^{(2)},\ t \in \{T_1+1,\cdots,T\}, \\ 0.9, & i \in G_3^{(2)},\ t \in \{T_1+1,\cdots,T\}. \end{cases}$$

For DGP 4.3, $\Theta_{2,it}$ is defined same as other DGPs X.3 for $X \in \{1,2,3\}$, and

$$\Theta_{1,it} = \begin{cases} 0.1, & i \in G_1^{(1)},\ t \in \{1,\cdots,T_1\}, \\ 0.7, & i \in G_2^{(1)},\ t \in \{1,\cdots,T_1\}, \\ 0.1, & i \in G_1^{(2)},\ t \in \{T_1+1,\cdots,T\}, \\ 0.4, & i \in G_2^{(2)},\ t \in \{T_1+1,\cdots,T\}, \\ 0.7, & i \in G_3^{(2)},\ t \in \{T_1+1,\cdots,T\}. \end{cases}$$

6.2 Results

Table 1 shows the frequency of correct rank estimation results for fixed effect and slope matrices estimates from SVT in Section 3.3. From Table 1, we notice that the true rank of both fixed effect matrix and slope matrix can be perfectly estimated for the sample sizes under investigation. Table 2 gives the results of break point estimation in Step 3 based on different (N,T) combinations. Clearly, the break points can be perfectly estimated even when the break size is small as in DGP X.1 for X$= 1,2,3$, and 4.

Table 3 shows the group membership estimation results. With known number of groups, the STK algorithm degenerates to the traditional K-means algorithm. The "Infeasible" part gives the frequency of correct group membership estimation before and after the estimated break point, G_B and G_A, based on the known true number of groups and K-means algorithm. Obviously, K-means classification exhibits excellent performance in this case.

However, without prior information on the true number of groups, STK algorithm is able to estimate the group membership and the number of groups simultaneously. In this case, the frequencies of correct estimation of the group membership and that the number of groups are shown in the "Feasible" part in Table 3 and in Table 4, respectively. Table 5 presents more results for the estimation of the number of groups. For DGPs 1.X and DGP 2.X where we have static panels with independent errors, the results show that the group membership and the number of groups can be well estimated with nearly 100% correct rate under different (N,T) combinations. For DGPs 3.X and 4.X where we have static panels with serial correlated errors and dynamic panels, respectively, the frequency of correct estimation of the group membership and the number of groups estimation are not great when T is small, but they are gradually approaching 1 when as T increases. One reason for this is that we need to use HAC estimates of certain long-run variance object in the STK algorithm and it is well known that large T is required in order for the HAC estimates to be reasonably well behaved.

7 Empirical Study

Foreign direct investment (FDI), the inflow of the investment from one economy to the other, is an important indicator to stimulate the economic growth. According to the literature, however,

Table 1: Frequency of correct rank estimation

N		100		200		N		100		200	
T		100	200	100	200	T		100	200	100	200
DGP 1.1	$r_0 = 1$	1.00	1.00	1.00	1.00	DGP 3.1	$r_0 = 1$	1.00	1.00	1.00	1.00
	$r_1 = 2$	1.00	1.00	1.00	1.00		$r_1 = 2$	1.00	1.00	1.00	1.00
	$r_2 = 2$	1.00	1.00	1.00	1.00		$r_2 = 2$	1.00	1.00	1.00	1.00
DGP 1.2	$r_0 = 1$	1.00	1.00	1.00	1.00	DGP 3.2	$r_0 = 1$	1.00	1.00	1.00	1.00
	$r_1 = 2$	1.00	1.00	1.00	1.00		$r_1 = 2$	1.00	1.00	1.00	1.00
	$r_2 = 2$	1.00	1.00	1.00	1.00		$r_2 = 2$	1.00	1.00	1.00	1.00
DGP 1.3	$r_0 = 1$	1.00	1.00	1.00	1.00	DGP 3.3	$r_0 = 1$	1.00	1.00	1.00	1.00
	$r_1 = 2$	1.00	1.00	1.00	1.00		$r_1 = 2$	1.00	1.00	1.00	1.00
	$r_2 = 2$	1.00	1.00	1.00	1.00		$r_2 = 2$	0.998	1.00	1.00	1.00
DGP 2.1	$r_0 = 1$	1.00	1.00	1.00	1.00	DGP 4.1	$r_0 = 1$	1.00	1.00	1.00	1.00
	$r_1 = 2$	1.00	1.00	1.00	1.00		$r_1 = 2$	1.00	1.00	1.00	1.00
	$r_2 = 2$	1.00	1.00	1.00	1.00		$r_2 = 2$	1.00	1.00	1.00	1.00
DGP 2.2	$r_0 = 1$	1.00	1.00	1.00	1.00	DGP 4.2	$r_0 = 1$	1.00	1.00	1.00	1.00
	$r_1 = 2$	1.00	1.00	1.00	1.00		$r_1 = 2$	1.00	1.00	1.00	1.00
	$r_2 = 2$	1.00	1.00	1.00	1.00		$r_2 = 2$	1.00	1.00	1.00	1.00
DGP 2.3	$r_0 = 1$	1.00	1.00	1.00	1.00	DGP 4.3	$r_0 = 1$	1.00	1.00	1.00	1.00
	$r_1 = 2$	1.00	1.00	1.00	1.00		$r_1 = 2$	1.00	1.00	1.00	1.00
	$r_2 = 2$	1.00	1.00	1.00	1.00		$r_2 = 2$	1.00	1.00	1.00	1.00

Table 2: Frequency of correct break point estimation

N	100		200		N	100		200	
T	100	200	100	200	T	100	200	100	200
DGP 1.1	1.00	1.00	1.00	1.00	DGP 3.1	1.00	1.00	1.00	1.00
DGP 1.2	0.999	1.00	1.00	1.00	DGP 3.2	1.00	1.00	1.00	1.00
DGP 1.3	1.00	1.00	1.00	1.00	DGP 3.3	1.00	1.00	1.00	1.00
DGP 2.1	1.00	1.00	1.00	1.00	DGP 4.1	1.00	1.00	1.00	1.00
DGP 2.2	1.00	1.00	1.00	1.00	DGP 4.2	1.00	1.00	1.00	1.00
DGP 2.3	1.00	1.00	1.00	1.00	DGP 4.3	1.00	1.00	1.00	1.00

Table 3: Frequency of correct group membership estimation

	N		100		200			N		100		200	
	T		100	200	100	200		T		100	200	100	200
Infeasible	DGP 1.1	G_B	1.00	1.00	1.00	1.00	Feasible	DGP 1.1	G_B	1.00	1.00	1.00	1.00
		G_A	1.00	1.00	1.00	1.00			G_A	1.00	1.00	1.00	1.00
	DGP 1.2	G_B	1.00	1.00	1.00	1.00		DGP 1.2	G_B	1.00	1.00	1.00	1.00
		G_A	1.00	1.00	1.00	1.00			G_A	1.00	1.00	1.00	1.00
	DGP 1.3	G_B	1.00	1.00	1.00	1.00		DGP 1.3	G_B	1.00	1.00	1.00	1.00
		G_A	0.989	0.999	0.978	0.999			G_A	0.989	0.999	0.978	0.999
	DGP 2.1	G_B	1.00	1.00	1.00	1.00		DGP 2.1	G_B	0.991	0.999	0.983	0.999
		G_A	1.00	1.00	1.00	1.00			G_A	0.993	0.998	0.985	0.998
	DGP 2.2	G_B	1.00	1.00	1.00	1.00		DGP 2.2	G_B	0.989	0.999	0.992	0.999
		G_A	1.00	1.00	1.00	1.00			G_A	0.992	0.999	0.977	0.998
	DGP 2.3	G_B	1.00	1.00	1.00	1.00		DGP 2.3	G_B	0.992	0.999	0.961	0.999
		G_A	0.998	1.00	0.999	1.00			G_A	0.989	0.999	0.992	0.999
	DGP 3.1	G_B	1.00	1.00	1.00	1.00		DGP 3.1	G_B	0.989	0.997	0.967	0.998
		G_A	1.00	1.00	1.00	1.00			G_A	0.976	0.997	0.977	0.994
	DGP 3.2	G_B	1.00	1.00	1.00	1.00		DGP 3.2	G_B	0.985	0.996	0.962	0.993
		G_A	1.00	1.00	1.00	1.00			G_A	0.985	0.994	0.973	0.998
	DGP 3.3	G_B	1.00	1.00	1.00	1.00		DGP 3.3	G_B	0.985	0.998	0.973	0.995
		G_A	0.981	0.997	0.982	0.999			G_A	0.971	0.994	0.968	0.998
	DGP 4.1	G_B	1.00	1.00	1.00	1.00		DGP 4.1	G_B	0.975	0.998	0.957	0.998
		G_A	1.00	1.00	1.00	1.00			G_A	0.992	0.999	0.987	0.998
	DGP 4.2	G_B	1.00	1.00	1.00	1.00		DGP 4.2	G_B	0.994	0.998	0.952	0.997
		G_A	1.00	1.00	1.00	1.00			G_A	0.977	0.999	0.985	0.999
	DGP 4.3	G_B	1.00	1.00	1.00	1.00		DGP 4.3	G_B	0.983	0.998	0.948	0.999
		G_A	1.00	1.00	1.00	1.00			G_A	0.982	0.998	0.983	0.998

Table 4: Frequency of correct estimation of the number of groups

N		100		200		N		100		200	
T		100	200	100	200	T		100	200	100	200
DGP 1.1	$K^{(1)} = 2$	0.999	1.00	1.00	1.00	DGP 3.1	$K^{(1)} = 2$	0.924	0.980	0.788	0.983
	$K^{(2)} = 2$	0.998	1.00	0.999	1.00		$K^{(2)} = 2$	0.823	0.979	0.832	0.960
DGP 1.2	$K^{(1)} = 2$	1.00	1.00	1.00	1.00	DGP 3.2	$K^{(1)} = 2$	0.868	0.985	0.759	0.940
	$K^{(2)} = 2$	1.00	1.00	1.00	0.999		$K^{(2)} = 2$	0.897	0.971	0.829	0.987
DGP 1.3	$K^{(1)} = 2$	0.999	1.00	1.00	1.00	DGP 3.3	$K^{(1)} = 2$	0.889	0.988	0.802	0.965
	$K^{(2)} = 3$	1.00	0.999	1.00	1.00		$K^{(2)} = 3$	0.932	0.977	0.907	0.988
DGP 2.1	$K^{(1)} = 2$	0.933	0.990	0.864	0.989	DGP 4.1	$K^{(1)} = 2$	0.786	0.980	0.679	0.984
	$K^{(2)} = 2$	0.936	0.987	0.901	0.990		$K^{(2)} = 2$	0.938	0.991	0.895	0.983
DGP 2.2	$K^{(1)} = 2$	0.919	0.995	0.940	0.994	DGP 4.2	$K^{(1)} = 2$	0.933	0.988	0.630	0.975
	$K^{(2)} = 2$	0.930	0.993	0.809	0.982		$K^{(2)} = 2$	0.758	0.988	0.870	0.989
DGP 2.3	$K^{(1)} = 2$	0.940	0.989	0.724	0.990	DGP 4.3	$K^{(1)} = 2$	0.877	0.991	0.657	0.991
	$K^{(2)} = 3$	0.946	0.995	0.952	0.992		$K^{(2)} = 3$	0.900	0.987	0.874	0.980

Table 5: Determination of the number of groups

DGP	N	T	$\hat{K}^{(1)}$				$\hat{K}^{(2)}$			
			2	3	4	>4	2	3	4	>4
DGP 1.1	100	100	**0.999**	0.001	0.00	0.00	**0.998**	0.002	0.00	0.00
		200	**1.00**	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
	200	100	**1.00**	0.00	0.00	0.00	**0.999**	0.001	0.00	0.00
		200	**1.00**	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
DGP 1.2	100	100	**1.00**	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
		200	**1.00**	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
	200	100	**1.00**	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
		200	**1.00**	0.00	0.00	0.00	**0.999**	0.001	0.00	0.00
DGP 1.3	100	100	**0.999**	0.001	0.00	0.00	0.00	**1.00**	0.00	0.00
		200	**1.00**	0.00	0.00	0.00	0.00	**0.999**	0.001	0.00
	200	100	**1.00**	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
		200	**1.00**	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
DGP 2.1	100	100	**0.933**	0.058	0.009	0.00	**0.936**	0.060	0.003	0.001
		200	**0.990**	0.010	0.00	0.00	**0.987**	0.013	0.00	0.00
	200	100	**0.864**	0.126	0.010	0.00	**0.901**	0.090	0.009	0.00
		200	**0.989**	0.011	0.00	0.00	**0.990**	0.010	0.000	0.00
DGP 2.2	100	100	**0.919**	0.074	0.007	0.00	**0.930**	0.067	0.003	0.00
		200	**0.995**	0.003	0.00	0.002	**0.993**	0.006	0.00	0.001
	200	100	**0.940**	0.056	0.004	0.00	**0.809**	0.164	0.027	0.00
		200	**0.994**	0.006	0.00	0.00	**0.982**	0.018	0.00	0.00
DGP 2.3	100	100	**0.940**	0.055	0.005	0.00	0.00	**0.946**	0.039	0.015
		200	**0.989**	0.011	0.00	0.00	0.00	**0.995**	0.002	0.003
	200	100	**0.724**	0.230	0.046	0.00	0.00	**0.952**	0.031	0.017
		200	**0.990**	0.010	0.00	0.00	0.00	**0.992**	0.006	0.002
DGP 3.1	100	100	**0.924**	0.064	0.012	0.00	**0.823**	0.144	0.033	0.00
		200	**0.980**	0.019	0.001	0.00	**0.979**	0.020	0.001	0.00
	200	100	**0.788**	0.189	0.023	0.00	**0.832**	0.143	0.025	0.00
		200	**0.983**	0.017	0.00	0.00	**0.960**	0.038	0.002	0.00
DGP 3.2	100	100	**0.868**	0.109	0.023	0.00	**0.897**	0.099	0.004	0.00
		200	**0.985**	0.008	0.003	0.004	**0.971**	0.021	0.006	0.002
	200	100	**0.759**	0.198	0.042	0.001	**0.829**	0.147	0.024	0.00
		200	**0.940**	0.055	0.005	0.00	**0.987**	0.013	0.000	0.00
DGP 3.3	100	100	**0.889**	0.100	0.011	0.00	0.00	**0.932**	0.055	0.013
		200	**0.988**	0.009	0.003	0.00	0.00	**0.977**	0.013	0.010
	200	100	**0.802**	0.175	0.023	0.00	0.00	**0.907**	0.073	0.020
		200	**0.965**	0.035	0.000	0.000	0.000	**0.988**	0.010	0.002
DGP 4.1	100	100	**0.786**	0.116	0.086	0.012	**0.938**	0.991	0.895	0.983
		200	**0.980**	0.011	0.007	0.002	**0.054**	0.008	0.091	0.016
	200	100	**0.679**	0.160	0.152	0.009	**0.007**	0.001	0.014	0.001
		200	**0.984**	0.009	0.006	0.001	**0.001**	0.000	0.000	0.000
DGP 4.2	100	100	**0.933**	0.051	0.012	0.004	**0.758**	0.141	0.089	0.012
		200	**0.988**	0.006	0.004	0.002	**0.988**	0.005	0.006	0.001
	200	100	**0.630**	0.158	0.196	0.016	**0.870**	0.080	0.048	0.002
		200	**0.975**	0.013	0.012	0.000	**0.989**	0.009	0.002	0.000
DGP 4.3	100	100	**0.877**	0.076	0.042	0.005	0.000	**0.900**	0.055	0.045
		200	**0.991**	0.006	0.002	0.001	0.000	**0.987**	0.010	0.003
	200	100	**0.657**	0.191	0.129	0.023	0.000	**0.874**	0.072	0.054
		200	**0.991**	0.005	0.004	0.000	0.000	**0.980**	0.012	0.008

the effect of the FDI to the economic growth is different across economies. For instance, Adewumi (2007) discovers that the contribution of FDI to economic growth is positive in most of the African countries but not significant during the time period 1970-2003, and Juma (2012) shows the positive and significant effect of the FDI to the economic growth in Sub-Saharan Africa. For the relationship of FDI and economic growth, in general, some argue that FDI leads to economic growth and productivity increases in the economy as a whole and hence contributes to the differences in economic growth and development performance across countries, but others stress the risk of FDI destroying local capabilities and extracting natural resources without adequately compensating poor countries. Moreover, some argue that the impact of FDI is not only positive or negative, but also depends on the type of FDI, economic conditions and policies; see, e.g., Dirk Willem (2006).

Inspired by the above observation, we aim to study the relationship of FDI and economic growth rate. We consider the following linear panel data model with IFEs:

$$Growth_{it} = \lambda_i' f_t + \Theta_{1,it} Growth_{i,t-1} + \Theta_{2,it} FDI_{it} + e_{it}, \tag{7.1}$$

where λ_i is the individual fixed effect, f_t is the time fixed effect, $Growth_{it}$ is the economic growth measured by the growth rate of real GDP. FDI_{it} is the ratio of foreign direct investment to GDP for country i at year t.

We obtain the data from the World Bank Development Indicators (WDI) historical database for 166 countries and regions from 2000-2019. Based on model (7.1) and the rank estimation result, $\hat{r}_1 = 2$ and $\hat{r}_2 = 1$, we want to detect the break point and recover the group structure before and after the break point.

Based on our estimation algorithm, result shows that the break point takes place at year 2006. It's well known that the Great Recession was observed in 2008, which is owing to the subprime mortgage taken place in 2006. We conjecture that our estimated break point is owing to the the subprime mortgage. Furthermore, group structure estimated before and after the break is shown in Table 6, Figures 2 and 3. We notice that there are five different groups from year 2000 to 2005 and four groups from year 2006-2019. After the break, not only the group membership changes but also one group is vanishing compared to the time period before the group. Consequently, we may conclude that the effects of FDI and previous growth rate on the economic growth rate are different across time and countries. Especially, the effects across different countries do not simply depend on whether they are developed or developing countries or they belong to the high income and low income countries. Indeed, the last group in the left hand side of Table 6 includes both developed countries like United States, United Kingdom, and Singapore and developing countries like Thailand. We conjecture that it may due to some unobservable heterogeneity like the geographic or the spatial correlation.

Table 6: Classification before and after the break date

Before the break
Australia, Austria, Bahrain, Belize, Benin, Bosnia and Herzegovina, Botswana, The Democratic Republic of the Congo, Costa Rica, Denmark, El Salvador, Finland, France, Gabon, Ghana, Guatemala, Haiti, Israel, Kiribati, Lebanon, Malaysia, Maldives, Malta, Mexico, Netherlands, Paraguay, Sierra Leone, Togo, United Arab Emirates
Angola, Azerbaijan, Chad, Comoros, Dominican Republic, Equatorial Guinea, Guyana, Papua New Guinea, Rwanda, Sudan, Uruguay
Barbados, Albania, Algeria, Argentina, Bangladesh, Bermuda, Cambodia, Cameroon, China, Cote d'Ivoire, Cyprus, Czech Republic, Dominica, Ethiopia, Germany, Iceland, India, Indonesia, Iran, Ireland, Jamaica, Jordan, Kazakhstan, Kenya, Laos, Lesotho, Mauritania, Myanmar, New Zealand, Niger, Norway, Pakistan, Samoa, Senegal, Seychelles, Solomon Islands, Suriname, Tanzania, Trinidad and Tobago, Uzbekistan, Vanuatu, Vietnam, West Bank and Gaza, Zambia
Antigua and Barbuda, Belarus, Bolivia, Brazil, Brunei Darussalam, Bulgaria, Cabo Verde, Colombia, The Republic of Congo, Croatia, Ecuador, Egypt, Eswatini, Georgia, Greece, Hungary, Iraq, Italy, Korea, Kuwait, Latvia, Malawi, Mauritius, Mongolia, Morocco, Mozambique, Namibia, Nepal, Nicaragua, Nigeria, North, Macedonia, Oman, Panama, Peru, Philippines, Poland, Portugal, Romania, Slovenia, Spain, Sri Lanka, St. Vincent and the Grenadines, Tajikistan, Tunisia, Uganda, Zimbabwe
Armenia, The Bahamas, Belgium, Burkina Faso, Canada, Central African Republic, Chile, Estonia, Fiji, The Gambia, Grenada, Guinea-Bissau, Honduras, Hong Kong (China), Japan, Kyrgyz Republic, Lithuania, Macao (China), Madagascar, Mali, Moldova, Russian Federation, Saudi Arabia, Singapore, Slovak Republic, South Africa, St. Kitts and Nevis, St. Lucia, Sweden, Switzerland, Thailand, Tonga, Turkey, Ukraine, United Kingdom, United States

After the break
Algeria, Angola, Antigua and Barbuda, Bahrain, Benin, Cambodia, Cameroon, Cote d'Ivoire, Cyprus, Equatorial Guinea, Guyana, Kazakhstan, Kuwait, Malawi, Mauritius, Nepal, New Zealand, Pakistan, Panama, Romania, Rwanda, Solomon Islands, Spain, Sri Lanka, St. Vincent and the Grenadines, Togo, Uganda, Zimbabwe
Albania, Azerbaijan, Bangladesh, Bolivia, China, The Democratic Republic of the Congo, Egypt, Ethiopia, Greece, India, Indonesia, Jordan, Laos, Lebanon, Myanmar, Nigeria, Philippines, Poland, Tajikistan, Tanzania, United Arab Emirates, Uruguay, Uzbekistan, Vietnam, Zambia
Belarus, Belize, Bermuda, Bosnia and Herzegovina, Brunei Darussalam, Bulgaria, Cabo Verde, Colombia, The Republic of Congo, Croatia, Dominican Republic, Ecuador, Eswatini, Georgia, Ghana, Guatemala, Hungary, Iceland, Ireland, Israel, Italy, Jamaica, Korea, Latvia, Lesotho, Mongolia, Morocco, Mozambique, Nicaragua, North Macedonia, Oman, Papua New Guinea, Peru, Portugal, Samoa, Senegal, Suriname, Trinidad and Tobago, Tunisia
Barbados, Argentina, Armenia, Australia, Austria, Bahamas, Belgium, Botswana, Brazil, Burkina Faso, Canada, Central African Republic, Chad, Chile, Comoros, Costa Rica, Czech Republic, Denmark, Dominica, El Salvador, Estonia, Fiji, Finland, France, Gabon, Gambia, Germany, Grenada, Guinea-Bissau, Haiti, Honduras, Hong Kong (China), Iran, Iraq, Japan, Kenya, Kiribati, Kyrgyz Republic, Lithuania, Macao (China), Madagascar, Malaysia, Maldives, Mali, Malta, Mauritania, Mexico, Moldova, Namibia, Netherlands, Niger, Norway, Paraguay, Russian Federation, Saudi Arabia, Seychelles, Sierra Leone, Singapore, Slovak Republic, Slovenia, South Africa, St. Kitts and Nevis, St. Lucia, Sudan, Sweden, Switzerland, Thailand, Tonga, Turkey, Ukraine, United Kingdom, United States, Vanuatu, West Bank and Gaza

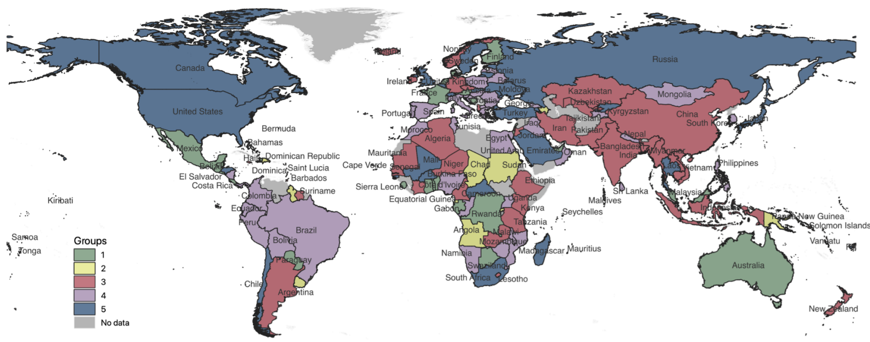

Figure 2: Group classification by countries using data 2000-2005

8 Conclusion

This paper considers the linear panel model with IFEs and two-way heterogeneity such that the heterogeneity across individuals is captured by latent group structures and the heterogeneity across time is captured by an unknown structural break. We allow the model to have different group numbers, or different group membership, or just the value of slope coefficient changing for some specific groups before and after the break. To estimate the unknown structural break, the number of groups and group membership before and after the break point, we propose an estimation algorithm with initial estimates from nuclear norm regularization followed by row- and column-wise linear regressions. Then, the break point estimator is obtained by binary segmentation and the group structure together with the number of groups are estimated simultaneously by sequential testing K-means algorithm. Asymptotic theory shows that the structural break estimator, the number of groups estimators and group membership estimates before and after the break point are consistent, and the final slope coefficient estimators have the oracle property.

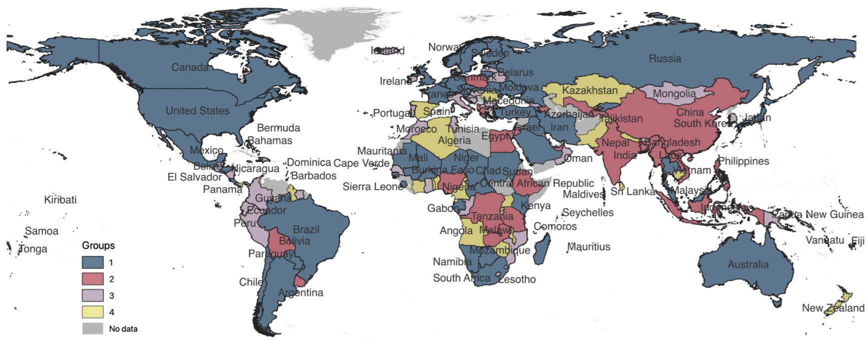

Figure 3: Group classification by countries using data 2006-2019

With latent group structure across individuals in the cross-sectional span, we also discuss a test for the null that the slope coefficient is homogeneous across t against the alternative that the slope coefficient shows one structural break. Following this, we propose a way to generalize our model having multiple structural breaks, which opens up the opportunity to have the comprehensive analyses of multiple structural breaks together with the time-varying latent group structure for the future research.

References

Adewumi, S. (2007). The impact of fdi on growth in developing countries: An african experience.

Ando, T. and Bai, J. (2015). A simple new test for slope homogeneity in panel data models with interactive effects. *Economics Letters*, 136:112–117.

Ando, T. and Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, 31(1):163–191.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, pages 817–858.

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

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics*, 157(1):78–92.

Bai, J. and Ng, S. (2019). Rank regularized estimation of approximate factor models. *Journal of Econometrics*, 212(1):78–96.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78.

Baltagi, B. H., Kao, C., and Liu, L. (2017). Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term. *Econometric Reviews*, 36(1-3):85–102.

Belloni, A., Chen, M., Padilla, O. H. M., and Wang, Z. (2019). High dimensional latent panel quantile regression with an application to asset pricing. *arXiv preprint arXiv:1912.02151*.

Ben-David, D. (1998). Convergence clubs and subsistence economies. *Journal of Development Economics*, 55(1):155–171.

Berthelemy, J.-C. and Varoudakis, A. (1996). Economic growth, convergence clubs, and the role of financial development. *Oxford economic papers*, 48(2):300–328.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Chernozhukov, V., Hansen, C. B., Liao, Y., and Zhu, Y. (2020). Inference for heterogeneous effects using low-rank estimations. Technical report, CEMMAP working paper.

Chu, Z. (2012). Logistics and economic growth: a panel data approach. *The Annals of regional science*, 49(1):87–102.

Dirk Willem, T. V. (2006). *Foreign direct investment and development: An historical perspective.* Overseas Development Institute ODI London.

Durlauf, S. N. and Johnson, P. A. (1995). Multiple regimes and cross-country growth behaviour. *Journal of applied econometrics*, 10(4):365–384.

Feng, J. (2019). Regularized quantile regression with interactive fixed effects. *arXiv preprint arXiv:1911.00166*.

Golub, G. H. and Van Loan, C. F. (1996). Matrix computations. edition.

Hong, S., Su, L., and Jiang, T. (2022). Profile gmm estimation of panel data models with interactive fixed effects. *Journal of Econometrics*, forthcoming.

Huang, W., Jin, S., and Su, L. (2020). Identifying latent grouped patterns in cointegrated panels. *Econometric Theory*, 36(3):410–456.

Jin, J., Ke, Z. T., Luo, S., and Wang, M. (2022). Optimal estimation of the number of network communities. *Journal of the American Statistical Association*, 0(0):1–16.

Juma, M.-A. (2012). The effect of foreign direct investment on growth in sub-saharan africa.

Ke, Y., Li, J., and Zhang, W. (2016). Structure identification in panel data analysis. *The Annals of Statistics*, 44(3):1193–1233.

Ke, Z. T., Fan, J., and Wu, Y. (2015). Homogeneity pursuit. *Journal of the American Statistical Association*, 110(509):175–194.

Keane, M. and Neal, T. (2020). Climate change and us agriculture: Accounting for multidimensional slope heterogeneity in panel data. *Quantitative Economics*, 11(4):1391–1429.

Kim, D. (2011). Estimating a common deterministic time trend break in large panels with cross sectional dependence. *Journal of Econometrics*, 164(2):310–330.

Kim, D. (2014). Common breaks in time trends for large panel data with a factor structure. *The Econometrics Journal*, 17(3):301–337.

Klapper, L. and Love, I. (2011). The impact of the financial crisis on new firm registration. *Economics Letters*, 113(1):1–4.

Li, D., Qian, J., and Su, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association*, 111(516):1804–1819.

Lin, C.-C. and Hsu, C.-C. (2011). Change-point estimation for nonstationary panel data. Technical report, Working Paper, Department of Economics, National Central University.

Lin, C.-C. and Ng, S. (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, 1(1):42–55.

Long, W., Li, N., Wang, H., and Cheng, S. (2012). Impact of us financial crisis on different countries: based on the method of functional analysis of variance. *Procedia Computer Science*, 9:1292–1298.

Lu, X. and Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 190(1):148–175.

Lu, X. and Su, L. (2017). Determining the number of groups in latent panel structures with an application to income and democracy. *Quantitative Economics*, 8(3):729–760.

Lu, X. and Su, L. (2022). Uniform inference in linear panel data models with two-dimensional heterogeneity. *Journal of Econometrics*, forthcoming.

Lumsdaine, R. L., Okui, R., and Wang, W. (2022). Estimation of panel group structure models with structural breaks in group memberships and coefficients. *Journal of Econometrics*, forthcoming.

Ma, S., Su, L., and Zhang, Y. (2020). Detecting latent communities in network formation models. *arXiv preprint arXiv:2005.03226*.

Merlevède, F., Peligrad, M., and Rio, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High dimensional probability V: the Luminy volume*, pages 273–292. Institute of Mathematical Statistics.

Miao, K., Phillips, P. C., and Su, L. (2022). High-dimensional vars with common factors. *Journal of Econometrics*, forthcoming.

Moon, H. R. and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195.

Moon, H. R. and Weidner, M. (2018). Nuclear norm regularized estimation of panel regression models. *arXiv preprint arXiv:1810.10987*.

Okui, R. and Wang, W. (2021). Heterogeneous structural breaks in panel data models. *Journal of Econometrics*, 220(2):447–473.

Pesaran, M. H. and Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of econometrics*, 142(1):50–93.

Prakasa Rao, B. L. (2009). Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics*, 61(2):441–460.

Qian, J. and Su, L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics*, 191(1):86–109.

Sarafidis, V. and Weber, N. (2015). A partially heterogeneous framework for analyzing panel data. *Oxford Bulletin of Economics and Statistics*, 77(2):274–296.

Song, M. (2013). *Essays on Large Panel Data Analysis*. Columbia University.

Su, L. and Chen, Q. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory*, 29(6):1079–1135.

Su, L. and Ju, G. (2018). Identifying latent grouped patterns in panel data models with interactive fixed effects. *Journal of Econometrics*, 206(2):554–573.

Su, L., Shi, Z., and Phillips, P. C. (2016). Identifying latent structures in panel data. *Econometrica*, 84(6):2215–2264.

Su, L. and Ullah, A. (2013). A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. *Econometric Theory*, 29(1):187–212.

Su, L., Wang, W., and Zhang, Y. (2020). Strong consistency of spectral clustering for stochastic block models. *IEEE Transactions on Information Theory*, 66(1):324–338.

Su, L. and Wang, X. (2020). Testing for structural changes in factor models via a nonparametric regression. *Econometric Theory*, 36(6):1127–1158.

Su, L., Wang, X., and Jin, S. (2019). Sieve estimation of time-varying panel data models with latent structures. *Journal of Business & Economic Statistics*, 37(2):334–349.

Tropp, J. A. (2011). User-friendly tail bounds for matrix martingales. Technical report, CALIFORNIA INST OF TECH PASADENA.

Vaart, A. W. and Wellner, J. A. (1996). Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*.

Wang, W., Phillips, P. C., and Su, L. (2018). Homogeneity pursuit in panel data models: Theory and application. *Journal of Applied Econometrics*, 33(6):797–815.

Wang, W., Phillips, P. C., and Su, L. (2019). The heterogeneous effects of the minimum wage on employment across states. *Economics Letters*, 174:179–185.

Wang, W. and Su, L. (2021). Identifying latent group structures in nonlinear panels. *Journal of Econometrics*, 220(2):272–295.

Wang, Y., Su, L., and Zhang, Y. (2022). Low-rank panel quantile regression: Estimation and inference. *https://arxiv.org/abs/2210.11062*.

White, H. (2014). *Asymptotic theory for econometricians*. Academic press.

Yu, Y., Wang, T., and Samworth, R. J. (2015). A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323.

Zhang, J. and Cheng, L. (2019). Threshold effect of tourism development on economic growth following a disaster shock: Evidence from the wenchuan earthquake, pr china. *Sustainability*, 11(2):371.

Zhang, Y., Wang, H. J., and Zhu, Z. (2019). Quantile-regression-based clustering for panel data. *Journal of Econometrics*, 213(1):54–67.

Online Supplement for
"Panel Data Models with Time-Varying Latent Group Structures"

Yiren Wang[a], Peter C.B. Phillips[b] and Liangjun Su[c]

[a]School of Economics, Singapore Management University, Singapore

[b]Yale University, University of Auckland,

University of Southampton, Singapore Management University

[c]School of Economics and Management, Tsinghua University, China

This online supplement contains four sections. Section A contains the proofs of the main results by calling upon some technical lemmas in Section B and Section D. Section B states and proves the technical lemmas used in Section A. Section C contains the estimation procedure for the panel model with interactive fixed effects (IFEs) and slope heterogeneity, and proposes the test statistics for the slope homogeneity. Section D shows the uniformly asymptotic theories for the slope estimators and test statistics proposed in Section C.

A Proofs of the Main Results

A.1 Proof of Lemma 2.1

(i) Recall that $\mathcal{G}_j^{(\ell)} = \left\{G_{1,j}^{(\ell)}, \cdots, G_{K_\ell,j}^{(\ell)}\right\}$. For the special case when $\mathcal{G}_j^{(1)} = \mathcal{G}_j^{(2)}$ and $\alpha_{k,j}^{(1)} = \mu\alpha_{k,j}^{(2)}$ such that μ is a constant, the group structure does not change, the break size is same for all groups, and $r_j = 1$. Except for this case, below we will show that $r_j = 2$.

Let $A_{j,i}^{(\ell)} = \sum_{k=1}^{K_\ell} \alpha_{k,j}^{(\ell)} \mathbf{1}\{i \in G_{k,j}^{(\ell)}\}$, $A_{j,i} = \left(A_{j,i}^{(1)}, A_{j,i}^{(2)}\right)'$ and $A_j = (A_{j,1}, \cdots, A_{j,N})' \in R^{N\times 2}$. Define the 2×2 symmetric matrix $B_j = A_j'A_j$. Let $B_j^{\frac{1}{2}}$ be the symmetric square root of B_j. By the Singular value decomposition (SVD), $B_j^{\frac{1}{2}} \begin{bmatrix} \sqrt{\tau_T} & 0 \\ 0 & \sqrt{1-\tau_T} \end{bmatrix} = L_j S_j R_j'$, where $L_j'L_j = R_j'R_j = I_2$ and S_j is diagonal. Then

$$\begin{aligned}
\Theta_j^0 &= \begin{bmatrix} A_{j,1}^{(1)}\iota_{T_1}' & A_{j,1}^{(2)}\iota_{T-T_1}' \\ \vdots & \vdots \\ A_{j,N}^{(1)}\iota_{T_1}' & A_{j,N}^{(2)}\iota_{T-T_1}' \end{bmatrix} = A_j \begin{bmatrix} \iota_{T_1} & \mathbf{0}_{T_1} \\ \mathbf{0}_{T-T_1} & \iota_{T-T_1} \end{bmatrix}' \\
&= A_j B_j^{-1/2} L_j S_j R_j' \begin{bmatrix} \frac{1}{\sqrt{\tau_T}} & 0 \\ 0 & \frac{1}{\sqrt{1-\tau_T}} \end{bmatrix} \begin{bmatrix} \iota_{T_1} & \mathbf{0}_{T_1} \\ \mathbf{0}_{T-T_1} & \iota_{T-T_1} \end{bmatrix}' \\
&= A_j B_j^{-\frac{1}{2}} L_j S_j R_j' \begin{bmatrix} \frac{1}{\sqrt{\tau_T}}\iota_{T_1} & \mathbf{0}_{T_1} \\ \mathbf{0}_{T-T_1} & \frac{1}{\sqrt{1-\tau_T}}\iota_{T-T_1} \end{bmatrix}' \\
&= \left(A_j B_j^{-\frac{1}{2}} L_j\right)\left(\sqrt{T} S_j\right)\left\{\frac{1}{\sqrt{T}} R_j' \begin{bmatrix} \frac{1}{\sqrt{\tau_T}}\iota_{T_1} & \mathbf{0}_{T_1} \\ \mathbf{0}_{T-T_1} & \frac{1}{\sqrt{1-\tau_T}}\iota_{T-T_1} \end{bmatrix}'\right\} := \mathcal{U}_j \Sigma_j \mathcal{V}_j',
\end{aligned}$$

where $\mathcal{U}_j = A_j B_j^{-\frac{1}{2}} L_j \in R^{N\times 2}$, $\mathcal{V}_j = \frac{1}{\sqrt{T}} \begin{bmatrix} \frac{1}{\sqrt{\tau_T}} \iota_{T_1} & \mathbf{0}_{T_1} \\ \mathbf{0}_{T-T_1} & \frac{1}{\sqrt{1-\tau_T}} \iota_{T-T_1} \end{bmatrix} R_j \in R^{T\times 2}$, and $\Sigma_j = \sqrt{T} S_j \in R^{2\times 2}$. It is easy to verify that

$$\begin{aligned} \mathcal{U}_j' \mathcal{U}_j &= L_j' B_j^{-\frac{1}{2}} A_j' A_j B_j^{-\frac{1}{2}} L_j = L_j' B_j^{-\frac{1}{2}} B_j B_j^{-\frac{1}{2}} L_j = L_j' L_j = I_2 \quad \text{and} \\ \mathcal{V}_j' \mathcal{V}_j &= R_j' R_j = I_2. \end{aligned}$$

Now, let $U_j = \mathcal{U}_j \Sigma_j / \sqrt{T}$ and $V_j = \sqrt{T} \mathcal{V}_j$. We have $\Theta_j^0 = U_j V_j^\top$ and $V_j = \begin{bmatrix} \frac{1}{\sqrt{\tau_T}} \iota_{T_1} & \mathbf{0}_{T_1} \\ \mathbf{0}_{T-T_1} & \frac{1}{\sqrt{1-\tau_T}} \iota_{T-T_1} \end{bmatrix} R_j = D_j R_j$. This proves (i).

(ii) Given R_j is an orthonormal matrix, this follows from (i) automatically. ■

A.2 Proof of Theorem 4.1

A.2.1 Proof of Statement (i).

Let $\mathcal{R}(C_1) := \left\{ \{\Delta_{\Theta_j}\}_{j\in[p]\cup\{0\}} \in \mathbb{R}^{N\times T} : \sum_{j\in[p]\cup\{0\}} \left\| \mathcal{P}_j^{\perp}(\Delta_{\Theta_j}) \right\|_* \leq C_1 \sum_{j\in[p]\cup\{0\}} \left\| \mathcal{P}_j(\Delta_{\Theta_j}) \right\|_* \right\}$. By Lemma B.4, we notice that

$$\mathbb{P}\left\{ \left\{ \tilde{\Delta}_{\Theta_j} \right\}_{j\in[p]\cup\{0\}} \in \mathcal{R}(3) \right\} \to 1.$$

Recall from (4.1) that

$$\mathcal{R}(C_1, C_2) := \left\{ (\{\Delta_{\Theta_j}\}_{j\in[p]\cup\{0\}}) : \sum_{j\in[p]\cup\{0\}} \left\| \mathcal{P}_j^{\perp}(\Delta_{\Theta_j}) \right\|_* \leq C_1 \sum_{j\in[p]\cup\{0\}} \left\| \mathcal{P}_j(\Delta_{\Theta_j}) \right\|_* \sum_{j\in[p]\cup\{0\}} \left\| \Theta_j \right\|_F^2 \geq C_2 \sqrt{NT} \right\}.$$

When $\left\{ \tilde{\Delta}_{\Theta_j} \right\}_{j\in[p]\cup\{0\}} \in \mathcal{R}(3)$ and $\left\{ \tilde{\Delta}_{\Theta_j} \right\}_{j\in[p]\cup\{0\}} \notin \mathcal{R}(3, C_2)$, we have

$$\sum_{j\in[p]\cup\{0\}} \left\| \tilde{\Delta}_{\Theta_j} \right\|_F^2 < C_2 \sqrt{NT},$$

which gives

$$\frac{1}{\sqrt{NT}} \left\| \tilde{\Delta}_{\Theta_j} \right\|_F < \frac{C_2}{\sqrt{N \wedge T}}, \quad \forall j \in [p] \cup \{0\}.$$

So it suffices to focus on the case that $\left\{ \tilde{\Delta}_{\Theta_j} \right\}_{j\in[p]\cup\{0\}} \in \mathcal{R}(3, C_2)$.

Define the event

$$\mathscr{A}_{1,N}(c_3) = \left\{ \|E\|_{op} \leq c_3 \left(\sqrt{N} \vee \sqrt{T \log T} \right), \|X_j \odot E\|_{op} \leq c_3 \left(\sqrt{N} \vee \sqrt{T \log T} \right), \forall j \in [p] \right\}$$

for some positive constant c_3. By Lemma B.3, we have $\mathbb{P}(\mathscr{A}_{1,N}^c(c_3)) \leq \epsilon$ for any $\epsilon > 0$. By the definition of $\left\{ \tilde{\Theta}_j \right\}_{j\in[p]\cup\{0\}}$ in (3.1), we have

$$\begin{aligned} \sum_{j\in[p]\cup\{0\}} \nu_j \left(\left\| \Theta_j^0 \right\|_* - \left\| \tilde{\Theta}_j \right\|_* \right) &\geq \frac{1}{NT} \left\| Y - \tilde{\Theta}_0 - \sum_{j\in[p]} X_j \odot \tilde{\Theta}_j \right\|_F^2 - \frac{1}{NT} \left\| Y - \Theta_0^0 - \sum_{j\in[p]} X_j \odot \Theta_j^0 \right\|_F^2 \\ &= \frac{1}{NT} \left\| \tilde{\Delta}_{\Theta_0} + \sum_{j\in[p]} \tilde{\Delta}_{\Theta_j} \odot X_j \right\|_F^2 - \frac{2}{NT} tr\left(\tilde{\Delta}_{\Theta_0}' E \right) - \frac{2}{NT} \sum_{j\in[p]} tr\left[\tilde{\Delta}_{\Theta_0}' (E \odot X_j) \right]. \end{aligned}$$

Then conditioning on the event $\mathscr{A}_{1,N}(c_3)$, we have

$$
\begin{aligned}
&\frac{1}{NT}\left\|\tilde{\Delta}_{\Theta_0}+\sum_{j\in[p]}\tilde{\Delta}_{\Theta_j}\odot X_j\right\|_F^2\\
&\leq \frac{2}{NT}tr\left(\tilde{\Delta}_{\Theta_0}'E\right)+\frac{2}{NT}\sum_{j\in[p]}tr\left[\tilde{\Delta}_{\Theta_0}'\left(E\odot X_j\right)\right]+\sum_{j\in[p]\cup\{0\}}\nu_j\left(\left\|\Theta_j^0\right\|_*-\left\|\tilde{\Theta}_j\right\|_*\right)\\
&\leq 2c_3\frac{\sqrt{N}\vee\sqrt{T\log T}}{NT}\sum_{j\in[p]\cup\{0\}}\left\|\tilde{\Delta}_{\Theta_j}\right\|_*+\sum_{j\in[p]\cup\{0\}}\nu_j\left(\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_*-\left\|\mathcal{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_*\right)\\
&=2c_3\frac{\sqrt{N}\vee\sqrt{T\log T}}{NT}\sum_{j\in[p]\cup\{0\}}\left(\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_*+\left\|\mathcal{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_*\right)\\
&+4c_3\frac{\sqrt{N}\vee\sqrt{T\log T}}{NT}\sum_{j\in[p]\cup\{0\}}\left(\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_*-\left\|\mathcal{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_*\right)\\
&=6c_3\frac{\sqrt{N}\vee\sqrt{T\log T}}{NT}\sum_{j\in[p]\cup\{0\}}\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_*-2c_3\frac{\sqrt{N}\vee\sqrt{T\log T}}{NT}\sum_{j\in[p]\cup\{0\}}\left\|\mathcal{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_*\\
&\leq 6c_3\frac{\sqrt{N}\vee\sqrt{T\log T}}{NT}\sum_{j\in[p]\cup\{0\}}\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_*, \qquad \text{(A.1)}
\end{aligned}
$$

where the second inequality holds by the definition of event $\mathscr{A}_1(c_3)$, the fact that $|\mathrm{tr}\,(AB)|\leq\|A\|_*\|B\|_{op}$, and (B.9), the first equality holds by the fact that $\left\|\tilde{\Delta}_{\Theta_j}\right\|_*=\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_*+\left\|\mathcal{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_*$ (see, e.g., Lemma D.2(i) in Chernozhukov et al. (2020)) and that $\nu_j=\frac{4c_3\left(\sqrt{N}\vee\sqrt{T\log T}\right)}{NT}$. It follows that

$$
\begin{aligned}
C_3\sum_{j\in[p]\cup\{0\}}\left\|\tilde{\Delta}_{\Theta_j}\right\|_F^2 &\leq \left\|\tilde{\Delta}_{\Theta_0}+\sum_{j\in[p]}\tilde{\Delta}_{\Theta_j}\odot X_j\right\|_F^2+C_4\left(N+T\right)\\
&\leq 6c_3\left(\sqrt{N}\vee\sqrt{T\log T}\right)\sum_{j\in[p]\cup\{0\}}\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_*+C_4\left(N+T\right)\\
&\leq 12\bar{r}c_3\left(\sqrt{N}\vee\sqrt{T\log T}\right)\sum_{j\in[p]\cup\{0\}}\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_F+C_4\left(N+T\right)\\
&\leq 12\bar{r}c_3\left(\sqrt{N}\vee\sqrt{T\log T}\right)\sum_{j\in[p]\cup\{0\}}\left\|\tilde{\Delta}_{\Theta_j}\right\|_F+C_4\left(N+T\right)\\
&\leq 12\bar{r}c_3\left(\sqrt{N}\vee\sqrt{T\log T}\right)\sqrt{p+1}\sqrt{\sum_{j\in[p]\cup\{0\}}\left\|\tilde{\Delta}_{\Theta_j}\right\|_F^2}+C_4\left(N+T\right),
\end{aligned}
$$

where the first inequality holds by Assumption 4, the second inequality is by (A.1), the third inequality is by the fact that $\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_*\leq rank(\mathcal{P}_j(\tilde{\Delta}_{\Theta_j}))\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_F$ with $rank(\mathcal{P}_j(\tilde{\Delta}_{\Theta_j}))\leq 2\bar{r}$ by Lemma D.2(ii) in Chernozhukov et al. (2020), the fourth inequality is by the fact that $\left\|\tilde{\Delta}_{\Theta_j}\right\|_F=\left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_F+\left\|\mathcal{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_F$ (see, e.g., Lemma D.2(ii) in Chernozhukov et al. (2020)), and the last inequality holds by Jensen inequality. Consequently, we can conclude that

$$
\sum_{j\in[p]\cup\{0\}}\left\|\tilde{\Delta}_{\Theta_j}\right\|_F^2=O_p\left(N\vee\left(T\log T\right)\right).
$$

In addition, $\max_{k\in[r_j]}|\tilde{\sigma}_{k,j}-\sigma_{k,j}|=O_p(\eta_{N,1})$ by the Weyl's inequality with $\eta_{N,1}=\frac{\sqrt{\log T}}{\sqrt{N\wedge T}}$.

Now, we show the convergence rate for the singular vector estimates. For $\forall j\in[p]\cup\{0\}$, let $\tilde{D}_j=\frac{1}{NT}\tilde{\Theta}_j'\tilde{\Theta}_j=\hat{\tilde{\mathcal{V}}}_j\hat{\tilde{\Sigma}}_j\hat{\tilde{\mathcal{V}}}_j'$, and $D_j^0=\frac{1}{NT}\Theta_j^{0\prime}\Theta_j^0=\mathcal{V}_j^0\Sigma_j^0\mathcal{V}_j^{0\prime}$. Define the event

$$\mathscr{A}_{2,N}(M)=\left\{\frac{1}{\sqrt{NT}}\left\|\tilde{\Theta}_j-\Theta_j^0\right\|_F\leq M\eta_{N,1},\quad\forall j\in\{0,\cdots,p\}\right\}$$

for a large enough constant M. By the above analyses, we have $\mathbb{P}\left(\mathscr{A}_{2,N}^c(M)\right)\leq\epsilon$ for any $\epsilon>0$. On the event $\mathscr{A}_{2,N}(M)$, we observe that

$$\left\|\tilde{D}_j-D_j^0\right\|_F\leq\frac{1}{NT}\left(\left\|\tilde{\Theta}_j\right\|_F+\|\Theta_j^0\|_F\right)\left\|\tilde{\Theta}_j-\Theta_j^0\right\|_F\leq 2M^2\eta_{N,1}.$$

With Lemma C.1 of Su et al. (2020) and Davis-Kahan sinΘ theorem in Yu et al. (2015), we are ready to show that for some orthogonal matrix O_j,

$$\begin{aligned}\left\|\mathcal{V}_j^0-\hat{\tilde{\mathcal{V}}}_jO_j\right\|_F&\leq\sqrt{r_j}\left\|\mathcal{V}_j^0-\hat{\tilde{\mathcal{V}}}_jO_j\right\|_{op}\leq\sqrt{r_j}\frac{2\sqrt{2}M^2\eta_{N,1}}{\sigma_{K_j,1}^2-2M^2\eta_{N,1}}\\&\leq\sqrt{r_j}\frac{2\sqrt{2}M^2\eta_{N,1}}{c_\sigma^2-2M^2\eta_{N,1}}\leq\sqrt{r_j}\frac{2\sqrt{2}M^2\eta_{N,1}}{C_6c_\sigma^2}\leq C_7\eta_{N,1},\end{aligned}\tag{A.2}$$

for $C_7=\frac{2\sqrt{2}M^2\sqrt{\bar{r}}}{C_6c_\sigma^2}$, where the second inequality in line two is due to the fact that $\eta_{N,1}$ is sufficiently small and C_6 is some positive constant.

Then $\left\|V_j^0-\tilde{V}_jO_j\right\|_F\leq C_7\sqrt{T}\eta_{N,1}$ by the definition of $\tilde{V}_j$ and V_j. Together with the fact that $\mathbb{P}(\mathscr{A}_{2,N}^c(M))\to 0$ by Theorem 4.1(i), it implies $\frac{1}{\sqrt{T}}\left\|V_j^0-\tilde{V}_jO_j\right\|_F=O_p\left(T\eta_{N,1}\right).$

A.2.2 Proof of Statement (ii).

Define

$$u_i^0=\left[u_{i,0}^{0\prime},\cdots,u_{i,p}^{0\prime}\right]',\quad\dot{\Delta}_{i,j}=O_j'\dot{u}_{i,j}-u_{i,j}^0,\quad\dot{\Delta}_{i,u}=\left(\dot{\Delta}_{i,0}',\cdots,\dot{\Delta}_{i,p}'\right)',$$

$$\tilde{\phi}_{it}=\left[\left(O_0'\tilde{v}_{t,0}\right)',\left(O_1'\tilde{v}_{t,1}X_{1,it}\right)',\cdots,\left(O_p'\tilde{v}_{t,p}X_{p,it}\right)'\right]',\quad\text{and}\quad\tilde{\Phi}_i=\frac{1}{T}\sum_{t=1}^T\tilde{\phi}_{it}\tilde{\phi}_{it}'.$$

Let $\tilde{Y}_{it}:=Y_{it}-\left(O_0u_{i,0}^0\right)'\tilde{v}_{t,0}-\sum_{j=1}^p\left(O_ju_{i,j}^0\right)'\tilde{v}_{t,j}X_{j,it}$. By the definition of $\{\dot{u}_{i,j}\}$ in (3.2), we have

$$\begin{aligned}0&\geq\frac{1}{T}\sum_{t\in[T]}\left(Y_{it}-\dot{u}_{i,0}'\tilde{v}_{t,0}-\sum_{j=1}^p\dot{u}_{i,j}'\tilde{v}_{t,j}X_{j,it}\right)^2-\frac{1}{T}\sum_{t\in[T]}\tilde{Y}_{it}^2\\&=\frac{1}{T}\sum_{t\in[T]}\left(\tilde{Y}_{it}-\left(\dot{u}_{i,0}-O_0u_{i,0}^0\right)'\tilde{v}_{t,0}-\sum_{j\in[p]}\left(\dot{u}_{i,j}-O_ju_{i,j}^0\right)'\tilde{v}_{t,j}X_{j,it}\right)^2-\frac{1}{T}\sum_{t\in[T]}\tilde{Y}_{it}^2\\&=\frac{1}{T}\sum_{t\in[T]}\left[\left(\dot{\Delta}_{i,u}'\tilde{\phi}_{it}\right)^2-2\left(\dot{\Delta}_{i,u}'\tilde{\phi}_{it}\right)\left(Y_{it}-u_i^{0\prime}\tilde{\phi}_{it}\right)\right],\end{aligned}$$

which implies

$$\left\|\dot{\Delta}_{i,u}\right\|_2^2\lambda_{\min}\left(\frac{1}{T}\sum_{t\in[T]}\tilde{\phi}_{it}\tilde{\phi}_{it}'\right)\leq\frac{1}{T}\sum_{t\in[T]}\left(\dot{\Delta}_{i,u}'\tilde{\phi}_{it}\right)^2\leq\frac{2}{T}\sum_{t\in[T]}\dot{\Delta}_{i,u}'\tilde{\phi}_{it}\left(Y_{it}-u_i^{0\prime}\tilde{\phi}_{it}\right)$$

$$= \frac{2}{T} \sum_{t \in [T]} \dot{\Delta}'_{i,u} \tilde{\phi}_{it} \left[e_{it} - u_i^{0\prime} \left(\tilde{\phi}_{it} - \phi_{it}^0 \right) \right]$$

$$= 2 \left\{ \frac{1}{T} \sum_{t \in [T]} \phi_{it}^0 e_{it} \right\}' \dot{\Delta}_{i,u} + 2 \left\{ \frac{1}{T} \sum_{t \in [T]} \left(\tilde{\phi}_{it} - \phi_{it}^0 \right) e_{it} \right\}' \dot{\Delta}_{i,u} - \frac{2}{T} \sum_{t \in [T]} \tilde{\phi}'_{it} \dot{\Delta}_{i,u} \left[u_i^{0\prime} \left(\tilde{\phi}_{it} - \phi_{it}^0 \right) \right]$$

$$:= 2G'_{i,1} \dot{\Delta}_{i,u} + 2G_{i,2} - 2G_{i,3}. \tag{A.3}$$

We first deal with $G_{1,i}$. Conditional on $\mathscr{D}$, the randomness in $G_{1,i}$ comes from $\{e_{it}, X_{it}\}_{t \in [T]}$, which is the strong mixing sequence by Assumption 1(iii). Besides, we observe that

$$\max_{i \in [N], t \in [T]} \left\| Var \left(\phi_{it}^0 e_{it} \middle| \mathscr{D} \right) \right\|_F \lesssim \max_{i \in [N], t \in [T]} \left[\mathbb{E} \left(e_{it}^2 \middle| \mathscr{D} \right) + \sum_{j \in [p]} \mathbb{E} \left(X_{j,it}^2 e_{it}^2 \middle| \mathscr{D} \right) \right] = O(1) \ a.s$$

combining Lemma B.7(ii), Assumption 1(v). Following similar arguments, we have

$$\max_{i \in [N], t \in [T]} \sum_{s=t+1}^{T} \left\| Cov \left(\phi_{it}^0 e_{it}, \phi_{is}^0 e_{is} \middle| \mathscr{D} \right) \right\|_F$$

$$\leq 8 \max_{t \in [T]} \sum_{s=t+1}^{T} \left[\mathbb{E} \left(\left\| \phi_{it}^0 e_{it} \right\|_2^q \middle| \mathscr{D} \right) \right]^{1/q} \left[\left\| \mathbb{E} \left(\phi_{is}^0 e_{is} \right\|_2^q \middle| \mathscr{D} \right) \right]^{1/q} (\alpha(t-s))^{1-2/q}$$

$$= O(1) \ a.s,$$

where the first inequality is by Davydov's inequality, saying that

$$|Cov\left[a\left(x_t\right), a\left(x_s\right)\right]| \leq 8 \left[\mathbb{E} \left\|a\left(x_t\right)\right\|^p\right]^{\frac{1}{p}} \left[\mathbb{E} \left\|a\left(x_s\right)\right\|^q\right]^{\frac{1}{q}} \alpha(t-s)^{\frac{1}{r}}$$

for any strong mixing sequence $(x_t, t \in [T])$ with mixing coefficient $\alpha(\cdot)$ and $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$. See also in Lemma A.4, Su and Chen (2013).

Following this, for some constant C_8, we have

$$\max_{i \in [N], t \in [T]} \left[\left\| Var \left(\phi_{it}^0 e_{it} \middle| \mathscr{D} \right) \right\|_F + 2 \sum_{s=t+1}^{T} \left\| Cov \left(\phi_{it}^0 e_{it}, \phi_{is}^0 e_{is} \middle| \mathscr{D} \right) \right\|_F \right] \leq C_8,$$

and $\max_{i \in [N], t \in [T]} \left\| \phi_{it}^0 e_{it} \right\|_{\max} \leq C_8 (NT)^{1/q}$ by Lemma B.7(i) and Assumption 1(iv). Define $\mathscr{A}_{3,N}(M) = \left\{ \max_{i \in [N], t \in [T]} \left\| \phi_{it}^0 e_{it} \right\|_2 \leq M(NT)^{1/q} \right\}$ and $\mathscr{A}_{3,N,i}(M) = \left\{ \max_{t \in [T]} \left\| \phi_{it}^0 e_{it} \right\|_2 \leq M(NT)^{1/q} \right\}$ for a large enough constant M. For a positive constant C_9, it yields that

$$\mathbb{P} \left(\max_{i \in [N]} \frac{1}{T} \left\| \sum_{t \in [T]} \phi_{it}^0 e_{it} \right\|_2 > C_9 \sqrt{\frac{\log N}{T}} (NT)^{\frac{1}{q}} \right)$$

$$\leq \mathbb{P} \left(\max_{i \in [N]} \frac{1}{T} \left\| \sum_{t \in [T]} \phi_{it}^0 e_{it} \right\|_2 > C_9 \sqrt{\frac{\log N}{T}} (NT)^{\frac{1}{q}}, \mathscr{A}_{3,N}(M) \right) + \mathbb{P} \left(\mathscr{A}_{3,N}^c(M) \right)$$

$$\leq \sum_{i \in [N]} \mathbb{P} \left(\frac{1}{T} \left\| \sum_{t \in [T]} \phi_{it}^0 e_{it} \right\|_2 > C_9 \sqrt{\frac{\log N}{T}} (NT)^{\frac{1}{q}}, \mathscr{A}_{3,N}(M) \right) + \mathbb{P} \left(\mathscr{A}_{3,N}^c(M) \right)$$

$$
\begin{aligned}
&\leq \sum_{i\in[N]} \mathbb{P}\left(\frac{1}{T}\left\|\sum_{t\in[T]}\phi_{it}^0 e_{it}\right\|_2 > C_9\sqrt{\frac{\log N}{T}}(NT)^{\frac{1}{q}}, \mathscr{A}_{3,N,i}(M)\right) + \mathbb{P}\left(\mathscr{A}_{3,N}^c(M)\right) \\
&\leq \sum_{i\in[N]} \exp\left(-\frac{c_4 C_9^2 T\log N (NT)^{2/q}}{TC_8 + C_8^2(NT)^{2/q} + C_8C_9(NT)^{2/q}\sqrt{T\log N}(\log T)^2}\right) + o(1) \\
&= o(1)
\end{aligned}
$$

where the last inequality holds by Bernstein's inequality in Lemma B.6(ii) and the fact that $\mathbb{P}\left(\mathscr{A}_{3,N}^c(M)\right) = o(1)$. It follows that

$$
\max_{i\in[N]}\frac{\left|G_{i,1}'\dot{\Delta}_{i,u}\right|}{\left\|\dot{\Delta}_{i,u}\right\|_2} \leq \max_{i\in[N]}\|G_{i,1}\|_2 = O_p\left(\sqrt{\frac{\log N}{T}}(NT)^{\frac{1}{q}}\right). \tag{A.4}
$$

For $G_{i,2}$, we notice that

$$
\begin{aligned}
\max_{i\in[N]}\frac{|G_{i,2}|}{\left\|\dot{\Delta}_{i,u}\right\|_2} &= \max_{i\in[N]}\frac{\left|\left\{\frac{1}{T}\sum_{t\in[T]}\left(\tilde{\phi}_{it}-\phi_{it}^0\right)e_{it}\right\}'\dot{\Delta}_{i,u}\right|}{\left\|\dot{\Delta}_{i,u}\right\|_2} \leq \max_{i\in[N]}\left\|\frac{1}{T}\sum_{t\in[T]}\left(\tilde{\phi}_{it}-\phi_{it}^0\right)e_{it}\right\|_2 \\
&\leq \max_{i\in[N]}\sqrt{\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\phi}_{it}-\phi_{it}^0\right\|_2^2}\max_{i\in[N]}\sqrt{\frac{1}{T}\sum_{t\in[T]}|e_{it}|^2} = O_p(\eta_{N,1}(NT)^{1/q}),
\end{aligned} \tag{A.5}
$$

where the second inequality holds by Cauchy's inequality and the last equality is by Lemma B.7(iv) and Assumption 1(iv)

For $G_{i,3}$, we have

$$
\begin{aligned}
\max_{i\in[N]}\frac{|G_{i,3}|}{\left\|\dot{\Delta}_{i,u}\right\|_2} &= \max_{i\in[N]}\frac{\left|\frac{1}{T}\sum_{t\in[T]}\tilde{\phi}_{it}'\dot{\Delta}_{i,u}\left[u_i^{0\prime}\left(\tilde{\phi}_{it}-\phi_{it}^0\right)\right]\right|}{\left\|\dot{\Delta}_{i,u}\right\|_2} \\
&\leq \max_{i\in[N]}\sqrt{\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\phi}_{it}\right\|_2^2}\max_{i\in[N]}\left\|u_i^0\right\|_2\max_{i\in[N],t\in[T]}\sqrt{\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\phi}_{it}-\phi_{it}^0\right\|_2^2} \\
&= O_p(\eta_{N,1}(NT)^{1/q}),
\end{aligned} \tag{A.6}
$$

where the inequality holds by Cauchy's inequality and the last line is by Lemma B.7(i) and (iv).

Combining (A.3)-(A.6) and Lemma B.8 yields

$$
\max_{i\in[N]}\left\|\dot{u}_{i,j}-O_{i,j}^{(1)}u_{i,j}^0\right\|_2 \leq \max_{i\in[N]}\left\|\dot{\Delta}_{i,u}\right\|_2 = O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}(NT)^{1/q}\right).
$$

The union bound of $\dot{v}_{t,j}$ can be obtained in the same manner and we sketch the proof here. Define

$$
\begin{aligned}
&v_t^0 = \left(v_{t,0}^{0\prime},\cdots,v_{t,p}^{0\prime}\right)', \quad \dot{\Delta}_{t,j} = O_j'\dot{v}_{t,j} - v_{t,j}^0, \quad \dot{\Delta}_{t,v} = \left(\dot{\Delta}_{t,0}',\cdots,\dot{\Delta}_{t,p}'\right)', \\
&\dot{\psi}_{it} = \left[\left(O_0'\dot{u}_{i,0}\right)',\left(O_1'\dot{u}_{i,1}X_{1,it}\right)',\cdots,\left(O_p'\dot{u}_{i,p}X_{p,it}\right)'\right]', \quad \text{and} \quad \dot{\Psi}_t = \frac{1}{N}\sum_{i\in[N]}\dot{\psi}_{it}\dot{\psi}_{it}'.
\end{aligned}
$$

Following the steps to derive (A.3), we can also obtain

$$\left\|\dot{\Delta}_{t,v}\right\|_2^2 \lambda_{\min}\left(\frac{1}{N}\sum_{i\in[N]}\dot{\psi}_{it}\dot{\psi}_{it}'\right)$$
$$= 2\left\{\frac{1}{N}\sum_{i\in[N]}\psi_{it}^0 e_{it}\right\}'\dot{\Delta}_{t,v} + \frac{2}{N}\sum_{i\in[N]}\left(\dot{\psi}_{it}-\psi_{it}^0\right)'\dot{\Delta}_{t,v}e_{it} - \frac{2}{N}\sum_{i\in[N]}\dot{\psi}_{it}'\dot{\Delta}_{t,v}\left[v_t^{0\prime}\left(\dot{\psi}_{it}-\psi_{it}^0\right)\right]. \tag{A.7}$$

By the fact that

$$\max_{t\in[T]}\frac{1}{N}\sum_{i\in[N]}\left\|\dot{\psi}_{it}\right\|_2^2 = \max_{t\in[T]}\frac{1}{N}\sum_{i\in[N]}\left(\|\dot{u}_{i,0}\|_2^2 + \sum_{j\in[p]}\|\dot{u}_{i,j}\|_2^2 X_{j,it}^2\right)$$
$$\leq \max_{i\in[N]}\|\dot{u}_{i,0}\|_2^2 + \max_{i\in[N],j\in[p]}\|\dot{u}_{i,j}\|_2^2\sum_{j\in[p]}\max_{t\in[T]}\frac{1}{N}\sum_{i\in[N]}X_{j,it}^2 = O_p(1),$$
$$\max_{t\in[T]}\frac{1}{N}\sum_{i\in[N]}\left\|\dot{\psi}_{it}-\psi_{it}^0\right\|_2^2 \leq \max_{i\in[N]}\left\|\dot{u}_{i,0}-u_{i,0}^0\right\|_2^2 + \max_{i\in[N],j\in[p]}\left\|\dot{u}_{i,j}-u_{i,j}^0\right\|_2^2\sum_{j\in[p]}\max_{t\in[T]}\frac{1}{N}\sum_{i\in[N]}X_{j,it}^2 = O_p(\eta_{N,2}),$$

where $\eta_{N,2} = \sqrt{\frac{\log N\vee T}{N\wedge T}}(NT)^{1/q}$ and the first inequality is by Lemma B.7(i), we obtain that

$$\max_{t\in[T]}\frac{\left|\left\{\frac{1}{N}\sum_{i\in[N]}\psi_{it}^0 e_{it}\right\}'\dot{\Delta}_{t,v}\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2} = O_p\left(\sqrt{\frac{\log T}{N}}(NT)^{\frac{1}{q}}\right),$$
$$\max_{t\in[T]}\frac{\left|\frac{1}{N}\sum_{i\in[N]}\dot{\psi}_{it}'\dot{\Delta}_{t,v}\left[v_t^{0\prime}\left(\dot{\psi}_{it}-\psi_{it}^0\right)\right]\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2} = O_p(\eta_{N,2}), \text{ and}$$
$$\max_{t\in[T]}\frac{\left|\frac{1}{N}\sum_{i\in[N]}\left(\dot{\psi}_{it}-\psi_{it}^0\right)'\dot{\Delta}_{t,v}e_{it}\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2} = O_p(\eta_{N,2}),$$

where the first line is by conditional Bernstein's inequality for i.i.d. sequence and the last two lines are by the analogous arguments in (A.5) and (A.6). It follows that

$$\max_{t\in[T]}\left\|\dot{v}_{t,j}-O_j v_{t,j}^0\right\|_2 = O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}(NT)^{1/q}\right). \quad \blacksquare$$

A.2.3 Proof of Statement (iii).

For $\forall j\in[p]\cup\{0\}$, $i\in[N]$ and $t\in[T]$, we can show that

$$\dot{\Theta}_{j,it}-\Theta_{j,it}^0 = \dot{u}_{i,j}'\dot{v}_{t,j} - u_{i,j}^{0\prime}v_{t,j}^0$$
$$= \left(\dot{u}_{i,j}-O_j u_{i,j}^0\right)'\left(\dot{v}_{t,j}-O_j v_{t,j}^0\right) + O_j u_{i,j}^{0\prime}\left(\dot{v}_{t,j}-O_j v_{t,j}^0\right) + O_j v_{t,j}^{0\prime}\left(\dot{u}_{i,j}-O_j u_{i,j}^0\right),$$

which implies

$$\max_{i\in[N],t\in[T]}\left|\dot{\Theta}_{j,it}-\Theta_{j,it}^0\right| \leq \max_{i\in[N]}\left\|\dot{u}_{i,j}-O_j u_{i,j}^0\right\|_2 \max_{t\in[T]}\left\|\dot{v}_{t,j}-O_j v_{t,j}^0\right\|_2$$

$$
\begin{aligned}
&+ \max_{i\in[N]} \left\|O_j u^0_{i,j}\right\|_2 \max_{t\in[T]} \left\|\dot{v}_{t,j} - O_j v^0_{t,j}\right\|_2 + \max_{i\in[N]} \left\|\dot{u}_{i,j} - O_j u^0_{i,j}\right\|_2 \max_{t\in[T]} \left\|O_j v^0_{t,j}\right\|_2 \\
&= O_p(\eta_{N,2}),
\end{aligned}
$$

where the last equality combines results from Theorem 4.1(ii) and Lemma B.7(i). ■

A.3 Proof of Theorem 4.2

To prove $\mathbb{P}\left(\hat{T}_1 = T_1\right) \to 1$, it suffices to show: (i) $\mathbb{P}\left(\hat{T}_1 < T_1\right) \to 0$ and (ii) $\mathbb{P}\left(\hat{T}_1 > T_1\right) \to 0$.

First, we focus on (i). Let $\Delta_{it}(j) = \dot{\Theta}_{j,it} - \Theta^0_{j,it}$, $\bar{\Delta}_{s,i}(j) = \frac{1}{s}\sum_{t=1}^{s}\left(\dot{\Theta}_{j,it} - \Theta^0_{j,it}\right)$ and $\bar{\Delta}_{s_+,i}(j) = \frac{1}{T-s}\sum_{t=s+1}^{T}\left(\dot{\Theta}_{j,it} - \Theta^0_{j,it}\right)$. When $s < T_1$, we have

$$
\begin{aligned}
\bar{\dot{\Theta}}^{(1s)}_{j,i} &= \frac{1}{s}\sum_{t=1}^{s}\dot{\Theta}_{j,it} = \frac{1}{s}\sum_{t=1}^{s}\left[\Theta^0_{j,it} + \left(\dot{\Theta}_{j,it} - \Theta^0_{j,it}\right)\right] = \alpha^{(1)}_{g_i^{(1)},j} + \bar{\Delta}_{s,i}(j), \\
\bar{\dot{\Theta}}^{(2s)}_{j,i} &= \frac{1}{T-s}\sum_{t=s+1}^{T}\dot{\Theta}_{j,it} = \frac{1}{T-s}\sum_{t=s+1}^{T}\left[\Theta^0_{j,it} + \left(\dot{\Theta}_{j,it} - \Theta^0_{j,it}\right)\right] \\
&= \frac{T_1 - s}{T-s}\alpha^{(1)}_{g_i^{(1)},j} + \frac{T-T_1}{T-s}\alpha^{(2)}_{g_i^{(2)},j} + \bar{\Delta}_{s_+,i}(j),
\end{aligned}
$$

with $\alpha^{(1)}_{g_i^{(1)},j}$ and $\alpha^{(2)}_{g_i^{(2)},j}$ being the j-th element of $\alpha^{(1)}_{g_i^{(1)}}$ and $\alpha^{(2)}_{g_i^{(2)}}$, respectively. It yields

$$
\begin{aligned}
\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}^{(1s)}_{j,i} =& \dot{\Theta}_{j,it} - \alpha^{(1)}_{g_i^{(1)},j} - \bar{\Delta}_{s,i}(j) = \Delta_{it}(j) - \bar{\Delta}_{s,i}(j), \quad t \le s, \text{ and} \\
\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}^{(2s)}_{j,i} =& \dot{\Theta}_{j,it} - \frac{T_1 - s}{T-s}\alpha^{(1)}_{g_i^{(1)},j} - \frac{T-T_1}{T-s}\alpha^{(2)}_{g_i^{(2)},j} - \bar{\Delta}_{s_+,i}(j) \\
=& \begin{cases} \frac{T-T_1}{T-s}\left(\alpha^{(1)}_{g_i^{(1)},j} - \alpha^{(2)}_{g_i^{(2)},j}\right) + \Delta_{it}(j) - \bar{\Delta}_{s_+,i}(j) & \text{if } s+1 \le t \le T_1 \\ \frac{T_1-s}{T-s}\left(\alpha^{(2)}_{g_i^{(2)},j} - \alpha^{(1)}_{g_i^{(1)},j}\right) + \Delta_{it}(j) - \bar{\Delta}_{s_+,i}(j) & \text{if } T_1+1 \le t \le T \end{cases}.
\end{aligned}
$$

Then, we have

$$
\sum_{t=1}^{s}\left[\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}^{(1s)}_{j,i}\right]^2 = \sum_{t=1}^{s}\left[\Delta_{it}(j) - \bar{\Delta}_{s,i}(j)\right]^2,
$$

and

$$
\begin{aligned}
\sum_{t=s+1}^{T}\left[\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}^{(2s)}_{j,i}\right]^2 =& \sum_{t=s+1}^{T_1}\left[\frac{T-T_1}{T-s}\left(\alpha^{(1)}_{g_i^{(1)},j} - \alpha^{(2)}_{g_i^{(2)},j}\right) + \Delta_{it}(j) - \bar{\Delta}_{s_+,i}(j)\right]^2 \\
&+ \sum_{t=T_1+1}^{T}\left[\frac{T_1-s}{T-s}\left(\alpha^{(2)}_{g_i^{(2)},j} - \alpha^{(1)}_{g_i^{(1)},j}\right) + \Delta_{it}(j) - \bar{\Delta}_{s_+,i}(j)\right]^2 \\
=& \frac{(T_1-s)(T-T_1)^2}{(T-s)^2}\left(\alpha^{(1)}_{g_i^{(1)},j} - \alpha^{(2)}_{g_i^{(2)},j}\right)^2 + \sum_{t=s+1}^{T_1}\left[\Delta_{it}(j) - \bar{\Delta}_{s_+,i}(j)\right]^2 \\
&+ 2\frac{T-T_1}{T-s}\left(\alpha^{(1)}_{g_i^{(1)},j} - \alpha^{(2)}_{g_i^{(2)},j}\right)\sum_{t=s+1}^{T_1}\left[\Delta_{it}(j) - \bar{\Delta}_{s_+,i}(j)\right] \\
&+ \frac{(T-T_1)(T_1-s)^2}{(T-s)^2}\left(\alpha^{(2)}_{g_i^{(2)},j} - \alpha^{(1)}_{g_i^{(1)},j}\right)^2 + \sum_{t=T_1+1}^{T}\left[\Delta_{it}(j) - \bar{\Delta}_{s_+,i}(j)\right]^2
\end{aligned}
$$

$$
\begin{aligned}
&+2\frac{T_1-s}{T-s}\left(\alpha^{(2)}_{g_i^{(2)},j}-\alpha^{(1)}_{g_i^{(1)},j}\right)\sum_{t=T_1+1}^{T}\left[\Delta_{it}(j)-\bar{\Delta}_{s_+,i}(j)\right] \\
&=\frac{(T_1-s)\,(T-T_1)}{T-s}\left(\alpha^{(1)}_{g_i^{(1)},j}-\alpha^{(2)}_{g_i^{(2)},j}\right)^2+\sum_{t=s+1}^{T}\left[\Delta_{it}(j)-\bar{\Delta}_{s_+,i}(j)\right]^2 \\
&+2\frac{T-T_1}{T-s}\left(\alpha^{(1)}_{g_i^{(1)},j}-\alpha^{(2)}_{g_i^{(2)},j}\right)\sum_{t=s+1}^{T_1}\left[\Delta_{it}(j)-\bar{\Delta}_{s_+,i}(j)\right] \\
&+2\frac{T_1-s}{T-s}\left(\alpha^{(2)}_{g_i^{(2)},j}-\alpha^{(1)}_{g_i^{(1)},j}\right)\sum_{t=T_1+1}^{T}\left[\Delta_{it}(j)-\bar{\Delta}_{s_+,i}(j)\right]. \qquad \text{(A.8)}
\end{aligned}
$$

Define $L(s)=\frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\left\{\sum_{t=1}^{s}\left[\dot{\Theta}_{j,it}-\bar{\dot{\Theta}}^{(1s)}_{j,i}\right]^2+\sum_{t=s+1}^{T}\left[\dot{\Theta}_{j,it}-\bar{\dot{\Theta}}^{(2s)}_{j,i}\right]^2\right\}$. Then we have

$$
\begin{aligned}
L(s)&=\frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\frac{(T_1-s)\,(T-T_1)}{T-s}\left(\alpha^{(1)}_{g_i^{(1)},j}-\alpha^{(2)}_{g_i^{(2)},j}\right)^2 \\
&+\frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\sum_{t=1}^{s}\left[\Delta_{it}(j)-\bar{\Delta}_{s,i}(j)\right]^2+\frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\sum_{t=s+1}^{T}\left[\Delta_{it}(j)-\bar{\Delta}_{s_+,i}(j)\right]^2 \\
&+\frac{2}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\frac{T-T_1}{T-s}\left(\alpha^{(1)}_{g_i^{(1)},j}-\alpha^{(2)}_{g_i^{(2)},j}\right)\sum_{t=s+1}^{T_1}\left[\Delta_{it}(j)-\bar{\Delta}_{s_+,i}(j)\right] \\
&+\frac{2}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\frac{T_1-s}{T-s}\left(\alpha^{(2)}_{g_i^{(2)},j}-\alpha^{(1)}_{g_i^{(1)},j}\right)\sum_{t=T_1+1}^{T}\left[\Delta_{it}(j)-\bar{\Delta}_{s_+,i}(j)\right]:=\sum_{\ell=1}^{5}L_\ell(s). \qquad \text{(A.9)}
\end{aligned}
$$

Obviously,

$$
L(T_1)=L_2(T_1)+L_3(T_1). \qquad \text{(A.10)}
$$

Note that the event $\hat{T}_1<T$ implies that there exists an $s<T_1$ such that $L(s)-L(T_1)<0$, which means we can prove (i) by showing that $\mathbb{P}\left(\exists s<T_1, L(s)-L(T_1)<0\right)\to 0$. By (A.9) and (A.10), we observe that

$$
\begin{aligned}
L(s)-L(T_1)&=L_1(s)+[L_2(s)-L_2(T_1)]+[L_3(s)-L_3(T_1)]+L_4(s)+L_5(s) \\
&:=A_1(s)+A_2(s)+A_3(s)+A_4(s)+A_5(s). \qquad \text{(A.11)}
\end{aligned}
$$

Recall that $\eta_{N,2}=\sqrt{\frac{\log N\vee T}{N\wedge T}}(NT)^{1/q}$. Let $\frac{T_1-s}{T}=\kappa_s$ and note that $0<\frac{1}{T}\le\kappa_s\le\frac{T_1-2}{T}\asymp 1$. We analyze the five terms in (A.11) in turn.

For $A_1(s)$, we have

$$
\begin{aligned}
A_1(s)&=\frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\frac{(T_1-s)\,(T-T_1)}{T-s}\left(\alpha^{(1)}_{g_i^{(1)},j}-\alpha^{(2)}_{g_i^{(2)},j}\right)^2 \\
&=\frac{T_1-s}{T-s}\left(1-\tau_T\right)\frac{1}{pN}\sum_{j\in[p]}\sum_{i\in[N]}\left(\alpha^{(1)}_{g_i^{(1)},j}-\alpha^{(2)}_{g_i^{(2)},j}\right)^2 \\
&=\frac{T_1-s}{T-s}\left(1-\tau_T\right)\frac{1}{pN}\sum_{i\in[N]}\left\|\alpha^{(1)}_{g_i^{(1)}}-\alpha^{(2)}_{g_i^{(2)}}\right\|_2^2
\end{aligned}
$$

$$= \kappa_s \frac{(1-\tau_T)}{1-\frac{s}{T}} D_{N\alpha} = \kappa_s O(\zeta_{NT}^2), \tag{A.12}$$

where $D_{N\alpha} := \frac{1}{pN}\sum_{i\in[N]} \left\| \alpha^{(1)}_{g_i^{(1)}} - \alpha^{(2)}_{g_i^{(2)}} \right\|_2^2$ and the last equality holds by Assumption 6.

For $A_2(s)$, noting that

$$\bar{\Delta}_{T_1,i}(j) - \bar{\Delta}_{s,i}(j) = \frac{1}{T_1}\sum_{t=1}^{T_1} \Delta_{it}(j) - \frac{1}{s}\left[\sum_{t=1}^{T_1}\Delta_{it}(j) - \sum_{t=s+1}^{T_1}\Delta_{it}(j)\right] = \frac{s-T_1}{T_1 s}\sum_{t=1}^{T_1}\Delta_{it}(j) + \frac{1}{s}\sum_{t=s+1}^{T_1}\Delta_{it}(j),$$

we have

$$\begin{aligned} T_1\bar{\Delta}^2_{T_1,i}(j) - s\bar{\Delta}^2_{s,i}(j) &= (T_1-s)\bar{\Delta}^2_{T_1,i}(j) + s\left[\bar{\Delta}^2_{T_1,i}(j) - \bar{\Delta}^2_{s,i}(j)\right] \\ &= (T_1-s)\bar{\Delta}^2_{T_1,i}(j) + s\left[\bar{\Delta}_{T_1,i}(j) + \bar{\Delta}_{s,i}(j)\right]\left[\bar{\Delta}_{T_1,i}(j) - \bar{\Delta}_{s,i}(j)\right] \\ &= (T_1-s)\bar{\Delta}^2_{T_1,i}(j) + \left[\bar{\Delta}_{T_1,i}(j) + \bar{\Delta}_{s,i}(j)\right]\left[\sum_{t=s+1}^{T_1}\Delta_{it}(j) - \frac{T_1-s}{T_1}\sum_{t=1}^{T_1}\Delta_{it}(j)\right]. \end{aligned}$$

It follows that

$$\begin{aligned} A_2(s) &= \frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\left\{\sum_{t=1}^{s}\left[\Delta_{it}(j) - \bar{\Delta}_{s,i}(j)\right]^2 - \sum_{t=1}^{T_1}\left[\Delta_{it}(j) - \bar{\Delta}_{T_1,i}(j)\right]^2\right\} \\ &= \frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\left\{\sum_{t=1}^{s}\Delta^2_{it}(j) - s\bar{\Delta}^2_{s,i}(j) - \sum_{t=1}^{T_1}\Delta^2_{it}(j) + T_1\bar{\Delta}^2_{T_1,i}(j)\right\} \\ &= \frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\left\{-\sum_{t=s+1}^{T_1}\Delta^2_{it}(j) + T_1\bar{\Delta}^2_{T_1,i}(j) - s\bar{\Delta}^2_{s,i}(j)\right\} \\ &= -\kappa_s\frac{1}{pN(T_1-s)}\sum_{j\in[p]}\sum_{i\in[N]}\sum_{t=s+1}^{T_1}\Delta^2_{it}(j) + \kappa_s\frac{1}{pN}\sum_{j\in[p]}\sum_{i\in[N]}\bar{\Delta}^2_{T_1,i}(j) \\ &+ \kappa_s\frac{1}{pN}\sum_{j\in[p]}\sum_{i\in[N]}\left[\bar{\Delta}_{T_1,i}(j) + \bar{\Delta}_{s,i}(j)\right]\frac{1}{T_1-s}\sum_{t=s+1}^{T_1}\Delta_{it}(j) \\ &- \kappa_s\frac{1}{pN}\sum_{j\in[p]}\sum_{i\in[N]}\left[\bar{\Delta}_{T_1,i}(j) + \bar{\Delta}_{s,i}(j)\right]\frac{1}{T_1}\sum_{t=1}^{T_1}\Delta_{it}(j) \\ &= \kappa_s O_p\left(\eta^2_{N,2}\right) = \kappa_s o_p(\zeta^2_{NT}), \end{aligned} \tag{A.13}$$

where the second last equality holds by the fact that

$$\max_{i\in[N],t\in[T],j\in[p]} \left|\Delta_{it}(j)\right| = O_p\left(\eta_{N,2}\right) \tag{A.14}$$

from Theorem 4.1(iii) and

$$\begin{aligned} \max_{i\in[N],j\in[p]}\left|\bar{\Delta}_{s,i}(j)\right| &= \max_{i\in[N],j\in[p]}\left|\frac{1}{s}\sum_{t=1}^{s}\Delta_{it}(j)\right| \le \max_{i\in[N],t\in[T],j\in[p]}\left|\Delta_{it}(j)\right| = O_p\left(\eta_{N,2}\right), \\ \max_{i\in[N],j\in[p]}\left|\bar{\Delta}_{T_1,i}(j)\right| &= \max_{i\in[N],j\in[p]}\left|\frac{1}{T_1}\sum_{t=1}^{T_1}\Delta_{it}(j)\right| = O_p\left(\eta_{N,2}\right). \end{aligned} \tag{A.15}$$

Similarly, noting that

$$\begin{aligned}\bar{\Delta}_{T_1+,i}(j)-\bar{\Delta}_{s+,i}(j) &= \frac{1}{T-T_1}\sum_{t=T_1+1}^{T}\Delta_{it}(j)-\frac{1}{T-s}\sum_{t=s+1}^{T}\Delta_{it}(j)\\ &= \left(\frac{1}{T-T_1}-\frac{1}{T-s}\right)\sum_{t=T_1+1}^{T}\Delta_{it}(j)+\frac{1}{T-s}\left[\sum_{t=T_1+1}^{T}\Delta_{it}(j)-\sum_{t=s+1}^{T}\Delta_{it}(j)\right]\\ &= \frac{T_1-s}{(T-T_1)(T-s)}\sum_{t=T_1+1}^{T}\Delta_{it}(j)-\frac{1}{T-s}\sum_{t=s+1}^{T_1}\Delta_{it}(j)\end{aligned}$$

and

$$\begin{aligned}&(T-T_1)\bar{\Delta}^2_{T_1+,i}(j)-(T-s)\bar{\Delta}^2_{s+,i}(j)\\ &=(s-T_1)\bar{\Delta}^2_{T_1+,i}(j)+(T-s)\left[\bar{\Delta}^2_{T_1+,i}(j)-\bar{\Delta}^2_{s+,i}(j)\right]\\ &=(s-T_1)\bar{\Delta}^2_{T_1+,i}(j)+(T-s)\left[\bar{\Delta}_{T_1+,i}(j)+\bar{\Delta}_{s+,i}(j)\right]\left[\bar{\Delta}_{T_1+,i}(j)-\bar{\Delta}_{s+,i}(j)\right]\\ &=(s-T_1)\bar{\Delta}^2_{T_1+,i}(j)+\left[\bar{\Delta}_{T_1+,i}(j)+\bar{\Delta}_{s+,i}(j)\right]\left[\frac{T_1-s}{T-T_1}\sum_{t=T_1+1}^{T}\Delta_{it}(j)-\sum_{t=s+1}^{T_1}\Delta_{it}(j)\right],\end{aligned}$$

we have

$$\begin{aligned}A_3(s) &= \frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\left\{\sum_{t=s+1}^{T}\left[\Delta_{it}(j)-\bar{\Delta}_{s+,i}(j)\right]^2-\sum_{t=T_1+1}^{T}\left[\Delta_{it}(j)-\bar{\Delta}_{T_1+,i}(j)\right]^2\right\}\\ &= \frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\left\{\sum_{t=s+1}^{T}\Delta^2_{it}(j)-(T-s)\bar{\Delta}^2_{s+,i}(j)-\sum_{t=T_1+1}^{T}\Delta^2_{it}(j)+(T-T_1)\bar{\Delta}_{T_1+,i}(j)\right\}\\ &= \frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\left\{\sum_{t=s+1}^{T_1}\Delta^2_{it}(j)+(T-T_1)\bar{\Delta}^2_{T_1+,i}(j)-(T-s)\bar{\Delta}_{s+,i}(j)\right\}\\ &= \kappa_s\frac{1}{pN(T_1-s)}\sum_{j\in[p]}\sum_{i\in[N]}\sum_{t=s+1}^{T_1}\Delta^2_{it}(j)-\kappa_s\frac{1}{pN}\sum_{j\in[p]}\sum_{i\in[N]}\bar{\Delta}^2_{T_1+,i}(j)\\ &+\kappa_s\frac{1}{pN}\sum_{j\in[p]}\sum_{i\in[N]}\left[\bar{\Delta}_{T_1+,i}(j)+\bar{\Delta}_{s+,i}(j)\right]\frac{1}{T-T_1}\sum_{t=T_1+1}^{T}\Delta_{it}(j)\\ &-\kappa_s\frac{1}{pN}\sum_{j\in[p]}\sum_{i\in[N]}\left[\bar{\Delta}_{T_1+,i}(j)+\bar{\Delta}_{s+,i}(j)\right]\frac{1}{T_1-s}\sum_{t=s+1}^{T_1}\Delta_{it}(j)\\ &=\kappa_s O_p\left(\eta^2_{N,2}\right)=\kappa_s o_p(\zeta^2_{NT}),\end{aligned}\tag{A.16}$$

where the second last equality holds by (A.14) and the fact that

$$\begin{aligned}\max_{i\in[N],j\in[p]}\left|\bar{\Delta}_{s+,i}(j)\right| &= \max_{i\in[N],j\in[p]}\left|\frac{1}{T-s}\sum_{t=s+1}^{T}\Delta_{it}(j)\right| = O_p\left(\eta_{N,2}\right),\\ \max_{i\in[N],j\in[p]}\left|\bar{\Delta}_{T_1+,i}(j)\right| &= \max_{i\in[N],j\in[p]}\left|\frac{1}{T-T_1}\sum_{t=T_1+1}^{T}\Delta_{it}(j)\right| = O_p\left(\eta_{N,2}\right).\end{aligned}\tag{A.17}$$

Last, we notice that

$$
\begin{aligned}
&A_4(s) + A_5(s) \\
&= \frac{2}{pNT} \sum_{j\in[p]} \sum_{i\in[N]} \left(\alpha^{(1)}_{g_i^{(1)},j} - \alpha^{(2)}_{g_i^{(2)},j} \right) \left\{ \frac{T-T_1}{T-s} \sum_{t=s+1}^{T_1} \left[\Delta_{it}(j) - \bar{\Delta}_{s_+,i}(j) \right] - \frac{T_1-s}{T-s} \sum_{t=T_1+1}^{T} \left[\Delta_{it}(j) - \bar{\Delta}_{s_+,i}(j) \right] \right\} \\
&= \frac{2}{pNT} \sum_{j\in[p]} \sum_{i\in[N]} \left(\alpha^{(1)}_{g_i^{(1)},j} - \alpha^{(2)}_{g_i^{(2)},j} \right) \left[\frac{T-T_1}{T-s} \sum_{t=s+1}^{T_1} \Delta_{it}(j) - \frac{T_1-s}{T-s} \sum_{t=T_1+1}^{T} \Delta_{it}(j) \right] \\
&= 2\kappa_s \frac{1-\tau_T}{1-\frac{s}{T}} \frac{1}{pN} \sum_{j\in[p]} \sum_{i\in[N]} \left(\alpha^{(1)}_{g_i^{(1)},j} - \alpha^{(2)}_{g_i^{(2)},j} \right) \frac{1}{T_1-s} \sum_{t=s+1}^{T_1} \Delta_{it}(j) \\
&- 2\kappa_s \frac{1-\tau_T}{1-\frac{s}{T}} \frac{1}{pN} \sum_{j\in[p]} \sum_{i\in[N]} \left(\alpha^{(1)}_{g_i^{(1)},j} - \alpha^{(2)}_{g_i^{(2)},j} \right) \frac{1}{T-T_1} \sum_{t=T_1+1}^{T} \Delta_{it}(j) \\
&\leq 2\kappa_s \frac{1-\tau_T}{1-\frac{s}{T}} \sqrt{\frac{1}{N} \sum_{i\in[N]} \left\| \alpha^{(1)}_{g_i^{(1)}} - \alpha^{(2)}_{g_i^{(2)}} \right\|_2^2} \sqrt{\frac{1}{N} \sum_{i\in[N]} \left[\frac{1}{p(T_1-s)} \sum_{j\in[p]} \sum_{t=s+1}^{T_1} \Delta_{it}(j) \right]^2} \\
&+ 2\kappa_s \frac{1-\tau_T}{1-\frac{s}{T}} \sqrt{\frac{1}{N} \sum_{i\in[N]} \left\| \alpha^{(1)}_{g_i^{(1)}} - \alpha^{(2)}_{g_i^{(2)}} \right\|_2^2} \sqrt{\frac{1}{N} \sum_{i\in[N]} \left[\frac{1}{p(T-T_1)} \sum_{j\in[p]} \sum_{t=T_1+1}^{T} \Delta_{it}(j) \right]^2} \\
&= \kappa_s \frac{1-\tau_T}{1-\frac{s}{T}} \zeta_{NT} O_p(\eta_{N,2}) = \kappa_s o_p(\zeta_{NT}^2), \qquad \text{(A.18)}
\end{aligned}
$$

where the first inequality holds by Cauchy-Schwarz inequality.

Combining (A.11), (A.12), (A.13), (A.16), (A.18) and Assumption 6(i) yields that

$$
L(s) - L(T_1) = \kappa_s \frac{(1-\tau_T)}{1-\frac{s}{T}} D_{N\alpha} + \kappa_s o_p(\zeta_{NT}^2).
$$

Then for any $s < T_1$,

$$
\text{plim}_{(N,T)\to\infty} \frac{1}{\kappa_s \zeta_{NT}^2} \left[L(s) - L(T_1) \right] = \text{plim}_{(N,T)\to\infty} \frac{1-\tau_T}{1-\frac{s}{T}} \frac{1}{\zeta_{NT}^2} D_{N\alpha} \geq (1-\tau) D_\alpha > 0,
$$

where $D_\alpha :=\text{plim}_{(N,T)\to\infty} \frac{1}{\zeta_{NT}^2} D_{N\alpha} > 0$ by Assumption 6(i). This implies that

$$
\mathbb{P}\left(\hat{T}_1 < T_1 \right) \leq \mathbb{P}\left(\exists s < T_1, L(s) - L(T_1) < 0 \right) \to 0. \qquad \text{(A.19)}
$$

By analogous arguments, we prove (ii) in the following part. When $s > T_1$, we have

$$
\begin{aligned}
\bar{\dot{\Theta}}_{j,i}^{(1s)} &= \frac{1}{s} \sum_{t=1}^{s} \dot{\Theta}_{j,it} = \frac{1}{s} \sum_{t=1}^{s} \left[\Theta^0_{j,it} + \left(\dot{\Theta}_{j,it} - \Theta^0_{j,it} \right) \right] = \frac{T_1}{s} \alpha^{(1)}_{g_i^{(1)},j} + \frac{s-T_1}{s} \alpha^{(2)}_{g_i^{(2)},j} + \bar{\Delta}_{s,i}(j), \\
\bar{\dot{\Theta}}_{j,i}^{(2s)} &= \frac{1}{T-s} \sum_{t=s+1}^{T} \dot{\Theta}_{j,it} = \frac{1}{T-s} \sum_{t=s+1}^{T} \left[\Theta^0_{j,it} + \left(\dot{\Theta}_{j,it} - \Theta^0_{j,it} \right) \right] = \alpha^{(2)}_{g_i^{(2)},j} + \bar{\Delta}_{s_+,i}(j).
\end{aligned}
$$

It follows that

$$
\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}_{j,i}^{(1s)} = \begin{cases} \frac{s-T_1}{s} \left(\alpha^{(1)}_{g_i^{(1)},j} - \alpha^{(2)}_{g_i^{(2)},j} \right) + \Delta_{it}(j) - \bar{\Delta}_{s,i}(j) & \text{if } 1 \leq t \leq T_1 \\ \frac{T_1}{s} \left(\alpha^{(2)}_{g_i^{(2)},j} - \alpha^{(1)}_{g_i^{(1)},j} \right) + \Delta_{it}(j) - \bar{\Delta}_{s,i}(j) & \text{if } T_1 + 1 \leq t \leq s \end{cases}, \text{ and}
$$

$$\dot{\Theta}_{j,it}-\bar{\dot{\Theta}}_{j,i}^{(2s)}=\Delta_{it}(j)-\bar{\Delta}_{s_{+},i}(j),\quad s<t\leq T.$$

As in (A.8), we obtain that

$$\begin{aligned}
\sum_{t=1}^{s}\left[\dot{\Theta}_{j,it}-\bar{\dot{\Theta}}_{j,i}^{(1s)}\right]^2 &= \sum_{t=1}^{T_1}\left[\frac{s-T_1}{s}\left(\alpha_{g_i^{(1)},j}^{(1)}-\alpha_{g_i^{(2)},j}^{(2)}\right)+\Delta_{it}(j)-\bar{\Delta}_{s,i}(j)\right]^2\\
&+\sum_{t=T_1+1}^{s}\left[\frac{T_1}{s}\left(\alpha_{g_i^{(2)},j}^{(2)}-\alpha_{g_i^{(1)},j}^{(1)}\right)+\Delta_{it}(j)-\bar{\Delta}_{s,i}(j)\right]^2\\
&=\frac{T_1\left(s-T_1\right)^2}{s^2}\left(\alpha_{g_i^{(1)},j}^{(1)}-\alpha_{g_i^{(2)},j}^{(2)}\right)^2+\sum_{t=1}^{T_1}\left[\Delta_{it}(j)-\bar{\Delta}_{s,i}(j)\right]^2\\
&+2\frac{s-T_1}{s}\left(\alpha_{g_i^{(1)},j}^{(1)}-\alpha_{g_i^{(2)},j}^{(2)}\right)\sum_{t=1}^{T_1}\left[\Delta_{it}(j)-\bar{\Delta}_{s,i}(j)\right]\\
&+\frac{\left(s-T_1\right)T_1^2}{s^2}\left(\alpha_{g_i^{(2)},j}^{(2)}-\alpha_{g_i^{(1)},j}^{(1)}\right)^2+\sum_{t=T_1+1}^{s}\left[\Delta_{it}(j)-\bar{\Delta}_{s,i}(j)\right]^2\\
&+2\frac{T_1}{s}\left(\alpha_{g_i^{(2)},j}^{(2)}-\alpha_{g_i^{(1)},j}^{(1)}\right)\sum_{t=T_1+1}^{s}\left[\Delta_{it}(j)-\bar{\Delta}_{s,i}(j)\right]\\
&=\frac{\left(s-T_1\right)T_1}{s}\left(\alpha_{g_i^{(1)},j}^{(1)}-\alpha_{g_i^{(2)},j}^{(2)}\right)^2+\sum_{t=1}^{s}\left[\Delta_{it}(j)-\bar{\Delta}_{s,i}(j)\right]^2\\
&+2\frac{T_1\left(s-T_1\right)}{s}\left(\alpha_{g_i^{(1)},j}^{(1)}-\alpha_{g_i^{(2)},j}^{(2)}\right)\left[\frac{1}{T_1}\sum_{t=s+1}^{T_1}\Delta_{it}(j)-\frac{1}{s-T_1}\sum_{t=T_1+1}^{T}\Delta_{it}(j)\right]
\end{aligned}$$

and

$$\sum_{t=s+1}^{T}\left[\dot{\Theta}_{j,it}-\bar{\dot{\Theta}}_{j,i}^{(2s)}\right]^2=\sum_{t=s+1}^{T}\left[\Delta_{it}(j)-\bar{\Delta}_{s_{+},i}(j)\right]^2.$$

It follows that

$$\begin{aligned}
L(s)-L(T_1)&=\frac{T_1\left(s-T_1\right)}{sT}\frac{1}{pN}\sum_{i\in[N]}\left\|\alpha_{g_i^{(1)}}^{(1)}-\alpha_{g_i^{(2)}}^{(2)}\right\|_2^2\\
&+\frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\left\{\sum_{t=1}^{s}\left[\Delta_{it}(j)-\bar{\Delta}_{s,i}(j)\right]^2-\sum_{t=1}^{T_1}\left[\Delta_{it}(j)-\bar{\Delta}_{T_1,i}\left(j\right)\right]^2\right\}\\
&+\frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\left\{\sum_{t=s+1}^{T}\left[\Delta_{it}(j)-\bar{\Delta}_{s_{+},i}(j)\right]^2-\sum_{t=T_1+1}^{T}\left[\Delta_{it}(j)-\bar{\Delta}_{T_{1+},i}\left(j\right)\right]^2\right\}\\
&+2\frac{T_1\left(s-T_1\right)}{sT}\frac{1}{pN}\sum_{j\in[p]}\sum_{i\in[N]}\left(\alpha_{g_i^{(1)},j}^{(1)}-\alpha_{g_i^{(2)},j}^{(2)}\right)\left[\frac{1}{T_1}\sum_{t=s+1}^{T_1}\Delta_{it}(j)-\frac{1}{s-T_1}\sum_{t=T_1+1}^{T}\Delta_{it}(j)\right]\\
&:=B_1\left(s\right)+B_2\left(s\right)+B_3\left(s\right)+B_4\left(s\right),
\end{aligned}$$

where $B_4\left(s\right)$ parallels $A_4\left(s\right)+A_5\left(s\right)$ in (A.11). Let $\bar{\kappa}_s=\frac{s-T_1}{T}\in\left[\frac{1}{T},1-\tau_T\right]$. Following the analyses of $A_\ell\left(s\right)$'s, we can readily show that

$$B_1\left(s\right)=\bar{\kappa}_s\frac{T}{s}\frac{T_1}{T}D_{N\alpha}=\bar{\kappa}_sO_p(\zeta_{NT}^2),\ B_\ell\left(s\right)=\bar{\kappa}_sO_p\left(\eta_{N,2}^2\right)=\bar{\kappa}_so_p(\zeta_{NT}^2)\text{ for }\ell=2,3,$$

and $B_4(s) = \bar{\kappa}_s O_p(\eta_{N,2}\zeta_{NT}) = \bar{\kappa}_s o_p(\zeta_{NT}^2)$. It follows that for any $s > T_1$,

$$\text{plim}_{(N,T)\to\infty} \frac{1}{\bar{\kappa}_s \zeta_{NT}^2} [L(s) - L(T_1)] = \text{plim}_{(N,T)\to\infty} \frac{T_1 T}{Ts} \frac{1}{\zeta_{NT}^2} D_{N\alpha} \geq \tau D_\alpha > 0.$$

This implies that

$$\mathbb{P}\left(\hat{T}_1 > T_1\right) \leq \mathbb{P}\left(\exists s > T_1, L(s) - L(T_1) < 0\right) \to 0. \tag{A.20}$$

Combining (A.19) and (A.20), we can conclude that $\mathbb{P}\left(\hat{T}_1 = T_1\right) \to 1$. ■

A.4 Proof of Theorem 4.4

By Theorem 4.2, $\mathbb{P}\left(\hat{T}_1 = T_1\right) \to 1$. It follows that we can prove Theorem 4.4 by conditioning on the event that $\left\{\hat{T}_1 = T_1\right\}$. Below we prove the theorem under the event that $\left\{\hat{T}_1 = T_1\right\}$.

Define $\dot{\Theta}_{j,i}^{0,(1)} = \left(\dot{\Theta}_{j,i1}, \cdots, \dot{\Theta}_{j,iT_1}\right)'$, $\dot{\Theta}_{j,i}^{0,(2)} = \left(\dot{\Theta}_{j,i,T_1+1}, \cdots, \dot{\Theta}_{j,iT}\right)'$, $\dot{\beta}_i^{0,(1)} = \frac{1}{\sqrt{T_1}}(\dot{\Theta}_{1,i}^{0,(1)\prime}, \cdots, \dot{\Theta}_{p,i}^{0,(1)\prime})'$, and $\dot{\beta}_i^{0,(2)} = \frac{1}{\sqrt{T_2}}(\dot{\Theta}_{1,i}^{0,(2)\prime}, \cdots, \dot{\Theta}_{p,i}^{0,(2)\prime})'$. Noted that in the definitions of $\dot{\beta}_i^{0,(1)}$ and $\dot{\beta}_i^{0,(2)}$ we use the true break date T_1 rather than the estimated one compared to $\dot{\beta}_i^{(1)}$ and $\dot{\beta}_i^{(2)}$ defined in Step 4. As in (3.5) and (3.6), we further define

$$\left\{\dot{a}_{k,m}^{0,(\ell)}\right\}_{k\in[m]} = \underset{\left\{a_k^{(\ell)}\right\}_{k\in[m]}}{\arg\min} \frac{1}{N} \sum_{i\in[N]} \min_{k\in[m]} \left\|\dot{\beta}_i^{0,(\ell)} - a_k^{(\ell)}\right\|_2^2, \tag{A.21}$$

$$\hat{g}_{i,m}^{0,(\ell)} = \underset{k\in[m]}{\arg\min} \left\|\dot{\beta}_i^{(\ell)} - \dot{a}_{k,m}^{0,(\ell)}\right\|_2, \quad \forall i \in [N]. \tag{A.22}$$

(i) Under the case when $m = K^{(\ell)}$, Theorem 4.4(i.a) is from the combination of Lemma B.9 for the consistency of the membership estimates via K-means algorithm and Theorem 4.2 for the consistency of the break point estimator.

Next, we show (i.c). Recall that z_α is the critical value at α significance level calculated from the maximum of m independent $\chi^2(1)$ random variables. By the definition of the STK algorithm, we observe that

$$\mathbb{P}\left(\hat{K}^{(\ell)} \leq K^{(\ell)}\right) \geq \mathbb{P}\left(\hat{\Gamma}_{K^{(\ell)}}^{(\ell)} \leq z_\alpha\right),$$

which leads to the fact that (i.c) holds as long as we can show (i.b). This is because, under (i.b), we have

$$\mathbb{P}\left(\hat{\Gamma}_{K^{(\ell)}}^{(\ell)} \leq z_\alpha\right) \geq 1 - \alpha + o(1).$$

Now, we focus on (i.b). Notice that $\hat{\Gamma}_{k,K^{(\ell)}}^{(\ell)}$ depends on the K-means classification result, i.e., the estimated group membership $\hat{G}_{k,K^{(\ell)}}^{(\ell)}$ for $k \in [K^{(\ell)}]$. From Theorem 4.4(i.1), we notice that the we can change the estimated group membership $\hat{G}_{k,K^{(\ell)}}^{(\ell)}$ to the true group membership $G_k^{(\ell)}$, and this replacement has only asymptotically negligible effect. Recall that $\mathcal{T}_1 = [T_1]$, $\mathcal{T}_2 = [T]\backslash[T_1]$. Define $\mathcal{T}_{1,-1} = \mathcal{T}_1\backslash\{T_1\}$, $\mathcal{T}_{2,-1} = \mathcal{T}_2\backslash\{T\}$, $\mathcal{T}_{1,j} = \{1+j, \cdots, T_1\}$ and $\mathcal{T}_{2,j} = \{T_1+1+j, \cdots, T\}$ for some specific $j \in \mathcal{T}_{\ell,-1}$, similarly as previous notation. Moreover, let

$$\left(\left\{\hat{\theta}_{i,k,K^{(\ell)}}^{0,(\ell)}\right\}_{i\in G_k^{(\ell)}}, \hat{\Lambda}_{k,K^{(\ell)}}^{0,(\ell)}, \left\{\hat{f}_{t,k,K^{(\ell)}}^{0,(\ell)}\right\}_{t\in\mathcal{T}_\ell}\right) = \underset{\{\theta_i,\lambda_i\}_{i\in G_k^{(\ell)}}, \{f_t\}_{t\in\mathcal{T}_\ell}}{\arg\min} \sum_{i\in G_k^{(\ell)}} \sum_{t\in\mathcal{T}_\ell} \left(Y_{it} - X_{it}'\theta_i - \lambda_i' f_t\right)^2,$$

$\hat{F}^{0,(1)}_{k,K^{(1)}} = \left(\hat{f}_{1,k,K^{(1)}}, \cdots, \hat{f}_{T_1,k,K^{(1)}}\right)'$, $\hat{F}^{0,(2)}_{k,K^{(2)}} = \left(\hat{f}_{T_1+1,k,K^{(1)}}, \cdots, \hat{f}_{T,k,K^{(2)}}\right)'$, $\hat{\Lambda}^{(\ell)}_{k,K^{(\ell)}} = \left\{\hat{\lambda}^{(\ell)}_{i,k,K^{(\ell)}}\right\}_{i\in G^{(\ell)}_k}$, and $\left(\hat{z}^{0,(\ell)}_{it}\right)'$ denote the t-th row of $M_{\hat{F}^{0,(\ell)}_{k,K^{(\ell)}}} X^{(\ell)}_i$. Further define

$$\hat{\bar{\theta}}^{0,(\ell)}_{k,K^{(\ell)}} = \frac{1}{\left|G^{(\ell)}_k\right|}\sum_{i\in G^{(\ell)}_k}\hat{\theta}^{0,(\ell)}_{i,k,K^{(\ell)}}, \qquad \hat{S}^{0,(\ell)}_{ii,k,K^{(\ell)}} = \frac{\left(X^{(\ell)}_i\right)' M_{\hat{F}^{0,(\ell)}_{k,K^{(\ell)}}} X^{(\ell)}_i}{T_\ell},$$

$$\hat{\Omega}^{0,(\ell)}_{i,k,K^{(\ell)}} = \frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\hat{z}^{0,(\ell)}_{it}\hat{z}^{0,(\ell)\prime}_{it}\hat{e}^2_{it} + \frac{1}{T_\ell}\sum_{j\in\mathcal{T}_{\ell,-1}}k(j,L)\sum_{t\in\mathcal{T}_{\ell,j}}\left(\hat{z}^{0,(\ell)}_{it}\hat{z}^{0,(\ell)\prime}_{i,t+j}\hat{e}_{it}\hat{e}_{i,t+j} + \hat{z}^{0,(\ell)}_{i,t-j}\hat{z}^{0,(\ell)\prime}_{it}\hat{e}_{i,t-j}\hat{e}_{it}\right),$$

$$\hat{a}^{0,(\ell)}_{ii,k,K^{(\ell)}} = \hat{\lambda}^{(\ell)\prime}_{i,k,K^{(\ell)}}\left(\frac{\hat{\Lambda}^{(\ell)\prime}_{k,K^{(\ell)}}\hat{\Lambda}^{(\ell)}_{k,K^{(\ell)}}}{\left|G^{(\ell)}_k\right|}\right)^{-1}\hat{\lambda}^{(\ell)}_{i,k,K^{(\ell)}}.$$

Then $\forall k \in [K^{(\ell)}]$, we can define

$$\hat{\Gamma}^{0,(\ell)}_{k,K^{(\ell)}} = \sqrt{\left|G^{(\ell)}_k\right|}\left(\frac{\frac{1}{\left|G^{(\ell)}_k\right|}\sum_{i\in G^{(\ell)}_k}\hat{\mathbb{S}}^{0,(\ell)}_{i,k,K^{(\ell)}} - p}{\sqrt{2p}}\right)$$

where

$$\hat{\mathbb{S}}^{0,(\ell)}_{i,k,K^{(\ell)}} = T_\ell\left(\hat{\theta}^{0,(\ell)}_{i,k,K^{(\ell)}} - \hat{\bar{\theta}}^{0,(\ell)}_{k,K^{(\ell)}}\right)'\hat{S}^{0,(\ell)}_{ii,k,K^{(\ell)}}\left(\hat{\Omega}^{0,(\ell)}_{i,k,K^{(\ell)}}\right)^{-1}\hat{S}^{0,(\ell)}_{ii,k,K^{(\ell)}}\left(\hat{\theta}^{0,(\ell)}_{i,k,K^{(\ell)}} - \hat{\bar{\theta}}^{0,(\ell)}_{k,K^{(\ell)}}\right)\left(1 - \hat{a}^{0,(\ell)}_{ii,k,K^{(\ell)}}/\left|G^{(\ell)}_k\right|\right)^2.$$

By Lemma D.8, we notice that $\hat{\Gamma}^{0,(\ell)}_{k,K^{(\ell)}} \rightsquigarrow \mathcal{N}(0,1)$ owing to the fact that the slope coefficient $\alpha^{(\ell)}_k$ is homogeneous across $i \in G^{(\ell)}_k$ $\forall k \in [K^{(\ell)}]$. Furthermore, $\left\{\hat{\Gamma}^{0,(\ell)}_{k,K^{(\ell)}},\ k \in [K^{(\ell)}]\right\}$ are asymptotically independent under Assumption 1(i). It follows that

$$\hat{\Gamma}^{(\ell)}_{K^{(\ell)}} = \max_{k\in[K^{(\ell)}]}\left(\hat{\Gamma}^{(\ell)}_{k,K^{(\ell)}}\right)^2 = \max_{k\in[m]}\left(\hat{\Gamma}^{0,(\ell)}_{k,K^{(\ell)}}\right)^2 + o_p(1) \rightsquigarrow \mathcal{Z},$$

where $\mathcal{Z}$ is the maximum of m independent $\chi^2(1)$ random variables. Then Theorem 4.4(i.b) follows.

(ii) When $m < K^{(\ell)}$, Theorem 4.4(i.1) does not hold and we can not change the estimated group membership $\hat{\mathcal{G}}^{(\ell)}_{K^{(\ell)}}$ to the true group membership $\mathcal{G}^{(\ell)}$. To get around of this issue, we define the "pseudo groups". For $m < K^{(\ell)}$, let $\mathbb{G}^{(\ell)}_m := \left\{G^{(\ell)}_{1,m}, \cdots, G^{(\ell)}_{m,m}\right\}$ such that $\{1,\cdots,N\} = G^{(\ell)}_{1,m}\cup\cdots\cup G^{(\ell)}_{m,m}$, which indicates one possible partition of the set $\{1,\cdots,N\}$. We further define $\mathcal{G}^{(\ell)}_m$ to be the collection of all possible $\mathbb{G}^{(\ell)}_m$.

By Theorem 4.4(i.c), we can conclude that $\mathbb{P}\left(\hat{K}^{(\ell)} \neq K^{(\ell)}\right) \leq \alpha + o(1)$ provided we can show that $\hat{\Gamma}^{(\ell)}_m \to \infty$ when $m < K^{(\ell)}$. By Lemma B.11, we notice that $\hat{\mathcal{G}}^{(\ell)}_m \in \mathbb{G}^{(\ell)}_m$ w.p.a.1. Conditioning on the event $\left\{\hat{\mathcal{G}}^{(\ell)}_m \in \mathbb{G}^{(\ell)}_m\right\} \cap \left\{\hat{T}_1 = T_1\right\}$, we have

$$\hat{\Gamma}^{(\ell)}_m > \min_{\mathcal{G}^{(\ell)}_m\in\mathbb{G}^{(\ell)}_m}\hat{\Gamma}^{0,(\ell)}_m\left(\mathcal{G}^{(\ell)}_m\right) := \min_{\mathcal{G}^{(\ell)}_m\in\mathbb{G}^{(\ell)}_m}\left\{\max_{k\in[m]}\left[\hat{\Gamma}^{0,(\ell)}_{k,m}\left(G^{(\ell)}_{k,m}\right)\right]^2\right\},$$

where

$$\hat{\Gamma}^{0,(\ell)}_{k,m}\left(G^{(\ell)}_{k,m}\right) = \sqrt{\left|G^{(\ell)}_{k,m}\right|}\left(\frac{\frac{1}{\left|G^{(\ell)}_{k,m}\right|}\sum_{i\in G^{(\ell)}_{k,m}}\hat{\mathbb{S}}^{0,(\ell)}_{i,k,m} - p}{\sqrt{2p}}\right),$$

and $\hat{\mathbb{S}}_{i,k,m}^{0,(\ell)}$ is defined similarly to $\hat{\mathbb{S}}_{i,k,K^{(\ell)}}^{0,(\ell)}$ in the proof of (i).

Owing to the fact that $\left|\mathbb{G}_m^{(\ell)}\right| = m^{K^{(\ell)}}$ which is a constant since $K^{(\ell)}$ is a constant, we can show that $\hat{\Gamma}_m^{(\ell)} \to \infty$ by showing that $\hat{\Gamma}_m^{0,(\ell)}\left(\mathcal{G}_m^{(\ell)}\right) \to \infty$ for any possible realization $\mathcal{G}_m^{(\ell)}$. Under the case when $m < K^{(\ell)}$, there exists at least one $k \in [m]$ such that the slope coefficient is not homogeneous across $i \in G_{k,m}^{(\ell)}$. Assume that $G_{k,m}^{(\ell)}$ contains n true groups, i.e., $G_{k,m}^{(\ell)} = G_{k_1}^{(\ell)} \cup \cdots \cup G_{k_n}^{(\ell)}$ for $k_1, \cdots, k_n \in [K^{(\ell)}]$ and $k_1 \neq \cdots \neq k_n$. Then for $i \in G_{k,m}^{(\ell)}$, we have

$$\begin{aligned}\theta_i^{0,(\ell)} &= \sum_{s=1}^{n} \alpha_{k_s}^{(\ell)} \mathbf{1}\left\{i \in G_{k_s}^{(\ell)}\right\} = \frac{1}{n}\sum_{s^*=1}^{n} \alpha_{k_{s^*}}^{(\ell)} + \sum_{s=1}^{n}\left(\frac{n-1}{n}\alpha_{k_s}^{(\ell)} - \frac{1}{n}\sum_{s^* \neq s}\alpha_{k_{s^*}}^{(\ell)}\right)\mathbf{1}\left\{i \in G_{k_s}^{(\ell)}\right\} \\ &= \frac{1}{n}\sum_{s^*=1}^{n} \alpha_{k_{s^*}}^{(\ell)} + \sum_{s=1}^{n}\frac{1}{n}\sum_{s^* \neq s}\left(\alpha_{k_s}^{(\ell)} - \alpha_{k_{s^*}}^{(\ell)}\right)\mathbf{1}\left\{i \in G_{k_s}^{(\ell)}\right\} \\ &:= \bar{\theta}_n^{0,(\ell)} + c_i^{(\ell)}\end{aligned}$$

such that

$$\frac{T_\ell}{\sqrt{N}}\sum_{i\in[N]}\left\|c_i^{(\ell)}\right\|_2^2 = \frac{T_\ell}{\sqrt{N}}\sum_{s=1}^{n}\frac{N_{k_s}^{(\ell)}}{n}\left\|\sum_{s^*\neq s}\left(\alpha_{k_s}^{(\ell)} - \alpha_{k_{s^*}}^{(\ell)}\right)\right\|_2^2 = \frac{T_\ell}{\sqrt{N}n}\sum_{s=1}^{n}N_{k_s}^{(\ell)}\left\|\frac{\sum_{s^*\neq s}\alpha_{k_{s^*}}^{(\ell)}}{n-1} - \alpha_{k_s}^{(\ell)}\right\|_2^2 \to \infty$$

by Assumption 7(iii). Following this, it yields that $\left|\hat{\Gamma}_{k,m}^{0,(\ell)}\left(G_{k,m}^{(\ell)}\right)\right| \to \infty$ for some $k \in [m]$ by Lemma D.9. By the definition of $\hat{\Gamma}_m^{0,(\ell)}\left(\mathcal{G}_m^{(\ell)}\right)$, we have $\hat{\Gamma}_m^{0,(\ell)}\left(\mathcal{G}_m^{(\ell)}\right) \to \infty$, which yields $\hat{\Gamma}_m^{(\ell)} \to \infty$ w.p.a.1 for $m < K^{(\ell)}$ and $\mathbb{P}\left(\hat{K}^{(\ell)} \neq K^{(\ell)}\right) \leq \alpha + o(1)$. ■

A.5 Proof of Theorem 4.5

To show Theorem 4.5, we can directly derive the asymptotic distribution for the oracle estimator $\hat{\alpha}_k^{*(\ell)}$ by combining Theorems 4.2 and 4.4.

The asymptotic distribution theory of the linear panel model with IFEs has already been studied in the literature; see Bai (2009), Moon and Weidner (2017) and Lu and Su (2016) for instance. However, Bai (2009) rules out dynamic panels. Moon and Weidner (2017) allow dynamic panels and assume the independence over both i and t for the error term. Under Assumptions 1* and 2-9, which is for the dynamic linear panel model, Theorem 4.5 is same as Theorem 4.3 in Moon and Weidner (2017).

Below, we follow the arguments in Moon and Weidner (2017) and sketch the proof to allow the serial correlation of error terms in non-dynamic panels.[1] To proceed, let $\mathbb{C}_{NT,k}^{(\ell)}$ be the p-vector with j-th entry being $\mathbb{C}_{NT,k,j}^{(\ell)} = \mathbb{C}_1\left(\Lambda_k^{0,(\ell)}, F^{0,(\ell)}, \mathbb{X}_{j,k}^{(\ell)}, E_k^{(\ell)}\right) + \mathbb{C}_2\left(\Lambda_k^{0,(\ell)}, F^{0,(\ell)}, \mathbb{X}_{j,k}^{(\ell)}, E_k^{(\ell)}\right)$, where

$$\mathbb{C}_1\left(\Lambda_k^{0,(\ell)}, F^{0,(\ell)}, \mathbb{X}_{j,k}^{(\ell)}, E_k^{(\ell)}\right) = \frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left(M_{F^{0,(\ell)}}E_k^{(\ell)\prime}M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}\right),$$

[1] It is well known that one cannot allow for both dynamics and serially correlated errors in the panel for the least-squares based PCA estimation to avoid the endogeneity issue.

$$\mathbb{C}_2\left(\Lambda_k^{0,(\ell)}, F^{0,(\ell)}, \mathbb{X}_{j,k}^{(\ell)}, E_k^{(\ell)}\right) = -\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left(E_k^{(\ell)} M_{F^{0,(\ell)}} E_k^{(\ell)\prime} M_{\Lambda_k^{0,(\ell)}} \mathbb{X}_{j,k}^{(\ell)} F^{0,(\ell)} \left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1} \left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1} \Lambda_k^{0,(\ell)\prime}\right)$$

$$-\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left(E_k^{(\ell)\prime} M_{\Lambda_k^{0,(\ell)}} E_k^{(\ell)} M_{F^{0,(\ell)}} \mathbb{X}_{j,k}^{(\ell)\prime} \Lambda_k^{0,(\ell)} \left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1} \left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1} \left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1} F^{0,(\ell)\prime}\right)$$

$$-\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left(E_k^{(\ell)\prime} M_{\Lambda_k^{0,(\ell)}} \mathbb{X}_{j,k}^{(\ell)} M_{F^{0,(\ell)}} E_k^{(\ell)\prime} \Lambda_k^{0,(\ell)} \left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1} \left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1} \left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1} F^{0,(\ell)\prime}\right).$$

By Lemma B.12, we have $\sqrt{N_k^{(\ell)}T_\ell}\left(\hat{\alpha}_k^{(\ell)} - \alpha_k^{(\ell)}\right) = \mathbb{W}_{NT,k}^{(\ell)-1}\mathbb{C}_{NT,k}^{(\ell)} + o_p(1)$.

In Moon and Weidner (2017), the asymptotic distribution is derived mainly relying on their Lemmas B.1 and B.2. Lemma B.2 is the standard central limit theorem, which also holds under our Assumption 1. For Lemma B.1, we need to extend it to allow for serially correlated errors in non-dynamic panels in Lemma B.13. Hence, by Lemma B.13 and following the analogous arguments in the proof of Theorem 4.3 (Moon and Weidner (2017)), for a specific $\ell \in \{1, 2\}$ and $k \in [K^{(\ell)}]$, we can readily show that

$$\mathbb{W}_{NT,k}^{(\ell)}\sqrt{N_k^{(\ell)}T_\ell}\left(\hat{\alpha}_k^{(\ell)} - \alpha_k^{(\ell)}\right) - \mathbb{B}_{NT,k}^{(\ell)} \rightsquigarrow \mathcal{N}\left(0, \Omega_k^{(\ell)}\right)$$

which yields the final distributional results in Theorem 4.5 by stacking all subgroups of parameter estimators into a large vector and resorting to the Cramer-Wold device.

A.6 Proof of Theorem 5.1

Recall that $\dot{v}_{t,j}^* := \frac{\dot{v}_{t,j}}{\|\dot{v}_{t,j}\|_2}$, $\dot{v}_t^* = \left(\dot{v}_{t,1}^{*\prime}, \cdots, \dot{v}_{t,p}^{*\prime}\right)'$, $v_{t,j}^* = \frac{O_j v_{t,j}^0}{\left\|O_j v_{t,j}^0\right\|_2}$ and $v_t^* = \left(v_{t,1}^{*\prime}, \cdots, v_{t,p}^{*\prime}\right)'$. With the fact that

$$\begin{aligned}\frac{\dot{v}_{t,j}}{\|\dot{v}_{t,j}\|_2} - \frac{O_j v_{t,j}^0}{\left\|O_j v_{t,j}^0\right\|_2} &= \frac{\dot{v}_{t,j}\left\|O_j v_{t,j}^0\right\|_2 - O_j v_{t,j}^0\|\dot{v}_{t,j}\|_2}{\|\dot{v}_{t,j}\|_2\left\|O_j v_{t,j}^0\right\|_2} \\ &= \frac{\left(\dot{v}_{t,j} - O_j v_{t,j}^0\right)\left\|O_j v_{t,j}^0\right\|_2 + O_j v_{t,j}^0\left(\left\|O_j v_{t,j}^0\right\|_2 - \|\dot{v}_{t,j}\|_2\right)}{\|\dot{v}_{t,j}\|_2\left\|O_j v_{t,j}^0\right\|_2},\end{aligned}$$

It follows that

$$\max_{t\in[T]}\|\dot{v}_t^* - v_t^*\|_2 \le p \max_{j\in[p],t\in[T]}\left\|\dot{v}_{t,j}^* - v_{t,j}^*\right\|_2 \le 2p \max_{j\in[p],t\in[T]}\frac{\left\|\dot{v}_{t,j} - O_j v_{t,j}^0\right\|_2}{\left\|\dot{v}_{t,j}\right\|_2} = O_p(\eta_{N,2}),$$

where the last line is by Lemma B.7(i) and Theorem 4.1(ii).

B Technical Lemmas

Lemma B.1 *Consider a matrix sequence $\{A_i, i = 1, \cdots, N\}$ whose values are symmetric matrices with dimension d. Suppose $\{A_i, i = 1, \cdots, N\}$ is independent with $\mathbb{E}(A_i) = 0$ and $\|A_i\|_{op} \le M$ almost surely. Let $\sigma^2 = \left\|\sum_{i\in[N]}\mathbb{E}\left(A_i^2\right)\right\|_{op}$, for all $t > 0$, we have*

$$\mathbb{P}\left(\left\|\sum_{i\in[N]} A_i\right\|_{op} > t\right) \le d\cdot\exp\left\{-\frac{t^2/2}{\sigma^2 + Mt/3}\right\}.$$

Proof. Lemma B.1 states a Matrix Bernstein inequality; see Theorem 1.3 in Tropp (2011). ■

Lemma B.2 *Consider a specific matrix* $A \in \mathbb{R}^{N\times T}$ *whose rows (denoted as* A_i'*) are independent random vectors in* $\mathbb{R}^T$ *with* $\mathbb{E}A_i = 0$ *and* $\Sigma_i = \mathbb{E}(A_iA_i')$*. Suppose* $\max_{i\in[N]}\|A_i\|_2 \le \sqrt{m}$ *almost surely and* $\max_{i\in[N]}\|\Sigma_i\|_{op} \le M$ *for some positive constant* M*. Then for every* $t > 0$*, with probability* $1 - 2T\exp\left(-c_1t^2\right)$*, we have*

$$\|A\|_{op} \le \sqrt{NM} + t\sqrt{m+M},$$

where c_1 *is an absolute constant.*

Proof. The proof follows the argument similar as used in the proof of Theorem 5.41 in Vershynin (2010). Define $Z_i := \frac{1}{N}(A_iA_i' - \Sigma_i) \in \mathbb{R}^{T\times T}$, and we notice that $(Z_1, \cdots, Z_N)$ is an independent sequence with $\mathbb{E}(Z_i) = 0$. To use the matrix Bernstein's inequality, we analyze $\|Z_i\|_{op}$ and $\left\|\sum_{i\in[N]}\mathbb{E}\left(X_i^2\right)\right\|_{op}$ as followings:

$$\|Z_i\|_{op} \le \frac{1}{N}\left(\|A_iA_i'\|_{op} + \|\Sigma_i\|_{op}\right) \le \frac{1}{N}\left(\|A_i\|_2^2 + \|\Sigma_i\|_{op}\right) \le \frac{m+M}{N}, \quad a.s. \tag{B.1}$$

uniformly over i. Moreover, note that

$$\mathbb{E}\left[(A_iA_i')^2\right] = \mathbb{E}\left[\|A_i\|_2 A_iA_i'\right] \le m\Sigma_i$$

and

$$Z_i^2 = \frac{1}{N^2}\left[(A_iA_i')^2 - A_iA_i'\Sigma_i - \Sigma_iA_iA_i' + \Sigma_i^2\right].$$

We then obtain that

$$\begin{aligned}\|\mathbb{E}\left(Z_i^2\right)\|_{op} &= \left\|\mathbb{E}\left\{\frac{1}{N^2}\left[(A_iA_i')^2 - \Sigma_i^2\right]\right\}\right\|_{op} \le \frac{1}{N^2}\left\{\left\|\mathbb{E}\left[(A_iA_i')^2\right]\right\|_{op} + \|\Sigma_i\|_{op}^2\right\} \\ &\le \frac{1}{N^2}\left(m\|\Sigma_i\|_{op} + \|\Sigma_i\|_{op}^2\right) \le \frac{mM+M^2}{N^2} \quad a.s.\end{aligned}$$

uniformly over i, and

$$\left\|\sum_{i\in[N]}\mathbb{E}\left(Z_i^2\right)\right\|_{op} \le N\max_{i\in[N]}\|\mathbb{E}\left(Z_i^2\right)\|_{op} \le \frac{mM+M^2}{N} \quad a.s. \tag{B.2}$$

Define $\varepsilon = \max\left(\sqrt{M}\delta, \delta^2\right)$ with $\delta = t\sqrt{\frac{m+M}{N}}$. Combining (B.1) and (B.2), by matrix Bernstein's inequality, we have

$$\begin{aligned}&\mathbb{P}\left\{\left\|\frac{1}{N}\left(A'A - \sum_{i\in[N]}\Sigma_i\right)\right\|_{op} \ge \varepsilon\right\} = \mathbb{P}\left(\left\|\sum_{i\in[N]}Z_i\right\|_{op} \ge \varepsilon\right) \\ &\le 2T\exp\left\{-c\min\left(\frac{\varepsilon^2}{\frac{mM+M^2}{N}}, \frac{\varepsilon}{\frac{m+M}{N}}\right)\right\} \le 2T\exp\left\{-c\min\left(\frac{\varepsilon^2}{M}, \varepsilon\right)\frac{N}{m+M}\right\} \\ &\le 2T\exp\left\{-\frac{c\delta^2N}{m+M}\right\} = 2T\exp\left\{-c_1t^2\right\},\end{aligned}$$

for some positive constant c, where the third inequality is due to the fact that

$$\min\left(\frac{\varepsilon^2}{M}, \varepsilon\right) = \min\left(\max\left(\delta^2, \delta^4/M\right), \max\left(\sqrt{M}\delta, \delta\right)\right)$$
$$= \begin{cases} \min\left(\delta^2, \sqrt{M}\delta\right) = \delta^2, & \text{if} \quad \delta^2 \geq \dfrac{\delta^4}{M}, \\ \min\left(\delta^4/M, \delta^2\right) = \delta^2, & \text{if} \quad \delta^2 < \dfrac{\delta^4}{M}. \end{cases}$$

It implies that

$$\left\| \frac{1}{N} A'A - \frac{1}{N} \sum_{i\in[N]} \Sigma_i \right\|_{op} \leq \max\left(\sqrt{M}\delta, \delta^2\right) \tag{B.3}$$

with probability $1 - \exp\left(-ct^2\right)$. Combining the fact that $\|\Sigma_i\| \leq M$ uniformly over i and (B.3), we show that

$$\begin{aligned} \frac{1}{N}\|A\|_{op}^2 &= \left\|\frac{1}{N}A'A\right\|_{op} \leq \left\|\frac{1}{N}\sum_{i\in[N]}\Sigma_i\right\|_{op} + \left\|\frac{1}{N}A'A - \frac{1}{N}\sum_{i\in[N]}\Sigma_i\right\|_{op} \\ &\leq \max_{i\in[N]}\|\Sigma_i\|_{op} + \sqrt{M}\delta + \delta^2 \leq M + \sqrt{M}t\sqrt{\frac{m+M}{N}} + t^2\frac{m+M}{N} \\ &\leq \left(\sqrt{M} + t\sqrt{\frac{m+M}{N}}\right)^2, \end{aligned}$$

and the result follows: $\|A\|_{op} \leq \sqrt{NM} + t\sqrt{m+M}$. ■

Lemma B.3 *Recall that* $X_j = \{X_{j,it}\}$ *and* $E = \{e_{it}\}$. *Under Assumption 1,* $\forall j \in [p]$, *we have* $\|X_j \odot E\|_{op} = O_p\left(\sqrt{N} + \sqrt{T\log T}\right)$, *and* $\|E\|_{op} = O_p\left(\sqrt{N} + \sqrt{T\log T}\right)$.

Proof. We focus on $\|X_j \odot E\|_{op}$ as the result for $\|E\|_{op}$ can be derived in the same manner. We first note that, conditional on $\left\{V_j^0\right\}_{j\in[p]\cup\{0\}}$, $X_j \odot E$ are independent across i. Denote the i-th row of $X_j \odot E$ as $A_i' = X_{j,i}' \odot E_i'$, where $X_{j,i}'$ and E_i' being the i-th row of matrix X_j and E, respectively. Recall that $\mathscr{D}$ is the minimum σ-field generated by $\left\{V_j^0\right\}_{j\in[p]\cup\{0\}}$. In addition, for the t-th element of A_i, we have

$$\mathbb{E}\left[X_{j,it}e_{it}|\mathscr{D}\right] = \mathbb{E}\left\{X_{j,it}\mathbb{E}\left[e_{it}|\mathscr{D}, X_{it}\}\right] \middle| \mathscr{D}\right\} = 0,$$

where the second equality holds by Assumption 1(ii). Therefore, to apply Lemma B.1 conditionally on $\mathscr{D}$, we only need to upper bound $\|A_i\|_2$ and $\mathbb{E}\left[A_iA_i'\middle|\mathscr{D}\right]$.

First, under Assumption 1, we have $\frac{1}{T}\sum_{t\in[T]}\left(X_{j,it}e_{it}\right)^2 \leq C$ *a.s.* by Assumption 1(iv), which implies

$$||A_i||_2 = \|X_{j,i} \odot a_i\|_2 \leq C\sqrt{T} \ a.s. \tag{B.4}$$

Second, let $\Sigma_i = \mathbb{E}\left\{\left[\left(X_{j,i} \odot E_i\right)\left(X_{j,i} \odot E_i\right)'\right] \middle| \mathscr{D}\right\}$ with (t,s) element being $\mathbb{E}\left(X_{j,it}X_{j,is}e_{it}e_{is}\middle|\mathscr{D}\right)$. Recall that $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are norms induced by 1- and ∞-norms, i.e.,

$$\|\Sigma_i\|_1 = \max_{s\in[T]}\sum_{t\in[T]}\left|\mathbb{E}\left(X_{j,it}X_{j,is}e_{it}e_{is}\middle|\mathscr{D}\right)\right|, \text{ and } \|\Sigma_i\|_\infty = \max_{t\in[T]}\sum_{s\in[T]}\left|\mathbb{E}\left(X_{j,it}X_{j,is}e_{it}e_{is}\middle|\mathscr{D}\right)\right|.$$

By Davydov's inequality for strong mixing sequence, see also in Lemma 4.3 in Su and Chen (2013), we can show that

$$\begin{aligned}
&\max_{s\in[T]}\sum_{t\in[T]}\left|\mathbb{E}\left(X_{j,it}X_{j,is}e_{it}e_{is}\middle|\mathscr{D}\right)\right| = \max_{s\in[T]}\sum_{t\in[T]}\left|Cov\left(X_{j,it}e_{it},X_{j,is}e_{is}\middle|\mathscr{D}\right)\right| \\
&\leq \max_{s\in[T]}\sum_{t\in[T]}\left\{\mathbb{E}\left[\left|X_{j,it}e_{it}\right|^q\middle|\mathscr{D}\right]\right\}^{1/q}\left\{\mathbb{E}\left[\left|X_{j,is}e_{is}\right\}\right|^q\middle|\mathscr{D}\right]\right\}^{1/q}\times\alpha\left(t-s\right)^{(q-2)/q} \\
&\leq \max_{i\in[N],t\in[T]}\left\{\mathbb{E}\left[\left|X_{j,it}e_{j,it}\right|^q\middle|\mathscr{D}\right]\right\}^{2/q}\max_{s\in[T]}\sum_{t\in[T]}\left[\alpha\left(t-s\right)\right]^{(q-2)/q} \\
&\leq c_2\ a.s.,
\end{aligned}$$

where c_2 is a positive constant which does not depend on i. Similarly, we have

$$\max_{t\in[T]}\sum_{s\in[T]}\left|\mathbb{E}\left(X_{j,it}X_{j,is}e_{it}e_{is}\middle|\mathscr{D}\right)\right| \leq c_2\ a.s.$$

Therefore, by Corollary 2.3.2 in Golub and Van Loan (1996), we have

$$\max_{i\in[N]}\left\|\Sigma_i\right\|_{op} \leq \sqrt{\left\|\Sigma_i\right\|_1\left\|\Sigma_i\right\|_\infty} \leq c_2\ a.s. \tag{B.5}$$

Combining (B.1), (B.2), and Lemma B.1 with $t=\sqrt{\log T}$, we obtain the desired result. ■

Recall that $\mathcal{R}(C_1) := \left\{\left\{\Delta_{\Theta_j}\right\}_{j\in[p]\cup\{0\}} \in \mathbb{R}^{N\times T\times(p+1)} : \sum_{j\in[p]\cup\{0\}}\left\|\mathcal{P}_j^{\perp}(\Delta_{\Theta_j})\right\|_* \leq C_1\sum_{j\in[p]\cup\{0\}}\left\|\mathcal{P}_j(\Delta_{\Theta_j})\right\|_*\right\}$.

Lemma B.4 *Suppose Assumptions 1-3 hold, then* $\left\{\tilde{\Delta}_{\Theta_j}\right\}_{j\in[p]\cup\{0\}} \in \mathcal{R}(3)$ *w.p.a.1.*

Proof. Recall that event

$$\mathscr{A}_{1,N}(c_3) = \left\{\left\|E\right\|_{op} \leq c_3\left(\sqrt{N}\vee\sqrt{T\log T}\right), \left\|X_j\odot E\right\|_{op} \leq c_3\left(\sqrt{N}\vee\sqrt{T\log T}\right), \forall j\in[p]\right\}$$

for any positive constant c_3 with $\mathbb{P}(\mathscr{A}_1^c(c_3)) = \epsilon$ for any $\epsilon > 0$ by Lemma B.3. Under event $\mathscr{A}_{1,N}(c_3)$, by the definition of $\tilde{\Theta}_j$ in (3.1), we notice that

$$0 \leq \frac{1}{NT}\left\|Y-\Theta_0^0-\sum_{j\in[p]}X_j\odot\Theta_j^0\right\|_F^2 - \frac{1}{NT}\left\|Y-\tilde{\Theta}_0-\sum_{j\in[p]}X_j\odot\tilde{\Theta}_j\right\|_F^2 + \sum_{j\in[p]\cup\{0\}}\nu_j\left(\left\|\Theta_j^0\right\|_* - \left\|\tilde{\Theta}_j\right\|_*\right) \tag{B.6}$$

and

$$\begin{aligned}
&\frac{1}{NT}\left\|Y-\Theta_0^0-\sum_{j\in[p]}X_j\odot\Theta_j^0\right\|_F^2 - \frac{1}{NT}\left\|Y-\tilde{\Theta}_0-\sum_{j\in[p]}X_j\odot\tilde{\Theta}_j\right\|_F^2 \\
&= \frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\left\{e_{it}^2-\left[e_{it}-\left(\tilde{\Delta}_{\Theta_0,it}+\sum_{j\in[p]}X_{j,it}\tilde{\Delta}_{\Theta_j,it}\right)\right]^2\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{2}{NT} tr\left(E'\tilde{\Delta}_{\Theta_0}\right) + \sum_{j\in[p]} \frac{2}{NT} tr\left((E\odot X_j)'\,\tilde{\Delta}_{\Theta_j}\right) - \frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\left(\tilde{\Delta}_{\Theta_0,it} + \sum_{j\in[p]} X_{j,it}\tilde{\Delta}_{\Theta_j,it}\right)^2 \\
&\le \frac{2}{NT}\left|tr\left(E'\tilde{\Delta}_{\Theta_0}\right)\right| + \sum_{j\in[p]}\frac{2}{NT}\left|tr\left((E\odot X_j)'\,\tilde{\Delta}_{\Theta_j}\right)\right| \\
&\le \frac{2}{NT}\left\|E\right\|_{op}\left\|\tilde{\Delta}_{\Theta_0}\right\|_* + \sum_{j\in[p]}\frac{2}{NT}\left\|E\odot X_j\right\|_{op}\left\|\tilde{\Delta}_{\Theta_j}\right\|_* \\
&\le 2c_3\sum_{j\in[p]\cup\{0\}}\frac{\sqrt{N}\vee\sqrt{T\log T}}{NT}\left\|\tilde{\Delta}_{\Theta_j}\right\|_*, \qquad \text{(B.7)}
\end{aligned}
$$

where the second inequality holds by the fact that $tr(AB) \le ||A||_{op}||B||_*$, and the last inequality is by the definition of event $\mathscr{A}_{1,N}$.

Combining (B.6) and (B.7), we have

$$
0 \le \sum_{j\in[p]\cup\{0\}}\left\{\frac{2c_3(\sqrt{N}\vee\sqrt{T\log T})}{NT}\left\|\tilde{\Delta}_{\Theta_j}\right\|_* + \nu_j\left(\|\Theta_j^0\|_* - \left\|\tilde{\Theta}_j\right\|_*\right)\right\}\ w.p.a.1. \qquad \text{(B.8)}
$$

Besides, we can show that

$$
\begin{aligned}
\left\|\tilde{\Theta}_j\right\|_* &= \left\|\tilde{\Delta}_{\Theta_j} + \Theta_j^0\right\|_* = \left\|\Theta_j^0 + \mathcal{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j}) + \mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_* \\
&\ge \left\|\Theta_j^0 + \mathcal{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_* - \left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_* = \|\Theta_j^0\|_* + \left\|\mathcal{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_* - \left\|\mathcal{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_*, \qquad \text{(B.9)}
\end{aligned}
$$

where the second equality holds by Lemma D.2(i) in Chernozhukov et al. (2020), the first inequality is by triangle inequality and the last equality is by the construction of the linear space $\mathcal{P}_j^{\perp}$ and $\mathcal{P}_j$. Then combining (B.8) and (B.9), w.p.a.1, we have

$$
\sum_{j\in[p]\cup\{0\}}\nu_j\left\|\tilde{\Theta}_j\right\|_* \le \sum_{j\in[p]\cup\{0\}}\left\{\nu_j\left\|\Theta_j^0\right\|_* + 2c_3\sum_{j\in[p]\cup\{0\}}\frac{(\sqrt{N}\vee\sqrt{T})}{NT}\left\|\tilde{\Delta}_{\Theta_j}\right\|_*\right\}
$$

and

$$
\begin{aligned}
\sum_{j\in[p]\cup\{0\}}\nu_j\left\{\left\|\mathcal{P}_j^{\perp}\left(\tilde{\Delta}_{\Theta_j}\right)\right\|_* - \left\|\mathcal{P}_j\left(\tilde{\Delta}_{\Theta_j}\right)\right\|_*\right\} &\le 2c_3\sum_{j\in[p]\cup\{0\}}\frac{(\sqrt{N}\vee\sqrt{T\log T})}{NT}\left\|\tilde{\Delta}_{\Theta_j}\right\|_* \\
&= 2c_3\sum_{j\in[p]\cup\{0\}}\frac{(\sqrt{N}\vee\sqrt{T\log T})}{NT}\left\{\left\|\mathcal{P}_j\left(\tilde{\Delta}_{\Theta_j}\right)\right\|_* + \left\|\mathcal{P}_j^{\perp}\left(\tilde{\Delta}_{\Theta_j}\right)\right\|_*\right\},
\end{aligned}
$$

If we set $\nu_j = \frac{4c_3(\sqrt{N}\vee\sqrt{T\log T})}{NT}$, we obtain the final result

$$
\sum_{j\in[p]\cup\{0\}}\left\|\mathcal{P}_j^{\perp}\left(\tilde{\Delta}_{\Theta_j}\right)\right\|_* \le 3\sum_{j\in[p]\cup\{0\}}\left\|\mathcal{P}_j\left(\tilde{\Delta}_{\Theta_j}\right)\right\|_*.
$$

■

Lemma B.5 *Consider a sequence of random variables $\{B_i, i = 1, \cdots, n\}$.*

(i) Suppose $\{B_i, i=1,\cdots,n\}$ is independent with $\mathbb{E}(B_i)=0$ and $\max_{i\in[n]}|B_i|\leq M$ a.s. Let $\sigma^2=\sum_{i\in[n]}\mathbb{E}\left(B_i^2\right)$. Then for all $t>0$, we have

$$\mathbb{P}\left(\left|\sum_{i\in[n]}B_i\right|>t\right)\leq\exp\left\{-\frac{t^2/2}{\sigma^2+Mt/3}\right\}.$$

(ii) Suppose $\{B_i, i=1,\cdots,n\}$ is sequence of martingale difference with $\mathbb{E}_{i-1}(B_i)=0$ and $\max_{i\in[n]}|B_i|\leq M$ a.s., where $\mathbb{E}_{i-1}$ denotes $\mathbb{E}(\cdot|\mathscr{F}_{i-1})$, where $\{\mathscr{F}_i: i\leq n\}$ denotes the filtration that is clear from the context. Let $\left|\sum_{i\in[n]}\mathbb{E}_{i-1}\left(B_i^2\right)\right|\leq\sigma^2$. Then for all $t>0$, we have

$$\mathbb{P}\left(\left|\sum_{i\in[n]}B_i\right|>t\right)\leq\exp\left\{-\frac{t^2/2}{\sigma^2+Mt/3}\right\}.$$

Proof. Lemma B.5(i) and (ii) are Bernstein inequality for i.i.d. sequence and Freedman inequality for m.d.s., which are respectively stated in Lemma 2.2.9 Vaart and Wellner (1996) and Theorem 1.1 Tropp (2011). ■

Lemma B.6 *Let $\{\Upsilon_t, t=1,\cdots,T\}$ be a zero-mean strong mixing process, not necessarily stationary, with the mixing coefficients satisfying $\alpha(z)\leq c_\alpha\gamma^z$ for some $c_\alpha>0$ and $\gamma\in(0,1)$. If $\sup_{t\in[T]}|\Upsilon_t|\leq M_T$, then there exists a constant c_4 depending on c_α and γ such that for any $T\geq 2$ and $\varepsilon>0$,*

(i) $\mathbb{P}\left\{\left|\sum_{t=1}^T\Upsilon_t\right|>\varepsilon\right\}\leq\exp\left\{-\frac{c_4\varepsilon^2}{M_T^2T+\varepsilon M_T(\log T)(\log\log T)}\right\}$,

(ii) $\mathbb{P}\left\{\left|\sum_{t=1}^T\Upsilon_t\right|>\varepsilon\right\}\leq\exp\left\{-\frac{c_4\varepsilon^2}{v_0^2T+M_T^2+\varepsilon M_T(\log T)^2}\right\}$,

where $v_0^2=\sup_{t\in[T]}\left[Var(\Upsilon_t)+2\sum_{s>t}|Cov(\Upsilon_t,\Upsilon_s)|\right]$.

Proof. The proof is the same as that of Theorems 1 and 2 in Merlevède et al. (2009) with the condition $\alpha(a)\leq\exp\{-2ca\}$ for some $c>0$. Here we can set $c=-\log\gamma$ if $c_\alpha\geq 1$ and $c=-\log(\gamma/c_\alpha)$ otherwise. ■

Lemma B.7 *Suppose Assumptions 1-4 hold, for $j\in\{0,\cdots,p\}$, we have*

(i) $\max_{i\in[N]}\left\|u_{i,j}^0\right\|_2\leq M$ *and* $\max_{t\in[T]}\left\|v_{t,j}^0\right\|_2\leq\frac{M}{\sigma_{K_j,j}}\leq\frac{M}{c_\sigma}$,

(ii) $\max_{t\in[T]}\left\|O_j'\tilde{v}_{t,j}\right\|_2\leq\frac{2M}{\sigma_{K_j,j}}\leq\frac{2M}{c_\sigma}$ *w.p.a.1,*

(iii) $\max_{i\in[N]}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\phi}_{it}\right\|_2^2\leq\frac{4M^2}{c_\sigma^2}(1+pC)$ *w.p.a.1,*

(iv) $\max_{i\in[N]}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\phi}_{it}-\phi_{it}^0\right\|^2=O_p(\eta_{N,1}^2(NT)^{2/q})$.

Proof. (i) Recall that $\frac{1}{\sqrt{NT}}\Theta_j^0=\mathcal{U}_j^0\Sigma_j^0\mathcal{V}_j^{0\prime}$ with $U_j^0=\sqrt{N}\mathcal{U}_j^0\Sigma_j^0$ and $V_j=\sqrt{T}\mathcal{V}_j$. Note that

$$\frac{1}{\sqrt{T}}\Theta_j^0\mathcal{V}_j^0=\sqrt{N}\mathcal{U}_j^0\Sigma_j^0=U_j^0,\text{ and }\frac{1}{\sqrt{N}}\mathcal{U}_j^{0\prime}\Theta_j^0=\sqrt{T}\Sigma_j^0\mathcal{V}_j^{0\prime}=\Sigma_j^0V_j^{0\prime}. \quad \text{(B.10)}$$

Hence, it's natural to see that

$$\left\|u_{i,j}^{0}\right\|_{2}=\frac{1}{\sqrt{T}}\left\|\left[\Theta_{j}^{0}\mathcal{V}_{j}^{0}\right]_{i.}\right\|_{2}\leq\frac{1}{\sqrt{T}}\left\|[\Theta_{j}]_{i.}\right\|_{2}\leq M,$$

where the first inequality is due to the fact that $\mathcal{V}_j$ is the unitary matrix and the last inequality holds by Assumption 2. Since the upper bound M is not dependent on i, this result holds uniformly. Analogously, we see that

$$\left\|v_{t,j}^{0}\right\|_{2}\leq\frac{1}{\sqrt{N}}c_{\sigma}^{-1}\left\|\left[\mathcal{U}_{j}^{0\prime}\Theta_{j}^{0}\right]_{.t}\right\|_{2}\leq\frac{1}{\sqrt{N}}c_{\sigma}^{-1}\left\|\left[\Theta_{j}^{0}\right]_{.t}\right\|_{2}\leq\frac{M}{c_{\sigma}}.$$

(ii) As in (B.10), we have

$$\frac{1}{\sqrt{N}}\tilde{\mathcal{U}}_{j}^{(1)\prime}\tilde{\Theta}_{j}=\sqrt{T}\tilde{\Sigma}_{j}^{(1)}\tilde{\mathcal{V}}_{j}^{(1)\prime}=\tilde{\Sigma}_{j}^{(1)}\tilde{V}_{j}^{(1)\prime},$$

and

$$\left\|O_{j}'\tilde{v}_{t,j}\right\|_{2}\leq\frac{1}{\sqrt{N}}\frac{1}{\tilde{\sigma}_{K_{j},j}^{(1)}}\left\|\left[\tilde{\mathcal{U}}_{j}^{(1)\prime}\tilde{\Theta}_{j}\right]_{.t}\right\|_{2}\leq\frac{1}{\sqrt{N}}\frac{1}{\tilde{\sigma}_{K_{j},j}^{(1)}}\left\|\left[\tilde{\Theta}_{j}\right]_{.t}\right\|_{2}\leq\frac{2M}{c_{\sigma}},$$

where the last inequality holds due to the constrained optimization in (3.1) and the fact that $\max_{k\in[K_j]}\left|\tilde{\sigma}_{k,j}^{-1}-\sigma_{k,j}^{-1}\right|\leq\sigma_{K_j,j}^{-1}$ w.p.a.1.

(iii) Note that

$$\begin{aligned}\max_{i\in[N]}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\phi}_{it}\right\|_{2}^{2} &\leq \max_{i\in[N]}\left\{\frac{1}{T}\sum_{t\in[T]}\left\|O_{0}'\tilde{v}_{t,0}\right\|_{2}^{2}+\sum_{j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left\|O_{j}'\tilde{v}_{t,j}\right\|_{2}^{2}\left|X_{j,it}\right|^{2}\right\}\\ &\leq \max_{t\in[T],j\in[p]\cup\{0\}}\left\|O_{j}'\tilde{v}_{t,j}\right\|_{2}^{2}\left\{1+\max_{i\in[N]}\sum_{j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left|X_{j,it}\right|^{2}\right\}\leq\frac{4M^{2}}{c_{\sigma}^{2}}\left(1+pC\right)\end{aligned}$$

where the last inequality holds by Lemma B.7(ii).

(iv) Note that

$$\begin{aligned}&\max_{i\in[N]}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\phi}_{it}-\phi_{it}^{0}\right\|^{2}\leq\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{O}_{0}^{(1)\prime}\tilde{v}_{t,0}-v_{t,0}^{0}\right\|_{2}^{2}+p\max_{t\in[T],j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left|X_{j,it}\right|^{2}\left\|\tilde{O}_{j}^{(1)\prime}\tilde{v}_{t,j}-v_{t,j}^{0}\right\|_{2}^{2}\\ &\lesssim\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{O}_{0}^{(1)\prime}\tilde{v}_{t,0}-v_{t,0}^{0}\right\|_{2}^{2}+p(NT)^{2/q}\max_{j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{O}_{j}^{(1)\prime}\tilde{v}_{t,j}-v_{t,j}^{0}\right\|_{2}^{2}\\ &=\frac{1}{T}\left\|O_{0}\tilde{V}_{0}-V_{0}^{0}\right\|_{F}^{2}+p(NT)^{2/q}\max_{j\in[p]}\frac{1}{T}\left\|O_{j}\tilde{V}_{j}-V_{j}^{0}\right\|_{F}^{2}=O_{p}\left(\eta_{N,1}^{2}(NT)^{2/q}\right),\end{aligned}$$

where the second inequality is by Assumption 1(v) and the last equality holds by Theorem 4.1(ii). ■

Lemma B.8 *Under Assumptions 1-5, we have* $\min_{i\in[N]}\lambda_{\min}\left(\tilde{\Phi}_i\right)\geq\frac{c_\phi}{2}$ *w.p.a.1, and* $\min_{t\in[T]}\lambda_{\min}\left(\tilde{\Psi}_t\right)\geq\frac{c_\phi}{2}$ *w.p.a.1.*

Proof. Recall that $\Phi_i=\frac{1}{T}\sum_{t=1}^{T}\phi_{it}^{0}\phi_{it}^{0\prime}$ and $\tilde{\Phi}_i=\frac{1}{T}\sum_{t=1}^{T}\tilde{\phi}_{it}\tilde{\phi}_{it}'$, where

$$\phi_{it}^{0}=\left(v_{t,0}^{0\prime},v_{t,1}^{0\prime}X_{1,it},\cdots,v_{t,p}^{0\prime}X_{p,it}\right)'\text{ and }\tilde{\phi}_{it}=\left[\left(O_{0}'\tilde{v}_{t,0}\right)',\left(O_{1}'\tilde{v}_{t,1}X_{1,it}\right)',\cdots,\left(O_{p}'\tilde{v}_{t,p}X_{p,it}\right)'\right]'.$$

Uniformly over $i \in [N]$, it is clear that

$$\begin{aligned}\left\|\tilde{\Phi}_i - \Phi_i\right\|_F &\lesssim \frac{4M}{c_\sigma T}\sum_{t=1}^{T}\left\|O_0'\tilde{v}_{t,0} - v_{t,0}^0\right\|_2 + \frac{4M}{c_\sigma T}\sum_{j=1}^{p}\sum_{t=1}^{T}\left\|O_j'\tilde{v}_{t,j} - v_{t,j}^0\right\|_2 |X_{j,it}| \\ &\leq \frac{4M}{c_\sigma}\frac{1}{\sqrt{T}}\left\|O_0'\tilde{V}_0 - V_0^0\right\|_F + \frac{4M^2}{c_\sigma}\sum_{j=1}^{p}\frac{1}{\sqrt{T}}\left\|O_j'\tilde{V}_j - V_j^0\right\|_F\left(\frac{1}{T}\sum_{t\in[T]}|X_{j,it}|^2\right)^{1/2} \\ &= O_p\left(\eta_{N,1}\right),\end{aligned}$$

where the third line holds by Lemma B.7(i) and Assumption 1(iv). It follows that

$$\min_{i\in[N]}\lambda_{\min}\left[\tilde{\Phi}_i\right] \geq \min_{i\in[N]}\lambda_{\min}\left[\Phi_i\right] - O\left(\eta_{N,1}\right) \geq \frac{c_\phi}{2}, \quad \text{w.p.a.1.}$$

Analogously, we can establish the lower bound of $\lambda_{\min}\left(\tilde{\Psi}_t\right)$. ■

Lemma B.9 *Under Assumptions 1-7, we have* $\max_{i\in[N]}\mathbf{1}\left\{\hat{g}_{i,K^{(\ell)}}^{0,(\ell)} \neq g_i^{(\ell)}\right\} = 0$ *w.p.a.1, where* $\hat{g}_{i,K^{(\ell)}}^{0,(\ell)}$ *is defined in* (A.21).

Proof. The above lemma holds by Theorem 2.3 in Su et al. (2020) provided we can verify the conditions in their Assumption 4. Let $\alpha_k^{(\ell)} = \left(\alpha_{k,1}^{(\ell)}, \cdots, \alpha_{k,p}^{(\ell)}\right)'$. Then we have

$$\beta_i^{0,(\ell)} = \frac{1}{\sqrt{T_\ell}}\sum_{k\in[K^{(\ell)}]}\alpha_k^{(\ell)} \otimes \iota_{T_\ell}\mathbf{1}\left\{g_i^{(\ell)} = k\right\}$$

and

$$\max_{k\in[K^{(\ell)}]}\left\|\frac{1}{\sqrt{T_\ell}}\alpha_k^{(\ell)} \otimes \iota_{T_\ell}\right\|_2 = \max_{k\in[K^{(\ell)}]}\frac{1}{\sqrt{T_\ell}}\sqrt{T_\ell\sum_{j=1}^{p}\left(\alpha_{k,j}^{(\ell)}\right)^2} \leq \sqrt{p}\max_{k\in[K^{(\ell)}],j\in[p]}\left|\alpha_{k,j}^{(\ell)}\right| \leq \sqrt{p}M, \tag{B.11}$$

where the last inequality is due to Assumption 2.

Second, with $\Theta_{j,i}^{0,(1)} = \left(\Theta_{j,i1}^0, \cdots, \Theta_{j,iT_1}^0\right)'$ and $\Theta_{j,i}^{0,(2)} = \left(\Theta_{j,i,T_1+1}^0, \cdots, \Theta_{j,iT}^0\right)'$, we notice that

$$\begin{aligned}\max_{i\in[N]}\left\|\dot{\beta}_i^{0,(\ell)} - \beta_i^{0,(\ell)}\right\|_2 &= \frac{1}{\sqrt{T_\ell}}\max_{i\in[N]}\left\|\dot{\Theta}_i^{(\ell)} - \Theta_i^{0,(\ell)}\right\|_2 \\ &= \frac{1}{\sqrt{T_\ell}}\max_{i\in[N]}\sqrt{\sum_{j=1}^{p}\sum_{\ell\in\{1,2\}}\sum_{t\in\mathcal{T}_\ell}\left(\dot{\Theta}_{j,it} - \Theta_{j,it}^0\right)^2\mathbf{1}\left\{\ell = 1\right\}} \\ &\leq \sqrt{p}\max_{j\in[p],i\in[N],t\in[T]}\left|\dot{\Theta}_{j,it} - \Theta_{j,it}^0\right| \leq c_5\eta_{N,2} \quad \text{w.p.a.1,}\end{aligned} \tag{B.12}$$

with c_5 being some positive large enough constant, and the last inequality holds by Theorem 4.1(iii).

Third, we also observe that

$$\min_{1\leq k_s<k_{s^*}\leq K^{(\ell)}}\frac{1}{\sqrt{T_\ell}}\left\|\alpha_{k_s}^{(\ell)} \otimes \iota_{T_\ell} - \alpha_{k_{s^*}}^{(\ell)} \otimes \iota_{T_\ell}\right\|_2 = \min_{1\leq k_s<k_{s^*}\leq K^{(\ell)}}\sqrt{\sum_{j=1}^{p}\left(\alpha_{k_s,j}^{(\ell)} - \alpha_{k_{s^*},j}^{(\ell)}\right)^2} \geq C_5, \tag{B.13}$$

where the last inequality is by Assumption 7(i).

Combining (B.11), (B.12) and (B.13), we obtain that $\mathbb{P}\left(\max_{i\in[N]}\mathbf{1}\left\{\hat{g}^{0,(\ell)}_{i,K^{(\ell)}}\neq g^{(\ell)}_i\right\}=0\right)\to 1$ once we can ensure Assumption 4.3 in Su et al. (2020) holds with $c_{1n}=C_5$, $c_{2n}=c_5\eta_{N,2}$, $K=K^{(1)}$, and with their c_1 and M being replaced by $\underline{c}$ and and $\sqrt{p}M$ here. Under Assumption 7, Assumption 4.3 in Su et al. (2020) holds. This completes the proof of the lemma. ■

To study the NSP property of our group structure estimator, we introduce some notation in the following definition.

Definition B.10 *Fix $K^{(\ell)}>1$ and $1<m\leq K^{(\ell)}$. Define a $K^{(\ell)}\times p$ matrix $\alpha^{(\ell)}=\left(\alpha^{(\ell)}_1,\cdots,\alpha^{(\ell)}_K\right)'$. Let $d_{K^{(\ell)}}(\alpha^{(\ell)})$ be the minimum pairwise distance of all $K^{(\ell)}$ rows and $\alpha^{(\ell)}_k$ and $\alpha^{(\ell)}_l$ be the pair that satisfies $\left\|\alpha^{(\ell)}_k-\alpha^{(\ell)}_l\right\|_2=d_{K^{(\ell)}}(\alpha^{(\ell)})$ (if this holds for multiple pairs, pick the first pair in the lexicographical order). Remove row l from matrix $\alpha^{(\ell)}$ and let $d_{K^{(\ell)}-1}(\alpha^{(\ell)})$ be the minimum pairwise distance for the remaining $(K^{(\ell)}-1)$ rows. Repeat this step and define $d_{K^{(\ell)}-2}(\alpha^{(\ell)}),\cdots,d_2(\alpha^{(\ell)})$ recursively.*

Lemma B.11 *Recall that $\hat{\mathcal{G}}^{(\ell)}_m$ is the estimated group structure from K-means algorithm with m groups. Under Assumptions 1-7 and the event $\left\{\hat{T}_1=T_1\right\}$, w.p.a.1, for each $1<m<K^{(\ell)}$, $\hat{\mathcal{G}}^{(\ell)}_m$ enjoys the NSP defined in Definition 4.3.*

Proof. By Theorem 4.1 in Jin et al. (2022), Lemma B.11 is proved if we ensure all conditions in their Theorem 4.1 hold. We now apply their Theorem 4.1 with $\hat{x}_i=\dot{\beta}^{0,(\ell)}_i$, $x_i=\beta^{0,(\ell)}_i$ and $u_k=\frac{1}{\sqrt{T_\ell}}\alpha^{(\ell)}_k\otimes\iota_{T_\ell}$ for $k\in[K^{(\ell)}]$. By the definition of $d_m\left(\alpha^{(\ell)}\right)$ in Definition B.10, we notice that $d_m\left(\alpha^{(\ell)}\right)\geq d_{K^{(\ell)}}\left(\alpha^{(\ell)}\right)$ such that $d_{K^{(\ell)}}\left(\alpha^{(\ell)}\right)\geq C_5$ by Assumption 7(i). With (B.12) shown above and Assumption 2, we have

$$\max_{k\in[K^{(\ell)}]}\|u_k\|_2\leq M,\quad \max_{i\in[N]}\|\hat{x}_i-x_i\|_2=O_p(\eta_{N,2}),$$

which satisfy the Theorem 4.1 in Jin et al. (2022), i.e. $\max_{k\in[K^{(\ell)}]}\|u_k\|_2\lesssim d_m\left(\alpha^{(\ell)}\right)$ and $\max_{i\in[N]}\|\hat{x}_i-x_i\|_2\lesssim d_m\left(\alpha^{(\ell)}\right)$. Consequently, it leads to the NSP of $\hat{\mathcal{G}}^{(\ell)}_m$ for $1<m<K^{(\ell)}$ w.p.a.1 under the event $\left\{\hat{T}_1=T_1\right\}$. ■

Lemma B.12 *Under Assumptions 1, 6(ii), 7(ii), 8 and 9(i)-(iii), for $\ell=\{1,2\}$ and $k\in[K^{(\ell)}]$, we have $\hat{\alpha}^{(\ell)}_k\xrightarrow{p}\alpha^{(\ell)}_k$ and $\sqrt{N^{(\ell)}_kT_\ell}\left(\hat{\alpha}^{(\ell)}_k-\alpha^{(\ell)}_k\right)=\mathbb{W}^{(\ell)-1}_{NT,k}\mathbb{C}^{(\ell)}_{NT,k}+o_p(1)$.*

Proof. This Lemma combines Theorem 1 and Corollary 4.2 in Moon and Weidner (2017) under their Assumptions 1-4. Hence, we only need to verify the conditions in their Assumptions 2 and 3 since Assumptions 8 and 9(ii)-(iii) are same as their Assumptions 1 and 4.

Notice that the Assumption 2 in Moon and Weidner (2017) holds if we can show that

$$\frac{1}{N^{(\ell)}_kT_\ell}\sum_{i\in G^{(\ell)}_k}\sum_{t\in\mathcal{T}_\ell}X_{j,it}e_{it}\xrightarrow{p}0,\ \forall k\in[K^{(\ell)}],\ \ell\in\{1,2\}.$$

Fix a specific k and ℓ. We can show that

$$\mathbb{E}\left(\frac{1}{N^{(\ell)}_kT_\ell}\sum_{i\in G^{(\ell)}_k}\sum_{t\in\mathcal{T}_\ell}X_{j,it}e_{it}\middle|\mathscr{D}\right)^2$$

$$
\begin{aligned}
&= \frac{1}{\left(N_k^{(\ell)}T_\ell\right)^2} \sum_{i_1\in G_k^{(\ell)}} \sum_{i_2\in G_k^{(\ell)}} \sum_{t_1\in\mathcal{T}_\ell} \sum_{t_2\in\mathcal{T}_\ell} \mathbb{E}\left(X_{j,i_1t_1}X_{j,i_2t_2}e_{i_1t_1}e_{i_2t_2}|\mathscr{D}\right) \\
&= \frac{1}{\left(N_k^{(\ell)}T_\ell\right)^2} \sum_{i\in G_k^{(\ell)}} \sum_{t_1\in\mathcal{T}_\ell} \sum_{t_2\in\mathcal{T}_\ell} \mathbb{E}\left(X_{j,it_1}X_{j,it_2}e_{it_1}e_{it_2}|\mathscr{D}\right) \\
&= \frac{1}{\left(N_k^{(\ell)}T_\ell\right)^2} \sum_{i\in G_k^{(\ell)}} \sum_{t\in\mathcal{T}_\ell} \mathbb{E}\left(X_{j,it}^2e_{it}^2|\mathscr{D}\right) + \frac{2}{\left(N_k^{(\ell)}T_\ell\right)^2} \sum_{i\in G_k^{(\ell)}} \sum_{t_1\in\mathcal{T}_\ell} \sum_{t_2\in\mathcal{T}_\ell, t_2>t_1} \mathbb{E}\left(X_{j,it_1}X_{j,it_2}e_{it_1}e_{it_2}|\mathscr{D}\right) \\
&\leq \frac{M}{N_k^{(\ell)}T_\ell} + \frac{16}{\left(N_k^{(\ell)}T_\ell\right)^2} \max_{i\in G_k^{(\ell)}} \max_{t\in\mathcal{T}_L} \left(\mathbb{E}\left|X_{j,it}e_{it}\right|^q\right)^{2/q} \sum_{i\in G_k^{(\ell)}} \sum_{t_1\in\mathcal{T}_\ell} \sum_{t_2\in\mathcal{T}_\ell, t_2>t_1} \left[\alpha(t_2-t_1)\right]^{1-2/q} \\
&= O\left(\frac{1}{NT}\right),
\end{aligned} \tag{B.14}
$$

where the second equality holds by Assumption 1(i) with the conditional independence sequence for $i_1 \neq i_2$, the first inequality combines Assumption 1(ii), (iii), (v), and the Davydov's inequality for strong mixing sequence in Lemma 4.3, Su and Chen (2013), and the last equality is by Assumption 1(iii), (v), Assumption 6(ii) and Assumption 7(ii). Following this, it yields that

$$
\frac{1}{N_k^{(\ell)}T_\ell} \sum_{i\in G_k^{(\ell)}} \sum_{t\in\mathcal{T}_\ell} X_{j,it}e_{it} = O_p\left((NT)^{-1/2}\right).
$$

By similar arguments as used in the proof of Lemma B.3, we can show that

$$
\left\| E_k^{(\ell)} \right\|_{op} = O_p\left(\sqrt{N} + \sqrt{T\log T}\right), \tag{B.15}
$$

which, in conjunction with Assumption 9(i), implies that Assumption 3* in Moon and Weidner (2017) is satisfied. ■

For $j \in [p]$, recall that $X_{j,i}^{(1)} = (X_{j,i1}, \cdots, X_{j,iT_1})'$, $X_{j,i}^{(2)} = \left(X_{j,i(T_1+1)}, \cdots, X_{j,iT}\right)'$, $e_i^{(1)} = (e_{i1}, \cdots, e_{iT_1})'$, $e_i^{(2)} = \left(e_{i(T_1+1)}, \cdots, e_{iT}\right)'$, $\tilde{X}_{j,it} = X_{j,it} - \mathbb{E}\left(X_{j,it}|\mathscr{D}\right)$. Besides, let $\mathbb{X}_{j,k}^{(\ell)} \in \mathbb{R}^{N_k^{(\ell)}\times T_\ell}$ and $E_k^{(\ell)} \in \mathbb{R}^{N_k^{(\ell)}\times T_\ell}$ denote the regressor and error matrix for subgroup $k \in [K^{(\ell)}]$ with a typical row being $X_{j,i}^{(\ell)}$ and $e_i^{(\ell)}$, respectively. For $\ell \in \{1,2\}$ and $k \in [K^{(\ell)}]$, we also define

$$
\bar{\mathbb{X}}_{j,k}^{(\ell)} = \mathbb{E}\left(\mathbb{X}_{j,k}^{(\ell)}|\mathscr{D}\right), \quad \tilde{\mathbb{X}}_{j,k}^{(\ell)} = \mathbb{X}_{j,k}^{(\ell)} - \bar{\mathbb{X}}_{j,k}^{(\ell)}, \quad \mathfrak{X}_{j,k}^{(\ell)} = M_{\Lambda_k^{0,(\ell)}}\bar{\mathbb{X}}_{j,k}^{(\ell)}M_{F^{0,(\ell)}} + \tilde{\mathbb{X}}_{j,k}^{(\ell)},
$$

with $\mathfrak{X}_{j,k,it}^{(\ell)}$ being each entry of $\mathfrak{X}_{j,k}^{(\ell)}$. Further let $\mathfrak{X}_{k,it}^{(\ell)} = \left(\mathfrak{X}_{1,k,it}^{(\ell)}, \cdots, \mathfrak{X}_{p,k,it}^{(\ell)}\right)'$.

Lemma B.13 *Under Assumptions 1, 2, 6(ii), 7(ii), 8 and 9, for $j \in [p]$, $\ell \in \{1,2\}$ and $k \in [K^{(\ell)}]$, we have*

(i) $\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left(P_{F^{0,(\ell)}}E_k^{(\ell)\prime}P_{\Lambda_k^{0,(\ell)}}\tilde{\mathbb{X}}_{j,k}^{(\ell)}\right) = o_p(1),$

(ii) $\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left(P_{\Lambda_k^{0,(\ell)}}E_k^{(\ell)\prime}\tilde{\mathbb{X}}_{j,k}^{(\ell)}\right) = o_p(1),$

(iii) $\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left\{P_{F^{0,(\ell)}}\left[E_k^{(\ell)\prime}\tilde{\mathbb{X}}_{j,k}^{(\ell)} - \mathbb{E}\left(E_k^{(\ell)\prime}\tilde{\mathbb{X}}_{j,k}^{(\ell)}|\mathscr{D}\right)\right]\right\} = o_p(1),$

(iv) $\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left[E_k^{(\ell)}P_{F^{0,(\ell)}}E_k^{(\ell)\prime}M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}F^{0,(\ell)}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\Lambda_k^{0,(\ell)\prime}\right] = o_p(1),$

(v) $\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left[E_k^{(\ell)\prime}P_{\Lambda_k^{0,(\ell)}}E_k^{(\ell)}M_{F^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)\prime}\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right] = o_p(1),$

(vi) $\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left[E_k^{(\ell)\prime}M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}M_{F^{0,(\ell)}}E_k^{(\ell)\prime}\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right] = o_p(1),$

(vii) $\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left\{\left[E_k^{(\ell)}E_k^{(\ell)\prime} - \mathbb{E}\left(E_k^{(\ell)}E_k^{(\ell)\prime}|\mathscr{D}\right)\right]M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}F^{0,(\ell)}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\Lambda_k^{0,(\ell)\prime}\right\} = o_p(1),$

(viii) $\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left\{\left[E_k^{(\ell)\prime}E_k^{(\ell)} - \mathbb{E}\left(E_k^{(\ell)\prime}E_k^{(\ell)}|\mathscr{D}\right)\right]M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)\prime}\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right\} = o_p(1),$

(ix) $\frac{1}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathcal{T}_\ell}\left[e_{it}^2\mathfrak{X}_{k,it}^{(\ell)}\mathfrak{X}_{k,it}^{(\ell)\prime} - \mathbb{E}\left(e_{it}^2\mathfrak{X}_{k,it}^{(\ell)}\mathfrak{X}_{k,it}^{(\ell)\prime}|\mathscr{D}\right)\right] = o_p(1),$

(x) $\frac{1}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathcal{T}_\ell}e_{it}^2\left(\mathfrak{X}_{k,it}^{(\ell)}\mathfrak{X}_{k,it}^{(\ell)\prime} - \mathcal{X}_{it}\mathcal{X}_{it}^\prime\right) = o_p(1).$

Proof. (i) We first show that $\left\|F^{0,(\ell)\prime}E_k^{(\ell)\prime}\Lambda_k^{0,(\ell)}\right\|_F = O_p(\sqrt{NT})$. Note that

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{\left\|F^{0,(\ell)\prime}E_k^{(\ell)\prime}\Lambda_k^{0,(\ell)}\right\|_F}{\sqrt{N_k^{(\ell)}T_\ell}}\right)^2\middle|\mathscr{D}\right] = \frac{1}{N_k^{(\ell)}T_\ell}\mathbb{E}\left[\left(\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathcal{T}_\ell}e_{it}f_t^{0\prime}\lambda_i^0\right)^2\middle|\mathscr{D}\right]\\
&= \frac{1}{N_k^{(\ell)}T_\ell}\sum_{i_1\in G_k^{(\ell)}}\sum_{i_2\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_\ell}\sum_{t_2\in\mathcal{T}_\ell}\mathbb{E}\left(e_{i_1t_1}e_{i_2t_2}|\mathscr{D}\right)f_{t_1}^{0\prime}\lambda_{i_1}^0\lambda_{i_2}^{0\prime}f_{t_2}^0\\
&\leq \max_{i\in G_k^{(\ell)}}\left\|\lambda_i^0\right\|_2^2\max_{t\in\mathcal{T}_\ell}\left\|f_t^0\right\|_2^2\frac{1}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_\ell}\sum_{t_2\in\mathcal{T}_\ell}\left|\mathbb{E}\left(e_{it_1}e_{it_2}|\mathscr{D}\right)\right|\\
&\lesssim \frac{1}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathcal{T}_\ell}\left|Var\left(e_{it}|\mathscr{D}\right)\right| + \frac{2}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_\ell}\sum_{t_2\in\mathcal{T}_\ell,t_2>t_1}\left|Cov\left(e_{it_1},e_{it_2}|\mathscr{D}\right)\right|\\
&= O(1)\ a.s.,
\end{aligned}$$

where the fourth line is by Lemma B.7(i) and the last line combines Assumption 1(v) and Davydov's inequality for conditional strong mixing sequences, similarly as (B.14). It follows that

$$\begin{aligned}
\left\|P_{F^{0,(\ell)}}E_k^{(\ell)\prime}P_{\Lambda_k^{0,(\ell)}}\right\|_F &\leq \left\|F^{0,(\ell)}\right\|_F\left\|\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)\right\|_F\left\|F^{0,(\ell)\prime}E_k^{(\ell)\prime}\Lambda_k^{0,(\ell)}\right\|_F\left\|\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\right\|_F\left\|\Lambda_k^{0,(\ell)\prime}\right\|_F\\
&= O(T^{1/2})O_p(T^{-1})O_p(\sqrt{NT})O_p(N^{-1})O(N^{1/2}) = O_p(1), \qquad \text{(B.16)}
\end{aligned}$$

where the last line combines Assumptions 2 and 8.

Moreover, we have

$$\left\|P_{\Lambda_k^{0,(\ell)}}\tilde{\mathbb{X}}_{j,k}^{(\ell)}\right\|_F \leq \left\|\Lambda_k^{0,(\ell)}\right\|_F\left\|\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\right\|_F\left\|\Lambda_k^{0,(\ell)\prime}\tilde{\mathbb{X}}_{j,k}^{(\ell)}\right\|_F = O(N^{1/2})O_p(N^{-1})O_p(\sqrt{NT}) = O_p(T^{1/2}), \qquad \text{(B.17)}$$

where the first equality combines Assumption 2, Assumption 8(i) and the fact that

$$\mathbb{E}\left(\left\|\Lambda_k^{0,(\ell)\prime}\tilde{\mathbb{X}}_{j,k}^{(\ell)}\right\|_F^2|\mathscr{D}\right) = \sum_{r=1}^{r_0}\sum_{t\in\mathcal{T}_\ell}\mathbb{E}\left[\left(\sum_{i\in G_k^{(\ell)}}\lambda_{i,r}^0\tilde{X}_{j,it}\right)^2\middle|\mathscr{D}\right]$$

$$= \sum_{r=1}^{r_0} \sum_{t\in\mathcal{T}_\ell} \sum_{i\in G_k^{(\ell)}} \sum_{i^*\in G_k^{(\ell)}} \lambda_{i,r}^0 \lambda_{i^*,r}^0 \mathbb{E}\left(\tilde{X}_{j,it}\tilde{X}_{j,i^*t}|\mathscr{D}\right)$$

$$= \sum_{r=1}^{r_0} \sum_{t\in\mathcal{T}_\ell} \sum_{i\in G_k^{(\ell)}} \left(\lambda_{i,r}^0\right)^2 \mathbb{E}\left[\left(\tilde{X}_{j,it}\right)^2 |\mathscr{D}\right] = O_p(NT).$$

Then we are ready to show that

$$\left|\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left(P_{F^{0,(\ell)}} E_k^{(\ell)\prime} P_{\Lambda_k^{0,(\ell)}} \tilde{\mathbb{X}}_{j,k}^{(\ell)}\right)\right| = \left|\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left(P_{F^{0,(\ell)}} E_k^{(\ell)\prime} P_{\Lambda_k^{0,(\ell)}} P_{\Lambda_k^{0,(\ell)}} \tilde{\mathbb{X}}_{j,k}^{(\ell)}\right)\right|$$
$$\leq \frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} \left\|P_{F^{0,(\ell)}} E_k^{(\ell)\prime} P_{\Lambda_k^{0,(\ell)}}\right\|_F \left\|P_{\Lambda_k^{0,(\ell)}} \tilde{\mathbb{X}}_{j,k}^{(\ell)}\right\|_F$$
$$= \frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} O_p(1) O_p(T^{1/2}) = o_p(1).$$

(ii) Let $[A]_{jl}$ denote the (j,l)-th element of A. Note that

$$\left|\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left(P_{\Lambda_k^{0,(\ell)}} E_k^{(\ell)\prime} \tilde{\mathbb{X}}_{j,k}^{(\ell)}\right)\right|$$
$$= \left|\sum_{j_1,j_2=1}^{r_0} \left[\left(\frac{\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}}{N_k^{(\ell)}}\right)^{-1}\right]_{j_1j_2} \frac{1}{N_k^{(\ell)}\sqrt{N_k^{(\ell)}T_\ell}} \sum_{t\in\mathcal{T}_\ell} \sum_{i_1\in G_k^{(\ell)}} \sum_{i_2\in G_k^{(\ell)}} \lambda_{i_1,j_1}^0 \lambda_{i_2,j_2}^0 e_{i_1t} \tilde{X}_{j,i_2t}^{(\ell)}\right|$$
$$\lesssim \max_{j_1,j_2\in[r_0]} \left|\frac{1}{N_k^{(\ell)}\sqrt{N_k^{(\ell)}T_\ell}} \sum_{t\in\mathcal{T}_\ell} \sum_{i_1\in G_k^{(\ell)}} \sum_{i_2\in G_k^{(\ell)}} \lambda_{i_1,j_1}^0 \lambda_{i_2,j_2}^0 e_{i_1t} \tilde{X}_{j,i_2t}^{(\ell)}\right| = O_p(N^{-1/2}),$$

where the last line is owing to the fact that

$$\mathbb{E}\left(\left|\frac{1}{N_k^{(\ell)}\sqrt{N_k^{(\ell)}T_\ell}} \sum_{t\in\mathcal{T}_\ell} \sum_{i_1\in G_k^{(\ell)}} \sum_{i_2\in G_k^{(\ell)}} \lambda_{i_1,j_1}^0 \lambda_{i_2,j_2}^0 e_{i_1t} \tilde{X}_{j,i_2t}^{(\ell)}\right|^2 \middle| \mathscr{D}\right)$$
$$= \frac{1}{\left(N_k^{(\ell)}\right)^3 T_\ell} \sum_{t\in\mathcal{T}_\ell} \sum_{s\in\mathcal{T}_\ell} \sum_{i_1\in G_k^{(\ell)}} \sum_{i_2\in G_k^{(\ell)}} \sum_{m_1\in G_k^{(\ell)}} \sum_{m_2\in G_k^{(\ell)}} \lambda_{i_1,j_1}^0 \lambda_{i_2,j_2}^0 \lambda_{m_1,j_1}^0 \lambda_{m_2,j_2}^0 \mathbb{E}\left(e_{i_1t}\tilde{X}_{j,i_2t}^{(\ell)} e_{m_1s}\tilde{X}_{j,m_2s}^{(\ell)}|\mathscr{D}\right)$$
$$= \frac{1}{\left(N_k^{(\ell)}\right)^3 T_\ell} \sum_{t\in\mathcal{T}_\ell} \sum_{s\in\mathcal{T}_\ell} \sum_{i_1\in G_k^{(\ell)}} \sum_{i_2\in G_k^{(\ell)}} \left(\lambda_{i_1,j_1}^0\right)^2 \left(\lambda_{i_2,j_2}^0\right)^2 \mathbb{E}\left(e_{i_1t}e_{i_1s}\tilde{X}_{j,i_2t}^{(\ell)}\tilde{X}_{j,i_2s}^{(\ell)}|\mathscr{D}\right)$$
$$= \frac{1}{\left(N_k^{(\ell)}\right)^3 T_\ell} \sum_{t\in\mathcal{T}_\ell} \sum_{i_1\in G_k^{(\ell)}} \sum_{i_2\in G_k^{(\ell)}} \left(\lambda_{i_1,j_1}^0\right)^2 \left(\lambda_{i_2,j_2}^0\right)^2 \mathbb{E}\left(e_{i_1t}^2 \left(\tilde{X}_{j,i_2t}^{(\ell)}\right)^2 \middle| \mathscr{D}\right)$$
$$+ \frac{2}{\left(N_k^{(\ell)}\right)^3 T_\ell} \sum_{t\in\mathcal{T}_\ell} \sum_{s\in\mathcal{T}_\ell, s>t} \sum_{i_1\in G_k^{(\ell)}} \sum_{i_2\in G_k^{(\ell)}} \left(\lambda_{i_1,j_1}^0\right)^2 \left(\lambda_{i_2,j_2}^0\right)^2 \mathbb{E}\left(e_{i_1t}e_{i_1s}\tilde{X}_{j,i_2t}^{(\ell)}\tilde{X}_{j,i_2s}^{(\ell)}|\mathscr{D}\right)$$
$$= O_p(N^{-1}),$$

where the second equality is by Assumption 1(i) and the last line holds by Assumption 1(iii), (v), and Davydov's inequality.

(iii) Define $\zeta_{j,its}^{(\ell)} := e_{it}\tilde{\mathbb{X}}_{j,is}^{(\ell)} - \mathbb{E}\left(e_{it}\tilde{\mathbb{X}}_{j,is}^{(\ell)}|\mathscr{D}\right)$. As above, we have

$$\begin{aligned}
&\mathbb{E}\left\{\left|\frac{1}{T_\ell\sqrt{N_k^{(\ell)}T_\ell}}\sum_{i_1\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_\ell}\sum_{t_2\in\mathcal{T}_\ell}f_{t_1,j_1}^0 f_{t_2,j_2}^0\zeta_{j,i_1t_1t_2}^{(\ell)}\right|^2\Bigg|\mathscr{D}\right\}\\
&=\frac{1}{T_\ell^3N_k^{(\ell)}}\sum_{i_1\in G_k^{(\ell)}}\sum_{i_2\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_\ell}\sum_{t_2\in\mathcal{T}_\ell}\sum_{s_1\in\mathcal{T}_\ell}\sum_{s_2\in\mathcal{T}_\ell}f_{t_1,j_1}^0f_{t_2,j_2}^0f_{s_1,j_1}^0f_{s_2,j_2}^0\mathbb{E}\left(\zeta_{j,i_1t_1t_2}^{(\ell)}\zeta_{j,i_2s_1s_2}^{(\ell)}|\mathscr{D}\right\}\\
&\lesssim\frac{1}{T_\ell^3N_k^{(\ell)}}\sum_{i_1\in G_k^{(\ell)}}\sum_{i_2\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_\ell}\sum_{t_2\in\mathcal{T}_\ell}\sum_{s_1\in\mathcal{T}_\ell}\sum_{s_2\in\mathcal{T}_\ell}\left|Cov\left(e_{i_1t_1}\tilde{X}_{j,i_1t_2}^{(\ell)},e_{i_2s_1}\tilde{X}_{j,i_2s_2}^{(\ell)}|\mathscr{D}\right)\right|\\
&=\frac{1}{T_\ell^3N_k^{(\ell)}}\sum_{i\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_\ell}\sum_{t_2\in\mathcal{T}_\ell}\sum_{s_1\in\mathcal{T}_\ell}\sum_{s_2\in\mathcal{T}_\ell}\left|Cov\left(e_{it_1}\tilde{X}_{j,it_2}^{(\ell)},e_{is_1}\tilde{X}_{j,is_2}^{(\ell)}|\mathscr{D}\right)\right|\\
&=O_p(T^{-1}),
\end{aligned}$$

where the last line is by Assumption 9(iv). It follows that

$$\begin{aligned}
&\left|\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}}tr\left\{P_{F^{0,(\ell)}}\left[E_k^{(\ell)\prime}\tilde{\mathbb{X}}_{j,k}^{(\ell)}-\mathbb{E}\left(E_k^{(\ell)\prime}\tilde{\mathbb{X}}_{j,k}^{(\ell)}|\mathscr{D}\right)\right]\right\}\right|\\
&=\left|\sum_{j_1,j_2=1}^{r_0}\left[\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\right]_{j_1j_2}\frac{1}{T_\ell\sqrt{N_k^{(\ell)}T_\ell}}\sum_{i_1\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_1}\sum_{t_2\in\mathcal{T}_1}f_{t_1,j_1}^0f_{t_2,j_2}^0\zeta_{j,i_1t_1t_2}^{(\ell)}\right|\\
&=O_p(T^{-1/2}).
\end{aligned}$$

(iv) As in Moon and Weidner (2017), it is clear that

$$\begin{aligned}
&\left|\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}}tr\left[E_k^{(\ell)}P_{F^{0,(\ell)}}E_k^{(\ell)\prime}M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}F^{0,(\ell)}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\Lambda_k^{0,(\ell)\prime}\right]\right|\\
&\lesssim\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}}\left\|P_{\Lambda_k^{0,(\ell)}}E_k^{(\ell)}P_{F^{0,(\ell)}}\right\|_F\left\|E_k^{(\ell)}\right\|_{op}\left\|\mathbb{X}_{j,k}^{(\ell)}\right\|_F\left\|F^{0,(\ell)}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\Lambda_k^{0,(\ell)\prime}\right\|_F\\
&=\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}}O_p(1)O_p\left(\sqrt{N}+\sqrt{T\log T}\right)O_p((NT)^{1/2})O_p((NT)^{-1/2})=o_p(1),
\end{aligned}$$

where the last line combines (B.15),(B.16), the fact that $\left\|\mathbb{X}_{j,k}^{(\ell)}\right\|_F=O_p((NT)^{1/2})$ by Assumption 8(ii), and $\left\|F^{0,(\ell)}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\Lambda_k^{0,(\ell)\prime}\right\|_F=O_p((NT)^{-1/2})$ by Assumptions 2 and 8(i).

(v) The proof of (v) is analogous to that of (iv) and omitted for brevity.

(vi) First, we note that

$$\begin{aligned}
\mathbb{E}\left(\left\|\Lambda_k^{0,(\ell)\prime}E_k^{(\ell)}\mathbb{X}_{j,k}^{(\ell)\prime}\right\|_F\Bigg|\mathscr{D}\right)&=\mathbb{E}\left[\sum_{j_1=1}^{r_0}\sum_{m\in G_k^{(\ell)}}\left(\sum_{i_1\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_\ell}\lambda_{i_1,j_1}^0e_{i_1t_1}X_{j,mt_1}\right)^2\Bigg|\mathscr{D}\right]\\
&=\sum_{j_1=1}^{r_0}\sum_{m\in G_k^{(\ell)}}\sum_{i_1\in G_k^{(\ell)}}\sum_{i_2\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_\ell}\sum_{t_2\in\mathcal{T}_\ell}\lambda_{i_1,j_1}^0\lambda_{i_2,j_1}^0\mathbb{E}\left(e_{i_1t_1}X_{j,mt_1}e_{i_2t_2}X_{j,mt_2}|\mathscr{D}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j_1=1}^{r_0} \sum_{m\in G_k^{(\ell)}} \sum_{i_1\in G_k^{(\ell)}} \sum_{t_1\in\mathcal{T}_\ell} \sum_{t_2\in\mathcal{T}_\ell} \left(\lambda_{i_1,j_1}^0\right)^2 \mathbb{E}\left(e_{i_1t_1}X_{j,mt_1}e_{i_1t_2}X_{j,mt_2}|\mathscr{D}\right) \\
&\lesssim \sum_{m\in G_k^{(\ell)}} \sum_{i_1\in G_k^{(\ell)}} \sum_{t\in\mathcal{T}_\ell} \mathbb{E}\left(e_{i_1t}^2X_{j,mt}^2|\mathscr{D}\right) \\
&+ 2\sum_{m\in G_k^{(\ell)}} \sum_{i_1\in G_k^{(\ell)}} \sum_{t_1\in\mathcal{T}_\ell} \sum_{t_2\in\mathcal{T}_\ell, t_2>t_1} \left|Cov\left(e_{i_1t_1}X_{j,mt_1}, e_{i_1t_2}X_{j,mt_2}|\mathscr{D}\right)\right| \\
&= O_p(N^2T),
\end{aligned}
$$

which leads to the result that

$$
\left\|P_{\Lambda_k^{0,(\ell)}}E_k^{(\ell)}\mathbb{X}_{j,k}^{(\ell)\prime}\right\|_F \le \left\|\Lambda_k^{0,(\ell)}\right\|_F \left\|\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\right\|_F \left\|\Lambda_k^{0,(\ell)\prime}E_k^{(\ell)}\mathbb{X}_{j,k}^{(\ell)\prime}\right\|_F = O(N^{-1/2})O_p\left(N\sqrt{T}\right) = O_p\left(\sqrt{NT}\right).
$$

As in the proof of part (iv), it yields that

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left[E_k^{(\ell)\prime}M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}M_{F^{0,(\ell)}}E_k^{(\ell)\prime}\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right]\right| \\
&\le \left|\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left[E_k^{(\ell)\prime}M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}E_k^{(\ell)\prime}P_{\Lambda_k^{0,(\ell)}}\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right]\right| \\
&+ \left|\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left[E_k^{(\ell)\prime}M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}P_{F^{0,(\ell)}}E_k^{(\ell)\prime}P_{\Lambda_k^{0,(\ell)}}\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right]\right| \\
&\lesssim \frac{1}{\sqrt{N_k^{(\ell)}T_\ell}}\left\|E_k^{(\ell)}\right\|_{op}\left\|P_{\Lambda_k^{0,(\ell)}}E_k^{(\ell)}\mathbb{X}_{j,k}^{(\ell)\prime}\right\|_F\left\|\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right\|_F \\
&+ \frac{1}{\sqrt{N_k^{(\ell)}T_\ell}}\left\|E_k^{(\ell)}\right\|_{op}\left\|\mathbb{X}_{j,k}^{(\ell)}\right\|_F\left\|\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right\|_F\left\|P_{\Lambda_k^{0,(\ell)}}E_k^{(\ell)}P_{F^{0,(\ell)}}\right\|_F \\
&= o_p(1).
\end{aligned}
$$

(vii) For this statement, we sketch the proof because Lu and Su (2016) have already proved a similar result.

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{N_k^{(\ell)}T_\ell}} tr\left\{\left[E_k^{(\ell)}E_k^{(\ell)\prime} - \mathbb{E}\left(E_k^{(\ell)}E_k^{(\ell)\prime}|\mathscr{D}\right)\right]M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}F^{0,(\ell)}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\Lambda_k^{0,(\ell)\prime}\right\}\right| \\
&\lesssim \frac{1}{\left(N_k^{(\ell)}\right)^{3/2}}\left\|\Lambda_k^{0,(\ell)\prime}\frac{1}{T_\ell}\left[E_k^{(\ell)}E_k^{(\ell)\prime} - \mathbb{E}\left(E_k^{(\ell)}E_k^{(\ell)\prime}|\mathscr{D}\right)\right]M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}\right\|_F = o_p(1),
\end{aligned}
$$

where the last equality holds by the fact that

$$
\left(N_k^{(\ell)}\right)^{-3/2}\mathbb{E}\left\{\left\|\Lambda_k^{0,(\ell)\prime}\frac{1}{T_\ell}\left[E_k^{(\ell)}E_k^{(\ell)\prime} - \mathbb{E}\left(E_k^{(\ell)}E_k^{(\ell)\prime}|\mathscr{D}\right)\right]M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}\right\|_F\middle|\mathscr{D}\right\} = o_p(1)
$$

which follows by similar arguments as used in the proof of Lemma D.3(vi) in Lu and Su (2016).

(viii) Analogously to the previous statement, we have

$$(T_\ell)^{-3/2}\,\mathbb{E}\left\{\left\|F^{0,(\ell)\prime}\frac{1}{N_k^{(\ell)}}\left[E_k^{(\ell)\prime}E_k^{(\ell)}-\mathbb{E}\left(E_k^{(\ell)\prime}E_k^{(\ell)}|\mathscr{D}\right)\right]M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)\prime}\right\|_F\Big|\mathscr{D}\right\}=o_p(1)$$

by similar arguments as used in the proof of Lemma D.4(iii) in Lu and Su (2016). Then we are ready to show that

$$\begin{aligned}&\left|\frac{1}{\sqrt{N_k^{(\ell)}}T_\ell}tr\left\{\left[E_k^{(\ell)\prime}E_k^{(\ell)}-\mathbb{E}\left(E_k^{(\ell)\prime}E_k^{(\ell)}|\mathscr{D}\right)\right]M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)\prime}\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right\}\right|\\&\lesssim(T_\ell)^{-3/2}\left\|F^{0,(\ell)\prime}\frac{1}{N_k^{(\ell)}}\left[E_k^{(\ell)\prime}E_k^{(\ell)}-\mathbb{E}\left(E_k^{(\ell)\prime}E_k^{(\ell)}|\mathscr{D}\right)\right]M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)\prime}\right\|_F=o_p(1).\end{aligned}$$

(ix) This statement can be proved owing to the fact that the second moment of the term on the left side of the equality conditioning on $\mathscr{D}$ is $O_p(N^{-1})$. See the proof of Lemma B.1(i) for detail.

(x) Similarly to (B.17), we can also show that $\left\|\tilde{\mathbb{X}}_{j,k}^{(\ell)}P_{F^{0,(\ell)}}\right\|_F=O_p(N^{1/2})$. Then, following the same arguments as used in the proof of Lemma B.1(j) in Moon and Weidner (2017), we can finish the proof. ■

C Estimation of Panels with IFEs and Heterogeneous Slopes

For $\forall i\in\{n_1,\cdots,n_n\}$ and $t\in[T]$, consider the model

$$Y_{it}=\begin{cases}\lambda_i^{0\prime}f_t^0+X_{it}'\theta_i^{0,(1)}+e_{it}, & t\in\{1,\cdots,T_1\},\\ \lambda_i^{0\prime}f_t^0+X_{it}'\theta_i^{0,(2)}+e_{it}, & t\in\{T_1+1,\cdots,T\}.\end{cases}\tag{C.1}$$

Here $\{n_1,\cdots,n_n\}$ is a subset of $[N]$ and $n\asymp N$. To distinguish from the notation Λ^0 in the paper, we define $\Lambda_n^0:=(\lambda_{n_1},\cdots,\lambda_{n_n})'$.

Define $X_i^{(1)}=(X_{i1},\cdots,X_{iT_1})'$, $X_i^{(2)}=\left(X_{i(T_1+1)},\cdots,X_{iT}\right)'$, $e_i^{(1)}=(e_{i1},\cdots,e_{iT_1})'$, $e_i^{(2)}=\left(e_{i(T_1+1)},\cdots,e_{iT}\right)'$, $F^{0,(1)}=\left(f_1^0,\cdots,f_{T_1}^0\right)'$, $F^{0,(2)}=\left(f_{T_1+1}^0,\cdots,f_T^0\right)'$. To estimate $\theta_i^{0,(\ell)}$, λ_i^0 and f_t^0, we follow the lead of Bai (2009) and consider the PCA for heterogeneous panels. For $\forall\ell\in\{1,2\}$, let

$$\left(\left\{\hat{\theta}_i^{(\ell)}\right\}_{i\in\{n_1,\cdots,n_n\}},\hat{F}^{(\ell)}\right)=\underset{F^{(\ell)},\{\theta_i\}_{i\in\{n_1,\cdots,n_n\}}}{\arg\min}\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\left(Y_i^{(\ell)}-X_i^{(\ell)}\theta_i\right)'M_{F^{(\ell)}}\left(Y_i^{(\ell)}-X_i^{(\ell)}\theta_i\right),\tag{C.2}$$

where $T_2=T-T_1$, $W_i^{(1)}=(W_{i1},\cdots,W_{iT_1})'$, $W_i^{(2)}=\left(W_{i(T_1+1)},\cdots,W_{iT}\right)'$ for W_i denotes Y_i or X_i, $F^{(\ell)}$ is any $T_\ell\times r_0$ matrix such that $\frac{F^{(\ell)\prime}F^{(\ell)}}{T_\ell}=I_{r_0}$ and $M_{F^{(\ell)}}=I_{T_\ell}-\frac{F^{(\ell)}F^{(\ell)\prime}}{T_\ell}$. Note that we consider the concentrated objective function here by concentrating out the factor loadings. The solutions to the minimization problem in (C.2) solve the following nonlinear system of equations:

$$\hat{\theta}_i^{(\ell)}=\left(X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}X_i^{(\ell)}\right)^{-1}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}Y_i^{(\ell)},\tag{C.3}$$

$$\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\left(Y_i^{(\ell)}-X_i^{(\ell)}\hat{\theta}_i^{(\ell)}\right)'\left(Y_i^{(\ell)}-X_i^{(\ell)}\hat{\theta}_i^{(\ell)}\right)\right]\hat{F}^{(\ell)}=\hat{F}^{(\ell)}\hat{V}_{NT}^{(\ell)},\tag{C.4}$$

where $\hat{V}_{NT}^{(\ell)}$ is a diagonal matrix that contains the r_0 largest eigenvalues of the matrix in the square brackets in (C.4). Let $\hat{\lambda}_i^{(\ell)} = \frac{1}{T}\hat{F}^{(\ell)\prime}\left(Y_i^{(\ell)} - X_i^{(\ell)}\hat{\theta}_i^{(\ell)}\right)$, which are estimates of λ_i^0. Let $\hat{\Lambda}_n^{(\ell)} := \left(\hat{\lambda}_{n_1}^{(\ell)}, \cdots, \hat{\lambda}_{n_n}^{(\ell)}\right)'$, and $\hat{a}_{ii}^{(\ell)} := \hat{\lambda}_i^{(\ell)\prime}\left(\frac{\hat{\Lambda}_n^{(\ell)\prime}\hat{\Lambda}_n^{(\ell)}}{n}\right)^{-1}\hat{\lambda}_i^{(\ell)}$.

Let $\theta_i^{0,(\ell)} = \bar{\theta}^{0,(\ell)} + c_i^{(\ell)}$, where $\bar{\theta}^{0,(\ell)} = \frac{1}{n}\sum_{i\in\{n_1,\cdots,n_n\}}\theta_i^{0,(\ell)}$. Here, we consider testing the slope homogeneity for $i \in \{n_1, \cdots, n_n\}$. The null and alternative hypotheses are respectively given by

$$\begin{aligned} H_0 &: c_i^{(\ell)} = 0 \ \forall i \in \{n_1 \cdots, n_n\} \text{ and} \\ H_1 &: c_i^{(\ell)} \neq 0 \text{ for some } i \in \{n_1 \cdots, n_n\}. \end{aligned}$$

Following Pesaran and Yamagata (2008) and Ando and Bai (2016), we define

$$\hat{\Gamma}^{(\ell)} = \sqrt{n}\left(\frac{\frac{1}{n}\sum_{i\in\{n_1\cdots,n_n\}}\hat{\mathbb{S}}_i^{(\ell)} - p}{\sqrt{2p}}\right) \tag{C.5}$$

where

$$\hat{\mathbb{S}}_i^{(\ell)} = T_\ell\left(\hat{\theta}_i^{(\ell)} - \hat{\bar{\theta}}^{(\ell)}\right)'\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\left(\hat{\theta}_i^{(\ell)} - \hat{\bar{\theta}}^{(\ell)}\right)\left(1 - \hat{a}_{ii}^{(\ell)}/n\right)^2, \quad \hat{\bar{\theta}}^{(\ell)} = \frac{1}{n}\sum_{i\in\{n_1\cdots,n_n\}}\hat{\theta}_i^{(\ell)},$$

$$M_{\hat{F}^{(\ell)}} = I_{T_\ell} - \frac{\hat{F}^{(\ell)}\hat{F}^{(\ell)\top}}{T_\ell}, \quad \hat{S}_{ii}^{(\ell)} = \frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}X_i^{(\ell)}}{T_\ell}, \quad \left(\hat{\mathfrak{x}}_{it}^{(\ell)}\right)' \text{ is the } t\text{-th row of } M_{\hat{F}^{(\ell)}}X_i^{(\ell)},$$

$$\hat{\Omega}_i^{(\ell)} = \frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\hat{\mathfrak{x}}_{it}^{(\ell)}\hat{\mathfrak{x}}_{it}^{(\ell)\prime}\hat{e}_{it}^2 + \frac{1}{T_\ell}\sum_{j\in\mathcal{T}_{\ell,-1}}k(j/S_T)\sum_{t\in\mathcal{T}_{\ell,j}}\left(\hat{\mathfrak{x}}_{it}^{(\ell)}\hat{\mathfrak{x}}_{i,t+j}^{(\ell)\prime}\hat{e}_{it}\hat{e}_{i,t+j} + \hat{\mathfrak{x}}_{i,t-j}^{(\ell)}\hat{\mathfrak{x}}_{it}^{(\ell)\prime}\hat{e}_{i,t-j}\hat{e}_{it}\right)$$

and recall that $\mathcal{T}_1 = [T_1]$, $\mathcal{T}_2 = [T]\backslash[T_1]$, $\mathcal{T}_{1,-1} = \mathcal{T}_1\backslash\{T_1\}$, $\mathcal{T}_{2,-1} = \mathcal{T}_2\backslash\{T\}$, $\mathcal{T}_{1,j} = \{1+j, \cdots, T_1\}$, and $\mathcal{T}_{2,j} = \{T_1+1+j, \cdots, T\}$ for some specific $j \in \mathcal{T}_{\ell,-1}$.

In the next section, we study the asymptotic distribution of $\hat{\theta}_i^{(\ell)}$, the uniform convergence rates for the estimators of factors and factor loadings, and the asymptotic behavior for $\hat{\Gamma}^{(\ell)}$ under H_0 and H_1, respectively.

D Lemmas for Panel IFEs Model with Heterogeneous Slope

Below we derive the asymptotic distribution for the slope estimators in our heterogeneous panel models which allow for dynamics. To allow the dynamic panel, we focus on the Assumption 1* where the error term is the martingale difference sequence and conditional serial independence. If we focus on the Assumption 1, we can obtain the similar result by using the Davydov's inequality to derive the similar result for the non-dynamic panels with strong mixing distributed errors. Here we skip the proof for the non-dynamic panels with serial correlated errors for brevity. Let M be a generic large positive constant and $\mathcal{F}^{(\ell)} := \left\{F^{(\ell)} \in \mathbb{R}^{T_\ell\times r_0} : \frac{F^{(\ell)\prime}F^{(\ell)}}{T_\ell} = I_{r_0}\right\}$.

Lemma D.1 *Under Assumptions 1*, 2 and 8, we have*

(i) $\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)\prime}P_{F^{0,(\ell)}}e_i^{(\ell)}\right| = o_p(1)$,

(ii) $\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)\prime}P_{F^{(\ell)}}e_i^{(\ell)}\right| = o_p(1)$,

(iii) $\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}M_{F^{(\ell)}}e_i^{(\ell)}\right| = o_p(1)$,

(iv) $\sup_{\left\{\max_i\|\theta_i\|_{\max}\leq M\right\},F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left|\frac{1}{n}\sum_{i=n_1}^{n_n}\left(\theta_i-\theta_i^{0,(\ell)}\right)'\frac{X_i^{(\ell)\prime}M_{F^{(\ell)}}e_i^{(\ell)}}{T_\ell}\right| = o_p(1)$.

Proof. (i) We notice that

$$\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}e_i^{(\ell)\prime}P_{F^{0,(\ell)}}e_i^{(\ell)}\right| \leq \frac{1}{T_\ell}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathcal{T}_\ell}f_t^0e_{it}\right\|_2^2\right)\left\|\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\right\|_F$$
$$\lesssim \frac{1}{T_\ell}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathcal{T}_\ell}f_t^0e_{it}\right\|_2^2\right).$$

Recall that $\mathscr{D}$ denotes the minimum σ-fields generated by $\left\{V_j^0\right\}_{j\in[p]\cup\{0\}}$. Furthermore, we observe that

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathcal{T}_\ell}f_t^0e_{it}\right\|_2^2\middle|\mathscr{D}\right) = \frac{1}{n}\sum_{i=n_1}^{n_n}\mathbb{E}\left(\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathcal{T}_\ell}f_t^0e_{it}\right\|_2^2\middle|\mathscr{D}\right)$$
$$\leq \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in\mathcal{T}_\ell}\left\|\mathbb{E}\left(f_t^0f_s^{0\prime}e_{it}e_{is}|\mathscr{D}\right)\right\|_F$$
$$\leq \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in\mathcal{T}_\ell}\left\|f_t^0\right\|_2\left\|f_s^{0\prime}\right\|_2\left|\mathbb{E}\left(e_{it}e_{is}|\mathscr{D}\right)\right|$$
$$\lesssim \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\left|\mathbb{E}\left(e_{it}^2|\mathscr{D}\right)\right| \leq M\ a.s., \tag{D.1}$$

where the fourth line holds by the boundedness of factors shown in Lemma B.7(i) and the conditional independence of e_{it} under Assumption 1*(i), (iii). It follows that $\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathcal{T}_\ell}f_t^0e_{it}\right\|_F^2\right) = O_p(1)$ and $\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}e_i^{(\ell)\prime}P_{F^{0,(\ell)}}e_i^{(\ell)}\right| = O_p(T^{-1})$.

(ii) Noting that $P_{F^{(\ell)}} = F^{(\ell)}\left(F^{(\ell)\prime}F^{(\ell)}\right)^{-1}F^{(\ell)\prime} = T^{-1}F^{(\ell)}F^{(\ell)\prime}$ for $F^{(\ell)}\in\mathcal{F}^{(\ell)}$, we have

$$\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}e_i^{(\ell)\prime}P_{F^{(\ell)}}e_i^{(\ell)}\right| \leq \sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}f_te_{it}\right\|_2^2.$$

Next,

$$\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}f_te_{it}\right\|_2^2 = \sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}tr\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\frac{1}{T_\ell^2}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in\mathcal{T}_\ell}f_tf_s'e_{it}e_{is}\right)$$
$$= \sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}tr\left\{\frac{1}{T_\ell^2}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in\mathcal{T}_\ell}f_tf_s'\frac{1}{n}\sum_{i=n_1}^{n_n}\left[e_{it}e_{is}-\mathbb{E}\left(e_{it}e_{is}|\mathscr{D}\right)\right]\right\} + \sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}tr\left\{\frac{1}{T_\ell^2}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in\mathcal{T}_\ell}f_tf_s'\frac{1}{n}\sum_{i=n_1}^{n_n}\mathbb{E}\left(e_{it}e_{is}|\mathscr{D}\right)\right\}. \tag{D.2}$$

For the first term on the second line of (D.2), we have

$$\max_{t,s}\left|\frac{1}{n}\sum_{i=n_1}^{n_n}\left[e_{it}e_{is}-\mathbb{E}\left(e_{it}e_{is}|\mathscr{D}\right)\right]\right| = O_p\left(\sqrt{\frac{\log T}{N}}\right)$$

by conditional Bernstein's inequality for independent sequence combining the fact that $e_{it}e_{is}$ is independent across i given $\mathscr{D}$ by Assumption 1*(i). Then

$$\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}} tr\left\{\frac{1}{T_\ell^2}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in\mathcal{T}_\ell} f_t f_s' \frac{1}{n}\sum_{i=n_1}^{n_n}\left[e_{it}e_{is}-\mathbb{E}\left(e_{it}e_{is}\middle|\mathscr{D}\right)\right]\right\}$$
$$= O_p\left(\sqrt{\frac{\log T}{N}}\right)\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left(\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\|f_t\|_2\right)^2 = O_p\left(\sqrt{\frac{\log T}{N}}\right). \tag{D.3}$$

For the second term on the second line of (D.2), we have

$$\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}} tr\left\{\frac{1}{T_\ell^2}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in\mathcal{T}_\ell} f_t f_s' \frac{1}{n}\sum_{i=n_1}^{n_n}\mathbb{E}\left(e_{it}e_{is}\middle|\mathscr{D}\right)\right\}$$
$$\leq \sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in\mathcal{T}_\ell}\|f_t\|_2\|f_s\|_2\,\mathbb{E}\left(e_{it}e_{is}\middle|\mathscr{D}\right)$$
$$\lesssim \frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\left|\mathbb{E}\left(e_{it}^2\middle|\mathscr{D}\right)\right| = O_p(T^{-1}), \tag{D.4}$$

where the first inequality is by Cauchy's inequality, the third line is by the definition of $\mathcal{F}^{(\ell)}$ and similar arguments as in (D.1).

Combining (D.2)-(D.4), we have shown that $\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)\prime}P_{F^{(\ell)}}e_i^{(\ell)}\right| = O_p\left(\sqrt{\frac{\log T}{N}}\right)$.

(iii) Owing to the fact that

$$\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}M_{F^{(\ell)}}e_i^{(\ell)}\right| \leq \left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}e_i^{(\ell)}\right| + \sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}P_{F^{(\ell)}}e_i^{(\ell)}\right|,$$

we show the convergence rate for the two terms on the right side of above inequality. For the first term, we note that $\mathbb{E}\left(\lambda_i^{0\prime}f_t^0 e_{it}\middle|\mathscr{D}\right)=0$ and e_{it} is independent across i and strong mixing across t given $\mathscr{D}$. Then we have

$$\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}e_i^{(\ell)}\right| = \left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\lambda_i^{0\prime}f_t^0 e_{it}\right| = O_p\left(\frac{1}{\sqrt{NT}}\right)$$

by Lemma B.6(ii). For the second term, we note that

$$\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}P_{F^{(\ell)}}e_i^{(\ell)}\right|$$
$$= \sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}F^{(\ell)}\left(F^{(\ell)\prime}F^{(\ell)}\right)^{-1}\sum_{t\in\mathcal{T}_\ell}f_t e_{it}\right|$$
$$\leq \sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\sqrt{\frac{1}{n}\sum_{i=n_1}^{n_n}\|\lambda_i^0\|_2^2}\sqrt{\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}f_t e_{it}\right\|_2^2}\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left\|\left(\frac{F^{(\ell)\prime}F^{(\ell)}}{T_\ell}\right)^{-1}\right\|_2\left\|\frac{F^{0,(\ell)\prime}F^{(\ell)}}{T_\ell}\right\|_2$$
$$\lesssim \sqrt{\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}f_t e_{it}\right\|_2^2} = O_p\left[\left(\frac{\log T}{N}\right)^{1/4}\right] = o_p(1),$$

where the third is by Cauchy's inequality and the last line is by arguments in (D.3) and (D.4). Combining arguments above, we show the desired result.

(iv) We first observe that

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{it}\right\|_2^2\right)=\frac{1}{n}\sum_{i=n_1}^{n_n}\mathbb{E}\left(\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{it}\right\|_2^2\right)$$
$$\leq\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in\mathcal{T}_\ell}\left\|\mathbb{E}\left(X_{it}X_{is}'e_{it}e_{is}\right)\right\|_F=\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\left\|\mathbb{E}\left(X_{it}X_{it}'e_{it}^2\right)\right\|_F\leq M\ a.s.,\tag{D.5}$$

where the second equality is by Assumption 1*(ii) and the law of iterated expectations, and the last inequality is by Assumption 1*(v). It follows that

$$\sup_{\left\{\max_i\|\theta_i\|_{\max}\leq M\right\},F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left\|\frac{1}{n}\sum_{i=n_1}^{n_n}\left(\theta_i-\theta_i^{0,(\ell)}\right)'\frac{X_i^{(\ell)\prime}M_{F^{(\ell)}}e_i^{(\ell)}}{T_\ell}\right\|_F$$
$$\leq\sup_{\left\{\max_i\|\theta_i\|_{\max}\leq M\right\}}\left|\frac{1}{n}\sum_{i=n_1}^{n_n}\left(\theta_i-\theta_i^{0,(\ell)}\right)'\frac{X_i'e_i^{(\ell)}}{T_\ell}\right|+\sup_{\left\{\max_i\|\theta_i\|_{\max}\leq M\right\},F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left|\frac{1}{n}\sum_{i=n_1}^{n_n}\left(\theta_i-\theta_i^{0,(\ell)}\right)'\frac{X_i^{(\ell)\prime}P_{F^{(\ell)}}e_i^{(\ell)}}{T_\ell}\right|$$
$$\leq\sup_{\left\{\max_i\|\theta_i\|_{\max}\leq M\right\}}\frac{1}{\sqrt{T_\ell}}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\theta_i-\theta_i^{0,(\ell)}\right\|_2^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{it}\right\|_2^2\right)^{1/2}$$
$$+\sup_{\left\{\max_i\|\theta_i\|_{\max}\leq M\right\},F^{(\ell)}\in\mathcal{F}^{(\ell)}}\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\theta_i-\theta_i^{0,(\ell)}\right\|_2\left\|\frac{X_i^{(\ell)\prime}F^{(\ell)}}{T_\ell}\right\|_2\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}f_te_{it}\right\|_2\left\|\left(\frac{F^{(\ell)\prime}F^{(\ell)}}{T_\ell}\right)^{-1}\right\|_2$$
$$\lesssim O_p(T^{-1/2})+\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{X_i^{(\ell)\prime}F^{(\ell)}}{T_\ell}\right\|_2^2\right)^{1/2}\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}f_te_{it}\right\|_2^2\right)^{1/2}$$
$$=O_p(T^{-1/2})+O_p\left[\left(\frac{\log T}{N}\right)^{1/4}\right]=o_p(1),$$

where the second inequality is by Cauchy's inequality, the fifth line combines facts that both θ_i and $\theta_i^{0,(\ell)}$ are bounded, $\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{it}\right\|_2^2=O_p(1)$ by (D.5), the definition for $\mathcal{F}^{(\ell)}$ and Cauchy's inequality. The last line is due to (D.2) and the fact that

$$\sup_{F^{(\ell)}\in\mathcal{F}^{(\ell)}}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{X_i^{(\ell)\prime}F^{(\ell)}}{T_\ell}\right\|_2^2\right)^{1/2}\lesssim\max_{i\in\{n_1,\cdots,n_n\}}\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\|X_{it}\|_2=O_p(1).$$

by Assumption 8(ii). ■

Lemma D.2 *Under Assumptions 1*, 2 and 8, we have* $\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\xrightarrow{p}0$ *and* $\left\|P_{\hat{F}^{(\ell)}}-P_{F^{0,(\ell)}}\right\|_F\xrightarrow{p}0$.

Proof. Let

$$S_{NT}\left(\{\theta_i\},F^{(\ell)}\right)=\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\left(Y_i^{(\ell)}-X_i^{(\ell)}\theta_i\right)'M_{F^{(\ell)}}\left(Y_i^{(\ell)}-X_i^{(\ell)}\theta_i\right)-\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}e_i^{(\ell)\prime}M_{F^{0,(\ell)}}e_i^{(\ell)}.$$

Recall from (C.2) that $\left(\left\{\hat{\theta}_i^{(\ell)}\right\}, \hat{F}^{(\ell)}\right)$ is the minimizer of $S_{NT}\left(\{\theta_i\}, F^{(\ell)}\right)$. By (C.1) and Lemma D.1, we have

$$\begin{aligned}
S_{NT}\left(\{\theta_i\}, F^{(\ell)}\right) &= \tilde{S}_{NT}\left(\{\theta_i\}_{\forall i}, F^{(\ell)}\right) + \frac{2}{n}\sum_{i=n_1}^{n_n}\left(\theta_i - \theta_i^{0,(\ell)}\right)' \frac{X_i^{(\ell)\prime} M_{F^{(\ell)}} e_i^{(\ell)}}{T_\ell} \\
&+ \frac{2}{nT_\ell}\sum_{i=n_1}^{n_n} \lambda_i^{0\prime} F^{0,(\ell)\prime} M_{F^{(\ell)}} e_i^{(\ell)} + \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)\prime} P_{F^{(\ell)}} e_i^{(\ell)} - \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)\prime} P_{F^{0,(\ell)}} e_i^{(\ell)} \\
&= \tilde{S}_{NT}\left(\{\theta_i\}, F^{(\ell)}\right) + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{S}_{NT}\left(\{\theta_i\}, F^{(\ell)}\right) &= \frac{1}{n}\sum_{i=n_1}^{n_n}\left(\theta_i - \theta_i^{0,(\ell)}\right)' \frac{X_i^{(\ell)\prime} M_{F^{(\ell)}} e_i^{(\ell)}}{T_\ell}\left(\theta_i - \theta_i^{0,(\ell)}\right) + \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} \lambda_i^{0\prime} F^{0,(\ell)\prime} M_{F^{(\ell)}} F^{0,(\ell)} \lambda_i^0 \\
&+ \frac{2}{n}\sum_{i=n_1}^{n_n}\left(\theta_i - \theta_i^{0,(\ell)}\right)' \frac{X_i^{(\ell)\prime} M_{F^{(\ell)}} e_i^{(\ell)}}{T_\ell} F^{0,(\ell)} \lambda_i^0.
\end{aligned}$$

Following Song (2013) and Bai (2009), we can show that $\tilde{S}_{NT}\left(\{\theta_i\}_{\forall i}, F^{(\ell)}\right)$ is uniquely minimized at $\left(\left\{\theta_i^{0,(\ell)}\right\}, F^{0,(\ell)} H\right)$, where $H^{(\ell)}$ is a rotation matrix. Hence, we conclude that $\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)} \xrightarrow{p} 0$. Following the arguments as used in the proof in Proposition 1 of Bai (2009), we can show that $\left\| P_{\hat{F}^{(\ell)}} - P_{F^{0,(\ell)}} \right\|_F \xrightarrow{p} 0$. ■

Let B_N denote the uniform convergence rate for $\hat{\theta}_i^{(\ell)}$. That is, $\max_{i \in \{n_1, \cdots, n_n\}} \left\| \hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)} \right\|_2 = O_p(B_N)$.

Lemma D.3 *Under Assumptions 1*, 2 and 8, we have* $\frac{1}{\sqrt{T_\ell}} \left\| \hat{F}^{(\ell)} - F^{0,(\ell)} H^{(\ell)} \right\|_F = O_p\left(B_N + \frac{1}{\sqrt{N \wedge T}}\right)$, *where* $H^{(\ell)} := \left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)\left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right) V_{NT}^{(\ell)-1}$.

Proof. Recall that $V_{NT}^{(\ell)}$ is the diagonal matrix that contains the eigenvalues of

$$\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\left(Y_i^{(\ell)} - X_i^{(\ell)}\theta_i\right)'\left(Y_i^{(\ell)} - X_i^{(\ell)}\theta_i\right)$$

along its diagonal line. By inserting (C.1) into (C.4), we obtain that

$$\hat{F}^{(\ell)} V_{NT}^{(\ell)} - F^{0,(\ell)}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)\left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right) = \sum_{m\in[8]} J_m^{(\ell)}, \tag{D.6}$$

where

$$\begin{aligned}
J_1^{(\ell)} &= \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} X_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)' X_i^{(\ell)\prime}\hat{F}^{(\ell)}, \\
J_2^{(\ell)} &= \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} X_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)\lambda_i^{0\prime} F^{0,(\ell)\prime}\hat{F}^{(\ell)}, \\
J_3^{(\ell)} &= \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} X_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right) e_i^{(\ell)\prime}\hat{F}^{(\ell)},
\end{aligned}$$

$$J_4^{(\ell)} = \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} F^{0,(\ell)}\lambda_i^0\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)' X_i^{(\ell)\prime}\hat{F}^{(\ell)},$$

$$J_5^{(\ell)} = \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)' X_i^{(\ell)\prime}\hat{F}^{(\ell)}, \quad J_{6,t}^{(\ell)} = \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} F^{0,(\ell)}\lambda_i^0 e_i^{(\ell)\prime}\hat{F}^{(\ell)},$$

$$J_7^{(\ell)} = \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)}\lambda_i^{0\prime}F^{0,(\ell)\prime}\hat{F}^{(\ell)}, \quad \text{and} \quad J_8^{(\ell)} = \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)}e_i^{(\ell)\prime}\hat{F}^{(\ell)}.$$

We show the convergence rate for $J_m^{(\ell)}$ $\forall m \in [8]$ in the following.

For $J_1^{(\ell)}$, we notice that

$$\frac{\left\|J_1^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2^2 \frac{1}{\sqrt{T_\ell}}\left\|\hat{F}^{(\ell)}\right\|_F \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\left\|X_i^{(\ell)}\right\|_2^2 = O_p(B_N^2), \tag{D.7}$$

where the equality is by Assumption 8(ii) and normalization of the factor vector. Similarly, we have

$$\frac{\left\|J_2^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2 \max_{i\in\{n_1,\cdots,n_n\}}\left\|\lambda_i^0\right\|_2 \frac{\left\|F^{0,(\ell)}\right\|_F}{\sqrt{T_\ell}}\frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\frac{1}{n\sqrt{T_\ell}}\sum_{i=n_1}^{n_n}\left\|X_i^{(\ell)}\right\|_2 = O_p(B_N),$$

$$\frac{\left\|J_3^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2 \frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\sqrt{\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\left\|X_i^{(\ell)}\right\|_2^2}\sqrt{\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\left\|e_i^{(\ell)}\right\|_2^2} = O_p(B_N),$$

$$\frac{\left\|J_6^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \leq \frac{1}{\sqrt{n}}\frac{\left\|F^{0,(\ell)}\right\|_F}{\sqrt{T_\ell}}\frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\sqrt{\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left\|\frac{1}{\sqrt{n}}\sum_{i=n_1}^{n_n}\lambda_i^0 e_{it}\right\|_2^2} = O_p\left(\frac{1}{\sqrt{N}}\right)$$

$$\frac{\left\|J_8^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \leq \frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\frac{1}{nT_\ell}\left\|\sum_{i=n_1}^{n_n} e_i^{(\ell)}e_i^{(\ell)\prime}\right\|_F = O_p\left(\frac{1}{\sqrt{N\wedge T}}\right),$$

where the third line is by the fact that $\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left\|\frac{1}{\sqrt{n}}\sum_{i=n_1}^{n_n}\lambda_i^0 e_{it}\right\|_2 = O_p(1)$ by similar arguments as in (D.1) and the last line is due to the fact that

$$\begin{aligned}
&\mathbb{E}\left(\left\|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)}e_i^{(\ell)\prime}\right\|_F^2 \middle| \mathscr{D}\right) = \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{i^*=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\sum_{t^*\in\mathcal{T}_\ell}\mathbb{E}\left(e_{it}e_{it^*}e_{i^*t}e_{i^*t^*}\middle|\mathscr{D}\right)\\
&= \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\sum_{t^*\in\mathcal{T}_\ell}\mathbb{E}\left(e_{it}^2e_{it^*}^2\middle|\mathscr{D}\right) + \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{i^*\neq i}\sum_{t\in\mathcal{T}_\ell}\sum_{t^*\in\mathcal{T}_\ell}\mathbb{E}\left(e_{it}e_{it^*}\middle|\mathscr{D}\right)\mathbb{E}\left(e_{i^*t}e_{i^*t^*}\middle|\mathscr{D}\right)\\
&= \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\mathbb{E}\left(e_{it}^4|\mathscr{D}\right) + \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\sum_{t^*\in\mathcal{T}_\ell,t^*\neq t}\mathbb{E}\left(e_{it}^2|\mathscr{D}\right)\mathbb{E}\left(e_{it^*}^2|\mathscr{D}\right)\\
&+ \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{i^*\neq i}\sum_{t\in\mathcal{T}_\ell}\mathbb{E}\left(e_{it}^2|\mathscr{D}\right)\mathbb{E}\left(e_{i^*t}^2|\mathscr{D}\right)\\
&= O\left((N)^{-1} + (T)^{-1}\right) \ a.s.
\end{aligned} \tag{D.8}$$

with the application Assumptions 1*(i), (ii), (iii), (v). Besides, we have $\frac{\left\|J_4^{(\ell)}\right\|_F}{\sqrt{T_\ell}} = O_p(B_N)$, $\frac{\left\|J_5^{(\ell)}\right\|_F}{\sqrt{T_\ell}} = O_p(B_N)$, $\frac{\left\|J_7^{(\ell)}\right\|_F}{\sqrt{T_\ell}} = O_p\left(\frac{1}{\sqrt{N}}\right)$ by similar analyses as used for $J_2^{(\ell)}$, $J_3^{(\ell)}$ and $J_6^{(\ell)}$, respectively.

Combining the above arguments, premultiplying both sides of (D.6) by $\hat{F}^{(\ell)\prime}$ and using the fact that $\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)} = T_\ell I_r$, we have

$$\frac{1}{T_\ell}\left\|\hat{F}^{(\ell)}V_{NT}^{(\ell)} - F^{0,(\ell)}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)\left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)\right\|_F = O_p(B_N) + O_p\left(\frac{1}{\sqrt{N\wedge T}}\right), \tag{D.9}$$

and

$$\begin{aligned} V_{NT}^{(\ell)} &= \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)\left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right) + \frac{\hat{F}^{(\ell)\prime}}{\sqrt{T_\ell}}\frac{\sum_{m\in[8]} J_m^{(\ell)}}{\sqrt{T_\ell}} \\ &= \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)\left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right) + o_p(1). \end{aligned}$$

Then $V_{NT}^{(\ell)}$ is invertible and $\left\|V_{NT}^{(\ell)}\right\|_F = O_p(1)$. By the definition of $H^{(\ell)}$ and (D.9), we have $\frac{1}{\sqrt{T_\ell}}\left\|\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}\right\|_F = O_p\left(B_N + \frac{1}{\sqrt{N\wedge T}}\right)$. ■

Lemma D.4 *Under Assumptions 1*, 2 and 8, we have*

(i) $\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}X_{i^*}^{(\ell)}}{T_\ell} = \frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_{i^*}^{(\ell)}}{T_\ell} + O_p\left(B_N + \frac{1}{\sqrt{N\wedge T}}\right)$ *uniformly in* $i, i^* \in \{n_1, \cdots, n_n\}$,

(ii) $\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_i^{(\ell)}}{T_\ell} = \frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}e_i^{(\ell)}}{T_\ell} + O_p\left(B_N^2 + B_N\sqrt{\frac{\log N}{T}} + \frac{1}{N\wedge T} + \sqrt{\frac{\log N}{(N\wedge T)T}}\right)$ *uniformly in* $i \in \{n_1, \cdots, n_n\}$,

(iii) $\frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\left\|e_i^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_2^2 = O_p\left(B_N^2 + \frac{1}{N\wedge T}\right)$,

(iv) $\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}F^{0,(\ell)}\right\|_F = O_p\left(B_N + \frac{1}{\sqrt{N\wedge T}}\right)$,

(v) $\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}\left\|X_i^{(\ell)\prime}e_{i^*}^{(\ell)}\right\|_2^2 = O_p\left(\frac{\log N}{T}\right)$ *uniformly in* $i \in \{n_1, \cdots, n_n\}$,

(vi) $\left\|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^0 e_i^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_F = O_p\left(\frac{B_N}{\sqrt{N}} + \frac{1}{N} + \frac{1}{\sqrt{NT}}\right)$,

(vii) $\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}\right\|_F = O_p\left(\frac{B_N}{\sqrt{N}} + \frac{1}{N} + \sqrt{\frac{\log N}{NT}}\right)$,

(viii) $\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_{i^*}^{(\ell)}e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_F = O_p\left(B_N^2 + B_N\sqrt{\frac{\log N}{T}} + \frac{\sqrt{\log N}}{N\wedge T}\right)$.

Proof. (i) Noting that $M_{\hat{F}^{(\ell)}} = I_{T_\ell} - \hat{F}^{(\ell)}\left(\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}\right)^{-1}\hat{F}^{(\ell)\prime}$ and $M_{F^{0,(\ell)}} = I_{T_\ell} - F^{0,(\ell)}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}$, we can show that

$$\begin{aligned} M_{F^{0,(\ell)}} - M_{\hat{F}^{(\ell)}} &= \frac{\hat{F}^{(\ell)}}{\sqrt{T_\ell}}\left(\frac{\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1}\frac{\hat{F}^{(\ell)\prime}}{\sqrt{T_\ell}} - \frac{F^{0,(\ell)}}{\sqrt{T_\ell}}\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\frac{F^{0,(\ell)\prime}}{\sqrt{T_\ell}} \\ &= \frac{\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\left[\left(\frac{\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1} - \left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\right]\left(\frac{\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\right)' \\ &+ \frac{\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\left[\left(\frac{\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1} - \left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\right]\left(\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\right)' \end{aligned}$$

$$
\begin{aligned}
&+\frac{\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\left(\frac{\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\right)'\\
&+\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\left[\left(\frac{\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1}-\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\right]\left(\frac{\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\right)'\\
&+\frac{\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\left(\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\right)'\\
&+\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\left[\left(\frac{\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1}-\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\right]\left(\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\right)'\\
&+\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\left(\frac{\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\right)'. \qquad \text{(D.10)}
\end{aligned}
$$

With Lemma D.3, Assumption 8, the normalization for the factor space, and the fact that

$$
\begin{aligned}
\left\|\frac{\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}-\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right\|_F &\leq \frac{\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|_F^2}{T_\ell}+2\frac{\left\|F^{0,(\ell)}H^{(\ell)}\right\|\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|_F}{T_\ell}\\
&=O_p\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right), \qquad \text{(D.11)}
\end{aligned}
$$

it yields

$$
\left\|M_{F^{0,(\ell)}}-M_{\hat{F}^{(\ell)}}\right\|_F=O_p\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right). \qquad \text{(D.12)}
$$

Following this, we obtain that

$$
\begin{aligned}
&\max_{i,i^*\in\{n_1,\cdots,n_n\}}\left\|\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}X_{i^*}^{(\ell)}}{T_\ell}-\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_{i^*}^{(\ell)}}{T_\ell}\right\|_F\\
=&\max_{i,i^*\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}\left(M_{\hat{F}^{(\ell)}}-M_{F^{0,(\ell)}}\right)X_{i^*}^{(\ell)}\right\|_F\\
\leq&\max_{i\in\{n_1,\cdots,n_n\}}\frac{1}{T_\ell}\left\|X_i^{(\ell)}\right\|_F^2\left\|M_{F^{0,(\ell)}}-M_{\hat{F}^{(\ell)}}\right\|_F\\
=&\max_{i\in\{n_1,\cdots,n_n\}}\frac{1}{T_\ell}\left\|X_i^{(\ell)}\right\|_F^2O_p\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right)=O_p\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right).
\end{aligned}
$$

(ii) Combining (D.10), we notice that

$$
\begin{aligned}
&\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_i^{(\ell)}}{\sqrt{T_\ell}}-\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}e_i^{(\ell)}}{\sqrt{T_\ell}}\right\|_2\\
=&\sqrt{T_\ell}\max_{i\in\{n_1,\cdots,n_n\}}\frac{\left\|X_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\frac{\left\|e_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}}O_p\left[\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right)^3+\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right)^2\right]\\
&+\sqrt{T_\ell}\max_{i\in\{n_1,\cdots,n_n\}}\frac{\left\|X_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\frac{\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\left\|\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\right\|_F\frac{\left\|\left(F^{0,(\ell)}H^{(\ell)}\right)'e_i^{(\ell)}\right\|_F}{T_\ell}\\
&+\sqrt{T_\ell}\max_{i\in\{n_1,\cdots,n_n\}}\frac{\left\|X_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\frac{\left\|F^{0,(\ell)}H^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\left\|\left(\frac{\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1}-\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\right\|_F\frac{\left\|\left(F^{0,(\ell)}H^{(\ell)}\right)'e_i^{(\ell)}\right\|_F}{T_\ell}
\end{aligned}
$$

$$= \sqrt{T_\ell}\left[O_p\left(B_N^2 + \frac{1}{N \wedge T}\right) + O_p\left(B_N + \frac{1}{\sqrt{N \wedge T}}\right) O_p\left(\sqrt{\frac{\log N}{T}}\right)\right],$$

where the last line holds by combining Assumption 8(ii), (D.11), Lemma D.3 and the fact that

$$\max_{i\in\{n_1,\cdots,n_n\}} \frac{\left\|\left(F^{0,(\ell)}H^{(\ell)}\right)' e_i^{(\ell)}\right\|_F}{T_\ell} = \max_{i\in\{n_1,\cdots,n_n\}} \left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell} f_t^0 e_{it}\right\|_2 O_p(1) = O_p\left(\sqrt{\frac{\log N}{T}}\right).$$

We will show the last equality by using the Bernstein's inequality in Lemma B.5(i).

With the fact that

$$\begin{aligned}
&\max_{i\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell} \left\|Var\left(e_{it}f_t^0|\mathscr{D}\right)\right\|_F = \max_{i\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell} \left\|\mathbb{E}\left(e_{it}^2 f_t^0 f_t^{0\prime}|\mathscr{D}\right)\right\|_F \\
&= \max_{i\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell} \left\|\mathbb{E}\left(e_{it}^2|\mathscr{D}\right) f_t^0 f_t^{0\prime}\right\|_F = O_p(1) \quad \text{and} \qquad \text{(D.13)}\\
&\max_{i\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell} \left\|e_{it}f_t^0\right\|_2 = O_p\left((NT)^{1/q}\right),
\end{aligned}$$

where the second line is by Assumption 1*(v) and the last line is by Assumption 1*(v), we define events $\mathscr{A}_{4,N}(M) = \left\{\max_{i\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell} \left\|e_{it}f_t^0\right\|_2 \leq M(NT)^{1/q}\right\}$ and $\mathscr{A}_{4,N,i}(M) = \left\{\max_{t\in\mathcal{T}_\ell} \left\|e_{it}f_t^0\right\|_2 \leq M(NT)^{1/q}\right\}$ for a large enough constant M. Then for some large positive constants c_6 and c_7, we have $\mathbb{P}\left(\mathscr{A}_{4,N}^c(M)\right) \to 0$ and

$$\begin{aligned}
&\mathbb{P}\left(\max_{i\in\{n_1,\cdots,n_n\}} \left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell} e_{it}f_t^0\right\|_2 > c_6\sqrt{\frac{\log N}{T}}\right) \\
&\leq \mathbb{P}\left(\max_{i\in\{n_1,\cdots,n_n\}} \left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell} e_{it}f_t^0\right\|_2 > c_6\sqrt{\frac{\log N}{T}}, \mathscr{A}_{4,N}(M)\right) + \mathbb{P}\left(\mathscr{A}_{4,N}^c(M)\right) \\
&\leq \sum_{i=n_1}^{n_n} \mathbb{P}\left(\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell} e_{it}f_t^0\right\|_2 > c_6\sqrt{\frac{\log N}{T}}, \mathscr{A}_{4,N}(M)\right) + \mathbb{P}\left(\mathscr{A}_{4,N}^c(M)\right) \\
&\leq \sum_{i=n_1}^{n_n} \mathbb{P}\left(\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell} e_{it}f_t^0\right\|_2 > c_6\sqrt{\frac{\log N}{T}}, \mathscr{A}_{4,N,i}(M)\right) + \mathbb{P}\left(\mathscr{A}_{4,N}^c(M)\right) \\
&\leq \sum_{i=n_1}^{n_n} \mathbb{E}\mathbb{P}\left(\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell} e_{it}f_t^0\right\|_2 > c_6\sqrt{\frac{\log N}{T}}\middle|\mathscr{D}\right) \mathbf{1}\left\{\mathscr{A}_{4,N,i}(M)\right\} + \mathbb{P}\left(\mathscr{A}_{4,N}^c(M)\right) \\
&\leq \sum_{i=n_1}^{n_n} \exp\left\{-\frac{c_4 c_6^2 T\log N/2}{c_7 T + c_6 M\sqrt{T\log N}(NT)^{1/q}(\log T)^2/3}\right\} + o(1) \\
&= o(1), \qquad \text{(D.14)}
\end{aligned}$$

where the last inequality is by Lemma B.5(i) combining (D.13), and the definition of event $\mathscr{A}_{4,N,i}$, and the last line is by Assumption 1*(vi) and the fact that $q > 8$.

(iii) Owing to the fact that

$$\frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n} \left\|e_i^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_2^2 \leq \frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n} \left\|e_i^{(\ell)\prime}F^{0,(\ell)}H^{(\ell)}\right\|_2^2 + \frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n} \left\|e_i^{(\ell)\prime}\left(\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}\right)\right\|_2^2$$

$$\leq \frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\left\|e_i^{(\ell)\prime}F^{0,(\ell)}\right\|_2^2\left\|H^{(\ell)}\right\|_F^2+\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\left\|e_i^{(\ell)\prime}\right\|_2^2\frac{1}{T_\ell}\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|_F^2$$

$$\leq \frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\left\|e_i^{(\ell)\prime}F^{0,(\ell)}\right\|_2^2 O_p(1)+O_p\left(B_N^2+\frac{1}{N\wedge T}\right), \tag{D.15}$$

where the last inequality is by Assumption 8(ii) and Lemma D.3, we only need to show the rate for $\frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\left\|e_i^{(\ell)\prime}F^{0,(\ell)}\right\|_2^2$. Noting that

$$\frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\left\|e_i^{(\ell)\prime}F^{0,(\ell)}\right\|_2^2=\frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\left\|\sum_{t\in\mathcal{T}_\ell}e_{it}f_t^0\right\|_2^2=\frac{1}{T_\ell}\left[\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathcal{T}_\ell}e_{it}f_t^0\right\|_2^2\right]=\frac{1}{T_\ell}O_p(1), \tag{D.16}$$

by (D.1), which yields

$$\frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\left\|e_i^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_2^2=O_p\left(B_N^2+\frac{1}{N\wedge T}\right).$$

(iv) Noting that $M_{\hat{F}^{(\ell)}}\hat{F}^{(\ell)}=0$, we have

$$\begin{aligned}
&\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}F^{0,(\ell)}\right\|_F\\
&=\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\left(F^{0,(\ell)}-\hat{F}^{(\ell)}H^{(\ell)-1}\right)\right\|_F\\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{\sqrt{T_\ell}}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\right\|_F O_p\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right)=O_p\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right),
\end{aligned}$$

where the last inequality is by Lemma D.3 and the last equality is by the fact that

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{\sqrt{T_\ell}}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\right\|_F\leq\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{\sqrt{T_\ell}}X_i^{(\ell)\prime}\right\|_F\left(1+\frac{\left\|\hat{F}^{(\ell)}\hat{F}^{(\ell)\prime}\right\|_F}{T_\ell}\right)=O_p(1) \tag{D.17}$$

by Assumption 8(ii).

(v) Note that

$$\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}\left\|X_i^{(\ell)\prime}e_{i^*}^{(\ell)}\right\|_2^2=\frac{1}{n}\sum_{i^*=n_1}^{n_n}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\|_2^2\leq\max_{i,i^*\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\|_2^2.$$

Under Assumptions 1* and Assumption 8(ii)

$$\begin{aligned}
&\max_{i,i^*\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell}\|X_{it}e_{i^*t}\|_2=O_p\left((NT)^{1/q}\right),\\
&\max_{i,i^*\in\{n_1,\cdots,n_n\}}\left\|\sum_{t\in\mathcal{T}_\ell}\mathbb{E}\left(X_{it}X_{it}'e_{i^*t}^2\middle|\mathscr{G}_{t-1}\right)\right\|_F=\max_{i,i^*\in\{n_1,\cdots,n_n\}}\left\|\sum_{t\in\mathcal{T}_\ell}X_{it}X_{it}'\mathbb{E}\left(e_{i^*t}^2\middle|\mathscr{G}_{t-1}\right)\right\|_F\\
&\leq\max_{i^*\in\{n_1,\cdots,n_n\}}\mathbb{E}\left(e_{i^*t}^2\middle|\mathscr{G}_{t-1}\right)\max_{i\in\{n_1,\cdots,n_n\}}\sum_{t\in\mathcal{T}_\ell}\|X_{it}\|_2^2\leq c_8T\ a.s.
\end{aligned}$$

Define events $\mathscr{A}_{5,N}(M)=\left\{\max_{i\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell}\|X_{it}e_{i^*t}\|_2\leq M(NT)^{1/q}\right\}$ and $\mathscr{A}_{5,N,i,i^*}(M)=\left\{\max_{t\in\mathcal{T}_\ell}\|X_{it}e_{i^*t}\|_2\leq M(NT)\right.$
for a large enough constant M such that $\mathbb{P}\left(\mathscr{A}_{5,N}^c(M)\right)\to 0$. Then we have

$$\mathbb{P}\left(\max_{i,i^*\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\|_2>c_6\sqrt{\frac{\log N}{T}}\right)$$

$$
\begin{aligned}
&\leq \mathbb{P}\left(\max_{i,i^*\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\|_2 > c_6\sqrt{\frac{\log N}{T}}, \mathscr{A}_{5,N}(M)\right)+\mathbb{P}\left(\mathscr{A}_{5,N}^c(M)\right)\\
&\leq \sum_{i=n_1}^{n_n}\sum_{i^*=n_1}^{n_n}\mathbb{P}\left(\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\|_2 > c_6\sqrt{\frac{\log N}{T}}, \mathscr{A}_{5,N}(M)\right)+\mathbb{P}\left(\mathscr{A}_{5,N}^c(M)\right)\\
&\leq \sum_{i=n_1}^{n_n}\sum_{i^*=n_1}^{n_n}\mathbb{P}\left(\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\|_2 > c_6\sqrt{\frac{\log N}{T}}, \mathscr{A}_{5,N,i,i^*}(M)\right)+\mathbb{P}\left(\mathscr{A}_{5,N}^c(M)\right)\\
&\leq \sum_{i=n_1}^{n_n}\sum_{i^*=n_1}^{n_n}\exp\left\{\frac{-c_6^2T\log N/2}{c_8T+Mc_6(NT)^{1/q}\sqrt{T\log N}/3}\right\}+o(1)\\
&= o(1), \qquad\qquad \text{(D.18)}
\end{aligned}
$$

where the last inequality is by Lemma B.5(ii) and the last line is by Assumption 1*(vi).

(vi) Noted that

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{\sqrt{nT_\ell}}\sum_{i=n_1}^{n_n}\lambda_i^0e_i^{(\ell)\prime}F^{0,(\ell)}\right\|_F^2\middle|\mathscr{D}\right]=\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{i^*=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\sum_{t^*\in\mathcal{T}_\ell}\mathbb{E}\left(\lambda_i^0f_t^{0\prime}f_{t^*}^0\lambda_{i^*}^{0\prime}e_{it}e_{i^*t^*}|\mathscr{D}\right)\\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\lambda_i^0\right\|_2^2\max_{t\in\mathcal{T}_\ell}\left\|f_t^0\right\|_2^2\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{i^*=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\sum_{t^*\in\mathcal{T}_\ell}\left|\mathbb{E}\left(e_{it}e_{i^*t^*}|\mathscr{D}\right)\right|\\
&\lesssim \frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}\left|\mathbb{E}\left(e_{it}^2|\mathscr{D}\right)\right|=O(1)\ a.s.,
\end{aligned}
$$

where the last line combines Lemma B.7(i) and Assumption 1*(i), (ii), (iii), (v). Similarly as above, we can also show that $\mathbb{E}\left[\left\|\frac{1}{\sqrt{nT_\ell}}\sum_{i=n_1}^{n_n}\lambda_i^0e_i^{(\ell)\prime}\right\|_F^2\middle|\mathscr{D}\right]=O_p(1)$. Then

$$
\left\|\frac{1}{\sqrt{nT_\ell}}\sum_{i=n_1}^{n_n}\lambda_i^0e_i^{(\ell)\prime}F^{0,(\ell)}\right\|_F=O_p(1) \text{ and } \left\|\frac{1}{\sqrt{nT_\ell}}\sum_{i=n_1}^{n_n}\lambda_i^0e_i^{(\ell)\prime}\right\|_F=O_p(1).
$$

Furthermore, we have

$$
\begin{aligned}
\left\|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^0e_i^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_F &\leq \frac{1}{\sqrt{n}}\frac{1}{\sqrt{T_\ell}}\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|_F\left\|\frac{1}{\sqrt{nT_\ell}}\sum_{i=n_1}^{n_n}\lambda_i^0e_i^{(\ell)\prime}\right\|_F\\
&\quad+\frac{1}{\sqrt{nT_\ell}}\left\|\frac{1}{\sqrt{nT_\ell}}\sum_{i=n_1}^{n_n}\lambda_i^0e_i^{(\ell)\prime}F^{0,(\ell)}\right\|_F\left\|H^{(\ell)}\right\|_F\\
&=O_p\left(\frac{B_N}{\sqrt{N}}+\frac{1}{N+\sqrt{NT}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)=O_p\left(\frac{B_N}{\sqrt{N}}+\frac{1}{N}+\frac{1}{\sqrt{NT}}\right).
\end{aligned}
$$

(vii) We first notice that

$$
\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}\right\|_F=\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\lambda_{i^*}^{0\prime}\right\|_F=O_p\left(\sqrt{\frac{\log N}{NT}}\right) \quad \text{(D.19)}
$$

by similar arguments as used to obtain (D.18). This result, in conjunction with Lemma D.4(vi), implies that

$$
\begin{aligned}
&\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}\right\|_F \\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}\right\|_F+\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}\frac{\hat{F}^{(\ell)}\hat{F}^{(\ell)\prime}}{T_\ell}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}\right\|_F \\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}\right\|_F+\max_{i\in\{n_1,\cdots,n_n\}}\frac{\left\|X_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\frac{\hat{F}^{(\ell)}}{\sqrt{T_\ell}}\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}\hat{F}^{(\ell)\prime}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}\right\|_F \\
&=O_p\left(\sqrt{\frac{\log N}{NT}}\right)+O_p\left(\frac{B_N}{\sqrt{N}}+\frac{1}{N}+\frac{1}{\sqrt{NT}}\right)=O_p\left(\frac{B_N}{\sqrt{N}}+\frac{1}{N}+\sqrt{\frac{\log N}{NT}}\right).
\end{aligned}
$$

(viii) Combining (D.18)and Lemma D.4(iii), we have

$$
\begin{aligned}
&\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_{i^*}^{(\ell)}e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_F \\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}e_{i^*}^{(\ell)}e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_F+\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}\frac{\hat{F}^{(\ell)}\hat{F}^{(\ell)\prime}}{T_\ell}e_{i^*}^{(\ell)}e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_F \\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\sqrt{\frac{1}{n}\sum_{i^*=n_1}^{n_n}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\|_2^2}\sqrt{\frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\left\|e_i^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_2^2} \\
&+\frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\left\|e_i^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_2^2\max_{i\in\{n_1,\cdots,n_n\}}\frac{\left\|X_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \\
&=O_p\left(\sqrt{\frac{\log N}{T}}\right)O_p\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right)+O_p\left(B_N^2+\frac{1}{N\wedge T}\right) \\
&=O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{T}}+\frac{\sqrt{\log N}}{N\wedge T}\right).
\end{aligned}
$$

■

Define

$$
\begin{aligned}
&\xi_i^{0,(\ell)}:=\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}e_i^{(\ell)}}{T_\ell},\quad S_{ii^*}^{0,(\ell)}:=\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_{i^*}^{(\ell)}}{T_\ell},\quad a_{ii^*}^0:=\lambda_i^{0\prime}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\lambda_{i^*}^0, \\
&G_{ii^*}^{0,(\ell)}:=S_{ii^*}^{0,(\ell)}a_{ii^*}^0,\quad \text{and } \Omega_i^{0,(\ell)}:=Var(\xi_i^{0,(\ell)}).
\end{aligned}
$$

Lemma D.5 *Under Assumptions 1*, 2 and 8, we have*

(i) $\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right)\left(1-\frac{a_{ii}^0}{n}\right)=\xi_i^{0,(\ell)}+\mathcal{R}_i^{(\ell)}$ *such that* $\max_{i\in\{n_1,\cdots,n_n\}}\left\|\mathcal{R}_i^{(\ell)}\right\|_2=O_p\left(\frac{\log N}{N\wedge T}\right)$,

(ii) $\sqrt{T_\ell}\left(\Omega_i^{0,(\ell)}\right)^{-1/2}\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right)\left(1-\frac{a_{ii}^0}{n}\right)\rightsquigarrow\mathcal{N}(0,1)$,

(iii) $\max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right\|_2=O_p\left(\sqrt{\frac{\log N}{T}}\right)$.

Proof. (i) Recall from (C.3) that $\hat{\theta}_i^{(\ell)} = \left(X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} X_i^{(\ell)}\right)^{-1} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} Y_i^{(\ell)}$, which yields

$$
\begin{aligned}
\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)} &= \left(\frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} X_i^{(\ell)}}{T_\ell}\right)^{-1} \left[\frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} e_i^{(\ell)} + \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} F^{0,(\ell)} \lambda_i^0\right] \\
&= \left(\frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} X_i^{(\ell)}}{T_\ell}\right)^{-1} \frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} e_i^{(\ell)}}{T_\ell} \\
&+ \left(\frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} X_i^{(\ell)}}{T_\ell}\right)^{-1} \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \left[\hat{F}^{(\ell)} H^{(\ell)-1} - \sum_{m\in[8]} J_m^{(\ell)} \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1} \left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\right] \lambda_i^0 \\
&= \left(\frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} X_i^{(\ell)}}{T_\ell}\right)^{-1} \frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} e_i^{(\ell)}}{T_\ell} \\
&- \left(\frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} X_i^{(\ell)}}{T_\ell}\right)^{-1} \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \sum_{m\in[8]} J_m^{(\ell)} \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1} \left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1} \lambda_i^0, \qquad \text{(D.20)}
\end{aligned}
$$

where the second equality is from (D.6) and it's clear that

$$
\begin{aligned}
&\max_{i\in\{n_1,\cdots,n_n\}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \sum_{m\in[8]} J_m^{(\ell)} \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1} \left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1} \lambda_i^0 \right\|_2 \\
&\le \max_{i\in\{n_1,\cdots,n_n\}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \sum_{m\in[8]} J_m^{(\ell)} \right\|_F \left\| \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1} \right\|_F \left\| \left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1} \right\|_F \max_{i\in\{n_1,\cdots,n_n\}} \left\|\lambda_i^0\right\|_2 \\
&= \sum_{m\in[8]} \max_{i\in\{n_1,\cdots,n_n\}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} J_m^{(\ell)} \right\|_F O_p(1),
\end{aligned}
$$

where the last line combines Lemma B.7(i), Lemma D.3, and the normalization of factor and factor loading. Hence, it suffices to show the uniform convergence rate $\frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} J_m^{(\ell)}$ $\forall m \in [8]$ in the following.

First, we observe that

$$
\begin{aligned}
\max_{i\in\{n_1,\cdots,n_n\}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} J_1^{(\ell)} \right\|_F &\le \max_{i\in\{n_1,\cdots,n_n\}} \left\| \frac{1}{\sqrt{T_\ell}} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \right\|_F \left\| J_1^{(\ell)} \right\|_F \\
&= \max_{i\in\{n_1,\cdots,n_n\}} \left\| \frac{1}{\sqrt{T_\ell}} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \right\|_F O_p(B_N^2) = O_p(B_N^2),
\end{aligned}
$$

where the first equality is by (D.7) and the last equality is by (D.17)

Second, we have

$$
\begin{aligned}
&\frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} J_2^{(\ell)} \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1} \left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1} \lambda_i^0 \\
&= \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \frac{1}{nT_\ell} \sum_{i^*=n_1}^{n_n} X_{i^*}^{(\ell)} \left(\theta_{i^*}^{0,(\ell)} - \hat{\theta}_{i^*}^{(\ell)}\right) \lambda_{i^*}^{0\prime} F^{0,(\ell)\prime} \hat{F}^{(\ell)} \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1} \left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1} \lambda_i^0 \\
&= \frac{1}{n} \sum_{i^*=n_1}^{n_n} \frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} X_{i^*}^{(\ell)}}{T_\ell} \left(\theta_{i^*}^{0,(\ell)} - \hat{\theta}_{i^*}^{(\ell)}\right) \lambda_{i^*}^{0\prime} \left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1} \lambda_i^0
\end{aligned}
$$

which leads to $\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}J_2^{(\ell)}\left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\lambda_i^0\right\|_2 = O_p(B_N)$ and this term will be kept in the linear expansion for $\hat{\theta}_i^{(\ell)}$.

Third, it can be shown that

$$
\begin{aligned}
&\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}J_3^{(\ell)}\right\|_2 = \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}X_{i^*}^{(\ell)}\left(\theta_{i^*}^{0,(\ell)}-\hat{\theta}_{i^*}^{(\ell)}\right)e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_2\\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{\sqrt{T_\ell}}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\right\|_F\sqrt{\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}\left\|X_{i^*}^{(\ell)}\left(\theta_{i^*}^{0,(\ell)}-\hat{\theta}_{i^*}^{(\ell)}\right)\right\|_2^2}\sqrt{\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}\left\|e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_2^2}\\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{\sqrt{T_\ell}}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\right\|_F\max_{i\in\{n_1,\cdots,n_n\}}\left\|\theta_{i^*}^{0,(\ell)}-\hat{\theta}_{i^*}^{(\ell)}\right\|_2\sqrt{\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}\left\|X_{i^*}^{(\ell)}\right\|_2^2}\sqrt{\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}\left\|e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_2^2}\\
&= O_p\left(B_N^2 + B_N\frac{1}{\sqrt{N\wedge T}}\right),
\end{aligned}
$$

where the last line combines (D.17), Assumption 8(ii) and Lemma D.4(iii).

Next, for the term with $J_4^{(\ell)}$, we notice that

$$
\begin{aligned}
&\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}J_4^{(\ell)}\right\|_2 = \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}F^{0,(\ell)}\lambda_{i^*}^0\left(\theta_{i^*}^{0,(\ell)}-\hat{\theta}_{i^*}^{(\ell)}\right)'X_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_2\\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}F^{0,(\ell)}\right\|_F\max_{i\in\{n_1,\cdots,n_n\}}\|\lambda_i^0\|_2\max_{i\in\{n_1,\cdots,n_n\}}\left\|\theta_i^{0,(\ell)}-\hat{\theta}_i^{(\ell)}\right\|_2\frac{1}{n\sqrt{T_\ell}}\sum_{i^*=n_1}^{n_n}\left\|X_{i^*}^{(\ell)}\right\|_F\frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\\
&= O_p\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right)O_p(B_N) = O_p\left(B_N^2 + B_N\frac{1}{\sqrt{N\wedge T}}\right),
\end{aligned}
$$

where the last line is by Lemma D.4(iv), normalization of factor vectors and the fact that

$$
\frac{1}{n\sqrt{T_\ell}}\sum_{i=n_1}^{n_n}\left\|X_i^{(\ell)}\right\|_F = \frac{1}{n}\sum_{i=n_1}^{n_n}\sqrt{\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\|X_{it}\|_2^2} = O_p(1)
$$

by Assumption 8(ii).

Furthermore, it is clear that

$$
\begin{aligned}
&\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}J_5^{(\ell)}\right\|_F = \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}e_{i^*}^{(\ell)}\left(\theta_{i^*}^{0,(\ell)}-\hat{\theta}_{i^*}^{(\ell)}\right)'X_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_F\\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\frac{1}{n\sqrt{T_\ell}}\sum_{i^*=n_1}^{n_n}e_{i^*}^{(\ell)}\left(\theta_{i^*}^{0,(\ell)}-\hat{\theta}_{i^*}^{(\ell)}\right)'X_{i^*}^{(\ell)\prime}\right\|_F\frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\\
&\lesssim \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell\sqrt{T_\ell}}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}e_{i^*}^{(\ell)}\left(\theta_{i^*}^{0,(\ell)}-\hat{\theta}_{i^*}^{(\ell)}\right)'X_{i^*}^{(\ell)\prime}\right\|_F\\
&+ \max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}\hat{F}^{(\ell)}\frac{1}{n\sqrt{T_\ell}}\sum_{i^*=n_1}^{n_n}\frac{1}{T_\ell}\hat{F}^{(\ell)\prime}e_{i^*}^{(\ell)}\left(\theta_{i^*}^{0,(\ell)}-\hat{\theta}_{i^*}^{(\ell)}\right)'X_{i^*}^{(\ell)\prime}\right\|_F\\
&\leq \max_{i\in\{n_1,\cdots,n_n\}}\left\|\theta_i^{0,(\ell)}-\hat{\theta}_i^{(\ell)}\right\|_2\sqrt{\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}\left\|X_i^{(\ell)\prime}e_{i^*}^{(\ell)}\right\|_2^2}\sqrt{\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}\left\|X_{i^*}^{(\ell)\prime}\right\|_F^2}
\end{aligned}
$$

$$+\max_{i\in\{n_1,\cdots,n_n\}}\left\|\theta_i^{0,(\ell)}-\hat{\theta}_i^{(\ell)}\right\|_2\max_{i\in\{n_1,\cdots,n_n\}}\frac{\left\|X_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\sqrt{\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}\left\|e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_2^2}\sqrt{\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}\left\|X_{i^*}^{(\ell)\prime}\right\|_F^2}$$

$$=O_p(B_N)O_p\left(\sqrt{\frac{\log N}{T}}\right)+O_p(B_N)O_p\left(B_N+\frac{1}{\sqrt{N\wedge T}}\right)=O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{N\wedge T}}\right),$$

where the last line combines Lemma D.4(iii), D.4(v) and the fact that $\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}\left\|X_{i^*}^{(\ell)\prime}\right\|_F^2=O_p(1)$ by Assumption 8(ii).

Next, we can show that

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}J_6^{(\ell)}\right\|_F=\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}F^{0,(\ell)}\lambda_{i^*}^0e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_F$$

$$\le\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}F^{0,(\ell)}\right\|_F\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}\lambda_{i^*}^0e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_F$$

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

where the last line combines Lemma D.4(iv) and D.4(vi).

Moreover, for the term with $J_7^{(\ell)}$, with Lemma D.4(vii), we observe that

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}J_7^{(\ell)}\right\|_F=\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}F^{0,(\ell)\prime}\hat{F}^{(\ell)}\right\|_F$$

$$=\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}\right\|_F O_p(1)=O_p\left(\frac{B_N}{\sqrt{N}}+\frac{1}{N}+\sqrt{\frac{\log N}{NT}}\right).\tag{D.21}$$

At last, we can also show that

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}J_8^{(\ell)}\right\|_F=\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}e_{i^*}^{(\ell)}e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|_F$$

$$=O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{T}}+\frac{\sqrt{\log N}}{N\wedge T}\right)$$

by Lemma D.4(viii).

Combining (D.20) and the above arguments, we obtain that

$$\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}=\left(\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}X_i^{(\ell)}}{T_\ell}\right)^{-1}\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_i^{(\ell)}}{T_\ell}$$

$$+\left(\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}X_i^{(\ell)}}{T_\ell}\right)^{-1}\frac{1}{n}\sum_{i^*=n_1}^{n_n}\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}X_{i^*}^{(\ell)}}{T_\ell}\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)\lambda_{i^*}^{0\prime}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\lambda_i^0$$

$$+O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{N\wedge T}}+\frac{\sqrt{\log N}}{N\wedge T}\right)\tag{D.22}$$

uniformly in $i\in\{n_1,\cdots,n_n\}$. Combining (D.22) and Lemma D.4(i)-(ii), we have

$$\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}=\left(\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_i^{(\ell)}}{T_\ell}\right)^{-1}\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}e_i^{(\ell)}}{T_\ell}$$

$$+\left(\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_i^{(\ell)}}{T_\ell}\right)^{-1}\frac{1}{n}\sum_{i^*=n_1}^{n_n}\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_{i^*}^{(\ell)}}{T_\ell}\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)\lambda_{i^*}^{0\prime}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\lambda_i^0$$
$$+O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{N\wedge T}}+\frac{\sqrt{\log N}}{N\wedge T}\right).$$

Moreover, let $\mathfrak{x}_{it}^{0,(\ell)}$ be the t-th row of matrix $M_{F^{0,(\ell)}}X_i^{(\ell)}$ and note that $\mathfrak{x}_{it}^{0,(\ell)}$ is strong mixing across t and independent across i conditional on $\mathscr{D}$ by Assumption 1*(i), (iii). Then we can show that

$$\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_{i^*}^{(\ell)}}{T_\ell}-\mathbb{E}\left(\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_{i^*}^{(\ell)}}{T_\ell}\middle|\mathscr{D}\right)=\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left[\mathfrak{x}_{it}^{0,(\ell)}\mathfrak{x}_{i^*t}^{0,(\ell)\prime}-\mathbb{E}\left(\mathfrak{x}_{it}^{0,(\ell)}\mathfrak{x}_{i^*t}^{0,(\ell)\prime}\middle|\mathscr{D}\right)\right]=O_p\left(\sqrt{\frac{\log N}{T}}\right)$$

uniformly over $i,i^*\in\{n_1,\cdots,n_n\}$ by similar arguments in (D.14). Then by the fact that

$$\max_{i,i^*\in\{n_1,\cdots,n_n\}}\left\|\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}e_i^{(\ell)}}{T_\ell}\right\|_2=\max_{i,i^*\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{it}\right\|_2+\max_{i,i^*\in\{n_1,\cdots,n_n\}}\left\|\frac{X_i^{(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right\|_F\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}f_t^0e_{it}\right\|_2$$
$$=O_p\left(\sqrt{\frac{\log N}{T}}\right)$$

by (D.14) and (D.18), we obtain that

$$\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}=\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_i^{0,(\ell)}+\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\frac{1}{n}\sum_{i^*=n_1}^{n_n}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)$$
$$+O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{N\wedge T}}+\frac{\sqrt{\log N}}{N\wedge T}\right).\tag{D.23}$$

For the second term on the right side of (D.23), we observe that

$$\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\frac{1}{n}\sum_{i^*=n_1}^{n_n}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)$$
$$=\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\mathbb{E}\left(G_{ii}^{0,(\ell)}|\mathscr{D}\right)\frac{1}{n}\left(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right)+\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)$$
$$=\frac{a_{ii}^0}{n}\left(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right)+\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right).$$

By (D.23), it's clear that

$$\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)=\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left[\mathbb{E}\left(S_{i^*i^*}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_{i^*}^{0,(\ell)}$$
$$+\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left[\mathbb{E}\left(S_{i^*i^*}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\frac{1}{n}\sum_{j=n_1}^{n_n}\mathbb{E}\left(G_{i^*j}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_j^{(\ell)}-\theta_j^{0,(\ell)}\right)$$
$$+O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{N\wedge T}}+\frac{\sqrt{\log N}}{N\wedge T}\right)$$

where the second term on the right side of above equality gives the recursive form and shrinks to zero quickly owing to the $\frac{1}{n^k}$ term, and we only need to show the rate of the first term, i.e.,

$$\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left[\mathbb{E}\left(S_{i^*i^*}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_{i^*}^{0,(\ell)}=\frac{1}{nT_\ell}\sum_{i^*\neq i}\sum_{t\in\mathcal{T}_\ell}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left[\mathbb{E}\left(S_{i^*i^*}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\mathfrak{x}_{i^*t}e_{i^*t}$$

$$= O_p\left(\sqrt{\frac{\log N}{NT}}\right) \text{ uniformly over } i \in \{n_1, \cdots, n_n\},$$

similarly to the result in (D.19). This yields

$$\max_{i\in\{n_1,\cdots,n_n\}} \left\| \frac{1}{n}\sum_{i^*\neq i} \mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)} - \theta_{i^*}^{0,(\ell)}\right)\right\|_2 = O_p\left(B_N^2 + B_N\sqrt{\frac{\log N}{N\wedge T}} + \frac{\sqrt{\log N}}{N\wedge T}\right)$$

and further gives

$$\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right)\left(1 - \frac{a_{ii}^0}{n}\right) = \left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_i^{0,(\ell)} + O_p\left(B_N^2 + B_N\sqrt{\frac{\log N}{N\wedge T}} + \frac{\sqrt{\log N}}{N\wedge T}\right)$$

with $B_N = \max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2 = O_p\left(\sqrt{\frac{\log N}{T}}\right)$. Finally, we obtain that

$$\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right)\left(1 - \frac{a_{ii}^0}{n}\right) = \xi_i^{0,(\ell)} + \mathcal{R}_i^{(\ell)}$$

such that $\max_{i\in\{n_1,\cdots,n_n\}}\left\|\mathcal{R}_i^{(\ell)}\right\|_2 = O_p\left(\frac{\log N}{N\wedge T}\right)$.

(ii) Given the definition of $\Omega_i^{0,(\ell)}$ and by central limit theorem for m.d.s., we can easily obtain (ii).

(iii) The proof has already been done in the proof of (i). ∎

Lemma D.6 *Under Assumptions 1*, 2 and 8, we have*

(i) $\frac{1}{\sqrt{T_\ell}}\left\|\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}\right\|_F = O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right)$,

(ii) $\left\|M_{\hat{F}^{(\ell)}} - M_{F^{0,(\ell)}}\right\|_F = O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right)$,

(iii) $\max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{\lambda}_i^{(\ell)} - H^{(\ell)-1}\lambda_i^0\right\|_2 = O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right)$,

(iv) $\max_{t\in[T]}\left\|\hat{f}_t^{(\ell)} - H^{(\ell)\prime}f_t^0\right\|_2 = O_p\left(\sqrt{\frac{\log(N\vee T)}{N\wedge T}}\right)$.

Proof. (i) We obtain the result by combining Lemma D.3 and Lemma D.5(iii).

(ii) We obtain the result by combining (D.12) and Lemma D.5(iii).

(iii) Recall that

$$\begin{aligned}
\hat{\lambda}_i^{(\ell)} &= \left(\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}\right)^{-1}\hat{F}^{(\ell)\prime}\left(Y_i^{(\ell)} - X_i^{(\ell)\prime}\hat{\theta}_i^{(\ell)}\right) = \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left[Y_i^{(\ell)} - X_i^{(\ell)\prime}\theta_i^{0,(\ell)} - X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right)\right] \\
&= \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left[F^{0,(\ell)}\lambda_i^0 + e_i^{(\ell)} - X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right)\right] \\
&= \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left(F^{0,(\ell)} - \hat{F}^{(\ell)}H^{(\ell)-1}\right)\lambda_i^0 + \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}H^{(\ell)-1}\lambda_i^0 + \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}e_i^{(\ell)} - \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right) \\
&= H^{(\ell)-1}\lambda_i^0 + \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left(F^{0,(\ell)} - \hat{F}^{(\ell)}H^{(\ell)-1}\right)\lambda_i^0 + \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}e_i^{(\ell)} - \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right)
\end{aligned}$$

where the second and fifth equalities are by the normalization that $\frac{\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell} = I_{r_0}$. It follows that

$$\begin{aligned}
\hat{\lambda}_i^{(\ell)} - H^{(\ell)-1}\lambda_i^0 &= \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left(F^{0,(\ell)} - \hat{F}^{(\ell)}H^{(\ell)-1}\right)\lambda_i^0 + \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}e_i^{(\ell)} - \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right) \\
&= \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left(F^{0,(\ell)} - \hat{F}^{(\ell)}H^{(\ell)-1}\right)\lambda_i^0 + \frac{1}{T_\ell}\left(\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}\right)^{\prime}e_i^{(\ell)} \\
&+ H^{(\ell)\prime}\frac{1}{T_\ell}F^{0,(\ell)}e_i^{(\ell)} - \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right) \\
&:= I_{i,1}^{(\ell)} + I_{i,2}^{(\ell)} + I_{i,3}^{(\ell)} - I_{i,4}^{(\ell)}.
\end{aligned}$$

First, it is clear that

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|I_{i,1}^{(\ell)}\right\|_2 \leq \left\|\frac{\hat{F}^{(\ell)}}{\sqrt{T_\ell}}\right\|_F \frac{\left\|F^{0,(\ell)} - \hat{F}^{(\ell)}H^{(\ell)-1}\right\|_F}{\sqrt{T_\ell}} \max_{i\in\{n_1,\cdots,n_n\}}\left\|\lambda_i^0\right\|_2 = O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right),$$

where the equality is by Lemma D.6(i) and Lemma B.7(i). Second, we have

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|I_{i,2}^{(\ell)}\right\|_2 \leq \frac{1}{\sqrt{T_\ell}}\left\|\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}\right\|_F \frac{\left\|e_i^{(\ell)}\right\|_2}{\sqrt{T_\ell}} = O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right)$$

by Lemma D.6(i) and Assumption 8(ii). Next, we obtain that $\max_{i\in\{n_1,\cdots,n_n\}}\left\|I_{i,3}^{(\ell)}\right\|_2 = O_p\left(\sqrt{\frac{\log N}{T}}\right)$ by (D.14). And at last, we observe that

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|I_{i,4}^{(\ell)}\right\|_2 \leq \left\|\frac{\hat{F}^{(\ell)}}{\sqrt{T_\ell}}\right\|_F \frac{\left\|X_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2 = O_p\left(\sqrt{\frac{\log N}{T}}\right),$$

where the equality is by Lemma D.5(iii). Combining arguments above, we have shown that

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{\lambda}_i - H^{(\ell)-1}\lambda_i\right\|_2 = O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right).$$

(iv) Recall from (D.6) that $\hat{F}^{(\ell)\prime} - H^{(\ell)\prime}F' = V_{NT}^{(\ell)-1}\sum_{m\in[8]}J_m^{(\ell)\prime}$ with $J_m^{(\ell)}$ $\forall m\in[8]$ defined in the proof of Lemma D.3. Let $J_{m,t}^{(\ell)}$ be the t-th column of $V_{NT}^{(\ell)-1}J_{m,t}^{(\ell)}$ for $m\in[8]$. We observe that $\hat{f}_t - H^{(\ell)\prime}f_t^0$ is the t-th column of $\hat{F}^{(\ell)\prime} - H^{(\ell)\prime}F'$, which remains to show the convergence rate for $J_{m,t}^{(\ell)}, m\in[8]$. Below we examine the uniform convergence rate for $J_{m,t}^{(\ell)}$ one by one.

For $J_{1,t}^{(\ell)}$, we notice that

$$\begin{aligned}
\max_{t\in\mathcal{T}_\ell}\left\|J_{1,t}^{(\ell)}\right\|_2 &= \max_{t\in\mathcal{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}\left[\frac{1}{NT_\ell}\sum_{i=n_1}^{n_n}X_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)^{\prime}X_{it}\right]\right\|_2 \\
&\leq \left\|V_{NT}^{(\ell)-1}\right\|_F \frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2^2 \max_{i\in\{n_1,\cdots,n_n\}}\frac{\left\|X_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \max_{t\in\mathcal{T}_\ell}\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|X_{it}\right\|_2 \\
&= O_p\left(\frac{\log N}{T}\right),
\end{aligned}$$

where the last line combines Lemma D.5(iii) and Assumption 8(ii). Similarly, with Lemma B.7(i), we have

$$\max_{t\in\mathcal{T}_\ell}\left\|J_{2,t}^{(\ell)}\right\|_2 = \max_{t\in\mathcal{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}F^{0,(\ell)}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^0\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)^{\prime}X_{it}\right]\right\|_2$$

$$
\begin{aligned}
&\leq \left\|V_{NT}^{(\ell)-1}\right\|_F \left\|\frac{\hat{F}^{(\ell)}}{\sqrt{T_\ell}}\right\|_F \left\|\frac{F^{0,(\ell)}}{\sqrt{T_\ell}}\right\|_F \max_{i\in\{n_1,\cdots,n_n\}} \|\lambda_i^0\|_2 \max_{i\in\{n_1,\cdots,n_n\}} \left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2 \frac{1}{n}\sum_{i=n_1}^{n_n} \|X_{it}\|_2 \\
&= O_p\left(\sqrt{\frac{\log N}{T}}\right),
\end{aligned}
$$

$$
\begin{aligned}
\max_{t\in\mathcal{T}_\ell}\left\|J_{3,t}^{(\ell)}\right\|_2 &= \max_{t\in\mathcal{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)' X_{it}\right]\right\|_2 \\
&\leq \left\|V_{NT}^{(\ell)-1}\right\|_F \left\|\frac{\hat{F}^{(\ell)}}{\sqrt{T_\ell}}\right\|_F \max_{i\in\{n_1,\cdots,n_n\}} \left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2 \max_{i\in\{n_1,\cdots,n_n\}} \frac{\left\|e_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \max_{t\in\mathcal{T}_\ell}\frac{1}{n}\sum_{i=n_1}^{n_n}\|X_{it}\|_2 \\
&= O_p\left(\sqrt{\frac{\log N}{T}}\right),
\end{aligned}
$$

$$
\begin{aligned}
\max_{t\in\mathcal{T}_\ell}\left\|J_{4,t}^{(\ell)}\right\|_2 &= \max_{t\in\mathcal{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} X_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)\lambda_i^{0\prime}\right] f_t^0\right\|_2 \\
&\leq \left\|V_{NT}^{(\ell)-1}\right\|_F \left\|\frac{\hat{F}^{(\ell)}}{\sqrt{T_\ell}}\right\|_F \max_{i\in\{n_1,\cdots,n_n\}} \left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2 \max_{i\in\{n_1,\cdots,n_n\}} \frac{\left\|X_i^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \max_{i\in\{n_1,\cdots,n_n\}}\|\lambda_i^0\|_2 \max_{t\in\mathcal{T}_\ell}\|f_t^0\|_2 \\
&= O_p\left(\sqrt{\frac{\log N}{T}}\right).
\end{aligned}
$$

Besides, we have

$$
\max_{t\in\mathcal{T}_\ell}\left\|J_{5,t}^{(\ell)}\right\|_2 = \max_{t\in\mathcal{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} X_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right) e_{it}\right]\right\|_2 = O_p\left(\sqrt{\frac{\log N}{T}}\right)
$$

by similar arguments as for $J_{3,t}^{(\ell)}$. Moreover, we notice that

$$
\begin{aligned}
\max_{t\in\mathcal{T}_\ell}\left\|J_{6,t}^{(\ell)}\right\|_2 &= \max_{t\in\mathcal{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)}\lambda_i^{0\prime}\right] f_t^0\right\|_2 \\
&\leq \frac{1}{\sqrt{N}}\left\|V_{NT}^{(\ell)-1}\right\|_F \frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \max_{t\in\mathcal{T}_\ell}\|f_t^0\|_2 \frac{1}{\sqrt{nT_\ell}}\left\|\sum_{i=n_1}^{n_n} e_i^{(\ell)}\lambda_i^{0\prime}\right\|_F = O_p\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}
$$

where the equality follows from the fact that

$$
\mathbb{E}\left(\frac{1}{nT_\ell}\left\|\sum_{i=n_1}^{n_n} e_i^{(\ell)}\lambda_i^{0\prime}\right\|_F^2 \middle| \mathscr{D}\right) = \mathbb{E}\left(\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left\|\frac{1}{\sqrt{n}}\sum_{i\in[N]} e_{it}\lambda_i^0\right\|_2^2 \middle| \mathscr{D}\right) = O(1)
$$

with the same manner as (D.1).

Next, we show that

$$
\begin{aligned}
\max_{t\in\mathcal{T}_\ell}\left\|J_{7,t}^{(\ell)}\right\|_2 &= \max_{t\in\mathcal{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}F^{0,(\ell)}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^0 e_{it}\right]\right\|_2 \\
&\leq \left\|V_{NT}^{(\ell)-1}\right\|_F \frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}} \frac{\|F^{0,(\ell)}\|_F}{\sqrt{T_\ell}} \max_{t\in\mathcal{T}_\ell}\left\|\frac{1}{n}\sum_{i=n_1}^{n_n}\lambda_i^0 e_{it}\right\|_2 = O_p\left(\sqrt{\frac{\log T}{N}}\right),
\end{aligned}
$$

where the last equality is by the fact that $\max_{t\in\mathcal{T}_\ell}\left\|\frac{1}{n}\sum_{i=n_1}^{n_n}\lambda_i^0 e_{it}\right\|_2 = O_p\left(\sqrt{\frac{\log T}{N}}\right)$ by using the Bernstein's inequality for the independent sequence in Lemma B.5(i).

At last, we also obtain that

$$\begin{aligned}
\max_{t\in\mathcal{T}_\ell}\left\|J_{8,t}^{(\ell)}\right\|_2 &\le \frac{1}{\sqrt{n}}\left\|V_{NT}^{(\ell)-1}\right\|_F \frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\max_{t\in\mathcal{T}_\ell}\left\|\frac{1}{\sqrt{nT_\ell}}\sum_{i=n_1}^{n_n} e_i^{(\ell)}e_{it}\right\|_2 \\
&= \frac{1}{\sqrt{n}}\left\|V_{NT}^{(\ell)-1}\right\|_F \frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\sqrt{\max_{t\in\mathcal{T}_\ell}\frac{1}{T_\ell}\sum_{s\in[T_\ell]}\left(\frac{1}{\sqrt{n}}\sum_{i=n_1}^{n_n} e_{is}e_{it}\right)^2} \\
&\le \frac{1}{\sqrt{n}}\left\|V_{NT}^{(\ell)-1}\right\|_F \frac{\left\|\hat{F}^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\sqrt{\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in[T_\ell]}\left(\frac{1}{\sqrt{n}}\sum_{i=n_1}^{n_n} e_{is}e_{it}\right)^2} \\
&= O_p\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}$$

where the last line is given by the fact that $\mathbb{E}\left(\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in[T_\ell]}\left(\frac{1}{\sqrt{N}}\sum_{i=n_1}^{n_n} e_{is}e_{it}\right)^2\right) = O(1)$ by (D.8).

Combining the above arguments, we have derived that $\max_{t\in[T]}\left\|\hat{f}_t - H^{(\ell)\prime}f_t^0\right\|_2 = O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right)$. ■

Lemma D.7 *Under Assumptions 1*, 2 and 8, we have*

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{S}_{ii}^{(\ell)} - \mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right\|_F = O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right) \text{ and } \max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{\Omega}_i^{(\ell)} - \Omega_i^{0,(\ell)}\right\|_F = o_p(1).$$

Proof. Recall that $S_{ii}^{0,(\ell)} = \frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_i^{(\ell)}}{T_\ell}$ and $\hat{S}_{ii}^{(\ell)} = \frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}X_i^{(\ell)}}{T_\ell}$. Combining (D.21), Lemma D.4(i) and Lemma D.5(iii), it is clear that

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|\hat{S}_{ii}^{(\ell)} - \mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right\|_F = O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right).$$

Recall that $\mathfrak{x}_{it}^{(\ell)\prime}$ is the t-th row of $M_{F^{0,(\ell)}}X_i^{(\ell)}$ and let $\hat{\mathfrak{x}}_{it}^{(\ell)\prime}$ be the t-th row of $M_{\hat{F}^{(\ell)}}X_i^{(\ell)}$, respectively. Under Assumption 1*(iii), noted that

$$\hat{\Omega}_i^{(\ell)} = \frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\hat{\mathfrak{x}}_{it}^{(\ell)}\hat{\mathfrak{x}}_{it}^{(\ell)\prime}\hat{e}_{it}^2, \quad \Omega_i^{0,(\ell)} = \frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\mathbb{E}\left(\mathfrak{x}_{it}^{(\ell)}\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2\right),$$

and it remains to show

$$\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left[\hat{\mathfrak{x}}_{it}^{(\ell)}\hat{\mathfrak{x}}_{it}^{(\ell)\prime}\hat{e}_{it}^2 - \mathbb{E}\left(\mathfrak{x}_{it}^{(\ell)}\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2\right)\right]\right\|_F = o_p(1). \tag{D.24}$$

By the definitions of $\mathfrak{x}_{it}^{(\ell)}$ and $\hat{\mathfrak{x}}_{it}^{(\ell)}$, we notice that $\mathfrak{x}_{it}^{(\ell)} = X_{it} - \frac{1}{T_\ell}X_i^{(\ell)\prime}F^{0,(\ell)}f_t^0$ and $\hat{\mathfrak{x}}_{it}^{(\ell)} = X_{it} - \frac{1}{T_\ell}X_i^{(\ell)\prime}\hat{F}^{(\ell)}\hat{f}_t^0$, which gives

$$\mathfrak{x}_{it}^{(\ell)} - \hat{\mathfrak{x}}_{it}^{(\ell)} = \frac{1}{T_\ell}X_i^{(\ell)\prime}\left(\hat{F}^{(\ell)}\hat{f}_t^0 - F^{0,(\ell)}f_t^0\right) \tag{D.25}$$

such that

$$\max_{t\in\mathcal{T}_\ell}\frac{1}{\sqrt{T_\ell}}\left\|\hat{F}^{(\ell)}\hat{f}_t^0-F^{0,(\ell)}f_t^0\right\|_2\leq\frac{\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\max_{t\in[T]}\left\|\hat{f}_t^{(\ell)}-H^{(\ell)\prime}f_t^0\right\|_2$$
$$+\frac{\left\|F^{0,(\ell)}H^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\max_{t\in[T]}\left\|\hat{f}_t^{(\ell)}-H^{(\ell)\prime}f_t^0\right\|_2+\frac{\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|_F}{\sqrt{T_\ell}}\max_{t\in[T]}\left\|H^{(\ell)\prime}f_t^0\right\|_2$$
$$=O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right),$$

where the last line holds by Lemma D.6(i), (iv) and Lemma B.7(i). Together with (D.25) and Assumption 8(ii), it yields

$$\max_{i\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell}\left\|\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right\|_2\leq\frac{1}{\sqrt{T_\ell}}\left\|X_i\right\|_F O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right)=O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right).\tag{D.26}$$

Next, for $i\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell$, noted that

$$\hat{e}_{it}=Y_{it}-X_{it}'\hat{\theta}_i^{(\ell)}-\hat{\lambda}_i^{(\ell)\prime}\hat{f}_t^{(\ell)}$$
$$=e_{it}-\Bigg[X_{it}'\left(\theta_i^{0,(\ell)}-\hat{\theta}_i^{(\ell)}\right)+\left(\hat{\lambda}_i^{(\ell)}-H^{(\ell)-1}\lambda_i^0\right)'\left(\hat{f}_t^{(\ell)}-H^{(\ell)\prime}f_t^0\right)+\left(\hat{\lambda}_i^{(\ell)}-H^{(\ell)-1}\lambda_i^0\right)'H^{(\ell)\prime}f_t^0$$
$$+\left(H^{(\ell)-1}\lambda_i^0\right)'\left(\hat{f}_t^{(\ell)}-H^{(\ell)\prime}f_t^0\right)\Bigg],$$

which leads to the fact that

$$\max_{i\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell}\left|\hat{e}_{it}-e_{it}\right|=O_p\left(\sqrt{\frac{\log N}{T}}(NT)^{1/q}\right)+O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right),\quad\text{and}$$
$$\hat{e}_{it}^2-e_{it}^2=e_{it}\left(\hat{e}_{it}-e_{it}\right)+\left(\hat{e}_{it}-e_{it}\right)^2=e_{it}X_{it}'R_{1,it}+R_{2,it}\tag{D.27}$$
$$\text{s.t.}\max_{i\in\{n_1,\cdots,n_n\},t\in\mathcal{T}_\ell}\left\|R_{1,it}\right\|_2=O_p\left(\sqrt{\frac{\log N}{T}}\right),\ \max_{i\in[N],t\in\mathcal{T}_\ell}\left|R_{2,it}\right|=O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right)$$

by Lemma D.5(iii) and Lemma D.6(iii), D.6(iv). Then we obtain that

$$\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\hat{\mathfrak{x}}_{it}^{(\ell)}\hat{\mathfrak{x}}_{it}^{(\ell)\prime}\hat{e}_{it}^2=\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\mathfrak{x}_{it}^{(\ell)}\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)'\left(\hat{e}_{it}^2-e_{it}^2\right)$$
$$+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)'e_{it}^2+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\mathfrak{x}_{it}^{(\ell)}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)'\left(\hat{e}_{it}^2-e_{it}^2\right)$$
$$+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)\mathfrak{x}_{it}^{(\ell)\prime}\left(\hat{e}_{it}^2-e_{it}^2\right)+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2$$
$$+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\mathfrak{x}_{it}^{(\ell)}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)'e_{it}^2+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\mathfrak{x}_{it}^{(\ell)}\mathfrak{x}_{it}^{(\ell)\prime}\left(\hat{e}_{it}^2-e_{it}^2\right)$$
$$=\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\mathfrak{x}_{it}^{(\ell)}\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2+O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right)\quad\text{uniformly over } i\in\{n_1,\cdots,n_n\},\tag{D.28}$$

where the last line holds by (D.26), (D.27), Assumption 8(ii) and Assumption 1*(iv). Using similar arguments as used to derive (D.18) by the Bernstein's inequality for m.d.s., for a positive constant c_9, we have

$$\mathbb{P}\left\{\max_{i\in\{n_1,\cdots,n_n\}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left[\mathfrak{x}_{it}^{(\ell)}\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2-\mathbb{E}\left(\mathfrak{x}_{it}^{(\ell)}\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2\right)\right]\right\|_F>c_9\sqrt{\frac{\log N}{T}}\right\}=o(1). \tag{D.29}$$

Combining (D.28) and (D.29), (D.24) is obtained. ■

Lemma D.8 *Under Assumptions 1*, 2 and 8, we have* $\hat{\Gamma}^{(\ell)}\rightsquigarrow\mathcal{N}(0,1)$ *under* H_0.

Proof. Under the null that $\theta_i^{0,(\ell)}=\theta^{0,(\ell)}$ for $\forall i\in\{n_1,\cdots,n_n\}$, from Lemma D.5(i), we notice that

$$\begin{aligned}\hat{\bar{\theta}}^{(\ell)}-\theta^{0,(\ell)}&=\frac{1}{n}\sum_{i\in\{n_1,\cdots,n_n\}}\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_i^{(\ell)}\\&+\frac{1}{n^2}\sum_{i\in\{n_1,\cdots,n_n\}}a_{ii}^0\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\left(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right)+O_p\left(\frac{\log N}{N\wedge T}\right),\end{aligned} \tag{D.30}$$

such that

$$\left\|\frac{1}{n^2}\sum_{i\in\{n_1,\cdots,n_n\}}a_{ii}^0\left(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right)\right\|_2\leq\frac{1}{n}\max_{i\in\{n_1,\cdots,n_n\}}\left|a_{ii}^{(\ell)}\right|\left\|\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right\|_2=O_p\left(\frac{\sqrt{\log N}}{N\sqrt{T}}\right)$$

with Lemma D.5(iii) and the fact that $\max_{i\in\{n_1,\cdots,n_n\}}\left|a_{ii}^0\right|=O(1)$. For the first term on the right side of (D.30), we have

$$\frac{1}{n}\sum_{i\in\{n_1,\cdots,n_n\}}\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_i^{(\ell)}=\frac{1}{nT_\ell}\sum_{i\in\{n_1,\cdots,n_n\}}\sum_{t\in\mathcal{T}_\ell}\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\mathfrak{x}_{it}e_{it}=O_p\left(\frac{1}{\sqrt{nT}}\right)$$

by the central limit theorem for m.d.s., which yields that

$$\left\|\hat{\bar{\theta}}^{0,(\ell)}-\theta^{0,(\ell)}\right\|_2=O_p\left(\frac{\log N}{N\wedge T}\right). \tag{D.31}$$

Recall from (C.5) that $\hat{\Gamma}^{(\ell)}=\sqrt{n}\left(\frac{\frac{1}{n}\sum_{i\in\{n_1,\cdots,n_n\}}\hat{\mathbb{S}}_i^{(\ell)}-p}{\sqrt{2p}}\right)$ such that

$$\begin{aligned}\hat{\mathbb{S}}_i^{(\ell)}&=T_\ell\left(\hat{\theta}_i^{(\ell)}-\hat{\bar{\theta}}^{(\ell)}\right)'\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\left(\hat{\theta}_i^{(\ell)}-\hat{\bar{\theta}}^{(\ell)}\right)\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2\\&=T_\ell\left(\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)}\right)'\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\left(\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)}\right)\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2\\&+T_\ell\left(\hat{\bar{\theta}}^{(\ell)}-\theta^{0,(\ell)}\right)'\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\left(\hat{\bar{\theta}}^{(\ell)}-\theta^{0,(\ell)}\right)\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2\\&-2T_\ell\left(\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)}\right)'\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\left(\hat{\bar{\theta}}^{(\ell)}-\theta^{0,(\ell)}\right)\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2\\&:=\hat{\mathbb{S}}_{i,1}^{(\ell)}+\hat{\mathbb{S}}_{i,2}^{(\ell)}-\hat{\mathbb{S}}_{i,3}^{(\ell)}.\end{aligned}$$

Below we show that $\frac{1}{\sqrt{n}}\sum_{i\in\{n_1,\cdots,n_n\}}\hat{\mathbb{S}}_{i,2}^{(\ell)}$ and $\frac{1}{\sqrt{n}}\sum_{i\in\{n_1,\cdots,n_n\}}\hat{\mathbb{S}}_{i,3}^{(\ell)}$ are smaller terms and $\frac{1}{\sqrt{n}}\sum_{i\in\{n_1,\cdots,n_n\}}\hat{\mathbb{S}}_{i,1}^{(\ell)}\rightsquigarrow\mathcal{N}(0,1)$.

First, noted that

$$\left|\frac{1}{\sqrt{n}}\sum_{i\in\{n_1,\cdots,n_n\}}\hat{\mathbb{S}}_{i,2}^{(\ell)}\right|\leq\sqrt{n}T_\ell\left\|\hat{\bar{\theta}}^{(\ell)}-\theta^{0,(\ell)}\right\|_2^2\max_{i\in\{n_1,\cdots,n_n\}}\lambda_{\max}\left(\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\right)\max_{i\in\{n_1,\cdots,n_n\}}\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2$$

$$
\begin{aligned}
&= \sqrt{n} T_\ell O_p \left(\frac{(\log N)^2}{N^2 \wedge T^2} \right) \max_{i \in \{n_1, \cdots, n_n\}} \left\| \hat{S}_{ii}^{(\ell)} \right\|_F^2 \max_{i \in \{n_1, \cdots, n_n\}} \left\| \hat{\Omega}_i^{(\ell)} \right\|_F \max_{i \in \{n_1, \cdots, n_n\}} \left(1 - \hat{a}_{ii}^{(\ell)}/n \right)^2 \\
&= \sqrt{n} T_\ell O_p \left(\frac{(\log N)^2}{N^2 \wedge T^2} \right) \left[\max_{i \in \{n_1, \cdots, n_n\}} \left\| S_{ii}^{0,(\ell)} \right\|_F^2 \max_{i \in \{n_1, \cdots, n_n\}} \left\| \Omega_i^{0,(\ell)} \right\|_F \max_{i \in \{n_1, \cdots, n_n\}} \left(1 - a_{ii}^0/N \right)^2 + o_p(1) \right] \\
&= o_p(1),
\end{aligned}
$$

where the first equality is by (D.31), the second equality is by Lemma D.7 and the fact that $\max_{i \in \{n_1, \cdots, n_n\}} \left| \hat{a}_{ii}^{(\ell)} - a_{ii}^0 \right| = o_p(1)$ owing to Lemma D.6(iii), and the last line is due to the fact that $\max_{i \in \{n_1, \cdots, n_n\}} \left\| S_{ii}^{0,(\ell)} \right\|_F = O(1)$, $\max_{i \in \{n_1, \cdots, n_n\}} \left\| \Omega_i^{0,(\ell)} \right\|_F = O(1)$, and $\max_{i \in \{n_1, \cdots, n_n\}} \left| a_{ii}^{(\ell)} \right| = O(1)$ and Assumption 1*(vi).

Second, by analogous arguments as used above, we have

$$
\begin{aligned}
&\left| \frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,3}^{(\ell)} \right| \\
&\leq 2\sqrt{n} T_\ell \max_{i \in \{n_1, \cdots, n_n\}} \left\| \hat{\theta}_i^{(\ell)} - \theta^{0,(\ell)} \right\|_2 \left\| \hat{\bar{\theta}}^{(\ell)} - \theta^{0,(\ell)} \right\|_2 \max_{i \in \{n_1, \cdots, n_n\}} \lambda_{\max} \left(\hat{S}_{ii}^{(\ell)} \left(\hat{\Omega}_i^{(\ell)} \right)^{-1} \hat{S}_{ii}^{(\ell)} \right) \max_{i \in \{n_1, \cdots, n_n\}} \left(1 - \hat{a}_{ii}^{(\ell)}/n \right)^2 \\
&= \sqrt{n} T_\ell O_p \left(\sqrt{\frac{\log N}{T}} \right) O_p \left(\frac{\log N}{N \wedge T} \right) = o_p(1).
\end{aligned}
$$

At last for $\hat{\mathbb{S}}_{i,1}^{(\ell)}$, it's clear that

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,1}^{(\ell)} &= \frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} T_\ell \left(\hat{\theta}_i^{(\ell)} - \theta^{0,(\ell)} \right)' \hat{S}_{ii}^{(\ell)} \left(\hat{\Omega}_i^{(\ell)} \right)^{-1} \hat{S}_{ii}^{(\ell)} \left(\hat{\theta}_i^{(\ell)} - \theta^{0,(\ell)} \right) \left(1 - \hat{a}_{ii}^{(\ell)}/n \right)^2 \\
&= \frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{z}_i^{(\ell)} + o_p(1) \\
\text{s.t.} \quad \hat{z}_i^{(\ell)} &= T_\ell \left(\hat{\theta}_i^{(\ell)} - \theta^{0,(\ell)} \right)' S_{ii}^{0,(\ell)} \left(\Omega_i^{0,(\ell)} \right)^{-1} S_{ii}^{0,(\ell)} \left(\hat{\theta}_i^{(\ell)} - \theta^{0,(\ell)} \right) \left(1 - a_{ii}^0/n \right)^2.
\end{aligned}
$$

Then by the central limit theorem, we obtain the final result that

$$
\hat{\Gamma}^{(\ell)} = \frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \frac{\hat{z}_i^{(\ell)} - p}{\sqrt{2p}} + o_p(1) \rightsquigarrow \mathcal{N}(0, 1).
$$

■

Lemma D.9 *Under Assumptions 1*, 2 and 8, we have* $\left| \hat{\Gamma}^{(\ell)} \right| \to \infty$ *under* H_1 *if* $\frac{T_\ell}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \left\| c_i^{(\ell)} \right\|_2^2 \to \infty$.

Proof. With the fact that $\theta_i^{0,(\ell)} = \theta^{0,(\ell)} + c_i^{(\ell)}$, it is clear that

$$
\hat{\theta}_i^{(\ell)} - \hat{\bar{\theta}}^{(\ell)} = \left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)} \right) - \left(\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)} \right) + \theta_i^{0,(\ell)} - \theta^{0,(\ell)} = \left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)} \right) - \left(\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)} \right) + c_i^{(\ell)},
$$

which follows by

$$
\begin{aligned}
\hat{\mathbb{S}}_i^{(\ell)} &= T_\ell \left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)} \right)' \hat{S}_{ii}^{(\ell)} \left(\hat{\Omega}_i^{(\ell)} \right)^{-1} \hat{S}_{ii}^{(\ell)} \left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)} \right) \left(1 - \hat{a}_{ii}^{(\ell)}/n \right)^2 \\
&\quad + T_\ell \left(\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)} \right)' \hat{S}_{ii}^{(\ell)} \left(\hat{\Omega}_i^{(\ell)} \right)^{-1} \hat{S}_{ii}^{(\ell)} \left(\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)} \right) \left(1 - \hat{a}_{ii}^{(\ell)}/n \right)^2
\end{aligned}
$$

$$
\begin{aligned}
&- 2T_\ell \left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right)' \hat{S}_{ii}^{(\ell)} \left(\hat{\Omega}_i^{(\ell)}\right)^{-1} \hat{S}_{ii}^{(\ell)} \left(\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)}\right) \left(1 - \hat{a}_{ii}^{(\ell)}/n\right)^2 \\
&+ T_\ell c_i^{(\ell)\prime} \hat{S}_{ii}^{(\ell)} \left(\hat{\Omega}_i^{(\ell)}\right)^{-1} \hat{S}_{ii}^{(\ell)} c_i^{(\ell)} \left(1 - \hat{a}_{ii}^{(\ell)}/n\right)^2 + 2T_\ell \left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right)' \hat{S}_{ii}^{(\ell)} \left(\hat{\Omega}_i^{(\ell)}\right)^{-1} \hat{S}_{ii}^{(\ell)} c_i^{(\ell)} \left(1 - \hat{a}_{ii}^{(\ell)}/n\right)^2 \\
&- 2T_\ell \left(\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)}\right)' \hat{S}_{ii}^{(\ell)} \left(\hat{\Omega}_i^{(\ell)}\right)^{-1} \hat{S}_{ii}^{(\ell)} c_i^{(\ell)} \left(1 - \hat{a}_{ii}^{(\ell)}/n\right)^2 \\
&:= \sum_{m=4}^{9} \hat{\mathbb{S}}_{i,m}^{(\ell)}.
\end{aligned}
$$

In the proof of Lemma D.8, we have already shown that

$$
\frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \frac{\hat{\mathbb{S}}_{i,4}^{(\ell)} - p}{\sqrt{2p}} \rightsquigarrow \mathcal{N}(0,1) \quad \text{and} \quad \left| \frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,m}^{(\ell)} \right| = o_p(1), \; m = 5, 6.
$$

As for $\hat{\mathbb{S}}_{i,7}^{(\ell)}$, we can show that

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,7}^{(\ell)} &= \frac{T_\ell}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} c_i^{(\ell)\prime} \hat{S}_{ii}^{(\ell)} \left(\hat{\Omega}_i^{(\ell)}\right)^{-1} \hat{S}_{ii}^{(\ell)} c_i^{(\ell)} \left(1 - \hat{a}_{ii}^{(\ell)}/n\right)^2 \\
&= \frac{T_\ell}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \left[c_i^{(\ell)\prime} S_{ii}^{0,(\ell)} \left(\Omega_i^{0,(\ell)}\right)^{-1} S_{ii}^{0,(\ell)} c_i^{(\ell)} \left(1 - a_{ii}^0/n\right)^2 + o_p(1) \right] \\
&\geq \frac{T_\ell}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \left[\frac{\left\| c_i^{(\ell)} \right\|_2^2 \left(1 - a_{ii}^0/n\right)^2}{\lambda_{\max}\left(S_{ii}^{0,(\ell)-1} \Omega_i^{0,(\ell)} S_{ii}^{0,(\ell)-1}\right)} + o_p(1) \right] \\
&\geq \frac{T_\ell}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \left[\frac{\left\| c_i^{(\ell)} \right\|_2^2 \left(1 - a_{ii}^0/n\right)^2}{\left\| S_{ii}^{0,(\ell)-1} \right\|_F^2 \left\| \Omega_i^{0,(\ell)} \right\|_F} + o_p(1) \right] \\
&= \frac{T_\ell}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \left[\frac{\left\| c_i^{(\ell)} \right\|_2^2}{\left\| S_{ii}^{0,(\ell)-1} \right\|_F^2 \left\| \Omega_i^{0,(\ell)} \right\|_F} + o_p(1) \right] \\
&\geq \frac{1}{\max_{i \in \{n_1, \cdots, n_n\}} \left\| S_{ii}^{0,(\ell)-1} \right\|_F^2 \left\| \Omega_i^{0,(\ell)} \right\|_F} \frac{T_\ell}{\sqrt{n}} \left[\sum_{i \in \{n_1, \cdots, n_n\}} \left\| c_i^{(\ell)} \right\|_2^2 + o_p(1) \right] \to \infty,
\end{aligned}
$$

where the second line is by the uniform convergence of $\hat{S}_{ii}^{(\ell)}$, $\hat{\Omega}_i^{(\ell)}$ and $\hat{a}_{ii}^{(\ell)}$, the fifth line is by the fact that $\max_{i \in \{n_1, \cdots, n_n\}} \left| \frac{a_{ii}^0}{n} \right| = o_p(1)$ and the last line draws from the assumption that $\frac{T_\ell}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \left\| c_i^{(\ell)} \right\|_2^2 \to \infty$. Following this, by Cauchy's inequality, we also observe that

$$
\begin{aligned}
\left| \frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,8}^{(\ell)} \right| &\leq 2 \sqrt{\frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,4}^{(\ell)}} \sqrt{\frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,7}^{(\ell)}} = o_p\left(\frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,7}^{(\ell)} \right), \\
\left| \frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,9}^{(\ell)} \right| &\leq 2 \sqrt{\frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,5}^{(\ell)}} \sqrt{\frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,7}^{(\ell)}} = o_p\left(\frac{1}{\sqrt{n}} \sum_{i \in \{n_1, \cdots, n_n\}} \hat{\mathbb{S}}_{i,7}^{(\ell)} \right).
\end{aligned}
$$

Combining arguments above, we obtain that $\left| \hat{\Gamma}^{(\ell)} \right| \to \infty$. ■