Essays on large panel data models with two-way heterogeneity

Yiren WANG

Singapore Management University, yiren.wang.2018@phdecons.smu.edu.sg

ESSAYS ON LARGE PANEL DATA MODELS WITH TWO-WAY HETEROGENEITY

By

YIREN WANG

A DISSERTATION

In

ECONOMICS

Presented to the Singapore Management University in Partial Fulfilment

of the Requirements for the Degree of PhD in Economics

2023

____

Supervisor of Dissertation

____

PhD in Economics, Programme Director

ESSAYS ON LARGE PANEL DATA MODELS WITH TWO-WAY HETEROGENEITY

by
Yiren Wang

Submitted to the School of Economics in Partial Fulfilment of the
Requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Yicong Zhang (Supervisor/Chair)
Associate Professor of Economics
Singapore Management University

Liangjun Su
C.V. Starr Chair Professor of Economics
Tsinghua University

Peter C.B. Phillips
Sterling Professor of Economics & Professor of Statistics
Yale University
University of Auckland
University of Southampton
Singapore Management University

Jia Li
Lee Kong Chian Professor of Economics
Singapore Management University

Singapore Management University
2023

Abstract

This dissertation consists of two papers that contribute to the estimation and inference theory of the panel data models with two-way slope heterogeneity. The first paper considers the panel quantile regression model with slope heterogeneity along both individuals and time. By modelling this two-way heterogeneity with the low-rank slope matrix, the slope coefficient can be estimated via the nuclear norm regularization followed by sample-splitting, row- and column-wise quantile regression, and debiasing. The inferential theory for the final slope estimator along with its factor and factor loading is derived. Two specification tests are proposed: one tests whether the slope coefficient is a constant over one dimension (individual or time) without assuming the slope coefficient is homogeneous over the other dimension under the case that the true rank of the slope matrix equals one, and the other tests whether the slope coefficient follows the additive structure under the case that the true rank of slope matrix equals two. The second paper focuses on the estimation and inference of the linear panel model with interactive fixed effects and two-way slope heterogeneity. Specifically, individual coefficients are allowed to form by a latent group structure cross-sectionally, and such a structure can change after an unknown structural break. A multi-stage estimation algorithm is proposed, which involves nuclear norm regularization, break detection, and a K-means procedure, to estimate the break date, the number of groups, and the group structure. Under some regularity conditions, the break date estimator, number of groups estimator, and the group structure estimator can be shown to enjoy the oracle property. Monte Carlo studies and empirical applications are conducted to illustrate the finite sample performance of the proposed algorithms and estimators.

Table of Contents

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisors, Professor Liangjun Su and Associate Professor Yichong Zhang, who have shaped and illuminated my entire doctoral career. Their passion for research and unwavering dedication have inspired me, and their selfless and patient guidance have motivated me into the world of econometrics step by step. I consider myself incredibly fortunate to have them as my supervisors and role models for my whole life. This thesis would not have been possible without their strong support and encouragement.

I am also deeply indebted to my thesis committee, Professor Peter C.B. Phillips and Professor Jia Li. Their limitless help and warm encouragement during my job search have been invaluable, and their generous comments and advice have provided me with much-needed direction in the significant life choice. I am grateful for their guidance throughout these past five years.

I feel extremely fortunate to have joined the Ph.D. program at Singapore Management University, and I would like to extend my sincere thanks to our program directors, Associate Professor Nicolas L. Jacquet and Associate Professor Anthony Tay, for their outstanding leadership and unwavering support. I would also like to express my gratitude to Miss Qiu Ling Thor and Miss Amelia Tan, who have tirelessly assisted me with numerous administrative issues over the past years.

Thanks should also go to my friends and classmates, who provide me with warm and important companionship over the years. My Ph.D. life could not be such an enjoyable journey without their presence.

Finally, words cannot express my gratitude to my parents for their unconditional love, support, and sacrifice. Their optimism and encouragement have helped me overcome many challenges. This thesis is dedicated to my parents.

Chapter 1

Introduction

In recent years, high-dimensional panel data model is one of the most active and popular fields in modern econometrics research. Compared to the traditional econometric models, high-dimensional panel data models allow for different kinds of unobservable heterogeneity, which provides a more reliable and realistic inferential theory. However, technical difficulties arise with the rich heterogeneity arising in the model.

In high-dimensional statistics, reduced rank regression is a classical method for the estimation, which leads to the fast development for the study of the nuclear norm regularization (NNR) in statistics and econometrics, see Koltchinskii et al. (2011), Rohe et al. (2011), Negahban and Wainwright (2011), Fan et al. (2019), Moon and Weidner (2018), Chernozhukov et al. (2019), Belloni et al. (2023), Alidaee et al. (2020) among others. In this dissertation, we apply the NNR to study two important questions in high-dimensional panel data models: two-way heterogeneity issue in low-rank panel quantile models and time-varying latent group structure in the linear panel model.

In the second chapter, we consider a class of low-rank panel quantile regression models which allow for unobserved slope heterogeneity over both individuals and time. By modelling the two-way slope heterogeneity with the low-rank slope matrix, we estimate the heterogeneous intercept and slope matrices via NNR followed by sample splitting, row- and column-wise quantile regressions and debiasing. NNR is used as the first step to obtain the initial matrix estimators which are shown to converge to their truth in terms of the Frobenius norm on average. We then show that the estimators of the factors and factor loadings associated with the final slope matrices are asymptotically normally distributed. In addition, we develop two specification tests: one for the null hypothesis that the slope coefficient is a constant over time and/or individuals under the case that true rank of slope matrix equals one, and the other for the null hypothesis that the slope coefficient exhibits an additive structure under the case that the true rank of slope matrix equals two. We apply the estimation procedure to study the heterogeneity effect in the investment equation and the heterogeneous quantile effect of foreign direct investment on unemployment.

In the third chapter, we consider a linear panel model with interactive fixed effects such that individual heterogeneity is captured by some latent group structures 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 NNR 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. Monte Carlo simulations demonstrate excellent finite sample performance of the proposed estimation algorithm. An application to the empirical to the real house price growth across 377 Metropolitan Statistical Areas in the US from 1975 to 2014 suggests both structural breaks and group membership changes.

Chapter four concludes and all technical results are provided in the appendix.

Chapter 2

Low-rank Panel Quantile Regression: Estimation and Inference

2.1 Introduction

Panel quantile regressions are widely used to estimate the conditional quantiles, which can capture the heterogeneous effects that may vary across the distribution of the outcomes. Such effects are usually assumed to be homogeneous across individuals and over time periods. However, in empirical analyses, it is usually unknown whether the slope coefficients are homogeneous across individuals and/or time. Mistakenly forcing slopes to be homogeneous across time and individuals may lead to inconsistent estimation and misleading inferences. This prompts two questions to be answered: how can we estimate the true model at different quantiles when we allow for heterogeneous slopes across individuals and time at the same time? How to conduct specification tests for homogeneous effects over individuals or time and tests for the additive structure of the slope coefficients?

To answer the first question, we propose an estimation procedure for heterogeneous panel quantile regression models where we allow the fixed effects to be either additive or interactive, and the slope coefficients to be heterogeneous over both individuals and time. We impose a low-rank structure for both the intercept and slope coefficient matrices and estimate them via nuclear norm regularization (NNR) followed by the sample splitting, row- and column-wise quantile regressions and debiasing steps. The estimation algorithm is inspired by Chernozhukov et al. (2019), where the main difference is that we split the full sample into three subsamples rather than two because we need certain uniform results which require independence of regressors and regressand used in the debiasing step, and we do not have the closed form for the quantile regression estimates. At last, we derive the asymptotic distributions for the estimators of the factors and factor loadings associated with slope coefficient matrices.

To answer the second question, under the case when the rank of slope coefficient matrix equals one, we conduct sup-type specification tests for homogeneous effects over individuals or time following the lead of Castagnetti et al. (2015) and Lu and Su (2023). We show that our sup-test statistics follow the Gumbel distribution under the null, and the tests have non-trivial power against certain classes of local alternatives. Under the case when the rank of slope matrix equals two, our sup-type test statistic is also shown to follow the Gumbel distribution under the null that the slope coefficient exhibits an additive structure.

This paper relates to three bunches of literature. First, we contribute to the large literature on panel quantile regressions (PQRs). Since Koenker (2004) studied the PQRs with individual fixed effects, there has been an increasing number of papers on PQRs. Galvao and Montes-Rojas (2010), Kato et al. (2012), Galvao and Wang (2015), Galvao and Kato (2016), Machado and Silva (2019), and Galvao et al. (2020) study the asymptotics for PQRs with individual fixed effects. Chen et al. (2021) study quantile factor models and Chen (2022) considers PQRs with interactive fixed effects (IFEs). We complement the literature by allowing for unobserved heterogeneity in the slope coefficients of PQRs.

Second, our paper also pertains to slope heterogeneity in panel data models. Latent group structures across individuals and structural changes over time are two common types of slope heterogeneity that have received vast attention in the literature. To recover the unobserved group structures, various methods have been proposed. For example, Lin and Ng (2012), Bonhomme and Manresa (2015) and Ando and Bai (2016) use the K-means algorithm; Su et al. (2016) propose the C-lasso algorithm which is further studied and extended by Su and Ju (2018), Su et al. (2019) and Wang et al. (2019); Wang et al. (2018) propose an clustering algorithm in regression via data-driven segmentation called CARDS; Wang and Su (2021) propose a sequential binary segmentation algorithm to identify the latent group structures in nonlinear panels. Recent literature on the estimation with structural changes in panel data models includes, but is not limited to, Chen (2015), Cheng et al. (2016), Ma and Su (2018), Baltagi et al. (2021). In addition, Galvao et al. (2018) and Zhang et al. (2019) consider individual heterogeneity in PQRs while they assume homogeneity across time. To allow for both latent groups and structural breaks, Okui and Wang (2021) study a linear panel data model with individual fixed effects where each latent group has common breaks and the breaking points can be different across different groups, and they propose a grouped adaptive group fused lasso (GAGFL) approach to estimate slope coefficients. Lumsdaine et al. (2023) consider a linear panel data model with a grouped pattern of heterogeneity where the latent group membership structure and/or the values of slope coefficients can change at a breaking point, and they propose a K-means-type estimation algorithm and establish the asymptotic properties of the resulting estimators. Compared with the models studied above, our model combines both individual and time heterogeneity and only requires certain low-rank structure in the slope coefficient matrix. So the unobserved heterogeneity takes a more flexible form in our model than those in the literature such as Okui and Wang (2021) and Lumsdaine et al. (2023).

Last, our paper also connects with the burgeoning literature on nuclear norm regularization. Such a method has been widely adopted to study panel and network models. See, Alidaee et al. (2020), Athey et al. (2021), Bai and Ng (2019), Belloni et al. (2023), Chen et al. (2020), Chernozhukov et al. (2019), Feng (2019), Hong et al. (2022), Miao et al. (2023), among others. In the least squares panel framework, Moon and Weidner (2018) consider a homogeneous panel with IFEs by using NNR-based estimator as an initial estimator to construct iterative estimators that are asymptotically equivalent to the least squares estimators; Chernozhukov et al. (2019) study a heterogenous panel where both the intercept and slope coefficient matrices exhibit a low-rank structure and establish the asymptotic distribution theory based on NNR. In the presence of endogeneity, Hong et al. (2022) proposes a profile GMM method to estimate panel data models with IFEs. In the panel quantile regression setting, Feng (2019) develops error bounds for the low-rank estimates in terms of Frobenius norms under independence assumption; Belloni et al. (2023) relaxes the independence assumption to the β-mixing condition along the time dimension. Our paper extends Chernozhukov et al. (2019) from the least squares framework to the PQR framework, derives the asymptotic distribution theory and develops various specification tests under some strong mixing conditions along the time dimension that is weaker than the β-mixing condition. We also rely on the sequential symmetrization technique developed by Rakhlin et al. (2015) to obtain the convergence rates of the nuclear norm regularized estimators.

The rest of the paper is organized as follows. We first introduce the low-rank structure PQR model and the estimation algorithm in Section 2.2. We study the asymptotic properties of our estimators in Section 2.3. In Section 2.4, we propose two specification tests: one for the no-factor structure and one for the additive structure, and study the asymptotic properties of the test statistics. In Section 2.5, we show the finite sample performance of our method via Monte Carlo simulations. In Section 2.6, we apply our method to two datasets: one is to study how Tobin's q and cash flows affect corporate investment and whether firm's external investment to its internal financing exhibits heterogeneity structure, and the other is to study the relationship between economics growth, foreign direct investment and unemployment. Section 2.7 concludes. All proofs are related to the online supplement.

Notation. $\|\cdot\|_1$, $\|\cdot\|_{op}$, $\|\cdot\|_\infty$, $\|\cdot\|_{\max}$ $\|\cdot\|_2$, $\|\cdot\|_F$, $\|\cdot\|_*$ denote the matrix norm induced by 1-norms, the matrix norm induced by 2-norms, the matrix 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 max and the min of a and b, respectively. The symbol $\lesssim$ means "the left is bounded by a positive constant times the right". Let $A=\{A_{it}\}_{i\in[n],t\in[T]}$ be a matrix with its (i,t)-th entry denoted as A_{it}, where $[n]$ to denote the set $\{1,\cdots,n\}$ for any positive integer n. Let $\{A_j\}_{j=0}^p$ denote the collection of matrices A_j for all $j\in\{0,\cdots,p\}$. When A is symmetric, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote its largest and smallest eigenvalues, respectively. The operators $\rightsquigarrow$ and $\xrightarrow{p}$ denote convergence in distribution and in probability, respectively. Besides, we use w.p.a.1 and a.s. to abbreviate "with probability approaching 1" and "almost surely", respectively.

2.2 Model and Estimation

In this section, we introduce the PQR model and estimation algorithm.

2.2.1 Model

Consider the PQR model

$$\mathscr{Q}_\tau\left(Y_{it}\middle|\left\{X_{j,it}\right\}_{j\in[p],t\in[T]},\left\{\Theta^0_{j,it}(\tau)\right\}_{j\in[p]\cup\{0\},t\in[T]}\right)=\Theta^0_{0,it}(\tau)+\sum_{j=1}^{p}X_{j,it}\Theta^0_{j,it}(\tau), \tag{2.1}$$

where $i\in[N]$, $t\in[T]$, $\tau\in(0,1)$ is the quantile index, Y_{it} is the dependent variable, $X_{j,it}$ is the j-th regressor for individual i at time t, $\{\Theta^0_{j,it}\}_{j\in[p]}$ is the corresponding slope coefficient, $\Theta^0_{0,it}$ is the intercept, and

$$\mathscr{Q}_\tau\left(Y_{it}\middle|\left\{X_{j,it}\right\}_{j\in[p],t\in[T]},\left\{\Theta^0_{j,it}(\tau)\right\}_{j\in[p]\cup\{0\},t\in[T]}\right)$$

denotes the conditional τ-quantile of Y_{it} given the regressors $\left\{X_{j,it}\right\}_{j\in[p],t\in[T]}$ and the parameters $\left\{\Theta^0_{j,it}(\tau)\right\}_{j\in[p]\cup 0,t\in[T]}$.[1] Alternatively, we can rewrite the above model as

$$\begin{aligned}
&Y=\Theta^0_0(\tau)+\sum_{j=1}^{p}X_j\odot\Theta^0_j(\tau)+\varepsilon(\tau)\quad\text{and}\\
&\mathscr{Q}_\tau\left(\varepsilon_{it}(\tau)\middle|\left\{X_{j,it}\right\}_{j\in[p],t\in[T]},\left\{\Theta^0_{j,it}(\tau)\right\}_{j\in[p]\cup\{0\},t\in[T]}\right)=0,
\end{aligned}\tag{2.2}$$

where $\varepsilon(\tau)$ is the idiosyncratic error matrix with the (i,t)-th entry being $\varepsilon_{it}(\tau)$. Similarly, X_j, $\Theta_j(\tau)$, and Y are matrices with the (i,t)-th entry being $X_{j,it}$, $\Theta_{j,it}(\tau)$, and Y_{it}, respectively. In this model, we assume p, the number of regressors, is fixed and both N and T pass to infinity. In Assumption 2.1 below, we characterize the dependence of the data, under which (2.1) holds.

In the paper, we focus on the panel quantile regression for a fixed τ and thus suppress the dependence of $\Theta^0_j(\tau)$ and $\varepsilon(\tau)$ on τ for notation simplicity. In addition, we impose low-rank structures for the intercept and slope matrices, i.e., $\text{rank}(\Theta^0_j)=$

[1] We will assume that both the intercept term $\Theta^0_{0,it}$ and the slope coefficients $\{\Theta^0_{j,it}\}_{j\in[p]}$ have low-rank structures, and follow the convention in the panel data literature by treating the factors to be random. Therefore, $\{\Theta^0_{j,it}\}_{j\in[p]\cup\{0\}}$ are random as well.

K_j for some positive constant K_j and for each $j \in \{0, \cdots, p\}$. By the singular value decomposition (SVD), we have

$$\Theta_j^0 = \sqrt{NT}\mathscr{U}_j^0 \Sigma_j^0 \mathscr{V}_j^{0\prime} = U_j^0 V_j^{0\prime} \ \forall\, j = 0, \cdots, p,$$

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

The low-rank structure assumption includes several popular cases. For the intercept term, one commonly assumes that $\Theta_{0,it}^0$ to take the forms α_i^0, μ_t^0, or $\alpha_i^0 + \mu_t^0$ in classical PQRs. Then the matrix Θ_0^0 has rank 1, 1, and 2, respectively. It is also possible to assume $\Theta_{0,it}^0$ to take an interactive form, say, $\Theta_{0,it}^0 = \lambda_{0,i}^{0\prime} f_{0,t}^0$, where both $\lambda_{0,i}^0$ and $f_{0,t}^0$ are K_0-vectors. For the slope matrix Θ_j^0, $j \in [p]$, the early PQR models frequently assume that $\Theta_{j,it}^0$ is a constant across (i,t) to yield a homogenous PQR model. Obviously, such a model is very restrictive by assuming homogenous slope coefficients. It is possible to allow the slope coefficients to change over either i, or t, or both. See the following examples for different low-rank structures.

Example 1. When $\Theta_{j,it}^0 = \Theta_{j,i}^0 \ \forall t \in [T]$, or $\Theta_{j,it}^0 = \Theta_{j,t}^0 \ \forall i \in [N]$, or $\Theta_{j,it}^0 = \Theta_j^0$ $\forall (i,t) \in [N] \times [T]$, and this holds for all $j \in [p]$, we have the PQR models with only individual heterogeneity, with only time heterogeneity, and with homogeneity, respectively. We observe that $K_j = 1$ for these three cases.

Example 2. When $\Theta_{j,it}^0 = \lambda_{j,i}^0 + f_{j,t}^0$, we notice that

$$\frac{\Theta_j^0}{\sqrt{NT}} = \begin{bmatrix} \frac{1}{\sqrt{N}} & \frac{\lambda_{j,1}^0}{\sqrt{N}} \\ \vdots & \vdots \\ \frac{1}{\sqrt{N}} & \frac{\lambda_{j,N}^0}{\sqrt{N}} \end{bmatrix} \begin{bmatrix} \frac{f_{j,1}^0}{\sqrt{T}} & \cdots & \frac{f_{j,T}^0}{\sqrt{T}} \\ \frac{1}{\sqrt{T}} & \cdots & \frac{1}{\sqrt{T}} \end{bmatrix} := A_j B_j'.$$

Let $\Sigma_{A,j} := A_j' A_j$ and $\Sigma_{B,j} := B_j' B_j$. Let $\Sigma_{A,j}^{\frac{1}{2}}$ (resp. $\Sigma_{B,j}^{\frac{1}{2}}$) be the symmetric square root of $\Sigma_{A,j}$ (resp. $\Sigma_{B,j}$). By eigendecomposition, we have $\Sigma_{A,j}^{\frac{1}{2}} = P_{j,1} S_{j,1} P_{j,1}'$ and $\Sigma_{B,j}^{\frac{1}{2}} = P_{j,2} S_{j,2} P_{j,2}'$. Besides, we apply singular value decomposition to matrix $S_{j,1} P_{j,1}' P_{j,2} S_{j,2}$: $S_{j,1} P_{j,1}' P_{j,2} S_{j,2} = Q_{j,1} R_j Q_{j,2}'$. Then it follows that

$$\begin{aligned} \frac{\Theta_j^0}{\sqrt{NT}} &= A_j B_j' = A_j \Sigma_{A,j}^{-\frac{1}{2}} P_{j,1} S_{j,1} P_{j,1}' P_{j,2} S_{j,2} P_{j,2}' \Sigma_{B,j}^{-\frac{1}{2}} B_j' \\ &= A_j \Sigma_{A,j}^{-\frac{1}{2}} P_{j,1} Q_{j,1} R_j Q_{j,2}' P_{j,2}' \Sigma_{B,j}^{-\frac{1}{2}} B_j' := \mathscr{U}_j^0 \Sigma_j^0 \mathscr{V}_j^{0\prime}, \end{aligned}$$

where $\mathscr{U}_j^0 = A_j\Sigma_{A,j}^{-\frac{1}{2}}P_{j,1}Q_{j,1}$, $\Sigma_j^0 = R_j$ and $\mathscr{V}_j^0 = B_j\Sigma_{B,j}^{-\frac{1}{2}}P_{j,2}Q_{j,2}$. Given $P_{j,1}$, $P_{j,2}$, $Q_{j,1}$ and $Q_{j,2}$ are orthonormal matrices, it's easy to that $\mathscr{U}_j^0$ and $\mathscr{V}_j^0$ are also orthonormal so that $\mathscr{U}_j^{0\prime}\mathscr{U}_j^0 = \mathscr{V}_j^{0\prime}\mathscr{V}_j^0 = I_2$. When $j = 0$, $\left\{\lambda_{0,i}^0\right\}_{i=1}^N$ and $\left\{f_{0,t}^0\right\}_{t=1}^T$ are usually referred to as the individual and time fixed effects, respectively, so that the intercept term exhibits an additive fixed effects structure.

Example 3. Let $\Theta_{j,it}^0 = \sum_{k\in[K_{j,t}]}\alpha_{j,kt}\mathbf{1}\{i \in G_{j,kt}\}$, where $\left\{G_{j,kt}\right\}$ forms a partition of $[N]$ for each specific time t and $K_{j,t}$ is the number of groups at time t. Moreover, let

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

$$G_{j,kt} = \begin{cases} G_{j,k}^{(1)}, & \text{for} \quad t = 1,\ldots,T_b, k = 1,\ldots,K_j^{(1)}, \\ G_{j,k}^{(2)}, & \text{for} \quad t = T_b+1,...,T, k = 1,...,K_j^{(2)}, \end{cases}$$

where $K_j^{(1)}$ and $K_j^{(2)}$ are the number of groups before and after the break point T_b. If $K_j^{(1)} = K_j^{(2)}$, it is clear that $rank(\Theta_j^0) = 1$. If the group structure does not change after the break but $\alpha_{j,k}^{(1)} = c\alpha_{j,k}^{(2)}$ for some constant c, we also have $rank(\Theta_j^0) = 1$. Except for these two cases, we can show that

$$\Theta_j^0 = \begin{bmatrix} \sum\limits_{k\in[K^{(1)}]} \alpha_{j,k}^{(1)}\mathbf{1}\left\{1 \in G_{j,k}^{(1)}\right\}, & \sum\limits_{k\in[K^{(2)}]} \alpha_{j,k}^{(2)}\mathbf{1}\left\{1 \in G_{j,k}^{(2)}\right\} \\ \vdots & \vdots \\ \sum\limits_{k\in[K^{(1)}]} \alpha_{j,k}^{(1)}\mathbf{1}\left\{i \in G_{j,k}^{(1)}\right\}, & \sum\limits_{k\in[K^{(2)}]} \alpha_{j,k}^{(2)}\mathbf{1}\left\{i \in G_{j,k}^{(2)}\right\} \\ \vdots & \vdots \\ \sum\limits_{k\in[K^{(1)}]} \alpha_{j,k}^{(1)}\mathbf{1}\left\{N \in G_{j,k}^{(1)}\right\}, & \sum\limits_{k\in[K^{(2)}]} \alpha_{j,k}^{(2)}\mathbf{1}\left\{N \in G_{j,k}^{(2)}\right\} \end{bmatrix} \begin{bmatrix} \iota_{T_b} & \mathbf{0}_{T_b} \\ \mathbf{0}_{T-T_b} & \iota_{T-T_b} \end{bmatrix}'$$

where ι_{T_b} is a $T_b \times 1$ vector of ones and $\mathbf{0}_{T_b}$ is a $T_b \times 1$ vector of zeros. In this case, we notice that $rank(\Theta_j^0) = 2$.

Example 4. When $\Theta_{j,it}^0 = \lambda_{j,i}^{0\prime}f_{j,t}^0$ with $\lambda_{j,i}^0$ and $f_{j,t}^0$ being two K_j-vectors, we have the IFEs structure. This is the most general example without further restrictions.

Like Chernozhukov et al. (2019), we assume that for each $j \in [p]$, $X_{j,it}$ exhibits a factor structure: $X_{j,it} = \mu_{j,it} + e_{j,it} = l_{j,i}^{0\prime}w_{j,t}^0 + e_{j,it}$, where $w_{j,t}^0$ and $l_{j,i}^0$ are the factors and factor loadings of dimension r_j.

2.2.2 Estimation Algorithm

In this subsection we provide the estimation algorithm by assuming that K_j are all known for all j. In the next subsection, we will introduce a rank estimation method to estimate K_j consistently.

Define the check function $\rho_\tau(u) = u(\tau - \mathbf{1}\{u \leq 0\})$. The estimation procedure goes as follows:

Step 1: **Sample Splitting and Nuclear Norm Regularization.** Along the cross-section span, randomly split the sample into three subsets denoted as I_1, I_2 and I_3, where I_ℓ has N_ℓ individuals such that $N_1 \approx N_2 \approx N_3 \approx N/3$. Using the data with $(i,t) \in I_1 \times [T]$, we run the nuclear norm regularized quantile regression (QR) and obtain $\{\tilde{\Theta}_j^{(1)}\}_{j\in\{0,\cdots,p\}}$, i.e.,

$$\{\tilde{\Theta}_j^{(1)}\}_{j=0}^p = \underset{\{\Theta_j\}_{j=0}^p}{\arg\min} \frac{1}{N_1 T} \sum_{i\in I_1} \sum_{t=1}^T \rho_\tau(Y_{it} - \sum_{j=1}^p X_{j,it}\Theta_{j,it} - \Theta_{0,it}) + \sum_{j=0}^p \nu_j \left\|\Theta_j\right\|_*, \tag{2.3}$$

where ν_j is a tuning parameter. For each j, conduct the SVD: $\frac{1}{\sqrt{N_1 T}}\tilde{\Theta}_j^{(1)} = \hat{\tilde{\mathscr{U}}}_j^{(1)}\hat{\tilde{\Sigma}}_j^{(1)}\hat{\tilde{\mathscr{V}}}_j^{(1)\prime}$, where $\hat{\tilde{\Sigma}}_j^{(1)}$ is the diagonal matrix with the diagonal elements being the descending singular values of $\tilde{\Theta}_j^{(1)}$. Let $\tilde{\mathscr{V}}_j^{(1)}$ consists the first K_j columns of $\hat{\tilde{\mathscr{V}}}_j^{(1)}$. Let $\tilde{V}_j^{(1)} = \sqrt{T}\tilde{\mathscr{V}}_j^{(1)}$ and $\tilde{v}_{t,j}^{(1)\prime}$ be the t-th row of $\tilde{V}_j^{(1)}$ $\forall t \in [T]$.

Step 2: **Row- and Column-Wise Quantile Regression.** Using the data with $(i,t) \in I_2 \times [T]$, we first run the row-wise QR of Y_{it} on $\left(\tilde{v}_{t,0}^{(1)}, \left\{\tilde{v}_{t,j}^{(1)} X_{j,it}\right\}_{j\in[p]}\right)$ to obtain $\{\dot{u}_{i,j}^{(1)}\}_{j=0}^p$ for $i \in I_2$, and then run the column-wise QR of Y_{it} on $(\dot{u}_{i,0}^{(1)}, \{\dot{u}_{i,j}^{(1)} X_{j,it}\}_{j\in[p]})$ to obtain $\{\dot{v}_{t,j}^{(1)}\}_{j=0}^p$ for $t \in [T]$. That is,

$$\{\dot{u}_{i,j}^{(1)}\}_{j=0}^p = \underset{\{u_{i,j}\}_{j\in[p]\cup\{0\}}}{\arg\min} \frac{1}{T}\sum_{t\in[T]} \rho_\tau\left(Y_{it} - u_{i,0}'\tilde{v}_{t,0}^{(1)} - \sum_{j=1}^p u_{i,j}'\tilde{v}_{t,j}^{(1)} X_{j,it}\right), \forall i \in I_2, \tag{2.4}$$

$$\{\dot{v}_{t,j}^{(1)}\}_{j=0}^p = \underset{\{v_{t,j}\}_{j\in[p]\cup\{0\}}}{\arg\min} \frac{1}{N_2}\sum_{i\in I_2} \rho_\tau\left(Y_{it} - v_{t,0}'\dot{u}_{i,0}^{(1)} - \sum_{j=1}^p v_{t,j}'\dot{u}_{i,j}^{(1)} X_{j,it}\right), \forall t \in [T]. \tag{2.5}$$

Similarly, we run the row-wise QR of Y_{it} on $(\dot{v}_{t,0}^{(1)}, \{\dot{v}_{t,j}^{(1)} X_{j,it}\}_{j\in[p]})$ to obtain

$\{\dot{u}_{i,j}^{(1)}\}_{j=0}^{p}$ for $i \in I_3$, i.e.,

$$\{\dot{u}_{i,j}^{(1)}\}_{j=0}^{p} = \underset{\{u_{i,j}\}_{j\in[p]\cup\{0\}}}{\arg\min} \frac{1}{T}\sum_{t\in[T]} \rho_\tau \left(Y_{it} - u_{i,0}'\dot{v}_{t,0}^{(1)} - \sum_{j=1}^{p} u_{i,j}'\dot{v}_{t,j}^{(1)} X_{j,it} \right), \forall i \in I_3.$$

Step 3: **Debiasing.**

Step 3.1: For each $j \in [p]$, we conduct the principle component analysis (PCA) for $X_{j,it}$ with $(i,t) \in [N] \times [T]$ to obtain the factor and factor loading estimates as

$$\left\{\hat{l}_{j,i}, \hat{w}_{j,t}\right\}_{i\in[N],t\in[T]} = \underset{\left\{l_{j,i},w_{j,t}\right\}_{i\in[N],t\in[T]}}{\arg\min} \frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]} \left(X_{j,it} - l_{j,i}'w_{j,t}\right)^2, \tag{2.6}$$

subject to the normalizations: $\frac{1}{N}\sum_{i=1}^{N} l_{i,j}l_{i,j}' = I_{r_j}$ and $\frac{1}{T}\sum_{t=1}^{T} w_{j,t}w_{j,t}'$ is a diagonal matrix with descending diagonal elements. Then we define $\hat{\mu}_{j,it} = \hat{l}_{j,i}'\hat{w}_{j,t}$ and $\hat{e}_{j,it} = X_{j,it} - \hat{\mu}_{j,it}$.

Step 3.2: For $(i,t) \in I_3 \times [T]$, let $\tilde{Y}_{it} = Y_{it} - \sum_{j=1}^{p} \hat{\mu}_{j,it}\dot{u}_{i,j}^{(1)\prime}\dot{v}_{t,j}^{(1)}$. We run the row-wise QR of $\tilde{Y}_{it}$ on $(\dot{v}_{t,0}^{(1)}, \{\dot{v}_{t,j}^{(1)}\hat{e}_{j,it}\}_{j\in[p]})$ to obtain the final estimates $\hat{u}_{i,j}^{(3,1)}$, i.e.,

$$\{\hat{u}_{i,j}^{(3,1)}\}_{j=0}^{p} = \underset{\{u_{i,j}\}_{j=0}^{p}}{\arg\min} \frac{1}{T}\sum_{t\in[T]} \rho_\tau \left(\tilde{Y}_{it} - u_{i,0}'\dot{v}_{t,0}^{(1)} - \sum_{j=1}^{p} u_{i,j}'\dot{v}_{t,j}^{(1)}\hat{e}_{j,it} \right), \forall i \in I_3. \tag{2.7}$$

Updating $\hat{Y}_{it} = Y_{it} - \sum_{j=1}^{p} \hat{\mu}_{j,it}\hat{u}_{i,j}^{(3,1)\prime}\dot{v}_{t,j}^{(1)}$, we run the column-wise QR of $\hat{Y}_{it}$ on $(\hat{u}_{i,0}^{(3,1)}, \{\hat{u}_{i,j}^{(3,1)}\hat{e}_{j,it}\}_{j\in[p]})$ to obtain $\hat{v}_{t,j}^{(3,1)}$, i.e.,

$$\{\hat{v}_{t,j}^{(3,1)}\}_{j=0}^{p} = \underset{\{v_{t,j}\}_{j=0}^{p}}{\arg\min} \frac{1}{N_3}\sum_{i\in I_3} \rho_\tau \left(\hat{Y}_{it} - v_{t,0}'\hat{u}_{i,0}^{(3,1)} - \sum_{j=1}^{p} v_{t,j}'\hat{u}_{i,j}^{(3,1)}\hat{e}_{j,it} \right), \forall t \in [T]. \tag{2.8}$$

In order to obtain the final estimators for the full sample, we propose to switch the role of each subsample for the low-rank estimation, row- and column-wise QR and debiasing, then repeat Steps 1-3 to obtain $\left\{\hat{u}_{i,j}^{(a,b)}\right\}_{j=0}^{p}$ and $\left\{\hat{v}_{t,j}^{(a,b)}\right\}_{j=0}^{p}$ for $a \in [3]$ and $b \in [3] \setminus \{a\}$. Here (a,b) denotes the final estimates for subsample I_a obtained from the first step NNR estimates with subsample I_b. Table 2.1 shows the final estimators we obtain by using different combination of subsamples.

Several remarks are in order. First, we randomly split the full sample into three

Table 2.1: Estimators using different subsamples at different steps in algorithm.

Step 1 (b)	Step 2	Step 3 (a)	estimators (a,b)
I_1	I_2	I_3	$\hat{u}_{i,j}^{(3,1)}$, $\hat{v}_{t,j}^{(3,1)}$
I_2	I_1	I_3	$\hat{u}_{i,j}^{(3,2)}$, $\hat{v}_{t,j}^{(3,2)}$
I_1	I_3	I_2	$\hat{u}_{i,j}^{(2,1)}$, $\hat{v}_{t,j}^{(2,1)}$
I_3	I_1	I_2	$\hat{u}_{i,j}^{(2,3)}$, $\hat{v}_{t,j}^{(2,3)}$
I_2	I_3	I_1	$\hat{u}_{i,j}^{(1,2)}$, $\hat{v}_{t,j}^{(1,2)}$
I_3	I_2	I_1	$\hat{u}_{i,j}^{(1,3)}$, $\hat{v}_{t,j}^{(1,3)}$

subsamples, each playing a significant role in the algorithm. We use the first subsample for the low-rank estimation to obtain the preliminary NNR estimators of the submatrices of the intercept and slope matrices. But these estimators are only consistent in terms of Frobenius norm, and one cannot derive the pointwise or uniform convergence rates for them. With the low-rank estimates, we use the second subsample to do the row- and column-wise QRs and can now establish the uniform convergence rates for each row of factor and factor loading estimators. Then we use the remaining subsample to debias the second-stage estimator and to obtain the final estimators that have the desirable asymptotic properties.

Second, to reduce the randomness of sample splitting, one can run the estimation algorithm several times with different splittings in practice. Once one obtains factor and factor loading estimates, one can construct estimators for Θ_j^0 under different splittings and then choose the one specific splitting which yields the minimum quantile objective function.

Third, the bias in the second-stage estimator is inherent from the first-stage NNR estimator. We follow the lead of Chernozhukov et al. (2019) to assume that $X_{j,it}$ has a factor structure with an additive idiosyncratic term, and remove the bias by a QR with the demeaned $X_{j,it}$ as regressors. In the least squares panel regression framework, the objective function is smooth and one has closed-form solutions in the last stage so that Chernozhukov et al. (2019) only need to split the sample into two subsamples. In contrast, in the PQR framework, the objective function is nonsmooth, we do not have closed-form solutions in any stage. In order to remove the bias from the early stage estimation and to derive the distributional results, we need to split the sample into three subsamples.

To save space, we relegate the detailed algorithm for the nuclear norm regularization to the online supplement.

2.2.3 Rank Estimation

In this subsection we discuss how to estimate the ranks K_j consistently. To estimate the ranks, we consider the full sample NNR QR estimation:

$$\{\tilde{\Theta}_j\}_{j=0}^{p} = \arg\min_{\{\Theta_j\}_{j=0}^{p}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau(Y_{it} - \sum_{j=1}^{p} X_{j,it}\Theta_{j,it} - \Theta_{0,it}) + \sum_{j=0}^{p} \nu_j \left\|\Theta_j\right\|_*. \quad (2.9)$$

For $j \in \{0,\cdots,p\}$, we estimate K_j by the popular singular value thresholding (SVT) as follows

$$\hat{K}_j = \sum_m \mathbf{1}\left\{\lambda_m\left(\tilde{\Theta}_j\right) \geq 0.6\left(NT\nu_j\left\|\tilde{\Theta}_j\right\|_{op}\right)^{1/2}\right\}.$$

It is standard to show that $\mathbb{P}(\hat{K}_j = K_j) \to 1$ as $(N,T) \to \infty$ under some regularity conditions given in the next section; see also Proposition D.1 in Chernozhukov et al. (2019) and Theorem 2 in Hong et al. (2022). Since the ranks can be estimated consistently, we assume that they are known in the asymptotic theory below.

2.3 Asymptotic Theory

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

2.3.1 First Stage Estimator

Recall that

$$X_{j,it} = \mu_{j,it} + e_{j,it} = l_{j,i}^{0\prime} w_{j,t}^{0} + e_{j,it}$$

for each $j \in [p]$. Let $X_{it} = (X_{1,it},...,X_{p,it})'$ and $e_{it} = (e_{1,it},...,e_{p,it})'$. Define $\varepsilon_i = (\varepsilon_{i1},\cdots,\varepsilon_{it})'$, $e_{j,i} = \left(e_{j,i1},\cdots,e_{j,iT}\right)'$, W_j^0 as the $T \times r_j$ matrix with each row being $w_{j,t}^{0\prime}$, and V_j^0 as the $T \times K_j$ matrix with each row being $v_{t,j}^{0\prime}$. Further define $a_{it} = \tau - \mathbf{1}\{\varepsilon_{it} \leq 0\}$ with $a_i = (a_{i1},\cdots,a_{iT})'$ and $a = (a_1,\cdots,a_N)'$. Throughout the paper, we treat the factors $\{v_{t,j}^0\}_{t\in[T],j\in[p]\cup\{0\}}$ and $\{w_{j,t}^0\}_{t\in[T],j\in[p]}$ as random and their loadings $\{u_{i,j}^0\}_{i\in[N],j\in[p]\cup\{0\}}$ and $\{l_{j,i}^0\}_{i\in[N],j\in[p]}$ as deterministic.

Table 2.2 defines several σ-fields. We use $\mathscr{D}$ to denote the minimal σ-field generated by $\left\{V_j^0\right\}_{j\in[p]\cup\{0\}} \bigcup \left\{W_j^0\right\}_{j\in[p]}$; the superscripts I_1 and $I_1 \cup I_2$ are associated with the first subsample and the first two subsamples, respectively. For example, $\mathscr{D}_{e_i}^{I_1}$ denotes the minimal σ-field generated by $\mathscr{D}$, $\{e_{it}\}_{t\in[T]}$ and $\{\varepsilon_{it}, e_{it}\}_{i\in I_1, t\in[T]}$.

Table 2.2: Definition of various σ-fields

Notation	σ-fields generated by
$\mathscr{D}$	$\left\{V_j^0\right\}_{j\in[p]\cup\{0\}} \bigcup \left\{W_j^0\right\}_{j\in[p]}$
$\mathscr{D}_{e_{it}}$	$\mathscr{D} \bigcup e_{it}$
$\mathscr{D}_{e_i}$	$\mathscr{D} \bigcup \{e_{it}\}_{t\in[T]}$
$\mathscr{D}_{e}$	$\mathscr{D} \bigcup \{e_{it}\}_{i\in[N],t\in[T]}$
$\mathscr{D}^{I_1\cup I_2}$	$\mathscr{D} \bigcup \{\varepsilon_{it}, e_{it}\}_{i\in I_1\cup I_2, t\in[T]}$
$\mathscr{D}^{I_1}_{\{e_{is}\}_{s<t}}$	$\mathscr{D} \bigcup \{e_{is}\}_{s<t} \bigcup \{\varepsilon_{i^*t^*}, e_{i^*t^*}\}_{i^*\in I_1, t^*\in[T]}$
$\mathscr{D}^{I_1}_{e_i}$	$\mathscr{D} \bigcup \{e_{it}\}_{t\in[T]} \bigcup \{\varepsilon_{i^*t^*}, e_{i^*t^*}\}_{i^*\in I_1, t^*\in[T]}$
$\mathscr{D}^{I_1\cup I_2}_{e_i}$	$\mathscr{D} \bigcup \{e_{it}\}_{t\in[T]} \bigcup \{\varepsilon_{i^*t^*}, e_{i^*t^*}\}_{i^*\in I_1\cup I_2, t^*\in[T]}$
$\mathscr{D}^{I_1\cup I_2}_{e}$	$\mathscr{D} \bigcup \{e_{it}\}_{i\in[N,]t\in[T]} \bigcup \{\varepsilon_{it}, e_{it}\}_{i\in I_1\cup I_2, t\in[T]}$

Let M denote a generic bounded constant that may vary across places. Let $\mathscr{G}_{i,t-1}$ denote the minimal σ-field generated by $\mathscr{D} \cup \{e_{ls}\}_{l\leq i-1, s\in[T]} \cup \{e_{is}\}_{s\leq t} \cup \{\varepsilon_{ls}\}_{l\leq i-1, s\in[T]} \cup \{\varepsilon_{is}\}_{s\leq t-1}$. Let $\mathsf{F}_{it}(\cdot)$ and $\mathsf{f}_{it}(\cdot)$ be the conditional cumulative distribution function (CDF) and probability density function (PDF) of ε_{it} given $\mathscr{G}_{i,t-1}$, respectively. Similarly, let $\mathfrak{F}_{it}(\cdot)$ and $\mathfrak{f}_{it}(\cdot)$ denote the conditional CDF and PDF of ε_{it} given $\mathscr{D}_{e_i}$; $F_{it}(\cdot)$ and $f_{it}(\cdot)$ denote the conditional CDF and PDF of ε_{it} given $\mathscr{D}_e$. Let $\mathsf{f}'_{it}(\cdot)$, $\mathfrak{f}'_{it}(\cdot)$, and $f'_{it}(\cdot)$ denotes the first derivative of the density $\mathsf{f}_{it}(\cdot)$, $\mathfrak{f}_{it}(\cdot)$, and $f_{it}(\cdot)$, respectively.

We make the following assumptions.

Assumption 2.1. *(i)* $\{\varepsilon_{it}, e_{it}\}_{t\in[T]}$ *are conditionally independent across i given* $\mathscr{D}$.

(ii) $\mathbb{E}\left(a_{it} \middle| \mathscr{D}_e\right) = 0.$

(iii) For each i, $\{\varepsilon_{it}, t \geq 1\}$ *is strong mixing conditional on* $\mathscr{D}_{e_i}$*, and* $\{(\varepsilon_{it}, e_{it}), t \geq 1\}$ *is strong mixing conditional on* $\mathscr{D}$*. Both mixing coefficients are upper*

bounded by $\alpha_i(\cdot)$ *such that* $\max_{i\in[N]}\alpha_i(z)\le M\alpha^z$ *for some constant* $\alpha\in(0,1)$.

(iv) $\max_{i\in[N]}\frac{1}{T}\sum_{t\in[T]}\|X_{it}\|_2^3\le M$ *a.s.*, $\max_{t\in[T]}\frac{1}{N_2}\sum_{i\in I_2}\|X_{it}\|_2^4\le M$ *a.s.*, $\max_{(i,t)\in[N]\times[T]}\mathbb{E}\left[\|X_{it}\|_2^3\middle|\mathscr{D}\right]\le M$*a.s.*, $\max_{i\in[N]}\sqrt{\frac{1}{T}\sum_{t\in[T]}\left[\mathbb{E}\left(\varepsilon_{it}^2\middle|\mathscr{D}_{e_i}\right)\right]^2}\le M$ *a.s., and* $\max_{(i,t)\in[N]\times[T]}\mathbb{E}\left(\|X_{it}\|_2^2\middle|\mathscr{D}_{\{e_{is}\}_{s<t}}\right)\le M$ *a.s.*

(v) *For* $j\in[p]$*, there exists a positive sequence* ξ_N *such that* $\max_{(i,t)\in[N]\times[T]}\left|X_{j,it}\right|\le\xi_N$ *a.s.*

(vi) $\min_{(i,t)\in[N]\times[T]}\mathsf{f}_{it}(0)\ge\underline{\mathsf{f}}>0$ *and* $\max_{(i,t)\in[N]\times[T]}\sup_{\varepsilon}|\mathsf{f}_{it}'(\varepsilon)|\le\bar{\mathsf{f}}'$.

(vii) $\min_{(i,t)\in[N]\times[T]}\mathfrak{f}_{it}(0)\ge\underline{\mathfrak{f}}>0$ *and* $\max_{(i,t)\in[N]\times[T]}\sup_{\varepsilon}|\mathfrak{f}_{it}'(\varepsilon)|\le\bar{\mathfrak{f}}'$.

(viii) $\min_{(i,t)\in[N]\times[T]}f_{it}(0)\ge\underline{f}>0$ *and* $\max_{(i,t)\in[N]\times[T]}\sup_{\varepsilon}|f_{it}'(\varepsilon)|\le\bar{f}'$.

(ix) $\frac{\xi_N^4\log(N\vee T)\sqrt{N\vee T}}{N\wedge T}=o(1)$ *and* $\frac{\left(\frac{N}{T}\vee 1\right)^{1/2}}{(N\wedge T)^{\frac{1}{4+2\vartheta}}}(\log(N\vee T))^{\frac{3+\vartheta}{4+2\vartheta}}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}=o(1)$ *for any* $\vartheta>0$.

Assumptions 2.1(i) imposes conditional independence of the error terms and covariates $X_{j,it}$ given the fixed effects. Assumptions 2.1(ii) imposes the moment condition for QR. Assumptions 2.1(iii) imposes the weak dependence assumption along the time dimension via the use of the notion of conditional strong mixing. See Prakasa Rao (2009) for the definition of conditional strong mixing and Su and Chen (2013a) for an application in the panel setup. Assumptions 2.1(iv)-(v) essentially imposes some conditions on the moments and tail behavior of the both covariates and errors. Note that we allow $X_{j,it}$ to have an infinite support. Assumptions 2.1(vi)-(viii), which are used in the proofs of Theorems 2.1, 2.2 and 2.3, respectively, specify conditions on the conditional density of ε_{it} given different σ-fields. Assumption 2.1(ix) imposes some restrictions on N, T and ξ_N in order to obtain the error bound of NNR estimators and to achieve the unbiasedness. It allows not only the case that N and T diverge to infinity at the the same rate, but also the case that N diverges to infinity not too faster than T, and vice versa.

Assumption 2.2. *Θ_0^0 is the fixed effect matrix with fixed rank K_0 and $\left\|\Theta_0^0\right\|_{\max}\le M$. For each $j\in[p]$, Θ_j^0 is the slope matrix of regressor j with rank being K_j such that* $\max_{j\in[p]}\left\|\Theta_j\right\|_{\max}\le M$ *and* $\max_{j\in[p]}K_j\le\bar{K}$ *for some fixed finite* $\bar{K}$.

Assumption 2.2 is the low-rank assumption for the intercept and slope matrices, which is the key assumption for the NNR. The uniform boundedness of elements of these matrices facilitates the asymptotic analysis, but can be relaxed at the cost of more lengthy argument. See Ma et al. (2020) for a similar condition.

Assumption 2.3. *There exist some constants C_σ and c_σ such that*

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

Assumption 2.3 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. Such an assumption is common in the literature; see, e.g., Assumption 3 in Ma et al. (2020).

To introduce the next assumption, we need some notation. Let $\Theta_j^0 = R_j \Sigma_j S_j'$ be the SVD for Θ_j^0. Further decompose $R_j = \left(R_{j,r}, R_{j,0}\right)$ with $R_{j,r}$ being the singular vectors corresponding to the nonzero singular values, $R_{j,0}$ being the singular vectors corresponding to the zero singular values. Decompose $S_j = \left(S_{j,r}, S_{j,0}\right)$ with $S_{j,r}$ and $S_{j,0}$ defined analogously. For any matrix $W \in \mathbb{R}^{N\times T}$, we define

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

where $\mathscr{P}_j(W)$ and $\mathscr{P}_j^{\perp}(W)$ are the linear projection of matrix W onto the low-rank space and its orthogonal space, respectively. Let $\Delta_{\Theta_j} = \Theta_j - \Theta_j^0$ for any Θ_j. With some positive constants C_1 and C_2, we define the following cone-like restricted set:

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

Assumption 2.4. *Let $C_2 > 0$ be a sufficiently large but fixed constant. There are constants C_3, C_4, such that, uniformly over $(\{\Delta_{\Theta_j}\}_{j=0}^p) \in \mathscr{R}(3, C_2)$, we have*

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

The same condition holds when Θ_j^0 is replaced by $\{\Theta_{j,it}^0\}_{i\in I_a, t\in[T]}$ for $a = 1, 2, 3$.

Assumption 2.4 parallels the restricted strong convexity (RSC) condition in Assumption 3.1 of Chernozhukov et al. (2019) who also provide some sufficient primitive conditions.

For any $j \in \{0, \cdots, p\}$, define $\tilde{\Delta}_{\Theta_j} = \tilde{\Theta}_j - \Theta_j^0$ and $\tilde{\Delta}_{\Theta_j}^{(1)} = \tilde{\Theta}_j^{(1)} - \Theta_j^{0,(1)}$, where $\Theta_j^{0,(1)} = \left\{\Theta_{j,it}^0\right\}_{i \in I_1, t \in [T]}$. The following theorem establishes the convergence rates of the NNR estimators of the coefficient matrices.

Theorem 2.1. *If Assumptions 2.1-2.4 hold, for $\forall j \in \{0, \cdots, p\}$, we have*

(i) $\frac{1}{\sqrt{NT}} \left\|\tilde{\Delta}_{\Theta_j}\right\|_F = O_p\left(\sqrt{\frac{\log(N \vee T)}{N \wedge T}} \xi_N^2\right)$, $\frac{1}{\sqrt{NT}} \left\|\tilde{\Delta}_{\Theta_j}^{(1)}\right\|_F = O_p\left(\sqrt{\frac{\log(N \vee T)}{N \wedge T}} \xi_N^2\right)$,

(ii) $\max_{k \in [K_j]} \left|\tilde{\sigma}_{k,j} - \sigma_{k,j}\right| = O_p\left(\sqrt{\frac{\log(N \vee T)}{N \wedge T}} \xi_N^2\right)$,
$\max_{k \in [K_j]} \left|\tilde{\sigma}_{k,j}^{(1)} - \sigma_{k,j}\right| = O_p\left(\sqrt{\frac{\log(N \vee T)}{N \wedge T}} \xi_N^2\right)$,

(iii) $\frac{1}{\sqrt{T}} \left\|V_j^0 - \tilde{V}_j O_j\right\|_F = O_p\left(\sqrt{\frac{\log(N \vee T)}{N \wedge T}} \xi_N^2\right)$,
$\frac{1}{\sqrt{T}} \left\|V_j^0 - \tilde{V}_j^{(1)} O_j^{(1)}\right\|_F = O_p\left(\sqrt{\frac{\log N \vee T}{N \wedge T}} \xi_N^2\right)$,

where O_j and $O_j^{(1)}$ are some orthogonal rotation matrices defined in the proof.

Remark 1. Theorem 2.1(i) reports the "rough" convergence rates of the NNR estimators of the coefficient matrices in terms of Frobenius norm for both the full-sample and sub-sample estimators. Unlike the traditional $(N \wedge T)^{-1/2}$-rate in the least squares framework, NNR estimators' convergence rates in the PQR framework usually have an additional $\sqrt{\log(N \vee T)}$ term due to the use of some exponential inequalities. The extra term ξ_N^2 in our rate is due to the upper bound of $|X_{j,it}|$, and it disappears in case $X_{j,it}$'s are uniformly bounded. Theorem 2.1(ii)-(iii) report the convergence rates for the estimators of the factors and factor loadings of Θ_j^0, which are inherited from those in Theorem 2.1(i). To derive these results, we establish the symmetrization inequality and contraction principle for the sequential symmetrization developed by Rakhlin et al. (2015). See Lemmas A.8 and A.9 in the online supplement for more detail.

2.3.2 Second Stage Estimator

To study the asymptotic properties of the second-stage estimators, we add some notation. Define

$$\Phi_i = \frac{1}{T} \sum_{t=1}^{T} \Phi_{it}^0 \Phi_{it}^{0\prime} \quad \text{and } \Psi_t = \frac{1}{N_2} \sum_{i \in I_2} \Psi_{it}^0 \Psi_{it}^{0\prime},$$

where $\Phi_{it}^0 = (v_{t,0}^{0\prime}, v_{t,1}^{0\prime} X_{1,it}, \cdots, v_{t,p}^{0\prime} X_{p,it})'$ and $\Psi_{it}^0 = (u_{i,0}^{0\prime}, u_{i,1}^{0\prime} X_{1,it}, \cdots, u_{i,p}^{0\prime} X_{p,it})'$. Let $K = \sum_{j=0}^p K_j$. Note that Φ_i and Ψ_t are $K \times K$ matrices. We add the following two assumptions.

Assumption 2.5. *There exist constants C_ϕ and c_ϕ such that a.s.*

$$\infty > C_\psi \geq \limsup_T \max_{t\in[T]} \lambda_{\max}(\Psi_t) \geq \liminf_T \min_{t\in[T]} \lambda_{\min}(\Psi_t) \geq c_\psi > 0,$$

$$\infty > C_\phi \geq \limsup_N \max_{i\in I_2} \lambda_{\max}(\Phi_i) \geq \liminf_N \min_{i\in I_2} \lambda_{\min}(\Phi_i) \geq c_\phi > 0.$$

Assumption 2.5 is similar to Assumption 8 in Ma et al. (2020). To introduce Theorem 2.2, we define

$$\dot{\varpi}_{it} = \left(\dot{v}_{t,0}^{(1)\prime}, \dot{v}_{t,1}^{(1)\prime} X_{1,it}, \cdots, \dot{v}_{t,p}^{(1)\prime} X_{p,it}\right)',$$

$$\varpi_{it}^0 = \left(\left(O_0^{(1)} v_{t,0}^0\right)', \left(O_1^{(1)} v_{t,1}^0\right)' X_{1,it}, \cdots, \left(O_p^{(1)} v_{t,p}^0\right)' X_{p,it}\right)',$$

$$u_i^0 = (u_{i,0}^{0\prime}, \cdots, u_{i,p}^{0\prime})', \quad \dot{\Delta}_{t,j} = O_j^{(1)\prime} \dot{v}_{t,j}^{(1)} - v_{t,j}^0, \quad \dot{\Delta}_{t,v} = (\dot{\Delta}_{t,0}', \cdots, \dot{\Delta}_{t,p}')',$$

$$\dot{\Delta}_{i,j} = O_j^{(1)\prime} \dot{u}_{i,j}^{(1)} - u_{i,j}^0, \quad \dot{\Delta}_{i,u} = (\dot{\Delta}_{i,0}', \cdots, \dot{\Delta}_{i,p}')',$$

$$D_i^I = \frac{1}{T}\sum_{t=1}^T \mathfrak{f}_{it}(0) \varpi_{it}^0 \varpi_{it}^{0\prime}, \quad D_i^{II} = \frac{1}{T}\sum_{t=1}^T [\tau - \mathbf{1}\{\varepsilon_{it} \leq 0\}] \varpi_{it}^0,$$

$$\mathbb{J}_i\left(\{\dot{\Delta}_{t,v}\}_{t\in[T]}\right) = \frac{1}{T}\sum_{t=1}^T \left[\mathbf{1}\{\varepsilon_{it} \leq 0\} - \mathbf{1}\left\{\varepsilon_{it} \leq \dot{\Delta}_{t,v}' \Psi_{it}^0\right\}\right] \varpi_{it}^0.$$

Theorem 2.2 below gives the uniform convergence rate and linear expansion of the factor loading estimators from second stage estimation.

Theorem 2.2. *Suppose Assumptions 2.1-2.5 hold. Then for each $j \in \{0, \cdots, p\}$, we have*

(i) $\max_{i\in I_2 \cup I_3} \left\|\dot{u}_{i,j}^{(1)} - O_j^{(1)} u_{i,j}^0\right\|_2 = O_p\left(\sqrt{\frac{\log(N\vee T)}{N\wedge T}} \xi_N^2\right)$,

(ii) $\max_{t\in[T]} \left\|\dot{v}_{t,j}^{(1)} - O_j^{(1)} v_{t,j}^0\right\|_2 = O_p\left(\sqrt{\frac{\log(N\vee T)}{N\wedge T}} \xi_N^2\right)$,

(iii) $\dot{\Delta}_{i,u} = \left[D_i^I\right]^{-1} \left[D_i^{II} + \mathbb{J}_i\left(\{\dot{\Delta}_{t,v}\}_{t\in[T]}\right)\right] + o_p\left((N\vee T)^{-1/2}\right)$ *uniformly over $i \in I_3$.*

Remark 2. Theorem 2.2(i) reports the uniform convergence rate for the factor loading estimators of Θ_j^0 for $i \in I_2 \cup I_3$; Theorem 2.2(ii) reports the uniform convergence rate for the factor estimators of Θ_j^0 for $t \in [T]$; Theorem 2.2(iii) reports the linear expansion for the factor loading estimators of Θ_j^0 for $i \in I_3$. However, the $\mathbb{J}_i\left(\left\{\dot{\Delta}_{t,v}\right\}_{t\in[T]}\right)$ term is not mean-zero and represents the bias induced by the first stage NNR. In the third stage below, we aim to remove such a bias from the linear expansion.

2.3.3 Third Stage Estimator

In the debiasing stage, we first apply PCA to all independent variables $X_{j,it}$, and then run the row- and column-wise QRs to obtain the final estimators. Below we give Assumptions 2.6-2.8 for the PCA procedure and establish the asymptotic linear expansions of PCA estimates in the online supplement. Theorem 2.3 below gives the asymptotic distribution of our final factor and factor loading estimates.

Assumption 2.6. *For all $j \in [p]$, there exists a constant $M > 0$ such that*

(i) $\mathbb{E}\left(e_{j,it}|\mu_{j,it}\right) = 0$,

(ii) $\mathbb{E}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[e_{j,it}e_{j,is} - \mathbb{E}(e_{j,it}e_{j,is})\right]\right]^2 \leq M$,

(iii) for all $i \in [N]$, $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\mathbb{E}(e_{j,it}e_{j,is})\right| \leq M$,

(iv) $\max_{t\in[T]}\frac{1}{N\sqrt{T}}\left\|e'_{j,t}E_j\right\|_2 = O_p\left(\frac{\log N\vee T}{N\wedge T}\right)$,
$\max_{i\in[N]}\frac{1}{T\sqrt{N}}\left\|e'_{j,i}E'_j\right\|_2 = O_p\left(\frac{\log N\vee T}{N\wedge T}\right)$, *where* $e_{j,i} = \left(e_{j,i1},\ldots,e_{j,iT}\right)'$, $e_{j,t} = \left(e_{j,1t},\ldots,e_{j,Nt}\right)'$, *and* $E_j = \{e_{j,it}\}_{i\in[N],t\in[T]}$.

Assumption 2.7. *For all $j \in [p]$,*

(i) recall that $L_j^0 = \left(l_{j,1}^0,\cdots,l_{j,N}^0\right)'$ *and* $W_j^0 = \left(w_{j,1}^0,\cdots,w_{j,T}^0\right)'$. $\lim_{N\to\infty}\frac{L_j^{0\prime}L_j^0}{N} = \Sigma_{L_j} > 0$ *and* $\lim_{T\to\infty}\frac{W_j^{0\prime}W_j^0}{T} = \Sigma_{W_j} > 0$,

(ii) the r_j eigenvalues of $\Sigma_{L_j}\Sigma_{W_j}$ are distinct.

Assumption 2.8. *For all $j \in [p]$, there exists a constant $M > 0$ such that*

(i) $\max_{t\in[T]}\mathbb{E}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}l_{j,i}^0 e_{j,it}\right\|_2^2 \leq M$ *and* $\max_{t\in[T]}\frac{1}{NT}e'_{j,t}E'_jL_j^0 = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right)$,

(ii) $\max_{i\in[N]}\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_{j,t}^0 e_{j,it}\right\|_2^2 \leq M$ *and* $\max_{i\in[N]}\frac{1}{NT}e'_{j,t}E'_jW_j^0 = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right)$.

Assumption 2.9. *For $\forall j \in [p]$,* *(i)* $\mathbb{E}\left[f_{it}(0)e_{j,it}\middle|\mathscr{D}\right]=0,$

(ii) for each $i\in[N]$ *and* $j\in[p]$, $\{f_{it}(0),f_{it}(0)e_{j,it}\}$ *is stationary strong mixing across* t *conditional on* $\mathscr{D}$.

Assumptions 2.6-2.8 are stronger than those in Bai and Ng (2020) because we strengthen their Assumptions A1(c) and A3 to hold uniformly. Assumption 2.9 imposes some moment and mixing conditions. Even though $f_{it}(\cdot)$ (the PDF of ε_{it} given $\mathscr{D}_e$) is a function of $\{e_{j,it}\}_{j\in[p],i\in[N],t\in[T]}$, we can show that Assumption 2.9 holds under some reasonable conditions. For example, we consider the location scale model:

$$Y_{it}=\beta_{0,it}+\sum_{j\in[p]}X_{j,it}\beta_{j,it}+\left(\gamma_{0,it}+\sum_{j\in[p]}X_{j,it}\gamma_{j,it}\right)u_{it},\quad \text{with}\quad X_{j,it}=l^{0\prime}_{j,i}w^{0}_{j,t}+e_{j,it},$$

where u_{it} is independent of $\{w^0_{j,t},e_{j,it}\}_{j\in[p],t\in[T]}$ and $l^0_{j,i}$ and $\beta_{j,it}$ are nonrandom. In this case, $\Theta^0_{j,it}=\beta_{j,it}+\gamma_{j,it}\mathscr{Q}_\tau(u_{it})$, $\varepsilon_{it}=\left(\gamma_{0,it}+\sum_{j\in[p]}X_{j,it}\gamma_{j,it}\right)[u_{it}-\mathscr{Q}_\tau(u_{it})]$, where $\mathscr{Q}_\tau(u_{it})$ is the τ-quantile of u_{it}. It is clear that $f_{it}(\cdot)$ is the function of $\{e_{j,it}\}_{j\in[p],i\in[N],t\in[T]}$ and all factors. However, if u_{it} is independent of sequence $\{e_{j,it}\}_{j\in[p],t\in[T]}$, we observe that $f_{it}(0)$ is the PDF of $u_{it}-\mathscr{Q}_\tau(u_{it})$ evaluated at zero point, which is independent of $\{e_{j,it}\}_{j\in[p],t\in[T]}$. Therefore, Assumption 2.9(i) holds under mild conditions that u_{it} is independent of the sequence $\{e_{it}\}_{t\in[T]}$ and $\mathbb{E}\left(e_{it}\middle|\mathscr{D}\right)=0$.

Define

$$\hat{V}_{u_j,i}=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[f_{it}(0)e^2_{j,it}\middle|\mathscr{D}\right]v^0_{t,j}v^{0\prime}_{t,j},\quad V_{u_j,i}=\mathbb{E}\left(\hat{V}_{u_j}\right)$$

$$\Omega_{u_j,i}=Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_{j,it}v^0_{t,j}(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right],$$

$$\hat{V}^{(3)}_{v_j,t}=\frac{1}{N_3}\sum_{i\in I_3}f_{it}(0)e^2_{j,it}u^0_{i,j}u^{0\prime}_{i,j},\quad V_{v_j}=\mathbb{E}\left(\hat{V}^{(3)}_{v_j,t}\right),$$

$$\Omega^{(3)}_{v_j}=\tau(1-\tau)\frac{1}{N_3}\sum_{i\in I_3}\mathbb{E}\left(e^2_{j,it}u^0_{i,j}u^{0\prime}_{i,j}\right).$$

Let $\Sigma_{u_j,i}=O^{(1)}_jV^{-1}_{u_j,i}\Omega_{u_j,i}V^{-1}_{u_j,i}O^{(1)\prime}_j$, $\Sigma^{(3)}_{v_j}=O^{(1)}_j\left(V^{(3)}_{v_j}\right)^{-1}\Omega_{v_j}\left(V^{(3)}_{v_j}\right)^{-1}O^{(1)\prime}_j$, $b^0_{j,it}=e_{j,it}v^0_{t,j}(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})$ and $\xi^0_{j,it}=e_{j,it}u^0_{i,j}(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})$. The following theorem establishes the asymptotic properties of the third-stage estimators.

Theorem 2.3. *Suppose that Assumptions 2.1-2.9 hold. Suppose that Assump-*

tion A.1 in Appendix B.3 of the online supplement hold. Let $O_{u,j}^{(1)}$ be the bounded matrix defined in the appendix that is related to rotation matrix $O_j^{(1)}$. Then we have that $\forall j \in [p]$,

(i) $\hat{u}_{i,j}^{(3,1)} - O_{u,j}^{(1)} u_{i,j}^0 = O_j^{(1)} \hat{V}_{u_j,i}^{-1} \frac{1}{T} \sum_{t=1}^T b_{j,it}^0 + \mathscr{R}_{i,u}^j$ *and* $\sqrt{T}\left(\hat{u}_{i,j}^{(3,1)} - O_{u,j}^{(1)} u_{i,j}^0\right) \rightsquigarrow \mathbb{N}\left(0, \Sigma_{u_j,i}\right) \forall i \in I_3$,

(ii) $\hat{v}_{t,j}^{(3,1)} - \left(O_{u,j}^{(1)\prime}\right)^{-1} v_{t,j}^0 = O_j^{(1)} \left(\hat{V}_{v_j,t}^{(3)}\right)^{-1} \frac{1}{N_3} \sum_{i \in I_3} \xi_{j,it}^0 + \mathscr{R}_{t,v}^j$ *and* $\sqrt{N_3}\left(\hat{v}_{t,j}^{(3,1)} - O_{v,j}^{(1)} v_{t,j}^0\right) \rightsquigarrow \mathbb{N}\left(0, \Sigma_{v_j}^{(3)}\right) \forall t \in [T]$,
where $\max_{i \in I_3}\left|\mathscr{R}_{i,u}^j\right| = o_p\left((N \vee T)^{-1/2}\right)$, $\max_{t \in [T]}\left|\mathscr{R}_{t,v}^j\right| = o_p\left((N \vee T)^{-\frac{1}{2}}\right)$.

Remark 3. Theorem 2.3 reports the linear expansions for the factor and factor loading estimators for each slope matrix obtained in Step 3. Compared with Chernozhukov et al. (2019), Theorem 2.3 obtains the uniform convergence rate rather than the point-wise result for the reminder terms $\mathscr{R}_{i,u}^j$ and $\mathscr{R}_{t,v}^j$. In addition, since the regressors in the debiasing step are obtained from Step 2 instead of Step 1, we don't have independence between the regressors and error terms, which makes the proof more complex than that in Chernozhukov et al. (2019). See the proof in the appendix on how to handle the dependence. Assumption A.1 in the online supplement is a regularity condition on the density of ε_{it}.

Following Theorem 2.3 and estimators defined in Table 2.1, we have that $\forall j \in [p]$, $\forall i \in [N]$ and $\forall t \in [T]$,

$$\hat{u}_{i,j}^{(a,b)} - O_{u,j}^{(b)} u_{i,j}^0 = O_j^{(b)} \hat{V}_{u_j,i}^{-1} \frac{1}{T} \sum_{t=1}^T v_{t,j}^0 b_{j,it}^0 + \mathscr{R}_{i,u}^j,$$
$$\hat{v}_{t,j}^{(a,b)} - \left(O_{u,j}^{(b)\prime}\right)^{-1} v_{t,j}^0 = O_j^{(b)} \left(\hat{V}_{v_j,t}^{(a)}\right)^{-1} \frac{1}{N_a} \sum_{i \in I_a} u_{i,j}^0 \xi_{j,it}^0 + \mathscr{R}_{t,v}^j,$$

where $\hat{V}_{v_j,t}^{(a)} = \frac{1}{N_a} \sum_{i \in I_a} f_{it}(0) e_{j,it}^2 u_{i,j}^0 u_{i,j}^{0\prime}$, $a \in [3]$ and $b \in [3] \setminus \{a\}$.

Given the above estimates for the factors and factor loadings, we can estimate $\Theta_{j,it}^0$ by

$$\hat{\Theta}_{j,it} = \frac{1}{2} \sum_{a \in [3]} \sum_{b \in [3] \setminus \{a\}} \left\{\hat{u}_{i,j}^{(a,b)\prime} \hat{v}_{t,j}^{(a,b)}\right\} \mathbf{1}_{ia}$$

where $\mathbf{1}_{ia} = \mathbf{1}\{i \in I_a\}$ for $i \in [N]$. Let $\Xi_{j,it}^0 = \frac{1}{T} v_{t,j}^{0\prime} \Sigma_{u_j,i} v_{t,j}^0 + \sum_{a=1}^3 \frac{1}{N_a} \mathbf{1}_{ia} u_{i,j}^{0\prime} \Sigma_{v_j}^a u_{i,j}^0$. The following proposition studies the asymptotic properties of $\hat{\Theta}_{j,it}$.

Proposition 2.1. *Under Assumptions 2.1-2.9 and Assumption A.1, $\forall j \in [p]$ we have*

(i) $\hat{\Theta}_{j,it} - \Theta^0_{j,it} = \sum_{a=1}^{3} u^{0\prime}_{i,j} \left(\hat{V}^{(a)}_{v_j,t}\right)^{-1} \frac{1}{N_a} \sum_{i^* \in I_a} \xi_{j,i^*t} \mathbf{1}_{i^*a} + v^{0\prime}_{t,j} \hat{V}^{-1}_{u_j} \frac{1}{T} \sum_{t^*=1}^{T} b^0_{j,it^*} + \mathscr{R}^j_{it}$, *where* $\max_{i \in I_3, t \in [T]} \left|\mathscr{R}^j_{it}\right| = o_p\left((N \vee T)^{-1/2}\right)$,

(ii) $\max_{i \in [N], t \in [T]} \left|\hat{\Theta}_{j,it} - \Theta^0_{j,it}\right| = O_p\left(\sqrt{\frac{\log N \vee T}{N \wedge T}}\right)$,

(iii) $\left(\Xi^0_{j,it}\right)^{-1/2} \left(\hat{\Theta}_{j,it} - \Theta^0_{j,it}\right) \rightsquigarrow \mathbb{N}(0,1)$.

Remark 4. Proposition 2.1 establishes the distribution theory for the slope estimators. Recall that we remove the principle component from the independent variables $X_{j,it}$ which is the key point in the debiasing step and why we don't have the distribution theory result for the intercept estimates $\hat{\Theta}_{0,it}$ in the current framework. However, once we have the distribution theory for the slope estimates, we can follow Chen et al. (2021) and obtain a new estimator for $\Theta^0_{0,it}$ from the smoothed quantile regression and establish its distribution theory. We leave this for the further research.

To make inference for $u^0_{i,j}$, $v^0_{t,j}$, and $\Theta^0_{j,it}$, one needs to estimate their asymptotic variances $\Sigma_{u_j,i}$, Σ_{v_j} and $\Xi^0_{j,it}$ consistently. Let $k(\cdot)$ be a PDF-type kernel function and $K(\cdot)$ be its survival function such that $\int k(u)du = 1$ and $K(u) := \int_u^\infty k(v)dv$. Let h_N be the bandwidth such that $h_N \to 0$ with $N \to \infty$. Define $K_{h_N}(\cdot) = K(\frac{\cdot}{h_N})$, $k_{h_N}(\cdot) = \frac{1}{h_N} k(\frac{\cdot}{h_N})$. Let $\hat{\varepsilon}_{it} = Y_{it} - \hat{\Theta}_{0,it} - \sum_{j \in [p]} X_{j,it} \hat{\Theta}_{j,it}$, $\hat{\mathrm{v}}_{t,s,j} = \frac{1}{6} \sum_{a \in [3]} \sum_{b \in [3] \setminus \{a\}} \hat{v}^{(a,b)}_{t,j} \hat{v}^{(a,b)\prime}_{s,j}$, and $\hat{\mathrm{u}}_{i,i,j} = \frac{1}{2} \sum_{a \in [3]} \sum_{b \in [3] \setminus \{a\}} \hat{u}^{(a,b)}_{i,j} \hat{u}^{(a,b)\prime}_{i,j} \mathbf{1}_{ia}$. Define

$$\hat{\mathbb{V}}_{u_j} = \frac{1}{NT} \sum_{i \in [N]} \sum_{t \in [T]} k_{h_N}(\hat{\varepsilon}_{it}) \hat{e}^2_{j,it} \hat{\mathrm{v}}_{t,t,j}, \quad \hat{\mathbb{V}}_{v_j} = \frac{1}{NT} \sum_{i \in [N]} \sum_{t \in [T]} k_{h_N}(\hat{\varepsilon}_{it}) \hat{e}^2_{j,it} \hat{\mathrm{u}}_{i,i,j},$$

$$\hat{\Omega}_{u_j} = \frac{1}{NT} \sum_{i \in [N]} \left\{ \sum_{t \in [T]} \tau(1-\tau) \hat{e}^2_{j,it} \hat{\mathrm{v}}_{t,t,j} + \sum_{t=1}^{T-T_1} \sum_{s=t+1}^{t+T_1} S_{j,its} + \sum_{t=1+T_1}^{T} \sum_{s=t-T_1}^{t-1} S_{j,its} \right\},$$

$$\hat{\Omega}_{v_j} = \frac{\tau(1-\tau)}{NT} \sum_{i \in [N]} \sum_{t \in [T]} \hat{e}^2_{j,it} \hat{\mathrm{u}}_{i,i,j}, \quad \hat{\Sigma}_{u_j} = \hat{\mathbb{V}}^{-1}_{u_j} \hat{\Omega}_{u_j} \hat{\mathbb{V}}^{-1}_{u_j}, \quad \hat{\Sigma}_{v_j} = \hat{\mathbb{V}}^{-1}_{v_j} \hat{\Omega}_{v_j} \hat{\mathbb{V}}^{-1}_{v_j},$$

where $S_{j,its} = \hat{e}_{j,it} \hat{e}_{j,is} \hat{\mathrm{v}}_{t,s,j} \left[\tau - K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right)\right] \left[\tau - K\left(\frac{\hat{\varepsilon}_{is}}{h_N}\right)\right]$. We further define

$$\hat{\Xi}_{j,it} = \frac{1}{2} \sum_{a \in [3]} \sum_{b \in [3] \setminus \{a\}} \left(\frac{1}{T} \hat{v}^{(a,b)\prime}_{t,j} \hat{\Sigma}_{u_j} \hat{v}^{(a,b)}_{t,j} + \frac{1}{N_a} \mathbf{1}_{ia} \hat{u}^{(a,b)\prime}_{i,j} \hat{\Sigma}_{v_j} \hat{u}^{(a,b)}_{i,j} \right).$$

Let $F_{i,ts}(\cdot,\cdot)$ and $f_{i,ts}(\cdot,\cdot)$ denote the joint CDF and PDF of $(\varepsilon_{it},\varepsilon_{is})$ given $\mathscr{D}_e$, respectively. To justify the consistency of the variance estimators, we add the following assumption.

Assumption 2.10. *(i)* $\int_{-\infty}^{+\infty} k(u)du = 1$, $\int_{-\infty}^{+\infty} k(u)u^j du = 0$ *for* $j \in \{1,\cdots,m-1\}$ *and* $\int_{-\infty}^{+\infty} k(u)u^m du \neq 0$ *for* $m \geq 1$.

(ii) $h_N \to 0$ *and* $\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{1/4} \frac{\xi_N^2}{h_N} \to 0$.

(iii) $T_1 \to \infty$ *and* $\sqrt{\frac{\log(N\vee T)}{N\wedge T}} \frac{\xi_N^2 T_1}{h_N^2} \to 0$.

(iv) $f_{it}(c)$ *is m times continuously differentiable with respect to c and* $f_{i,ts}(c_1,c_2)$ *is m times continuously differentiable with respect to* (c_1,c_2).

(v) $\forall i \in [N]$, $V_{u_j,i} = V_{u_j}$ *and* $\Omega_{u_j,i} = \Omega_{u_j}$.

(vi) $\forall a \in [3]$, $V_{v_j}^{(a)} = \frac{1}{N} \sum_{i\in[N]} \mathbb{E}\left[f_{it}(0) e_{j,it}^2 u_{i,j}^0 u_{i,j}^{0\prime}\right] + o_p(1)$ *and*
$\Omega_{v_j}^{(a)} = \frac{\tau(1-\tau)}{N} \sum_{i\in[N]} \mathbb{E}\left(e_{j,it}^2 u_{i,j}^0 u_{i,j}^{0\prime}\right) + o_p(1)$.

Assumption 2.10(i)-(iv) are standard for consistent estimation of the asymptotic variance matrix; see, e.g., Chen (2022) and Galvao and Kato (2016). Assumption 2.10(v) imposes the homogeneity moment condition across individuals, and Assumption 2.10(vi) assumes the moments calculated from subsamples are close to those from the full sample given the random splitting. Under Assumption 2.10, following the idea of Chen (2022), we establish in Lemma A.33 of the online supplement the consistency of $\hat{\Sigma}_{u_j}$ and $\hat{\Sigma}_{v_j}$. Similar conclusions hold for the other estimates.

2.4 Specification Tests

In this section, we consider two specification tests under different rank conditions.

2.4.1 Testing for Homogeneity across Individuals or Time

When $K_j = 1$ for some $j \in [p]$, it is interesting to test whether the matrix Θ_j^0 is homogeneous across individuals (i.e., row-wise) or across time (i.e., column-wise).

For these two cases, we can write factors and factor loadings as

$$u_{i,j}^{0}=u_{j}+c_{i,j}^{u}\quad \text{and}\quad v_{t,j}^{0}=v_{j}+c_{t,j}^{v},\quad \text{respectively},$$

where $u_j=\frac{1}{N}\sum_{i=1}^{N}u_{i,j}^{0}$ and $v_j=\frac{1}{T}\sum_{t=1}^{T}v_{t,j}^{0}$. For the homogeneity across individuals, the null and alternative hypotheses can be written as

$$H_0^I: c_{i,j}^{u}=0\quad \forall i\in[N]\quad v.s.\quad H_1^I: c_{i,j}^{u}\neq 0 \text{ for some } i\in[N]. \tag{2.10}$$

Similarly, for the homogeneity across time, the null and alternative hypotheses can be written as

$$H_0^{II}: c_{t,j}^{v}=0\quad \forall t\in[T]\quad v.s.\quad H_1^{II}: c_{t,j}^{v}\neq 0 \text{ for some } t\in[T]. \tag{2.11}$$

Note that we aim to test the two null hypotheses separately. That is, we can test for homogeneous slope across individuals while allowing for heterogeneous slopes across time and vice versa. This is different from the majority of the literature which either tests for slope homogeneity across individuals while assuming the slopes are homogeneous across time or tests for structural breaks across time while assuming the slopes are homogeneous across individuals.

We first consider testing H_0^I. Following the lead of Castagnetti et al. (2015), we define[2]

$$\begin{aligned} S_{u_j}^{(a,b)} &= \max_{i\in I_a} T(\hat{u}_{i,j}^{(a,b)}-\hat{\bar{u}}_{j}^{(a,b)})'\hat{\Sigma}_{u_j}^{-1}(\hat{u}_{i,j}^{(a,b)}-\hat{\bar{u}}_{j}^{(a,b)})\quad \text{and}\\ S_{u_j} &= \max\left(S_{u_j}^{(3,1)},S_{u_j}^{(2,3)},S_{u_j}^{(1,2)}\right), \end{aligned} \tag{2.12}$$

where $\hat{\bar{u}}_{j}^{(a,b)}=\frac{1}{N_a}\sum_{i\in I_a}\hat{u}_{i,j}^{(a,b)}$. Similarly, to test for H_0^{II}, we construct

$$S_{v_j}=\max\left(\tilde{S}_{v_j}^{(3,1)},\tilde{S}_{v_j}^{(2,3)},\tilde{S}_{v_j}^{(1,3)}\right),$$

where

$$S_{v_j}^{(a,b)}=\max_{t\in[T]} N(\hat{v}_{t,j}^{(a,b)}-\hat{\bar{v}}_{j}^{(a,b)})'\hat{\Sigma}_{v_j}^{-1}(\hat{v}_{t,j}^{(a,b)}-\hat{\bar{v}}_{j}^{(a,b)}),\quad \tilde{S}_{v_j}^{(a,b)}=\frac{1}{2}S_{v_j}^{(a,b)}-\mathsf{b}(T),$$

$\hat{\bar{v}}_{j}^{(a,b)}=\frac{1}{T}\sum_{t=1}^{T}\hat{v}_{t,j}^{(a,b)}$, and $\mathsf{b}(n)=\log n-\frac{1}{2}\log\log n-\log\Gamma(\frac{1}{2})$ for $n\in\{N,T,NT\}$.

To proceed, we introduce some notation. Recall that $b_{j,it}^{0}=e_{j,it}v_{t,j}^{0}(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})$

[2] Alternatively, we can also define $S_{u_j}^{o}=\max\left(S_{u_j}^{(3,2)},S_{u_j}^{(2,1)},S_{u_j}^{(1,3)}\right)$. It is easy to show that this statistic shares the same asymptotic null distribution as S_{u_j}. But due to the unknown dependence structure between the two, we cannot take the maximum or the other continuous function of S_{u_j} and $S_{u_j}^{o}$ as a new test statistic.

and $\xi^0_{j,it} = e_{j,it} u^0_{i,j} (\tau - \mathbf{1}\{\varepsilon_{it} \le 0\})$. Define

$$\mathfrak{b}^{(1)}_{j,it} = \hat{V}^{-1}_{u_j} b^0_{j,it}, \quad \mathfrak{b}^{(2)}_{j,it} = \left(\hat{V}^{(3)}_{v_j,t}\right)^{-1} \xi^0_{j,it},\ \mathfrak{b}^{(3)}_{j,it} = \left(\hat{V}^{(2)}_{v_j,t}\right)^{-1} \xi^0_{j,it}, \text{ and } \mathfrak{b}^{(4)}_{j,it} = \left(\hat{V}^{(1)}_{v_j,t}\right)^{-1} \xi^0_{j,it}.$$

Let $\mathfrak{B}^{(\ell)}_{j,t} = \left(\mathfrak{b}^{(\ell)\prime}_{j,1t}, \cdots, \mathfrak{b}^{(\ell)\prime}_{j,Nt}\right)'$ for $\ell \in [4]$. Define

$$\Sigma^{(1)}_{\mathfrak{B},j} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathbb{E}\left(\mathfrak{B}^{(1)}_{j,t}\mathfrak{B}^{(1)\prime}_{j,s}\right), \quad \Sigma^{(2)}_{\mathfrak{B},j} = \frac{1}{N_3}\sum_{i\in I_3} \mathbb{E}\left(\mathfrak{B}^{(2)}_{j,i}\mathfrak{B}^{(2)\prime}_{j,i}\right),$$
$$\Sigma^{(3)}_{\mathfrak{B},j} = \frac{1}{N_2}\sum_{i\in I_2} \mathbb{E}\left(\mathfrak{B}^{(3)}_{j,i}\mathfrak{B}^{(3)\prime}_{j,i}\right), \quad \text{and } \Sigma^{(4)}_{\mathfrak{B},j} = \frac{1}{N_1}\sum_{i\in I_1} \mathbb{E}\left(\mathfrak{B}^{(4)}_{j,i}\mathfrak{B}^{(4)\prime}_{j,i}\right).$$

We add the following two assumptions.

Assumption 2.11. $\forall j \in [p]$, *we assume*

$$\bar{\lambda} \ge \lambda_{\max}\left(\Sigma_{u_j}\right) \ge \lambda_{\min}\left(\Sigma_{u_j}\right) \ge \underline{\lambda} > 0, \quad \bar{\lambda} \ge \lambda_{\max}\left(\Sigma_{v_j}\right) \ge \lambda_{\min}\left(\Sigma_{v_j}\right) \ge \underline{\lambda} > 0.$$

Assumption 2.12. *(i) There exists a high dimensional Gaussian vector* $\mathbb{Z}^{(1)}_{\mathfrak{B}} \sim N\left(0, \Sigma^{(1)}_{\mathfrak{B},j}\right)$ *such that* $\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathfrak{B}^{(1)}_{j,t} - \mathbb{Z}^{(1)}_{\mathfrak{B}}\right\|_{\max} = o_p(1)$.

(ii) There exists high dimensional Gaussian vectors $\mathbb{Z}^{(\ell)}_{\mathfrak{B}} \sim N\left(0, \Sigma^{(\ell)}_{\mathfrak{B},j}\right)$ *for* $\ell = 2,3,4$ *such that* $\left(\mathbb{Z}^{(2)}_{\mathfrak{B}}, \mathbb{Z}^{(3)}_{\mathfrak{B}}, \mathbb{Z}^{(4)}_{\mathfrak{B}}\right)$ *are independent,*

$$\left\|\frac{1}{\sqrt{N_3}}\sum_{i\in I_3}\mathfrak{B}^{(2)}_{j,i} - \mathbb{Z}^{(2)}_{\mathfrak{B}}\right\|_{\max} = o_p(1), \quad \left\|\frac{1}{\sqrt{N_2}}\sum_{i\in I_2}\mathfrak{B}^{(3)}_{j,i} - \mathbb{Z}^{(3)}_{\mathfrak{B}}\right\|_{\max} = o_p(1), \text{ and}$$
$$\left\|\frac{1}{\sqrt{N_1}}\sum_{i\in I_1}\mathfrak{B}^{(4)}_{j,i} - \mathbb{Z}^{(4)}_{\mathfrak{B}}\right\|_{\max} = o_p(1).$$

Assumption 2.11 implies that both Σ_{u_j} and Σ_{v_j} are well behaved. Assumption 2.12 imposes that we can approximate high dimensional vectors $\frac{1}{\sqrt{T}}\sum_{t\in[T]}\mathfrak{B}^{(1)}_{j,t}$, $\frac{1}{\sqrt{N_3}}\sum_{i\in I_3}\mathfrak{B}^{(2)}_{j,i}$, $\frac{1}{\sqrt{N_2}}\sum_{i\in I_2}\mathfrak{B}^{(3)}_{j,i}$ and $\frac{1}{\sqrt{N_1}}\sum_{i\in I_1}\mathfrak{B}^{(4)}_{j,i}$ by four Gaussian vectors. Similar conditions have been imposed in the literature; see, e.g., Assumption SA3 Lu and Su (2023).

The following theorem reports the asymptotic properties of S_{u_j} and S_{v_j} under the respective null and alternative hypotheses.

Theorem 2.4. *Suppose that Assumptions 2.1-2.12 and Assumptions A.1 in the online supplement hold and* $(N,T) \to \infty$. *Then*

(i) Under H^I_0, *we have* $\mathbb{P}\left(\frac{1}{2}S_{u_j} \le x + \mathsf{b}(N)\right) \to e^{-e^{-x}}$; *and under* H^{II}_0, *we have* $\mathbb{P}\left(S_{v_j} \le x\right) \to e^{-3e^{-x}}$.

(ii) Under H_1^I, *if* $\frac{T}{\log N}\max_{i\in[N]}\left\|c_{i,j}^u\right\|_2^2 \to \infty$, *we have* $\mathbb{P}\left(S_{u_j} > c_{\alpha,1\cdot N}\right) \to 1$ *with* $c_{\alpha,1\cdot N} = 2\mathrm{b}(N) - \log|\log(1-\alpha)|^2$ *and* α *is the significance level. Under* H_1^{II}, *if* $\frac{N}{\log T}\max_{t\in[T]}\left\|c_{t,j}^v\right\|_2^2 \to \infty$, *we have* $\mathbb{P}\left(S_{v_j} > c_{\alpha,2}\right) \to 1$ *with* $c_{\alpha,2} = -\log\left(-\frac{1}{3}\log(1-\alpha)\right)$.

Remark 5. Theorem 2.4 implies that our test statistics follow the Gumbel distributions asymptotically under the null, are consistent under the global alternatives, and have non-trivial power against the local alternatives. The power function of S_{u_j} approaches 1 as long as $\frac{T}{\log N}\max_{i\in[N]}\left\|c_{i,j}^u\right\|_2^2$ diverges to infinity as $(N,T)\to\infty$.

2.4.2 Test for an Additive Structure

When $K_j = 2$ for some $j\in[p]$, it is interesting to test whether $\Theta_{j,it}^0$ exhibits the additive structure which is widely assumed in a two-way fixed effects model. That is, one may test the following null hypothesis

$$H_0^{III}: \Theta_{j,it}^0 = \lambda_{j,i} + f_{j,t},\ \forall (i,t)\in[N]\times[T], \tag{2.13}$$

The alternative hypothesis H_1^{III} is the negation of H_0^{III}.

Let $\bar{\Theta}_{j,i\cdot} = \frac{1}{T}\sum_{t\in[T]}\Theta_{j,it}^0$, $\bar{\Theta}_{j,\cdot t}^{I_a} = \frac{1}{N_a}\sum_{i\in I_a}\Theta_{j,it}^0$, and $\bar{\Theta}_j^{I_a} = \frac{1}{N_aT}\sum_{i\in I_a}\sum_{t\in[T]}\Theta_{j,it}^0$ for $a\in[3]$. Define

$$\Theta_{j,it}^* = \Theta_{j,it}^0 - \bar{\Theta}_{j,i\cdot} - \bar{\Theta}_{j,\cdot t}^{I_a} + \bar{\Theta}_j^{I_a}, \quad \forall i\in I_a, t\in[T], j\in[p].$$

Note that $\Theta_{j,it}^* = 0\ \forall (i,t)\in[N]\times[T]$ under H_0^{III}. So we can propose a test for H_0^{III} based on estimates of $\Theta_{j,it}^*$. Define

$$\hat{\bar{\Theta}}_{j,i\cdot} = \frac{1}{T}\sum_{t\in[T]}\hat{\Theta}_{j,it},\ \hat{\bar{\Theta}}_{j,\cdot t}^{I_a} = \frac{1}{N_a}\sum_{i\in I_a}\hat{\Theta}_{j,it},\ \text{and } \hat{\bar{\Theta}}_j^{I_a} = \frac{1}{N_aT}\sum_{i\in I_a}\sum_{t\in[T]}\hat{\Theta}_{j,it}$$

for $a\in[3]$. Then, we can define the sample analogue of $\hat{\Theta}_{j,it}^*$ as

$$\hat{\Theta}_{j,it}^{*a} = \hat{\Theta}_{j,it} - \hat{\bar{\Theta}}_{j,i\cdot} - \hat{\bar{\Theta}}_{j,\cdot t}^{I_a} + \hat{\bar{\Theta}}_j^{I_a}, \quad \forall i\in I_a, t\in[T], j\in[p].$$

Its corresponding asymptotic variance can be estimated by $\hat{\Sigma}_{j,it}^*$ defined as

$$\begin{aligned}\hat{\Sigma}_{j,it}^* = &\frac{1}{2}\sum_{a\in[3]}\sum_{b\in[3]\setminus\{a\}}\frac{1}{N_a}\left(\hat{u}_{i,j}^{(a,b)} - \hat{\bar{u}}_j^{(a,b)}\right)'\hat{\Sigma}_{v_j}\left(\hat{u}_{i,j}^{(a,b)} - \bar{u}_j^{(a,b)}\right)\mathbf{1}_{ia}\\ &+\frac{1}{6}\sum_{a\in[3]}\sum_{b\in[3]\setminus\{a\}}\frac{1}{T}\left(\hat{v}_{t,j}^{(a,b)} - \hat{\bar{v}}_j^{(a,b)}\right)'\hat{\Sigma}_{u_j}\left(\hat{v}_{t,j}^{(a,b)} - \hat{\bar{v}}_j^{(a,b)}\right),\end{aligned}$$

where $\hat{\bar{u}}_j^{(a,b)} = \frac{1}{N_a}\sum_{i\in I_a}\hat{u}_{i,j}^{(a,b)}$ and $\hat{\bar{v}}_j^{(a,b)} = \frac{1}{T}\sum_{t\in[T]}\hat{v}_{t,j}^{(a,b)}$. Then, the final test statistic is

$$S_{NT} = \max_{i\in[N],t\in[T]} \left(\hat{\Theta}_{j,it}^*\right)^2 / \hat{\Sigma}_{j,it}^*.$$

The following theorem studies the asymptotic properties of S_{NT} under the null and alternatives.

Theorem 2.5. *Suppose that Assumptions 2.1-2.12 and Assumptions A.1 in the online supplement hold and $(N,T)\to\infty$. Under H_0^{III},*

$$\mathbb{P}\left(\frac{1}{2}S_{NT} \le x + \mathsf{b}(NT)\right) \to e^{-e^{-x}};$$

under H_1^{III}, if $\frac{N\wedge T}{\log NT}\max_{i\in[N],t\in[T]}\left|\Theta_{j,it}^*\right|^2 \to \infty$, then we have $\mathbb{P}\left(S_{NT} > c_{\alpha,3\cdot NT}\right) \to 1$ with $c_{\alpha,3\cdot NT} = 2\mathsf{b}(NT) - \log|\log(1-\alpha)|^2$.

Similar remark after Theorem 2.4 holds here. In particular, S_{NT} has the desired asymptotic Gumbel distribution under the null and is consistent under the global alternative.

2.5 Monte Carlo Simulations

In this section, we conduct a set of Monte Carlo simulations to show the finite sample performance of our low-rank quantile regression estimates and specification tests.

2.5.1 Data Generating Processes

Below we will consider the following data generating process (DGP):

$$Y_{it} = \Theta_{0,it} + X_{it}'\Theta_{it} + (1 + 0.1X_{1,it} + 0.1X_{2,it})u_{it},$$

where $X_{it} = (X_{1,it}, X_{2,it})'$, $\Theta_{it} = (\Theta_{1,it}, \Theta_{2,it})'$, $\Theta_{0,it}$ is the intercept term which will be specified via the IFEs.

First, we consider four DGPs where the rank of each slope matrix is 1:

DGP 1: **Constant slope with i.i.d. error.** Let $\Theta_{0,it} = \lambda_i f_t$, where $\lambda_i, f_t \sim N(2,5)$. Then let $\Theta_{1,it} = \Theta_{2,it} = 2\ \forall(i,t) \in [N]\times[T]$, and $X_{j,it} = l_{j,i}^0 w_{j,t}^0 + U(0,1)$ for $j \in \{1,2\}$ with $l_{1,i}^0$, $l_{2,i}^0$, $w_{1,t}^0$ and $w_{2,t}^0 \sim U(0,1)$. $u_{it} \overset{i.i.d}{\sim} \frac{t(3)}{\sqrt{3}}$.

DGP 2: **Factor slope with rank 1 and i.i.d. error.** Same as DGP 1 except that the slope coefficients follow the factor structure with one factor rather than homogeneous across both individuals and time, i.e., $\Theta_{1,it} = a_{1,i}g_{1,t}$, $\Theta_{2,it} = a_{2,i}g_{2,t}$, where $a_{1,i}$, $g_{1,t}$, $a_{2,i}$ and $g_{2,t} \sim N(0,2)$. Except these, all other settings remain the same as in DGP 1.

DGP 3: **Constant slope with serial correlation.** Same as DGP 1 except that we set $u_{it} = 0.2u_{i,t-1} + \varepsilon_{it}$, $\varepsilon_{it} \overset{i.i.d}{\sim} \frac{t(3)}{\sqrt{3}}$ and all other settings remain the same.

DGP 4: **Factor slope with rank 1 and serial correlation.** Same as DGP 2 except that we set $u_{it} = 0.2u_{i,t-1} + \varepsilon_{it}$, $\varepsilon_{it} \overset{i.i.d}{\sim} \frac{t(3)}{\sqrt{3}}$ and all other settings remain the same.

For the case that the rank of the slope matrix is 2, we consider two DGPs which have the additive structure for the slope coefficient of one regressor and the factor structure with two factors for the slope coefficient of another regressor. Specifically,

DGP 5: **Additive and factor slopes with i.i.d. error.** $\Theta_{0,it} = \lambda_i f_t$, $\Theta_{1,it} = a_{1,i} + g_{1,t}$ and $\Theta_{2,it} = a'_{2,i}g_{2,t}$ such that $a_{2,i} = (a_{2,i,1}, a_{2,i,2})'$, $g_{2,t} = (g_{2,t,1}, g_{2,t,2})'$, $\lambda_i, f_t, a_{1,i}, g_{1,i} \sim N(2,5)$ and $a_{2,i,1}, a_{2,i,2}, g_{2,i,1}, g_{2,i,2} \sim N(0,5)$. Moreover, $X_{1,it} = l^0_{1,i}w^0_{1,t} + U(0,4)$, $X_{2,it} = l^0_{2,i}w^0_{2,t} + Beta(2,5)$ with $l^0_{1,i}, w^0_{1,t} \sim U(0,4)$ and $l^0_{2,i}, w^0_{2,t} \sim Beta(2,5)$. $u_{it} \overset{i.i.d}{\sim} \frac{t(3)}{\sqrt{3}}$.

DGP 6: **Additive and factor slopes with serial correlation.** Same as DGP 5 except that the error u_{it} follows AR(1) process like in DGPs 3 and 4.

2.5.2 Estimation Results

For $\Theta \in \mathbb{R}^{N\times T}$, define $RMSE(\Theta) = \frac{1}{\sqrt{NT}}\left\|\Theta - \Theta^0\right\|_F$. Table 2.3 shows the RMSEs of the full-sample low rank matrix estimates under different quantiles for each DGP. As Theorem 2.1(i) predicts, the RMSEs decrease as both N and T increase. Given the fact that $N \wedge T = T$ in the simulations, the decrease of the RMSEs is largely driven by the increase of T.

Table 2.4 reports the frequency of correct rank estimation by the singular value thresholding (SVT) approach based on 1000 replications. Note that the true ranks of the intercept and slope matrices in DGPs 1-4 and 5-6 are 1 and 2, respectively.

The results show that the SVT can accurately determine the correct rank of the coefficient matrices in all DGPs for all three quantile indices under investigation.

Table 2.3: RMSEs of low rank estimates in the full sample

DGP	N	T	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
			$\tilde{\Theta}_0$	$\tilde{\Theta}_1$	$\tilde{\Theta}_2$	$\tilde{\Theta}_0$	$\tilde{\Theta}_1$	$\tilde{\Theta}_2$	$\tilde{\Theta}_0$	$\tilde{\Theta}_1$	$\tilde{\Theta}_2$
1	75	35	0.922	0.324	0.329	1.242	0.288	0.297	1.839	0.609	0.658
		70	0.707	0.280	0.275	0.819	0.220	0.203	1.266	0.519	0.523
	150	35	1.012	0.337	0.340	1.099	0.258	0.262	1.932	0.661	0.623
		70	0.745	0.272	0.265	0.825	0.205	0.206	1.324	0.522	0.504
2	75	35	0.871	0.521	0.505	0.881	0.704	0.680	1.278	1.055	0.970
		70	0.692	0.401	0.373	0.672	0.553	0.537	1.057	0.744	0.768
	150	35	0.877	0.507	0.480	1.022	0.790	0.815	1.334	1.018	1.040
		70	0.703	0.374	0.373	0.689	0.531	0.538	1.059	0.829	0.787
3	75	35	0.945	0.334	0.329	1.115	0.280	0.265	1.876	0.630	0.627
		70	0.682	0.286	0.279	0.809	0.230	0.214	1.244	0.486	0.492
	150	35	0.973	0.334	0.331	1.211	0.287	0.291	1.771	0.590	0.612
		70	0.757	0.274	0.272	0.801	0.208	0.195	1.360	0.494	0.527
4	75	35	0.885	0.515	0.519	0.915	0.693	0.723	1.382	1.125	1.037
		70	0.669	0.393	0.384	0.652	0.511	0.520	1.053	0.812	0.774
	150	35	0.889	0.513	0.483	0.905	0.761	0.686	1.409	1.118	1.133
		70	0.725	0.376	0.377	0.717	0.547	0.565	1.058	0.724	0.775
5	75	35	0.218	0.268	0.450	0.307	0.308	0.606	0.844	0.466	0.936
		70	0.174	0.226	0.414	0.213	0.200	0.493	0.610	0.388	0.838
	150	35	0.236	0.245	0.458	0.299	0.291	0.634	1.299	0.863	1.778
		70	0.174	0.214	0.423	0.216	0.203	0.450	0.629	0.377	0.679
6	75	35	0.253	0.267	0.293	0.382	0.227	0.421	1.293	0.609	0.892
		70	0.207	0.239	0.278	0.261	0.192	0.366	0.576	0.287	0.415
	150	35	0.225	0.254	0.269	0.363	0.225	0.422	1.486	0.695	0.992
		70	0.193	0.254	0.263	0.254	0.171	0.379	0.797	0.391	0.551

2.5.3 Test Results

In Section 4, we define S_{u_j} and S_{v_j} as the sup-type test statistics. Table 2.5 reports the empirical size and power at the 5% nominal level for the null hypothesis that the slope coefficient is homogeneous across either i or t. The results in DGPs 1 and 3 give the empirical size, and those in DGPs 2 and 4 give the empirical power. As the results in Table 2.5 indicate, our tests have reasonable size despite the fact that they are slightly conservative like most extreme-value based sup-tests in the literature. In terms of power, out tests have superb power in both DGPs across all three quantile indices.

Table 2.4: Frequency of correct rank estimation via the SVT approach

DGP	N	T	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
			$\hat{K}_0$	$\hat{K}_1$	$\hat{K}_2$	$\hat{K}_0$	$\hat{K}_1$	$\hat{K}_2$	$\hat{K}_0$	$\hat{K}_1$	$\hat{K}_2$
1	75	35	1.00	0.996	0.996	1.00	0.999	1.00	1.00	0.999	0.999
		70	1.00	0.994	0.996	1.00	1.00	1.00	1.00	1.00	1.00
	150	35	1.00	0.994	0.995	1.00	1.00	0.999	1.00	1.00	0.999
		70	1.00	0.995	0.996	1.00	0.999	1.00	1.00	0.999	1.00
2	75	35	1.00	0.993	0.999	1.00	0.999	1.00	1.00	1.00	1.00
		70	1.00	0.997	0.998	1.00	1.00	1.00	1.00	1.00	1.00
	150	35	1.00	0.995	0.996	1.00	0.998	1.00	1.00	1.00	1.00
		70	1.00	1.00	1.00	1.00	1.00	0.998	1.00	1.00	1.00
3	75	35	1.00	0.990	0.997	1.00	0.997	0.997	1.00	0.999	0.999
		70	1.00	0.994	0.994	1.00	0.999	0.999	1.00	0.999	0.998
	150	35	1.00	0.999	0.992	1.00	1.00	1.00	1.00	1.00	1.00
		70	1.00	0.996	0.994	1.00	0.998	1.00	1.00	1.00	1.00
4	75	35	1.00	0.992	0.991	1.00	0.999	0.999	1.00	0.999	1.00
		70	1.00	0.995	0.995	1.00	0.999	0.999	1.00	1.00	1.00
	150	35	1.00	0.996	0.997	1.00	0.999	1.00	1.00	1.00	1.00
		70	1.00	0.997	0.999	1.00	0.999	1.00	1.00	1.00	1.00
5	75	35	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	150	35	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	75	35	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	150	35	1.00	1.00	0.999	1.00	1.00	1.00	1.00	1.00	1.00
		70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2.6 shows the empirical size and power of our test for DGPs 5 and 6. The findings are similar to those in Table 2.5. In particular, our tests are a bit conservative under the null. The empirical power tends to 1 quickly as T increases.

2.6 Empirical Study

In this section we consider two empirical applications: the heterogeneous investment equation and the heterogeneous quantile effect of foreign direct investment on unemployment.

2.6.1 Investment Equation

In this subsection, we revisit the investment equation. Fazzari et al. (1988) point out that investment may show sensitivity to movements in cash flow when firms

Table 2.5: Empirical size and power of testing slope homogeneity across either i or t (nominal level: 0.05)

DGP	N	T	$\tau=0.25$				$\tau=0.5$				$\tau=0.75$			
			u_1	v_1	u_2	v_2	u_1	v_1	u_2	v_2	u_1	v_1	u_2	v_2
DGP 1	75	35	0.040	0.051	0.049	0.032	0.024	0.054	0.034	0.054	0.036	0.047	0.036	0.048
		70	0.040	0.055	0.050	0.044	0.020	0.056	0.017	0.068	0.025	0.037	0.029	0.029
	150	35	0.028	0.036	0.058	0.048	0.065	0.054	0.052	0.055	0.074	0.030	0.076	0.024
		70	0.034	0.025	0.030	0.023	0.035	0.048	0.028	0.040	0.035	0.025	0.039	0.025
DGP 2	75	35	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	150	35	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP 3	75	35	0.045	0.057	0.050	0.041	0.022	0.050	0.038	0.089	0.054	0.047	0.049	0.047
		70	0.048	0.031	0.031	0.033	0.028	0.086	0.023	0.069	0.041	0.046	0.032	0.038
	150	35	0.065	0.054	0.058	0.034	0.064	0.051	0.068	0.045	0.084	0.018	0.089	0.023
		70	0.046	0.030	0.044	0.025	0.022	0.037	0.037	0.030	0.046	0.022	0.048	0.015
DGP 4	75	35	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	150	35	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2.6: Empirical size and power for testing additive slopes (nomial level: 0.05)

DGP	N	T	$\tau=0.25$		$\tau=0.50$		$\tau=0.75$	
			size	power	size	power	size	power
DGP 5	75	35	0.027	0.807	0.034	1.00	0.026	0.979
		70	0.065	1.00	0.061	1.00	0.034	1.00
	150	35	0.018	1.00	0.026	1.00	0.011	1.00
		70	0.012	1.00	0.029	1.00	0.010	1.00
DGP 6	75	35	0.039	0.796	0.058	1.00	0.045	0.979
		70	0.024	1.00	0.06	1.00	0.032	1.00
	150	35	0.019	1.00	0.025	1.00	0.022	1.00
		70	0.01	1.00	0.022	1.00	0.014	1.00

face constraints for external finance. Since Fazzari et al. (1988), there has been a large literature on the effect of cash flow on the corporate investment; see Devereux and Schiantarelli (1990), Gilchrist and Himmelberg (1995), Kaplan and Zingales (1995), Cleary (1999), Rauh (2006), and Almeida and Campello (2007), among others. Using the panel dataset, we consider the scaled version of the investment equation as follows:

$$\frac{I_{it}}{K_{i,t-1}} = \Theta_{0,it} + \Theta_{1,it}\frac{CF_{it}}{K_{i,t-1}} + \Theta_{2,it}q_{i,t-1} + u_{it},$$

where I is the corporate investment, CF is the cash flow, q is the Tobin's q, K is the capital stock and u is the innovation. $\Theta_{0,it}$ refers to the fixed effects (FEs). Rather than the mean estimation, Galvao and Wang (2015) estimate the effects of the firm's cash flow and Tobin's q on investment at different quantiles. By using the panel quantile regression with individual FEs, they show that the slope estimates change across τ. However, they do not allow the slope coefficients, Θ_1 and Θ_2, to change either over i or t. Inspired by Galvao and Wang (2015), we estimate the following model

$$\mathscr{Q}_\tau \left(IK_{it} \middle| \{CFK_{it}, q_{i,t-1}\}_{t\in[T]}, \{\Theta_{j,it}\}_{t\in[T],j\in\{0,1,2\}} \right)$$
$$= \Theta_{0,it}(\tau) + \Theta_{1,it}(\tau)CFK_{it} + \Theta_{2,it}(\tau)q_{i,t-1}, \tag{2.14}$$

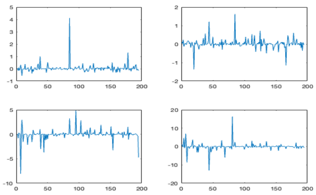
where $IK_{it} = \frac{I_{it}}{K_{i,t-1}}$, and $CFK_{it} = \frac{CF_{it}}{K_{i,t-1}}$. Here we don't restrict the specific structure on the FEs and they can be either additive or interactive.

The data are taken from the China Stock Market & Accounting Research (CSMAR) Database. We use quarterly data for 195 manufacturing firms in China from 2003 to 2020. Based on the model (2.14), we define corporate investment as $I_{it} = LI_{it} - LI_{i,t-1}$, where LI_{it} is the total value of long-term corporate investment as the sum of long-term equity investment, long-term bound investment, fixed assets and immaterial assets. The investment measures the change of firm's total investment compared to the last period. All these four variables can be easily obtained from the balance sheet. We directly use Tobin's q from the CSMAR database, where by definition $q = \frac{MV}{K}$ and MV is the market value of the firm. We obtain a balanced panel dataset with 195 firms and 72 time periods. The units of corporate investment, capital and cash flow are measured by billions of Chinese RMB.

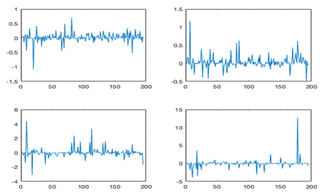
By using the SVT approach, we obtain the estimates of the ranks of Θ_1 and Θ_2 : $\hat{r}_1 = \hat{r}_2 = 1$ for each $\tau = \{0.25, 0.5, 0.75\}$. Consequently, we can consider the test that whether $\Theta_{j,it}$ is constant over i or constant over t for both $j = 1, 2$. Specifically, we want to test whether the effect of cash flow and Tobin's q on the firm's investment is homogeneous over i or across t with market imperfection. That is, for $j \in \{1, 2\}$, we shall test

- H_0^a: $\Theta_{j,it}$ is a constant over i,
- H_0^b: $\Theta_{j,it}$ is a constant over t.

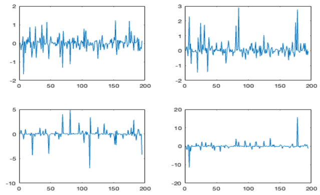
Figure 2.1 shows the estimation results for the factor and factor loadings of two slope coefficient matrices under different quantiles. In each sub-figure, the first and second rows report the results for Θ_1 and Θ_2, respectively. Specifically, the first row



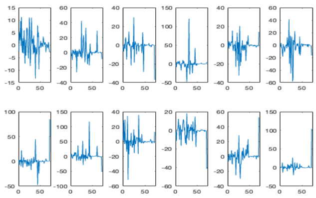
(a) factor loading estimates under $\tau = 0.25$



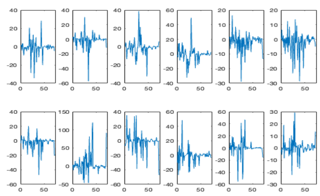
(b) factor loading estimates under $\tau = 0.5$



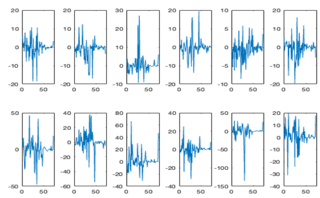
(c) factor loading estimates under $\tau = 0.75$



(d) factor estimates under $\tau = 0.25$



(e) factor estimates under $\tau = 0.5$



(f) factor estimates under $\tau = 0.75$

Figure 2.1: Factor loading and factor estimates under different quantiles

of Figure 2.1(a) gives the plot of $\{\hat{u}_{i,1}\}_{i\in[N]}$ as a catenation of $\{\hat{u}_{i,1}^{(1,2)}, \hat{u}_{i,1}^{(2,3)}, \hat{u}_{i,1}^{(3,1)}\}$ at the left and as a cantenation of $\{\hat{u}_{i,1}^{(1,3)}, \hat{u}_{i,1}^{(2,1)}, \hat{u}_{i,1}^{(3,2)}\}$ at the right in the first row, and similarly the plot of $\{\hat{u}_{i,2}\}_{i\in[N]}$ in the second row. Similarly, the first row of Figure 2.1(d) shows $\{\hat{v}_{t,1}^{(a,b)}\}_{t\in[T]}$ for $a \in [3]$, $b \in [3]\setminus\{a\}$ in the first row and $\{\hat{v}_{t,3}^{(a,b)}\}_{t\in[T]}$ for $a \in [3]$, $b \in [3]\setminus\{a\}$ in the second row.

Table 2.7 reports the test statistics, critical values, and p-values. Tobin's q can measure a firm's investment demand. After controlling the Tobin's q and the intercept FEs, the coefficient of cash flow captures a firm's potential for external investment with the variation of internal finance. It is clear that we can reject the homogeneous hypotheses for both i and t at the 1% significance level for each $\tau \in \{0.25, 0.5, 0.75\}$. This indicates that with high probability, the slope coefficient of both CFK and Tobin's q follow the factor structure with one factor.

The above study shows strong evidence that under imperfect market, the sensitivity of corporate investment to cash flow exhibits both individual heterogeneity and time heterogeneity across quantiles. It implies that neither the usual homogenous panel QR model nor the panel QR model with either cross-section or time heterogeneity alone in the slope coefficients fails to fully capture the unobserved heterogeneity in the investment equation.

Table 2.7: Test results under different quantiles for the investment equation

τ	Test	S	$cv_{\alpha=0.01}$	$cv_{\alpha=0.05}$	$cv_{\alpha=0.1}$	p-value
0.25	u_{CFK}	1.28×10^3	16.94	13.68	12.24	0.00
	u_q	4.16×10^4				0.00
	v_{CFK}	13.85	5.70	4.07	3.35	0.00
	v_q	870.85				0.00
0.50	u_{CFK}	148.28	16.94	13.68	12.24	0.00
	u_q	1.24×10^5				0.00
	v_{CFK}	49.57	5.70	4.07	3.35	0.00
	v_q	138.83				0.00
0.75	u_{CFK}	313.21	16.94	13.68	12.24	0.00
	u_q	2.03×10^4				0.00
	v_{CFK}	31.50	5.70	4.07	3.35	0.00
	v_q	58.29				0.00

Notes: S is the test statistics for the factor or factor loadings under different quantiles, $H_0^a(CFK)$ and $H_0^a(q)$ refer to the hypotheses that the slope of CFK and Tobin'q is homogeneous across i, respectively. $H_0^b(CFK)$ and $H_0^b(q)$ refer to the the hypotheses that the slope of CFK and Tobin'q is homogeneous across t, respectively. $\mathrm{cv}_{\alpha=a}$ is the critical value under the significance level a where a=0.1, 0.05, and 0.01.

2.6.2 Foreign Direct Investment and Unemployment

Investment is one of the major driving forces for economic growth and employment. Among the investment, foreign direct investment (FDI) is an important contributor to the employment. See Craigwell (2006), Aktar et al. (2009), Karlsson et al. (2009), Mucuk and Demirsel (2013), and Strat et al. (2015), among others. Controversially, Mucuk and Demirsel (2013) argue that FDI may have both positive and negative effects on employment. On the one hand, FDI adds to the net capital and creates jobs through forward and backward linkages and multiplier effects in local economy. On the other hand, acquisitions may rely on imports or displacement of existing firms which may result in job loss.

To study the relationship of FDI, economic growth rate and unemployment at the country level, we consider the following panel quantile regression model,

$$\begin{aligned}&\mathcal{Q}_\tau\left(U_{it}\middle|\left\{G_{i,t-1},FDI_{it}\right\}_{t\in[T]},\left\{\Theta_{j,it}\right\}_{t\in[T],j\in\{0,1,2\}}\right)\\&=\Theta_{0,it}(\tau)+\Theta_{1,it}(\tau)G_{i,t-1}+\Theta_{2,it}(\tau)FDI_{it},\end{aligned}$$

where U_{it} is the unemployment rate of country i at year t, $G_{i,t-1}$ is the economic growth measured by the growth of real GDP. $\Theta_{0,it}$ is the FEs of country i and year t, $\Theta_{1,it}$ is the elasticity of the economic growth in the previous year to the unemployment this year, and $\Theta_{2,it}$ is the elasticity of FDI to the unemployment.

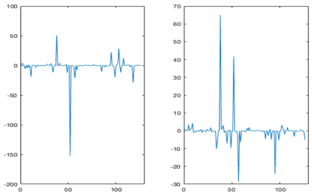
We draw the data for 126 countries from 1992-2019. The data for the unemployment rate are taken from International Labor Organization (ILO) and GDP growth and FDI are from the World Bank Development Indicators (WDI) historical database. The rank estimation procedure shows that $\hat{r}_1 = 2$ and $\hat{r}_2 = 1$. Consequently, we can test whether the elasticity of FDI to the unemployment rate is homogeneous across individual countries and over years 1992-2019, and whether the elasticity of growth rate to unemployment follows the additive structure, i.e.,

- H_0^c: $\Theta_{1,it} = \Theta_{1,i} + \Theta_{1,t}$,
- H_0^d: $\Theta_{2,it}$ is a constant over i,
- H_0^e: $\Theta_{2,it}$ is a constant over t.

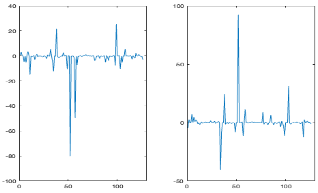
Table 2.8 reports the test results under quantiles 0.25, 0.5 and 0.75 for the above three null hypotheses. Figure 2.2 gives the estimation results for the factor and factor loading estimates of the slope coefficient Θ_2. As Table 2.8 suggests, we can reject all the above three null hypotheses safely at the conventional 5% significance level. This means that the effect of FDI on the unemployment rate is different across both countries and time even though the estimated rank of Θ_2 is one, and the effect of economic growth rate on the unemployment is heterogeneous across both countries and time and it does not exhibit an additive structure.

Table 2.8: Test results under different quantiles

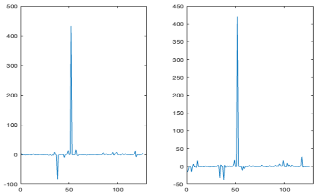
Test	τ	S	$cv_{\alpha=0.01}$	$cv_{\alpha=0.05}$	$cv_{\alpha=0.10}$	p-value
H_0^c	0.25	38.92	35.55	32.29	30.85	0.00
	0.50	80.84	22.29	19.03	17.59	0.00
	0.75	66.24	35.55	32.29	30.85	0.00
H_0^d	0.25	1.41×10^6	16.15	12.89	11.45	0.00
	0.50	6.39×10^6				0.00
	0.75	3.07×10^7				0.00
H_0^e	0.25	36.36	5.70	4.07	3.35	0.00
	0.50	164.03				0.00
	0.75	5.44				0.013



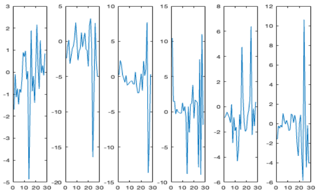
(a) factor loading estimates under $\tau = 0.25$



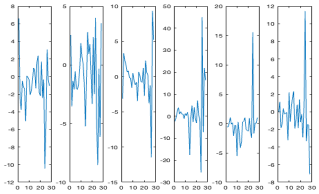
(b) factor loading estimates under $\tau = 0.5$



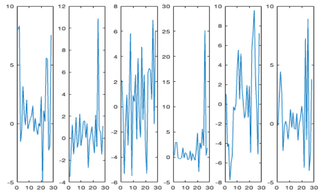
(c) factor loading estimates under $\tau = 0.75$



(d) factor estimates under $\tau = 0.25$



(e) factor estimates under $\tau = 0.5$



(f) factor estimates under $\tau = 0.75$

Figure 2.2: Factor loading and factor estimates of Θ_2 under different quantiles

2.7 Conclusion

This paper considers panel QR model with heterogeneous slopes over both i and t. Compared to Chernozhukov et al. (2019), to remove the bias from the nuclear norm regularization, we split the full sample into three subsamples. We then use the first subsample to compute initial estimators via NNR, the second sample to refine the convergence rate of the initial estimator, and the last subsample to debias the refined estimator. Our asymptotic theory shows that the factor estimates, factor loading estimates and the slope estimates all follow the normal distributions asymptotically. By constructing the consistent estimator for the asymptotic variance, we also conduct two specification tests: (1) the slope coefficient is constant over time or individuals under the case that true rank of slope matrix equals one and (2) the slope coefficient exhibits the additive structure under the case that true rank of the slope coefficient matrix equals two. Our test statistics are shown to follow the Gumbel distribution asymptotically under the null, consistent under the global alternative and have non-trivial power against local alternatives. Monte Carlo simulation and empirical studies illustrate the finite sample performance of our algorithm and test statistics.

Chapter 3

Panel Data Models with Time-Varying Latent Group Structures

3.1 Introduction

Heterogeneous panel data models have been widely used in empirical research in economics because they can capture a rich degree of unobserved heterogeneity. 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 and 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. Klapper and Love (2011), Chu (2012), and Zhang and Cheng (2019) show an exogenous shock like policy implementation has different impacts on different individuals, and Long et al. (2012) argue that 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 (2023) who consider linear panel data models with two-dimensional unobserved heterogeneity in the slope coefficients that are modelled via the usual additive structure, and Chernozhukov et al. (2019) 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. (2023) 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. (2023) 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 propose 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. (2023) 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 the cross section and time dimensions, 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 subsamples. In the fourth step, we follow the lead of Lin and Ng (2012) and Jin et al. (2022) to focus on each subsample 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 memberships 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, both 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 (2023). 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 attentions recently in panel data analyses. NNR has been used in recent researches in econometrics, see Bai and Ng (2019), Moon and Weidner (2018), Chernozhukov et al. (2019), Belloni et al. (2023), Miao et al. (2023), Feng (2019), and Hong et al. (2023), 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. (2023) 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. (2023) allow for at most one-way heterogeneity (individual fixed effects) 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. (2023) 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 3.2 and provide the estimation algorithm in Section 3.3. The asymptotic properties are given in Section

3.4. In Section 3.5, we propose an alternative approach to detect the break point, provide the test statistics for the null that the slope coefficients exhibit no structure change against the alternative with one break point, and discuss the estimation for the model with multiple breaks. In Sections 3.6 and 3.7, we show the finite sample performance of our method by Monte Carlo simulations and an empirical application, respectively. Section 3.8 concludes. All proofs are related to the online appendix.

Notation. Let $\|\cdot\|_{\max}$, $\|\cdot\|_{op}$, $\|\cdot\|$, and $\|\cdot\|_*$ denote the (elementwise) maximum norm, operator norm, Frobenius norm, and nuclear norm, respectively. Let $\odot$ denote the element-wise Hadamard product. $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor and ceiling functions, respectively. Let $a \vee b = \max(a,b)$ and $a \wedge b = \min(a,b)$. $a_n \lesssim b_n$ means $a_n/b_n = O_p(1)$ and $a_n \gg b_n$ means $b_n a_n^{-1} = o(1)$. Let $A = \{A_{it}\}$ be a matrix with its (i,t)-th entry denoted as A_{it}. Let denote $\{A_j\}_{j\in[p]\cup\{0\}}$ be the collection of matrices A_j, $j \in \{0,1,\cdots,p\}$. For a specific $A \in \mathbb{R}^{m\times n}$ with rank n, let $P_A = A(A'A)^{-1}A'$ 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. Let $[n] = \{1,\cdots,n\}$ for any positive integer n. Let $\mathbf{1}\{\cdot\}$ be the usual indicator function. Besides, w.p.a.1 and a.s. abbreviate "with probability approaching 1" and "almost surely", respectively.

3.2 Model Setup

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

$$Y_{it} = \Theta^0_{0,it} + X'_{it}\Theta^0_{it} + e_{it}, \tag{3.1}$$

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 $Y = \{Y_{it}\}$, $X_j = \left\{X_{j,it}\right\}$, $\Theta^0_j = \{\Theta^0_{j,it}\}$ and $E = \{e_{it}\}$, all of which are $N\times T$ matrices. Then we can rewrite (3.1) in terms of matrices as follows:

$$Y = \Theta_0^0 + \sum_{j=1}^{p} X_j \odot \Theta_j^0 + E. \tag{3.2}$$

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

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

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:

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

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 $\mathcal{U}_j^0 \in \mathbb{R}^{N\times r_j}$, $\mathcal{V}_j^0 \in \mathbb{R}^{T\times r_j}$, $\Sigma_j^0 = \operatorname{diag}(\sigma_{1,j},\cdots,\sigma_{r_j,j})$, $U_j^0 = \sqrt{N}\mathcal{U}_j^0\Sigma_j^0$ 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 $\{\Theta_{j,it}^0\}_{i=1}^N$. Let $\mathcal{G}_j^{(\ell)} = \{G_{1,j}^{(\ell)},\cdots,G_{K_j^{(\ell)},j}^{(\ell)}\}$, $\ell = 1,2$, denote the associated latent group structures. Define $N_{k,j}^{(\ell)} = |G_{k,j}^{(\ell)}|$ and $\pi_{k,j}^{(\ell)} = \frac{N_{k,j}^{(\ell)}}{N}$ for $\ell = 1,2$, where $|A|$ denotes the cardinality of set A. Further define $\tau_T := \frac{T_1}{T}$. We show that Θ_j^0 has a low-rank structure in all of the following cases:

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

In this case, $K_j^{(1)} = K_j^{(2)} = 1$, and $\Theta_{j,it}^0 = \alpha_j \; \forall (i,t) \in [N] \times [T]$. Without loss of generality, assume that $\alpha_j > 0$. Then 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 $\Theta_{j,it}^0 = \sum_{k \in [K_j]} \alpha_{k,j} \mathbf{1}\{i \in G_{k,j}\}$ 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 a structural break and a group structure.

(i) $K_j^{(1)} \neq K_j^{(2)}$, where we have different numbers of groups before and after the break;

(ii) $K_j^{(1)} = K_j^{(2)} = K_j$ and $G_{k,j}^{(1)} \neq G_{k,j}^{(2)}$, where we have the same number of groups before and after the break, but the group memberships change 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 $k \in [K_j]$, where 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 3.1. *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) $\Theta_j^0 = \mathscr{U}_j \Sigma_j \mathscr{V}_j' = U_j V_j'$ *where* $U_j = \mathscr{U}_j \Sigma_j / \sqrt{T}$, $V_j = \sqrt{T} \mathscr{V}_j = D_j R_j$, $D_j = \begin{bmatrix} \frac{1}{\sqrt{\tau_T}} \iota_{T_1} & \mathbf{0}_{T_1} \\ \mathbf{0}_{T-T_1} & \frac{1}{\sqrt{1-\tau_T}} \iota_{T-T_1} \end{bmatrix}$ *and* $R_j' R_j = I_2$;

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

By Lemma 3.1 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 3.1(ii) indicates that singular vectors of the slope matrix with rank 2 contain the structural break information.

3.3 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.

3.3.1 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 (NNR). We run the nuclear norm regularized regression and obtain the preliminary estimates as follows:

$$\{\tilde{\Theta}_j\}_{j \in [p] \cup \{0\}} = \underset{\{\Theta_j\}_{j=0}^p}{\arg\min} \frac{1}{NT} \left\| Y - \sum_{j=1}^p X_j \odot \Theta_j - \Theta_0 \right\|^2 + \sum_{j=0}^p \nu_j \left\| \Theta_j \right\|_*, \quad (3.3)$$

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

$$\left\{\dot{u}_{i,j}\right\}_{j\in[p]\cup\{0\}} = \underset{\{u_{i,j}\}_{j\in[p]\cup\{0\}}}{\arg\min} \frac{1}{T}\sum_{t\in[T]}\left(Y_{it} - u_{i,0}'\tilde{v}_{t,0} - \sum_{j=1}^{p} u_{i,j}'\tilde{v}_{t,j}X_{j,it}\right)^2 \quad i \in [N], \tag{3.4}$$

$$\left\{\dot{v}_{t,j}\right\}_{j\in[p]\cup\{0\}} = \underset{\{v_{t,j}\}_{j\in[p]\cup\{0\}}}{\arg\min} \frac{1}{N}\sum_{i\in[N]}\left(Y_{it} - v_{t,0}'\dot{u}_{i,0} - \sum_{j=1}^{p} v_{t,j}'\dot{u}_{i,j}X_{j,it}\right)^2 \quad t \in [T]. \tag{3.5}$$

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

$$\hat{T}_1 = \underset{s\in\{2,\cdots,T-1\}}{\arg\min} \frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\left\{\sum_{t=1}^{s}\left(\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}_{j,i}^{(1s)}\right)^2 + \sum_{t=s+1}^{T}\left(\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}_{j,i}^{(2s)}\right)^2\right\}, \tag{3.6}$$

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 also $\dot{\beta}_i^{(2)} = \frac{1}{\sqrt{\hat{T}_2}}(\dot{\Theta}_{1,i}^{(2)\prime}, \cdots, \dot{\Theta}_{p,i}^{(2)\prime})'$. Let z_ς be some predetermined value which will be specified in the next subsection. Given the subsample before and after the estimated break point, initialize $m = 1$ and classify each subsample into m groups by K-means algorithm with group membership obtained as $\hat{\mathscr{G}}_m^{(\ell)} := \{\hat{G}_{k,m}^{(\ell)}\}_{k\in[m]}$. Next, we construct test statistic $\hat{\Gamma}_m^{(\ell)}$, compare it to z_ς, set $m = m+1$ and go to the next iteration if $\hat{\Gamma}_m^{(\ell)} > z_\varsigma$ and stop the STK algorithm otherwise. At last, define $\hat{K}^{(\ell)} = m$ and $\hat{\mathscr{G}}^{(\ell)} = \hat{\mathscr{G}}_m^{(\ell)}$. In the next subsection, we will present each step of the STK algorithm in detail.

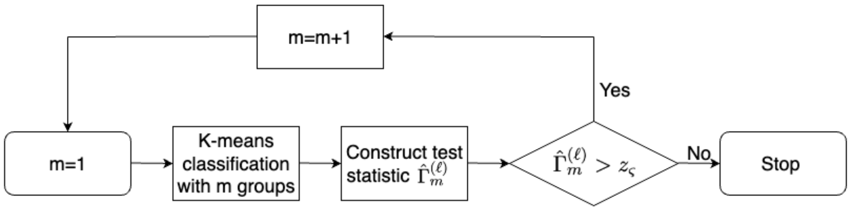


Figure 3.1: The flow chart of STK algorithm

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 NNR 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.

3.3.2 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:

$$\left\{\dot{a}_{k,m}^{(\ell)}\right\}_{k\in[m]} = \underset{\left\{a_{k,m}^{(\ell)}\right\}_{k\in[m]}}{\arg\min}\ \frac{1}{N}\sum_{i\in[N]}\min_{k\in[m]}\left\|\dot{\beta}_i^{(\ell)} - a_{k,m}^{(\ell)}\right\|^2, \tag{3.7}$$

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

$$\hat{g}_{i,m}^{(\ell)} = \underset{k\in[m]}{\arg\min}\left\|\dot{\beta}_i^{(\ell)} - \dot{a}_{k,m}^{(\ell)}\right\| \quad \forall i\in[N]. \tag{3.8}$$

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

Second, we discuss the construction of the test statistic based on the idea of homogeneity test for several subsamples. 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 $\hat{\mathscr{T}}_1 = [\hat{T}_1]$, $\hat{\mathscr{T}}_2 = [T]\backslash[\hat{T}_1]$, $\hat{\mathscr{T}}_{1,-1} = \hat{\mathscr{T}}_1\backslash\{\hat{T}_1\}$, $\hat{\mathscr{T}}_{2,-1} = \hat{\mathscr{T}}_2\backslash\{T\}$, $\hat{\mathscr{T}}_{1,j} = \{1+j,\cdots,\hat{T}_1\}$, and $\hat{\mathscr{T}}_{2,j} = \{\hat{T}_1+1+j,\cdots,T\}$ for some specific $j\in\hat{\mathscr{T}}_{\ell,-1}$. 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:

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

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

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

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

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

$$\hat{S}_{ii,k,m}^{(\ell)} = \frac{1}{\hat{T}_\ell}\hat{X}_i^{(\ell)\prime}M_{\hat{F}_{k,m}^{(\ell)}}\hat{X}_i^{(\ell)}, \quad \hat{a}_{ii,k}^{(\ell)} = \hat{\lambda}_{i,k,m}^{(\ell)\prime}\left(|\hat{G}_{k,m}^{(\ell)}|^{-1}\hat{\Lambda}_{k,m}^{(\ell)\prime}\hat{\Lambda}_{k,m}^{(\ell)}\right)^{-1}\hat{\lambda}_{i,k,m}^{(\ell)}.$$

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:

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

where

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

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

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

We will show 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 subsamples is homogeneous, while it diverges to infinity under the alternative. Let z_ς denote be the critical value at significance level ς, 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 ς if $\hat{\Gamma}_m^{(\ell)} > z_\varsigma$.

3.3.3 Rank Estimation

To obtain the rank estimator, we use the low-rank estimators from (3.3) and estimate r_j by the singular value thresholding (SVT):

$$\hat{r}_j = \sum_{i=1}^{N\wedge T} \mathbf{1}\left\{\sigma_i\left(\tilde{\Theta}_j\right) \geq 0.5\left(\nu_j \left\|\tilde{\Theta}_j\right\|_{op}\right)^{1/2}\right\} \quad \forall j \in \{0\}\cup[p],$$

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. (2019) and that of Theorem 3.2 in Hong et al. (2023), we can show that $\mathbb{P}(\hat{r}_j = r_j) \to 1$ for each j as $(N,T) \to \infty$.

3.3.4 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 $\{\alpha_k^{(\ell)}\}_{k\in[\hat{K}^{(\ell)}]}$ along with the factors and factor loadings as follows:

$$\left(\hat{\Lambda}^{(\ell)},\hat{F}^{(\ell)},\left\{\hat{\alpha}_k^{(\ell)}\right\}_{k\in[\hat{K}^{(\ell)}]}\right)=\arg\min\mathbb{L}\left(\Lambda,F,\left\{a_k^{(\ell)}\right\}_{k\in[\hat{K}^{(\ell)}]}\right) \tag{3.9}$$

where $\mathbb{L}\left(\Lambda,F,\left\{a_k^{(\ell)}\right\}_{k\in[\hat{K}^{(\ell)}]}\right)=\frac{1}{N\hat{T}_\ell}\sum_{k=1}^{\hat{K}^{(\ell)}}\sum_{i\in\hat{G}_k^{(\ell)}}\sum_{t\in\hat{\mathscr{T}}_\ell}\left(Y_{it}-\lambda_i' f_t-X_{it}' a_k^{(\ell)}\right)^2$. Here, we ignore the fact that the prior- and post-break regimes share the same set of factor loadings and estimate the 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:

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

where

$$\begin{aligned}&\mathbb{L}\left(\Lambda,F,\left\{a_k^{(1)}\right\}_{k\in[\hat{K}^{(1)}]},\left\{a_k^{(2)}\right\}_{k\in[\hat{K}^{(2)}]}\right)\\&=\mathbb{L}\left(\Lambda,F,\left\{a_k^{(1)}\right\}_{k\in[\hat{K}^{(1)}]}\right)+\mathbb{L}\left(\Lambda,F,\left\{a_k^{(2)}\right\}_{k\in[\hat{K}^{(2)}]}\right).\end{aligned} \tag{3.10}$$

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 (3.9).

3.4 Asymptotic Theory

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

3.4.1 Basic Assumptions

Define $e_i=(e_{i1},\cdots,e_{iT})'$ and $X_{j,i}=\left(X_{j,i1},\cdots,X_{j,iT}\right)'$. 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 $\{V_j^0\}_{j\in[p]\cup\{0\}}$ as random and their loadings $\{U_j^0\}_{j\in[p]\cup\{0\}}$ as deterministic. Let $\mathscr{D} := \sigma(\{V_j^0\}_{j\in[p]\cup\{0\}})$ denote the minimum σ-field generated by $\{V_j^0\}_{j\in[p]\cup\{0\}}$. Similarly, let

$$\mathscr{G}_t := \sigma(\mathscr{D}, \{X_{is}\}_{i\in[N],s\le t+1}, \{e_{is}\}_{i\in[N],s\le t}).$$

Let $\max_i = \max_{i\in[N]}$, $\max_t = \max_{t\in[T]}$ and $\max_{i,t} = \max_{i\in[N],t\in[T]}$. Let M and C be generic bounded positive constants which may vary across lines.

Assumption 3.1. *(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, $\{(e_{it}, X_{it}), t \ge 1\}$ is strong mixing conditional on $\mathscr{D}$ with the mixing coefficient $\alpha_i(\cdot)$ satisfying $\max_i \alpha_i(z) \le M\alpha^z$ for some constant $\alpha \in (0,1)$.

(iv) There exists a constant $C > 0$ such that $\max_i \frac{1}{T}\sum_{t\in[T]} \|\xi_{it}\|^2 \le C$ a.s. and $\max_t \frac{1}{N}\sum_{i\in[N]} \|\xi_{it}\|^2 \le C$ a.s. for $\xi_{it} = e_{it}$, X_{it} and $X_{it}e_{it}$.

(v) $\max_{i,t} \mathbb{E}[\|\xi_{it}\|^q \,|\mathscr{D}] \le M$ a.s. and $\max_{i,i^*,t} \mathbb{E}[\|X_{it}e_{i^*t}\|^q \,|\mathscr{D}] \le M$ a.s. for some $q > 8$ and $\xi_{it} = e_{it}$, X_{it} and $X_{it}e_{it}$.

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

Assumption 3.1*. *(i), (iv) and (v) are same as Assumption 3.1(i), (iv) and (v). Besides,*

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

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

Assumption 3.1(i) imposes conditional independence on $\{e_{it}, X_{it}\}_{t\in[T]}$ across the cross sectional units. Assumption 3.1(ii) imposes the moment condition. Assumption 3.1(iii) imposes conditional strong mixing conditions along the time dimension. See Prakasa Rao (2009) for the definition of conditional strong mixing and Su and Chen (2013b) for an application in the panel setup. Assumption 3.1(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 3.1(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 3.1* is for the study of dynamic panel data models. To be specific, Assumption 3.1*(ii) requires that the error sequence $\{e_{it}, t \geq 1\}$ be 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 3.1*(iii) imposes the conditional independence of error term across t. The presence of serially correlated errors in dynamic panels typically cause the endogeneity issue, which invalidates the least-squares-based PCA estimation.

Assumption 3.2. *$rank(\Theta_j^0) = r_j \leq \bar{r}$ for $j \in [p] \cup \{0\}$ and some fixed $\bar{r}$, and* $\max_{j \in [p] \cup \{0\}} ||\Theta_j^0||_{\max} \leq M$.

Assumption 3.2 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 3.2. 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 3.3. *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*

$$\infty > C_\sigma \geq \lim \sup_{(N,T) \to \infty} \max_{j \in [p]} \sigma_{1,j} \geq \lim \inf_{(N,T) \to \infty} \min_{j \in [p]} \sigma_{r_j,j} \geq c_\sigma > 0.$$

Assumption 3.3 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.

Consider the SVD for Θ_j^0: $\Theta_j^0 = \mathscr{U}_j \Sigma_j \mathscr{V}_j' \ \forall j \in [p] \cup \{0\}$. Decompose $\mathscr{U}_j = (\mathscr{U}_{j,r}, \mathscr{U}_{j,0})$ and $\mathscr{V}_j = (\mathscr{V}_{j,r}, \mathscr{V}_{j,0})$ with $(\mathscr{U}_{j,r}, \mathscr{V}_{j,r})$ being the singular vectors corresponding to nonzero singular values and $(\mathscr{U}_{j,0}, \mathscr{V}_{j,0})$ being the singular vectors corresponding to zero singular values. Hence, for any matrix $W \in \mathbb{R}^{N\times T}$, we define

$$\mathscr{P}_j^{\perp}(W) = \mathscr{U}_{j,0}\mathscr{U}_{j,0}'W\mathscr{V}_{j,0}\mathscr{V}_{j,0}', \quad \mathscr{P}_j(W) = W - \mathscr{P}_j^{\perp}(W),$$

where $\mathscr{P}_j(W)$ can be seen as the linear projection of matrix W into the low-rank space with $\mathscr{P}_j^{\perp}(W)$ 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:

$$\mathscr{R}(C_1, C_2) := \left\{ (\{\Delta_{\Theta_j}\}_{j\in[p]\cup\{0\}}) : \sum_{j\in[p]\cup\{0\}} \left\| \mathscr{P}_j^{\perp}(\Delta_{\Theta_j}) \right\|_* \le C_1 \sum_{j\in[p]\cup\{0\}} \left\| \mathscr{P}_j(\Delta_{\Theta_j}) \right\|_*, \right.$$
$$\left. \sum_{j\in[p]\cup\{0\}} \left\|\Theta_j\right\|^2 \ge C_2\sqrt{NT} \right\}. \tag{3.11}$$

Lemma A.4 in the online appendix shows that our nuclear norm estimators are in a restricted set larger than (3.11), which deprives of the restriction on the Frobenius norm in the definition of $\mathscr{R}(C_1, C_2)$. Intuitively, the first restriction in 3.11 means the projection onto the orthogonal low-rank space of the estimator error can be controlled by its projection onto the low-rank space. Theorem 3.1 below will greatly rely on this property.

Assumption 3.4. *For any $C_2 > 0$, there are constants C_3 and C_4 such that for any $(\{\Delta_{\Theta_j}\}_{j\in[p]\cup\{0\}}) \in \mathscr{R}(3, C_2)$, we have*

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

Assumption 3.4 imposes the restricted strong convexity (RSC) condition, which is similar to Assumption 3.1 in Chernozhukov et al. (2019). 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 $\phi_{it}^0 = (v_{t,0}^{0\prime}, v_{t,1}^{0\prime}X_{1,it}, \cdots, v_{t,p}^{0\prime}X_{p,it})'$, and $\psi_{it}^0 = (u_{i,0}^{0\prime}, u_{i,1}^{0\prime}X_{1,it}, \cdots, u_{i,p}^{0\prime}X_{p,it})'$.

Assumption 3.5. *There exist constants C_ϕ and c_ϕ such that*

$$\infty > C_\phi \ge \limsup_T \max_{t\in[T]} \lambda_{\max}(\Psi_t) \ge \liminf_T \min_{t\in[T]} \lambda_{\min}(\Psi_t) \ge c_\phi > 0,$$

$$\infty > C_\phi \ge \limsup_N \max_{i\in[N]} \lambda_{\max}(\Phi_i) \ge \liminf_N \min_{i\in[N]} \lambda_{\min}(\Phi_i) \ge c_\phi > 0.$$

Assumption 3.5 is similar to Assumption 8 in Ma et al. (2020) and it imposes some rank conditions.

3.4.2 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 3.1. *Suppose that Assumptions 3.1–3.4 hold. Then $\forall j \in [p] \cup \{0\}$, we have*

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

If in addition Assumption 3.5 is also satisfied, then we have

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

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

Theorem 3.1(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 (conditional) strong mixing processes. Theorem 3.1(ii)–(iii) reports the uniform convergence rate of the factor and factor loading estimators. The extra $(NT)^{1/q}$ term in the definition of $\eta_{N,2}$ is due to the nonboundedness of $X_{j,it}$ in Assumption 3.1(v), and it disappears when $X_{j,it}$ is assumed to be uniformly bounded.

3.4.3 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 3.6. *(i)* $\sqrt{\frac{1}{N}\sum_{i\in[N]}||\alpha_{g_i^{(1)}} - \alpha_{g_i^{(2)}}||^2} = C_5\zeta_{NT}$, *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* $T \to \infty$.

Assumption 3.6(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 structures, 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 3.2. *Suppose Assumptions 3.1–3.6 hold, with the true break point being* T_1 *and the estimator defined in (3.6). Then* $\mathbb{P}(\hat{T}_1 = T_1) \to 1$ *as* $(N,T) \to \infty$.

Theorem 3.2 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 a certain rate.

3.4.4 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 3.1. *Fix $K^{(\ell)} > 1$ and $m \le K^{(\ell)}$. The estimated group structure $\hat{\mathscr{G}}_m^{(\ell)}$ 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 3.1 describes the non-splitting property introduced by Jin et al. (2022). The latter authors show that the STK algorithm yields the estimated group structures enjoying the NSP.

To proceed, we add following assumptions.

Assumption 3.7. *(i) Let k_s and k_{s^*} be different group indices. Assume that $\min_{1\le k_s<k_{s^*}\le K^{(\ell)}} \left\|\alpha_{k_s}^{(\ell)} - \alpha_{k_{s^*}}^{(\ell)}\right\|_2 \ge C_5$ for $\ell \in \{1,2\}$.*

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

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

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

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

Remark 3. Assumption 3.7(i)–(ii) is the standard assumption for K-means algorithm, which is comparable to Assumption 4 in Su et al. (2020) and greatly facilitates the subsequent analyses. Assumption 3.7(i) assumes that the minimum distance of two distinct groups is bounded away from 0, and Assumption 3.7(ii) imposes that each group has asymptotically non-negligible number of units. For Assumption 3.7(iii), it can be shown to hold under mild conditions. Below we explain this assumption in detail. When $n = 2$, it's clear that

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

by Assumptions 3.6(ii) and 3.7(i)–(ii). When $n > 2$, we consider a special case such that $S_s =: ||\sum_{s^*\in[n],s^*\neq s} (\alpha_{k_{s^*}}^{(\ell)} - \alpha_{k_s}^{(\ell)})|| = 0$ for some specific $s = s_0 \in [n]$. Then it is easy to see S_s is non-zero for all $s \in [n]\backslash\{s_0\}$ under Assumption 3.7(i). Hence, if we assume S_s is lower bounded by a constant c for any $s \in [n] \backslash \{s_0\}$, Assumption 3.7(iii) will holds naturally. Similar arguments hold for the other general cases.

Assumption 3.8. *Let $\mathscr{T}_1 = [T_1]$ and $\mathscr{T}_2 = [T] \backslash [T_1]$.*

(i) $\frac{1}{T_\ell} \sum_{t \in \mathscr{T}_\ell} f_t^0 f_t^{0\prime} \xrightarrow{p} \Sigma_F^{(\ell)} > 0$ *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* $\max_i \frac{1}{T_\ell} \sum_{t \in \mathscr{T}_\ell} \|\xi_{it}\|^2 \leq C$ *a.s. for* $\xi_{it} = e_{it}$, X_{it} *and* $X_{it} e_{it}$.

Assumption 3.8(i) imposes some standard assumptions on the factors and factor loadings. Assumption 3.8(ii) is similar as Assumption 3.1(iv), which strengthen the Assumption 3.1(iv) to hold for two time regimes.

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

Theorem 3.3. *Fix* $\varsigma = \varsigma_N \in (0,1)$. *Suppose that Assumption 3.1* and Assumptions 3.2–3.8 hold. Then for* $\ell \in \{1,2\}$, *we have*

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

(a) $\max_{i \in [N]} \mathbf{1}\{\hat{g}_{i,K^{(\ell)}}^{(\ell)} \neq g_i^{(\ell)}\} = 0$ *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) $\mathbb{P}(\hat{K}^{(\ell)} \leq K^{(\ell)}) \geq 1 - \varsigma + o(1)$,

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

Theorem 3.3 studies the asymptotic properties of the STK algorithm. Since we allow $\varsigma = \varsigma_N$ to shrink to zero at rate N^{-c}, the critical value z_ς diverges to infinity at rate $\log N$ as $N \to \infty$ by the tail properties of $\chi^2(1)$ random variables. 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 the $(m+1)$-th iteration. At iteration m with $m = K^{(\ell)}$, if we set $\varsigma_N \to 0$ as $N \to \infty$, by the joint distribution function for $K^{(\ell)}$ independent $\chi^2(1)$ random variables, we can easily find that $z_\varsigma \to \infty$ while the test statistics $\hat{\Gamma}_m^{(\ell)}$ is stochastically bounded. As a result, the iteration stops w.p.a.1 and we have $\mathbb{P}(\hat{K}^{(\ell)}=K^{(\ell)})\to 1$ provided that $\varsigma=\varsigma_N\to 0$ as $N\to\infty$. As aforementioned, Theorem 3.3 ensures the application of K-means algorithm only for the under-fitting and just-fitting cases and it avoids the theoretical challenge in handling the over-fitting case in the classification.

For dynamic panels, we can focus on Assumption 3.1*, where the error term is a martingale difference sequence (m.d.s.). Under this assumption, the HAC estimator $\hat{\Omega}^{(\ell)}_{i,k,m}$ degenerates to $\frac{1}{\hat{T}_\ell}\sum_{t\in\hat{\mathscr{T}}_\ell}\hat{z}^{(\ell)}_{it}\hat{z}^{(\ell)\prime}_{it}\hat{e}^2_{it}$. For static panel, we typically allow for serially correlated errors and employ the HAC estimator, and the results in Theorem 3.3 continue to hold with Assumption 3.1* replaced by Assumption 3.1. For the kernel function and bandwidth, we can follow Andrews (1991) and let $k(\cdot)$ belongs to the following class of kernels:

$$\mathscr{K}=\Big\{k(\cdot):\mathbb{R}\mapsto[-1,1]\mid k(0)=1,\ k(u)=k(-u),\ \int|k(u)|du<\infty,$$
$$k(\cdot)\text{ is continuous at 0 and at all but a finite number of other points}\Big\}.$$

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

3.4.5 Distribution Theory for the Group-specific Slope Estimators

For $\ell\in\{1,2\}$, let $\{\hat{\alpha}^{*(\ell)}_k\}_{k\in K^{(\ell)}}$ 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 $\{\hat{\alpha}^{(\ell)}_k\}_{k\in K^{(\ell)}}$, $\ell\in\{1,2\}$, we only need to show that for the oracle estimators $\{\hat{\alpha}^{*(\ell)}_k\}_{k\in K^{(\ell)}}$ based on Theorems 3.2 and 3.3 by extending the result of Bai (2009) and Moon and Weidner (2017).

To proceed, we add some notation. For $\ell\in\{1,2\}$, we first define the matrix notation for each subgroup. For $j\in[p]$, let $X^{(1)}_{j,i}=\left(X_{j,i1},\cdots,X_{j,iT_1}\right)'$, $X^{(2)}_{j,i}=\left(X_{j,i(T_1+1)},\cdots,X_{j,iT}\right)'$, $e^{(1)}_i=(e_{i1},\cdots,e_{iT_1})'$ and $e^{(2)}_i=\left(e_{i(T_1+1)},\cdots,e_{iT}\right)'$. Then we use $\mathbb{X}^{(\ell)}_{j,k}\in\mathbb{R}^{N^{(\ell)}_k\times T_\ell}$ and $E^{(\ell)}_k\in\mathbb{R}^{N^{(\ell)}_k\times T_\ell}$ to denote the regressor and error matrix for subgroup $k\in[K^{(\ell)}]$ with each row being $X^{(\ell)}_{j,i}$ and $e^{(\ell)}_i$, respectively. Let $\mathscr{X}^{(\ell)}_{j,k}=M_{\Lambda^{0,(\ell)}_k}\mathbb{X}^{(\ell)}_{j,k}M_{F^{0,(\ell)}}\in\mathbb{R}^{N^{(\ell)}_k\times T_\ell}$ with the (i,t)th entry given by $\mathscr{X}^{(\ell)}_{j,k,it}$. Let $\mathscr{X}^{(\ell)}_{k,it}=(\mathscr{X}^{(\ell)}_{1,k,it},\cdots,\mathscr{X}^{(\ell)}_{p,k,it})'$. Further define

$$\mathbb{B}^{(\ell)}_{NT,1,j,k}=\frac{1}{N^{(\ell)}_k}tr\left[P_{F^{0,(\ell)}}\mathbb{E}\left(E^{(\ell)\prime}_k\mathbb{X}^{(\ell)}_{j,k}|\mathscr{D}\right)\right],$$

$$\mathbb{B}_{NT,2,j,k}^{(\ell)} = \frac{1}{T_\ell} tr\left[\mathbb{E}\left(E_k^{(\ell)} E_k^{(\ell)\prime} | \mathscr{D}\right) M_{\Lambda_k^{0,(\ell)}} \mathbb{X}_{j,k}^{(\ell)} F^{0,(\ell)} \left(F^{0,(\ell)\prime} F^{0,(\ell)}\right)^{-1} \left(\Lambda_k^{0,(\ell)\prime} \Lambda_k^{0,(\ell)}\right)^{-1} \Lambda_k^{0,(\ell)\prime}\right],$$

$$\mathbb{B}_{NT,3,j,k}^{(\ell)} = \frac{1}{N_k^{(\ell)}} tr\left[\mathbb{E}\left(E_k^{(\ell)} E_k^{(\ell)\prime} | \mathscr{D}\right) M_{F^{0,(\ell)}} \mathbb{X}_{j,k}^{(\ell)} \Lambda_k^{0,(\ell)} \left(\Lambda_k^{0,(\ell)\prime} \Lambda_k^{0,(\ell)}\right)^{-1} \left(F^{0,(\ell)\prime} F^{0,(\ell)}\right)^{-1} F^{0,(\ell)\prime}\right],$$

$$\mathbb{B}_{NT,m,k}^{(\ell)} = \left(\mathbb{B}_{NT,m,1,k}^{(\ell)}, \cdots, \mathbb{B}_{NT,m,p,k}^{(\ell)}\right)', \quad \forall m \in \{1,2,3\},$$

$$\Omega_k^{(\ell)} = \frac{1}{N_k^{(\ell)} T_\ell} \sum_{i \in G_k^{(\ell)}} \sum_{t \in \mathscr{T}_\ell} e_{it}^2 \mathscr{X}_{k,it}^{(\ell)} \mathscr{X}_{k,it}^{(\ell)\prime}.$$

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

$$\mathbb{B}_{NT,k}^{(\ell)} = -\rho_k^{(\ell)} \mathbb{B}_{NT,1,k}^{(\ell)} - (\rho_k^{(\ell)})^{-1} \mathbb{B}_{NT,2,k}^{(\ell)} - \rho_k^{(\ell)} \mathbb{B}_{NT,3,k}^{(\ell)}$$

where $\rho_k^{(\ell)} = \sqrt{\frac{N_k^{(\ell)}}{T_\ell}}$. To state the main result in this subsection, we add the following assumption.

Assumption 3.9. *(i) As $(N,T) \to \infty$, $T(\log T) N^{-4/3} \to 0$.*

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

(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 $b = \left(b_{p_1+1}, \cdots, b_p\right)'$. There exists a positive constant C_b such that

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

(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 $w_k^{(\ell)} = (w_{1,k}^{(\ell)}, \cdots, w_{p_1,k}^{(\ell)}) \in \mathbb{R}^{N \times p_1}$, $v^{(\ell)} = (v_1^{(\ell)}, \cdots, v_{p_1}^{(\ell)}) \in \mathbb{R}^{T_\ell \times p_1}$, $M_{w_k^{(\ell)}} = I_{N_k^{(\ell)}} - w_k^{(\ell)} (w_k^{(\ell)\prime} w_k^{(\ell)})^{-1} w_k^{(\ell)\prime}$ and $M_{v^{(\ell)}} = I_{T_\ell} - v^{(\ell)} \left(v^{(\ell)\prime} v^{(\ell)}\right)^{-1} v^{(\ell)\prime}$. There exists a positive constant C_B such that $(N_k^{(\ell)})^{-1} \Lambda_k^{0,(\ell)\prime} M_{w_k^{(\ell)}} \Lambda_k^{0,(\ell)} > C_B I_{r_0}$ 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 $\forall j \in [p]$, $\ell \in \{1,2\}$, *and* $k \in K^{(\ell)}$,

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

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

Assumption 3.9 imposes some conditions to derive the asymptotic distribution theory for the panel model with IFEs which allows for dynamics. Assumption 3.9(i) strengthens Assumption 3.1(vi) a bit. Assumption 3.9(ii) is the standard non-collinearity condition for regressors, which is analogous to Assumption 4(i) in Moon and Weidner (2017). Assumption 3.9(iii) is the identification assumption which is comparable to Assumption 4 in Moon and Weidner (2017). With the conditional strong mixing condition in Assumption 3.1(iii), we can verify Assumption 3.9(iv).

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

Theorem 3.4. *Suppose that Assumption 3.1 or 3.1* and Assumptions 3.2–3.9 hold. For* $\ell \in \{1,2\}$, *the estimators* $\{\hat{\alpha}_k^{(\ell)}\}_{k \in K^{(\ell)}}$ *are asymptotically equivalent to the oracle estimators* $\{\hat{\alpha}_k^{*(\ell)}\}_{k \in K^{(\ell)}}$, *and we have*

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

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* $\Omega^{(\ell)} = diag\left(\Omega_1^{(\ell)}, \cdots, \Omega_{K^{(\ell)}}^{(\ell)}\right)$.

Theorem 3.4 establishes the asymptotic distribution for the estimators of the group-specific slope coefficients before and after the break. It shows that the parameter estimators from our algorithm enjoy the oracle property given the results in Theorems 3.2 and 3.3. The proof of the above theorem can be done by following Moon and Weidner (2017) and Lu and Su (2016).

3.5 Alternatives and Extensions

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

3.5.1 Alternative for Break Point Detection

The algorithm proposed in Section 3.3 uses low-rank estimates of Θ_j^0 to find the break point estimates. However, by Lemma 3.1(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 where the maximum rank of the slope matrix in the model is 2. Let $\dot{v}_{t,j}^* := \frac{\dot{v}_{t,j}}{\|\dot{v}_{t,j}\|}$ and $\dot{v}_t^* := \left(\dot{v}_{t,1}^{*\prime}, \cdots, \dot{v}_{t,p}^{*\prime}\right)^{\prime}$, with the true values being $v_{t,j}^* := \frac{O_j v_{t,j}^0}{\|O_j v_{t,j}^0\|}$ and $v_t^* := \left(v_{t,1}^{*\prime}, \cdots, v_{t,p}^{*\prime}\right)^{\prime}$, respectively. Then Step 3 can be replaced by Step 3* below:

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

$$\tilde{T}_1 = \underset{s \in \{2, \cdots, T-1\}}{\arg\min} \frac{1}{T} \left\{ \sum_{t=1}^{s} \left\| \dot{v}_t^* - \bar{\dot{v}}^{*(1)s} \right\|^2 + \sum_{t=s+1}^{T} \left\| \dot{v}_t^* - \bar{\dot{v}}^{*(2)s} \right\|^2 \right\}, \tag{3.12}$$

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 3.5. *Suppose that Assumptions 3.1–3.5 hold. Then* $\max_t \|\dot{v}_t^* - v_t^*\| = O_p(\eta_{N,2})$.

Theorem 3.6. *Suppose that Assumptions 3.1–3.6 hold. Then* $\mathbb{P}(\tilde{T}_1 = T_1) \to 1$ *as* $(N, T) \to \infty$.

Since the singular vectors of the slope matrices contain the structural change information, Theorem 3.5 indicates that we can consistently estimate the break point by using the factor estimates instead of the slope matrix estimates in (3.6). Given Theorem 3.5 and Lemma 3.1(iii), we can prove Theorem 3.6 with arguments analogous to those used in the proof of Theorem 3.2.

3.5.2 Test for the Presence of a Structural Break

In Section 3.2, 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 3.2.

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

$$\begin{aligned} H_0 &: \; c_{it} = 0 \;\text{ for all } i \in [N], \quad \text{and} \\ H_1 &: \; c_{it} \neq 0 \;\text{ for some } i \in [N]. \end{aligned} \tag{3.13}$$

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

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

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 $\sup_{T_1\in\mathscr{T}_\varepsilon} F_i(T_1)$ is associated with the p-vector of Wiener processes on $[0,1]$, based on which one can derive 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 $\{\alpha_j^{(1)}(T_1)\}_{j\in K^{(1)}}$ prior to the potential break point T_1 and those of the group-specific parameters $\{\alpha_j^{(2)}(T_1)\}_{j\in K^{(2)}}$ after the potential break point T_1.

[1]See Section B.3 in the appendix for the detail of the PCA estimation in linear panels with IFEs.

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 $T_1 \in \mathscr{T}_\varepsilon$. As one can imagine, this approach is also quite 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 (2013b) and Su and Wang (2020). We leave this for future research.

3.5.3 The Case of Multiple Breaks

In Section 3.2, 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 3.2 to allow for multiple breaks. In this case, we have

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

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 (3.6). Second, for each regime before and after the estimated break point, we test the hypothesis in (3.13) 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 subsamples. At the end, we can obtain the break point estimates $\{\hat{T}_a\}_{a\in[\hat{b}]}$ where $\hat{b}$ is the estimated number of breaks. We conjecture that we can establish the consistency of $\hat{b}$ and $\{\hat{T}_a\}$.

After we obtain the estimated number of breaks and break points, for each subsample

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

$a \in [\hat{b}+1]$ with $\hat{T}_0 := 0$ and $\hat{T}_{\hat{b}+1} := T$, we can continue Step 4 in the estimation algorithm in Section 3.3 to obtain the estimated group structure for each subsample.

3.6 Monte Carlo Simulations

In this section, we show the simulation results for the low-rank estimates, break point estimates, group membership estimates and the group number estimates with 1000 replications, and we choose the tuning parameter ν_j by the similar procedure described in Chernozhukov et al. (2019). 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 $\Theta_{it} = (\Theta_{1,it}, \Theta_{2,it})'$.

3.6.1 Data Generating Process (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 $e_{it} \sim i.i.d.\ \mathbb{N}(0,1)$. 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 $e_{it} \sim i.i.d.\ \mathbb{N}(0,\sigma_{it}^2)$ such that $\sigma_{it}^2 \sim i.i.d.\ U(0.5,1)$. 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 $\eta_{it} \sim i.i.d.\ \mathbb{N}(0,1)$ 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 $Y_{i,0} \sim i.i.d.\ \mathbb{N}(0,1)$. $X_{2,it} \sim i.i.d.\ U(-2,2)$, and $e_{it} \sim i.i.d.\ \mathbb{N}(0,0.5)$.

For each DGP above, λ_i and f_t are each $i.i.d.\ \mathbb{N}(0,1)$ 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 DGPs 1.1, 2.1, and 3.1,

$$\Theta_{1,it} = \Theta_{2,it} = \begin{cases} 0.1, & i \in G_1,\ t \in \{1, \cdots, T_1\}, \\ 0.9, & i \in G_2,\ t \in \{1, \cdots, T_1\}, \\ 0.05, & i \in G_1,\ t \in \{T_1+1, \cdots, T\}, \\ 0.45, & i \in G_2,\ t \in \{T_1+1, \cdots, T\}. \end{cases}$$

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

$$\Theta_{1,it} = \begin{cases} 0.1, & i \in G_1,\ t \in \{1, \cdots, T_1\}, \\ 0.7, & i \in G_2,\ t \in \{1, \cdots, T_1\}, \\ 0.05, & i \in G_1,\ t \in \{T_1+1, \cdots, T\}, \\ 0.35, & i \in G_2,\ t \in \{T_1+1, \cdots, T\}. \end{cases}$$

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 as in DGP X.3 for X$\in \{1,2,3\}$, and

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

3.6.2 Results

Table 3.1 reports the proportion of correct rank estimation for the intercept (IFE) and slope matrices based on the SVT in Section 3.3. Note that r_0 denotes the true rank of the intercept matrix and r_1 and r_2 denote that of the two slope matrices. From Table 3.1, we notice that the true ranks of both the intercept and slope matrices can be almost perfectly estimated for the sample sizes under investigation.

Table 3.2 reports the results for the break point estimation in Step 3 based on different (N, T) combinations. We summarize some important findings from Table 3.2. First, when the group membership and the number of groups do not change as in DGP X.1 for X$\in [3]$, the frequency of correct break point estimation may not be 1 especially if N is not large. This suggests that the binary segmentation does not work perfectly in such a scenario. Second, the change of group membership or the number of groups help to identify the break point as reflected in the simulation results for DGP X.2 and X.3 for X$\in [4]$. In general, the binary segmentation works well in our setting.

Table 3.3 reports the results for the group membership estimation when the number of groups are either known (infeasible in practice) or estimated from the data (feasible). With known number of groups, the STK algorithm degenerates to the traditional K-means algorithm. The "Infeasible" part of Table 3.3 reports 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, the K-means classification exhibits excellent performance in this case. Nevertheless, 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 frequency of correct estimation of the group membership and that the number of groups are shown in the "Feasible" part in Table 3.3 and in Table 3.4, respectively. As expected, the performance of the STK algorithm is slightly worse than that of the K-means algorithm with true number of groups. But the performance improves when both N and T increase. Table 3.4 suggests that the number of groups can be nearly perfectly estimated in DGPs 1.1, 1.2, 1.3 and 2.1. For the more complicated DGPs (e.g., the dynamic case in DGPs 4.1, 4.2, and 4.3 or the static panel with serially correlated errors in DGPs 3.1, 3.2, and 3.3), the performance is not as good as that in the simple DGPs.

Table 3.5 presents more detailed 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 serially 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 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 a relatively large value of T is required in order for the HAC estimates to be reasonably well behaved in finite samples.

Table 3.6 shows the result for the post-classification estimator for the first slope coefficient. We follow Su et al. (2016) to define the evaluation criteria as bias and coverage. Specifically, we define the bias to be the weighted versions of bias for slope estimator from all estimated groups, i.e. $\text{Bias} = \sum_{k=1}^{K^{(1)}} \text{Bias}(\alpha_{k,1}^{(\ell)})$ for $\ell \in \{1,2\}$. Similarly, we define the weighted version of coverage ratio of the 95% confidence interval estimators. The "Infeasible" panel shows the result assuming the number of groups information is known, and the "Feasible" panel shows the result without knowing the number of groups information by STK algorithm. From Table 3.6, we notice that the coverage ratio for DGP 1 and 2 is close to 95% under different combination of N and T for both the "Infeasible" and "Feasible" panels, which is owing to the higher correct classification ratio. For DGP 3 and 4, by using the STK algorithm, although the coverage ratio is a bit lower when T equals 100, which is due to inaccuracy of the number of groups and group membership estimator, the coverage ratio becomes better with T approaching 100.

3.7 Empirical Study

In this section, we apply the proposed estimation methods to analyze the time-varying latent group structure of real house price changes at Metropolitan Statistical Areas (MSAs) in the United States. The studies of the U.S. house price changes are plentiful in the literature. Malpezzi (1999), Capozza et al. (2002), Gallin (2006), and Ortalo-Magne and Rady (2006) show that the house price changes are closely correlated with the real income in the long run. Su et al. (2023) consider a heterogenous spatial panel and show that real income growth affects the U.S. house prices in different ways for different MSAs. In this application, we examine whether there exist latent group structures for the real income growth elasticity of house price changes and whether they change over the time dimension.

3.7.1 Model

We consider the panel model with IFEs and two-way slope heterogeneity as in (3.14):

$$\pi_{it} = \lambda_i' f_t + \Theta_{1,it} ginc_{it} + \Theta_{2,it} ginc_{i,t-1} + e_{it}, \tag{3.14}$$

where the dependent variable π_{it} measures the percentage of real house price growth for MSA i at time period t. λ_i and f_t are the individual fixed effects and time fixed effects, respectively, the covariate $ginc_{it}$ denotes the percentage of income

Table 3.1: 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
	$r_1 = 1$	1.00	1.00	1.00	1.00		$r_1 = 1$	1.00	1.00	1.00	1.00
	$r_2 = 1$	1.00	1.00	1.00	1.00		$r_2 = 1$	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
	$r_1 = 1$	1.00	1.00	1.00	1.00		$r_1 = 1$	1.00	1.00	1.00	1.00
	$r_2 = 1$	1.00	1.00	1.00	1.00		$r_2 = 1$	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		$r_1 = 1$	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		$r_1 = 1$	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 3.2: 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	0.980	0.993	1.00	1.00	DGP 3.1	0.985	0.972	1.00	0.999
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	0.998	0.999	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 3.3: Frequency of correct group membership estimation

	N		100		200			N		100		200	
	T		100	200	100	200		T		100	200	100	200
	DGP 1.1	G_B	1.00	1.00	1.00	1.00		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	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 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
Infeasible	DGP 2.3	G_B	1.00	1.00	1.00	1.00	Feasible	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	0.981	0.999	0.949	0.999
		G_A	1.00	1.00	1.00	1.00			G_A	0.981	0.993	0.979	0.996
	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	0.999	0.984	0.999
		G_A	1.00	1.00	1.00	1.00			G_A	0.985	0.998	0.949	0.997
	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 3.4: 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	$K^{(1)}=2$	1.00	1.00	0.999	1.00	DGP 3.1	$K^{(1)}=2$	0.880	0.993	0.675	0.985
	$K^{(2)}=2$	1.00	1.00	1.00	1.00		$K^{(2)}=2$	0.890	0.960	0.873	0.971
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	$K^{(1)}=2$	1.00	1.00	1.00	1.00	DGP 4.1	$K^{(1)}=2$	0.807	0.981	0.825	0.982
	$K^{(2)}=2$	1.00	1.00	1.00	1.00		$K^{(2)}=2$	0.919	0.988	0.714	0.980
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 3.5: Determination of the number of groups

DGP	N	T	$\hat{K}^{(1)}$				$\hat{K}^{(2)}$			
			2	3	4	≥5	2	3	4	≥5
DGP 1.1	100	100	**1.00**	0	0	0	**1.00**	0	0	0
		200	**1.00**	0	0	0	**1.00**	0	0	0
	200	100	**0.999**	0.001	0	0	**1.00**	0	0	0
		200	**1.00**	0	0	0	**1.00**	0	0	0
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	**1.00**	0.00	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	**0.880**	0.097	0.022	0.001	**0.890**	0.062	0.031	0.017
		200	**0.993**	0.007	0	0	**0.960**	0.019	0.012	0.009
	200	100	**0.675**	0.224	0.099	0.002	**0.873**	0.081	0.041	0.005
		200	**0.985**	0.015	0	0	**0.971**	0.023	0.005	0.001
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	**0.807**	0.084	0.089	0.02	**0.919**	0.041	0.019	0.021
		200	**0.981**	0.013	0.004	0.002	**0.988**	0.004	0.005	0.003
	200	100	**0.825**	0.107	0.061	0.007	**0.714**	0.118	0.084	0.084
		200	**0.982**	0.011	0.006	0.001	**0.98**	0.010	0.004	0.006
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

Table 3.6: Point estimation of $\alpha^{(1)}_{\cdot,1}$ and $\alpha^{(2)}_{\cdot,1}$

DGP	N	T	Infeasible				Feasible			
			Before the break		After the break		Before the break		After the break	
			Bias($\times 10^{-6}$)	Coverage	Bias($\times 10^{-6}$)	Coverage	Bias($\times 10^{-6}$)	Coverage	Bias($\times 10^{-6}$)	Coverage
1.1	100	100	2.585	0.951	-2.869	0.946	2.585	0.951	-2.869	0.946
		200	-1.944	0.944	-8.920	0.945	-1.958	0.944	-8.920	0.945
	200	100	-1.096	0.943	1.407	0.947	-1.096	0.943	1.407	0.947
		200	-1.910	0.945	0.960	0.947	-1.910	0.945	0.960	0.947
1.2	100	100	-1.050	0.949	-27.398	0.941	-1.050	0.949	-27.398	0.941
		200	-5.449	0.930	7.616	0.953	-5.449	0.930	7.655	0.953
	200	100	4.770	0.949	1.866	0.951	4.770	0.949	1.866	0.951
		200	1.317	0.941	1.874	0.945	1.317	0.941	1.874	0.945
1.3	100	100	-0.961	0.943	11.417	0.944	-1.050	0.949	-27.398	0.941
		200	-4.213	0.951	-5.002	0.941	-5.449	0.930	7.655	0.953
	200	100	-1.571	0.938	-3.756	0.938	4.770	0.949	1.866	0.951
		200	0.403	0.941	-4.159	0.945	1.317	0.941	1.874	0.945
2.1	100	100	14.840	0.944	9.410	0.950	14.816	0.943	9.406	0.950
		200	-7.222	0.951	1.795	0.951	-7.222	0.951	1.795	0.951
	200	100	0.916	0.940	3.575	0.948	0.916	0.940	3.575	0.948
		200	0.452	0.948	-0.797	0.947	0.452	0.948	-0.797	0.947
2.2	100	100	-21.379	0.946	0.234	0.937	-21.379	0.946	0.234	0.937
		200	0.264	0.942	-15.542	0.953	0.264	0.942	-15.542	0.953
	200	100	-1.379	0.945	-1.489	0.951	-1.379	0.944	-1.489	0.951
		200	-1.101	0.950	1.127	0.949	-1.101	0.950	1.127	0.949
2.3	100	100	-8.610	0.945	5.254	0.952	-8.610	0.945	5.261	0.952
		200	0.927	0.949	5.840	0.949	0.927	0.949	5.840	0.949
	200	100	-1.560	0.943	-2.569	0.941	-1.560	0.943	-2.569	0.941
		200	-0.775	0.947	4.408	0.947	-0.775	0.947	4.386	0.947
3.1	100	100	-20.928	0.955	-73.947	0.945	-26.250	0.927	-77.613	0.920
		200	3.066	0.949	-12.443	0.937	2.884	0.940	-13.116	0.934
	200	100	-2.663	0.951	-8.742	0.944	-3.517	0.857	-7.730	0.888
		200	-3.747	0.949	-2.107	0.945	-3.642	0.939	-1.971	0.938
3.2	100	100	-55.980	0.952	-10.846	0.943	-58.714	0.926	-15.109	0.863
		200	-2.774	0.950	4.690	0.946	-3.218	0.945	4.913	0.942
	200	100	6.979	0.951	8.879	0.945	6.287	0.858	6.894	0.848
		200	-1.704	0.947	0.438	0.945	-2.122	0.928	0.381	0.940
3.3	100	100	-25.340	0.950	37.639	0.907	-29.905	0.924	37.016	0.890
		200	2.042	0.947	-6.431	0.960	1.667	0.940	-6.245	0.960
	200	100	-2.391	0.946	14.364	0.892	-2.735	0.891	13.680	0.840
		200	4.339	0.943	4.890	0.942	4.493	0.932	5.113	0.938
4.1	100	100	800.620	0.930	-466.590	0.929	777.650	0.928	-454.980	0.924
		200	126.760	0.931	550.210	0.942	126.220	0.942	548.160	0.943
	200	100	-224.960	0.931	-339.020	0.939	-214.900	0.904	-313.430	0.876
		200	417.320	0.938	412.500	0.947	415.110	0.941	410.700	0.944
4.2	100	100	1246.000	0.921	726.940	0.943	1205.600	0.918	709.260	0.903
		200	-440.880	0.943	83.433	0.944	-436.670	0.943	81.585	0.951
	200	100	-1600.500	0.930	1937.500	0.927	-1538.300	0.901	1781.600	0.819
		200	-1513.300	0.950	-272.500	0.946	-1502.000	0.935	-271.420	0.953
4.3	100	100	-2067.900	0.931	-505.100	0.940	-1951.700	0.866	-491.200	0.929
		200	317.360	0.946	411.770	0.945	316.550	0.951	407.820	0.935
	200	100	1279.900	0.930	3660.100	0.888	1246.900	0.906	3355.100	0.874
		200	-772.380	0.940	-335.250	0.948	-768.870	0.932	-334.190	0.940

growth for MSA i at time period t, and $ginc_{i,t-1}$ is the lagged value of $ginc_{it}$. Unlike Aquaro et al. (2021) and Su et al. (2023) who consider individual fixed effects and additive two-way fixed effects, respectively, we allow the model to have IFEs. In the above model, we allow the slope parameters $(\Theta_{1,it}, \Theta_{2,it})$ to be exhibit latent group structures along the cross-sectional dimension and an unknown break along the time dimension.

3.7.2 Data

The data we use is obtained from Aquaro et al. (2021), which is the quarterly data for 377 MSAs over 1975 to 2014. To construct the growth rate and the lagged term, we lose two periods of observations, which yields $T = 158$. Similarly to Su et al. (2023), we deseasonalize the growth rate of real house price and real income. We don't defactor the variables since our model contains the IFEs to control the common shocks.

3.7.3 Empirical Results

We first apply the singular value thresholding to estimate the ranks of $\Theta_0 = \{\lambda_i' f_t\}$, $\Theta_1 = \{\Theta_{1,it}\}$ and $\Theta_2 = \{\Theta_{2,it}\}$. The estimates are: $\hat{r}_0 = 1$, $\hat{r}_1 = 2$, and $\hat{r}_2 = 2$. Before applying the proposed estimation algorithm in Section 3.3, we first test the presence of the structural break showing in Section 3.5.2. As provided in Bai and Perron (2003), for each individual i, the critical value of test statistic $\sup_{T_1 \in \mathscr{T}_\varepsilon} F_i(T_1)$ is 15.37. We then construct the sup-F test statistic for each MSA. Results show that $\min_{i \in [N]} \sup_{T_1 \in \mathscr{T}_\varepsilon} F_i(T_1) = 0.0195$ and the final test statistic is $F_{NT}(1|0) = \max_{i \in [N]} \sup_{T_1 \in \mathscr{T}_\varepsilon} F_i(T_1) = 2161.65$. Based on this, we reject the null that there is no structural break for slope coefficient $\Theta_{1,it}$ and $\Theta_{2,it}$ at 1% significance level.

With the presence of a structural break, we apply the proposed multi-stage estimation result in Section 3.3 to estimate the break date and numbers of groups before and after the break. The estimated break date is given by $\hat{T}_1 = 51$, which suggests that the structural break happens at the first quarter in 1988. We conjecture that this is owing to the catastrophic stock market crash that occurred on October 1987, which is known to be the first contemporary global financial crisis.

By setting $\varsigma_N = N^{-2}$ for the STK algorithm as in the simulations, we obtain the estimated prior- and post-break numbers of groups given by $\hat{K}^{(1)} = 6$ and $\hat{K}^{(2)} = 2$, respectively. As for the group structure, Figures 3.2 and 3.3 use six and two colors to show the classification results for the 377 MSAs during 1975Q3 to 1987Q4 and 1988Q1 to 2014Q4, respectively. Table 3.7 reports the pooled regression results for the full sample in column (1), the subsample before the estimated break point in column (2), and the subsample after the estimated break point in column (3). All the slope estimators are bias-corrected. The pooled regression results in Table 3.7 show that the real income growth has positive and significant effect on the house price. Comparing the two subsamples before and after the estimated break, we observe that, with 1 percentage increase for the real income growth rate, the real house price growth rate will increase 0.09 percentage before year 1988, which is 0.02 percentage higher than that after year 1988. Besides, we notice that the slope estimates for the lagged term are similar for the two subsamples.

Table 3.7: Results for the pooled regressions

	Pooled (full sample) (1)	Pooled ($1975Q3-1987Q4$) (2)	Pooled ($1988Q3-2014Q4$) (3)
$ginc_{it}$	0.1021***	0.0904***	0.0702***
	(0.0067)	(0.0119)	(0.0065)
$ginc_{i,t-1}$	0.0590***	0.0392***	0.0401***
	(0.0067)	(0.0117)	(0.0066)
#individuals	377	377	377

Note: Column (1) reports the pooled regression results for the full sample. Columns (2) and (3) report the pooled regression results for the subsamples before and after the estimated break point, respectively. Slope estimators are all bias-corrected. Values in parentheses are standard errors and *** indicates significance at 1% level.

To examine the difference for each of the 6 estimated groups before the break, Table 3.8 reports the post-classification regression results for each estimated group before the estimated break. Even though the effects of the real income growth are positive for all estimated groups, they differ vastly across groups. The effect of the real income growth for Group 2 is the highest, followed by Groups 5 and 3, and the effects of real income growth on the real house price in all of these three groups are higher than 0.15 percentage. In contrast, the effects of the real income growth for the remaining three groups, viz., Groups 1, 4, and 6, are less than 0.07 percentage. Similarly, Table 3.9 reports the post-classification regression results for each estimated group after the estimated break. The estimated slope coefficient for both groups are statistically significant. Especially, during 1988Q1-2014Q1, the slope estimator for the lagged term in the first group is much higher than that for the second group.

We also apply the C-Lasso algorithm in Su et al. (2016) to estimate the group structure before and after the estimated break point. The C-Lasso approach together with the IC detects 2 groups both before and after the break. With the six groups shown by our algorithm, we conjecture that it may due to the smaller time periods before the break.

Table 3.8: Results for the post-classification regressions

	Group 1 (1)	Group 2 (2)	Group 3 (3)	Group 4 (4)	Group 5 (5)	Group 6 (6)
$ginc_{it}$	0.0301	0.3169***	0.1522***	0.0168	0.1877**	0.0661***
	(0.0345)	(0.0292)	(0.0408)	(0.0561)	(0.0775)	(0.0153)
$ginc_{i,t-1}$	0.1217***	−0.0191	−0.0089	−0.0298	−0.1269*	0.0331***
	(0.0348)	(0.0288)	(0.0407)	(0.0560)	(0.0754)	(0.0151)
#individuals	60	92	35	12	36	142

Note: Each column reports the regression results for each estimated group during 1977Q3-1987Q4. Slope estimators are all bias-corrected. Values in parentheses are standard errors. ***, **, and * indicate significance at 1% level, 5% level, 10% level, respectively.

Table 3.9: Results for the post-classification regressions after the break

	Group 1 (1)	Group 2 (2)
$ginc_{it}$	0.0670***	0.0714***
	(0.0130)	(0.0079)
$ginc_{i,t-1}$	0.0870***	0.0275***
	(0.0135)	(0.0079)
#individuals	103	274

Note: Each column reports the regression results for each estimated group during 1988Q1-2014Q4. Slope estimators are all bias-corrected. Values in parentheses are standard errors. *** indicates significance at 1% level.

3.8 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 memberships, or just changes in the slope coefficients for some specific groups before and after the break. To estimate the unknown structural break, the number of groups and group memberships before and after the break point, we propose an estimation algorithm with initial nuclear-norm-regularized estimates, 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. We show that the structural break estimator, the group number estimators, and the group membership estimators before and after the break point are all consistent, and the final post-classification slope coefficient estimators enjoy the oracle property.

There are several interesting topics for further research. First, even though we discuss a possible test for the existence of a break in the panel data models with latent group structures, we have not fully worked out the asymptotic theory. Second, we assume the presence of a single break in the data and it is interesting to extend our theory to allow for multiple breaks. Third, our theory rules out both unit-root-type nonstationarity and nonstochastic trending nonstationarity and it is interesting to extend our theory to allow for nonstationarity. We will pursue these topics in future research.

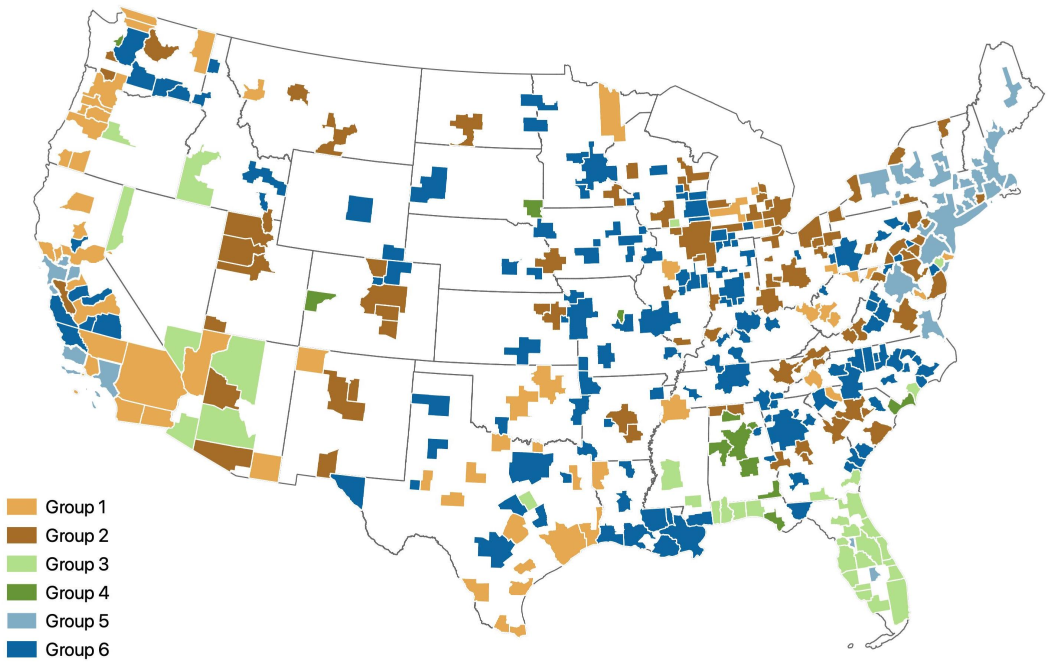


Figure 3.2: Group classification result before the break

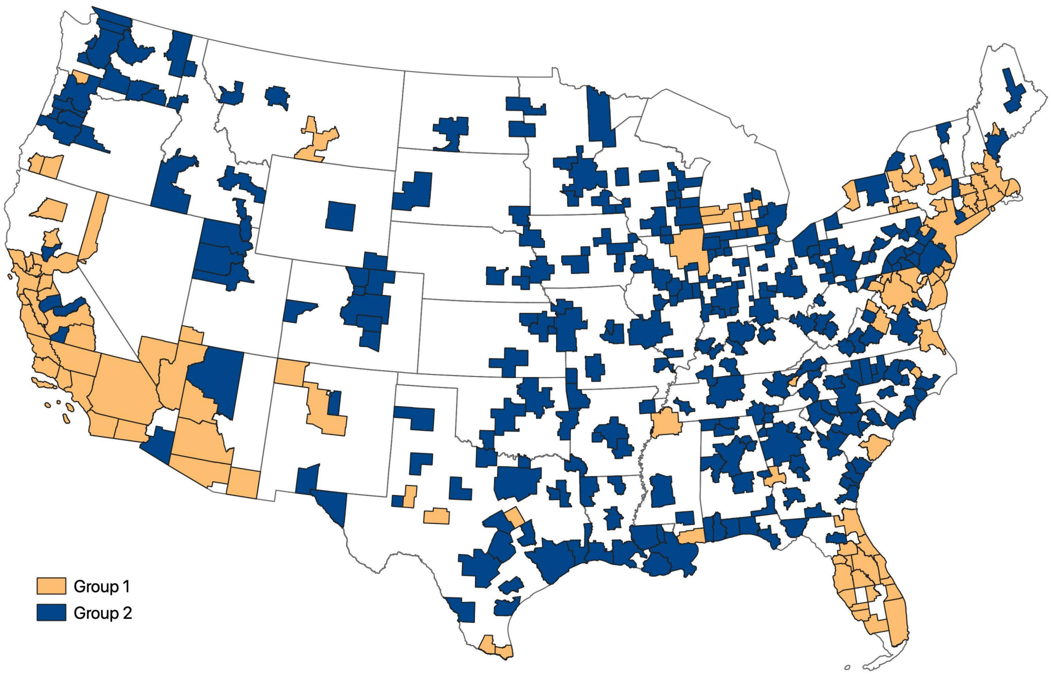


Figure 3.3: Group classification result before the break

Chapter 4

Conclusion

This study contributes to the estimation and inference of the high-dimensional panel models with two-way slope heterogeneity. We have considered the heterogeneity in two models: (1) two-way heterogeneity modeling as low-rank slope matrix in panel quantile model, (2) time-varying latent group structure in linear panel model. In the first model, the asymptotic theory for the debaised slope estimators is established, and we propose two specification tests for the slope coefficient under different rank conditions. In the second model, we recover the unknown break point, the number of groups, and the latent group structure before and after the unknown break point and show that proposed estimators enjoy the oracle property. Monte Carlo experiments are done to show the good performance of the proposed estimators and tests. In empirical applications, estimation algorithms are employed to study several problems in microeconomics and macroeconomics.

In the future research, it is interesting to extend models and estimation procedures in several directions. First, the inference theory for the low-rank panel quantile regressions is obtained by sample-splitting. One can try to change the sample-splitting approach to the full sample estimation, which obtains more stable estimates. Second, one can also generalize the time-varying latent group structure with multiple breaks by using the potential sequential testing approach. Third, one can also extend our model to the non-stationary and nonlinear framework.

Appendix A

Technical Results for Chapter 2

A.1 Proofs of the Main Results

A.1.1 Proof of Theorem 2.1

We focus on the full sample estimators $\tilde{\Delta}_{\Theta_j}$, $\tilde{\sigma}_{k,j}$, and $\tilde{V}_j$ in the proof. The results for their subsample counterparts can be established in the same manner, and we omit the detail for brevity.

Proof of Statement (i)

Recall that

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

Define $\mathscr{R}(C_1) := \left\{ \{\Delta_{\Theta_j}\}_{j=0}^p : \sum_{j=0}^p \left\| \mathscr{P}_j^{\perp}(\Delta_{\Theta_j}) \right\|_* \leq C_1 \sum_{j=0}^p \left\| \mathscr{P}_j(\Delta_{\Theta_j}) \right\|_* \right\}$. By Lemma A.4, $\mathbb{P}\{\{\tilde{\Delta}_{\Theta_j}(\tau)\}_{j=0}^p \in \mathscr{R}(3)\} \to 1$. When $\{\tilde{\Delta}_{\Theta_j}\}_{j=0}^p \in \mathscr{R}(C_1)$ and $\{\tilde{\Delta}_{\Theta_j}\}_{j=0}^p \notin \mathscr{R}(3,C_2)$, we have $\sum_{j=0}^p \left\| \tilde{\Delta}_{\Theta_j} \right\|_F^2 < C_2 \sqrt{NT}$, which implies

$$\frac{1}{\sqrt{NT}} \left\| \tilde{\Delta}_{\Theta_j} \right\|_F = O_p\left((N \wedge T)^{-1/2} \right), \quad \forall j \in [p] \cup \{0\}.$$

It suffices to consider the case that $\{\tilde{\Delta}_{\Theta_j}\}_{j=0}^p \in \mathscr{R}(3,C_2)$.

Define

$$\mathbb{Q}_\tau\left(\{\Theta_j\}_{j=0}^p\right) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \rho_\tau\left(Y_{it} - \Theta_{0,it} - \sum_{j=1}^p X_{j,it}\Theta_{j,it}\right), \text{ and}$$

$$Q_\tau(\{\Theta_j\}_{j=0}^p) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left[\rho_\tau\left(Y_{it} - \Theta_{0,it} - \sum_{j=1}^p X_{j,it}\Theta_{j,it}\right)\Big|\mathscr{G}_{i,t-1}\right],$$

where $\mathscr{G}_{i,t-1}$ is defined in Assumption 2.1. Then we have

$$\begin{aligned}
0 &\geq \mathbb{Q}_\tau\left(\{\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\}_{j=0}^p\right) - \mathbb{Q}_\tau\left(\{\Theta_j^0\}_{j=0}^p\right) + \sum_{j=0}^p \nu_j\left(\|\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\|_* - \|\Theta_j^0\|_*\right) \\
&= \left\{\mathbb{Q}_\tau\left(\{\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\}_{j=0}^p\right) - \mathbb{Q}_\tau\left(\{\Theta_j^0\}_{j=0}^p\right) - \left[Q_\tau\left(\{\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\}_{j=0}^p\right) - Q_\tau\left(\{\Theta_j^0\}_{j=0}^p\right)\right]\right\} \\
&+ \left[Q_\tau\left(\{\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\}_{j=0}^p\right) - Q_\tau\left(\{\Theta_j^0\}_{j=0}^p\right)\right] + \sum_{j=0}^p \nu_j\left(\|\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\|_* - \|\Theta_j^0\|_*\right),
\end{aligned} \tag{A.1}$$

where the first inequality holds by the definition of the estimator. Noted that

$$\nu_j\left|\|\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\|_* - \|\Theta_j^0\|_*\right| \leq \nu_j\|\tilde{\Delta}_{\Theta_j}\|_* \leq c_8\nu_j\sum_{j=0}^p\|\tilde{\Delta}_{\Theta_j}\|_F \tag{A.2}$$

where the first inequality is due to triangle inequality and the second inequality holds by Lemma A.7 with positive constant c_8 defined in the lemma.

Define

$$\rho_{it}\left(\{\Delta_{\Theta_j,it}, X_{j,it}\}_{j=0}^p, \varepsilon_{it}\right) = \rho_\tau\left(\varepsilon_{it} - \Delta_{0,it} - \sum_{j=1}^p X_{j,it}\Delta_{\Theta_j,it}\right) - \rho_\tau(\varepsilon_{it}), \tag{A.3}$$

$$\begin{aligned}
\bar{\rho}_{it}\left(\{\Delta_{\Theta_j,it}, X_{j,it}\}_{j=0}^p, \varepsilon_{it}\right) &= \mathbb{E}\left[\rho_\tau\left(\varepsilon_{it} - \Delta_{0,it} - \sum_{j=1}^p X_{j,it}\Delta_{\Theta_j,it}\right) - \rho_\tau(\varepsilon_{it})\Big|\mathscr{G}_{i,t-1}\right], \\
\tilde{\rho}_{it}\left(\{\Delta_{\Theta_j,it}, X_{j,it}\}_{j=0}^p, \varepsilon_{it}\right) &= \rho_{it}\left(\{\Delta_{\Theta_j,it}, X_{j,it}\}_{j=0}^p, \varepsilon_{it}\right) - \bar{\rho}_{it}\left(\{\Delta_{\Theta_j,it}, X_{j,it}\}_{j=0}^p, \varepsilon_{it}\right),
\end{aligned} \tag{A.4}$$

$$\mathscr{A}_1 = \left\{\sup_{\{\Delta_{\Theta_j}\}_{j=0}^p \in \mathscr{R}(3,C_2)} \frac{\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{\rho}_{it}\left(\{\Delta_{\Theta_j,it}, X_{j,it}\}_{j=0}^p, \varepsilon_{it}\right)\right|}{\sum_{j=0}^p\|\Delta_{\Theta_j}\|_F} \leq C_5 a_{NT}\right\},$$

with $a_{NT} = \frac{\sqrt{(N\vee T)\log(N\vee T)}}{NT}$ for some positive constant C_5, and $\mathscr{A}_1^c$ as the complement of $\mathscr{A}_1$.

On $\mathscr{A}_1$, following (A.1), we have w.p.a.1

$$\begin{aligned}
0 &\geq \left[Q_\tau\left(\{\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\}_{j=0}^p\right) - Q_\tau\left(\{\Theta_j^0\}_{j=0}^p\right)\right] \\
&- \left|\mathbb{Q}_\tau\left(\{\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\}_{j=0}^p\right) - \mathbb{Q}_\tau\left(\{\Theta_j^0\}_{j=0}^p\right) - \left[Q_\tau\left(\{\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\}_{j=0}^p\right) - Q_\tau\left(\{\Theta_j^0\}_{j=0}^p\right)\right]\right|
\end{aligned}$$

$$
\begin{aligned}
&-\sum_{j=0}^{p} \nu_j \left| \left\|\Theta_j^0 + \tilde{\Delta}_{\Theta_j}\right\|_* - \left\|\Theta_j^0\right\|_* \right| \\
&\geq \frac{c_7 C_3}{NT\xi_N^2} \sum_{j=0}^{p} \left\|\tilde{\Delta}_{\Theta_j}\right\|_F^2 - \frac{c_7 C_4}{NT\xi_N^2}(N+T) - \frac{C_5 \sum_{j=0}^{p} \left\|\tilde{\Delta}_{\Theta_j}\right\|_F \sqrt{(N\vee T)\log(N\vee T)}}{NT} \\
&- c_8 \sum_{j=0}^{p} \nu_j \sum_{j=0}^{p} \left\|\tilde{\Delta}_{\Theta_j}\right\|_F ,
\end{aligned}
$$

where the first inequality is by triangle inequality, the second inequality holds by (A.2) and Lemmas A.6 and A.11. It follows that

$$
\begin{aligned}
&\frac{c_7 C_3}{NT\xi_N^2} \sum_{j=0}^{p} \left\|\tilde{\Delta}_{\Theta_j}\right\|_F^2 - \frac{c_8 (p+1) c_0 \sqrt{N \vee (T\log T)} + C_5 \sqrt{(N\vee T)\log(N\vee T)}}{NT} \sum_{j=0}^{p} \left\|\tilde{\Delta}_{\Theta_j}\right\|_F \\
&- \frac{c_7 C_4}{NT\xi_N^2}(N+T) \leq 0,
\end{aligned}
$$

which implies

$$
\frac{1}{\sqrt{NT}} \left\|\tilde{\Delta}_{\Theta_j}\right\|_F = O\left(\frac{\sqrt{\log(N\vee T)}\xi_N^2}{\sqrt{N\wedge T}} \right)
$$

under the event $\mathscr{A}_1$.

By Lemma A.11, for any $\delta > 0$, we can choose a sufficiently large C_5 such that $\mathbb{P}\left\{\mathscr{A}_1^c\right\} \leq \delta$. This implies

$$
\frac{1}{\sqrt{NT}} \left\|\tilde{\Theta}_j - \Theta_j^0\right\|_F = O_p\left(\frac{\sqrt{\log(N\vee T)}\xi_N^2}{\sqrt{N\wedge T}} \right), \quad \forall j \in \{0, \cdots, p\}. \quad \blacksquare
$$

Proof of Statement (ii)

With the statement (i), the second statement holds by the Weyl's inequality. ■

Proof of Statement (iii)

For $\forall j \in \{0, \cdots, p\}$, let $\tilde{D}_j = \frac{1}{NT}\tilde{\Theta}_j'\tilde{\Theta}_j = \hat{\tilde{\mathscr{V}}}_j \hat{\tilde{\Sigma}}_j \hat{\tilde{\mathscr{V}}}_j'$, and recall that $D_j^0 = \frac{1}{NT}\Theta_j^{0\prime}\Theta_j^0 = \mathscr{V}_j^0 \Sigma_j^0 \mathscr{V}_j^{0\prime}$. Define the event $\mathscr{A}_2(M) = \left\{ \frac{1}{\sqrt{NT}} \left\| \tilde{\Theta}_j - \Theta_j^0 \right\|_F \leq M\eta_N, \forall j \in \{0, \cdots, p\} \right\}$ with $\eta_N = \frac{\sqrt{\log(N\vee T)}\xi_N^2}{\sqrt{N\wedge T}}$. On event $\mathscr{A}_2(M)$, for some positive constant C_6,

$$
\left\|\tilde{D}_j - D_j^0\right\|_F^2 \leq C_6 \eta_N.
$$

By Lemma C.1 of Su et al. (2020) and Davis-Kahan sinΘ theorem in Yu et al. (2015), there exists an orthogonal rotation matrix O_j such that

$$
\left\| \mathscr{V}_j^0 - \hat{\tilde{\mathscr{V}}}_j O_j \right\|_F \leq \sqrt{K_j} \left\| \mathscr{V}_j^0 - \hat{\tilde{\mathscr{V}}}_j O_j \right\|_{op} \leq \sqrt{K_j} \frac{\sqrt{2} C_6 \eta_N}{\Sigma_{K_j,1}^2 - C_6 \eta_N} \leq \sqrt{K_j} \frac{\sqrt{2} C_6 \eta_N}{c_\sigma^2 - C_6 \eta_N}
$$

$$\leq \sqrt{K_j}\frac{\sqrt{2}C_6\eta_N}{C_7c_\sigma^2} \leq C_8\eta_N,$$

for $C_8 = \frac{\sqrt{2}C_6\sqrt{\bar{K}}}{C_7c_\sigma^2}$. The second last inequality holds with some positive constant C_7 and the fact that η_N is sufficiently small.

Then $\left\|V_j^0 - \tilde{V}_jO_j\right\|_F \leq C_8\sqrt{T}\eta_N$ by the definition of $\tilde{V}_j$ and V_j. Let $\mathscr{A}_2^c(M)$ be the complement of event $\mathscr{A}_2(M)$. Combining the fact that $\mathbb{P}\left\{\mathscr{A}_2^c(M)\right\} \to 0$, it implies $\left\|V_j^0 - \tilde{V}_jO_j\right\|_F = O_p\left(\sqrt{T}\eta_N\right).$ ■

A.1.2 Proof of Theorem 2.2

Proof of Statement (i)

We prove that $\max_{i\in I_2}\left\|O_j^{(1)\prime}\dot{u}_{i,j}^{(1)} - u_{i,j}^0\right\|_2 = O_p(\eta_N)$ and

$$\max_{i\in I_3}\left\|O_j^{(1)\prime}\dot{u}_{i,j}^{(1)} - u_{i,j}^0\right\|_2 = O_p(\eta_N)$$

can be derived in the same manner once statement (ii) is satisfied. Define

$$\begin{aligned}
&\tilde{\mathbb{Q}}_{\tau i}\left(\left\{u_{i,j}\right\}_{j\in[p]\cup\{0\}}\right) = \frac{1}{T}\sum_{t=1}^T\rho_\tau\left(Y_{it} - u_{i,0}'\tilde{v}_{t,0}^{(1)} - \sum_{j=1}^p u_{i,j}'\tilde{v}_{t,j}^{(1)}X_{j,it}\right),\\
&u_i^0 = \left[u_{i,0}^{0\prime},\cdots,u_{i,p}^{0\prime}\right]', \quad \dot{\Delta}_{i,j} = O_j^{(1)\prime}\dot{u}_{i,j}^{(1)} - u_{i,j}^0, \quad \dot{\Delta}_{i,u} = \left(\dot{\Delta}_{i,0}',\cdots,\dot{\Delta}_{i,p}'\right)',\\
&\tilde{\Phi}_{it}^{(1)} = \left[\left(O_0^{(1)\prime}\tilde{v}_{t,0}^{(1)}\right)',\left(O_1^{(1)\prime}\tilde{v}_{t,1}^{(1)}X_{1,it}\right)',\cdots,\left(O_p^{(1)\prime}\tilde{v}_{t,p}^{(1)}X_{p,it}\right)'\right]',\\
&\tilde{\Phi}_i^{(1)} = \frac{1}{T}\sum_{t=1}^T\tilde{\Phi}_{it}^{(1)}\tilde{\Phi}_{it}^{(1)\prime},\\
&w_{1,it} = Y_{it} - \left(O_0^{(1)}u_{i0}^0\right)'\tilde{v}_{t,0}^{(1)} - \sum_{j=1}^p\left(O_j^{(1)}u_{i,j}^0\right)'\tilde{v}_{t,j}^{(1)}X_{j,it} = Y_{it} - u_i^{0\prime}\tilde{\Phi}_{it}^{(1)}\\
&= \varepsilon_{it} - u_i^{0\prime}(\tilde{\Phi}_{it}^{(1)} - \Phi_{it}^0),
\end{aligned}$$

and for $i \in I_2$, recall that $\mathscr{D}_{e_i}^{I_1}$ is the σ-field generated by

$$\left\{\varepsilon_{i^*t}, e_{i^*t}\right\}_{i^*\in I_1,t\in[T]}\bigcup\left\{e_{it}\right\}_{t\in[T]}\bigcup\left\{V_j^0\right\}_{j\in[p]\cup\{0\}}\bigcup\left\{W_j^0\right\}_{j\in[p]}.$$

By construction, we have

$$\begin{aligned}
0 &\geq \dot{Q}_{\tau i,u}\left(\left\{\dot{u}_{i,j}^{(1)}\right\}_{j\in[p]\cup\{0\}}\right) - \dot{Q}_{\tau i,u}\left(\left\{O_j^{(1)}u_{i,j}^0\right\}_{j\in[p]\cup\{0\}}\right)\\
&= \frac{1}{T}\sum_{t=1}^T\rho_\tau\left(Y_{it} - \dot{u}_{i,0}^{(1)\prime}O_0^{(1)}O_0^{(1)\prime}\tilde{v}_{t,0}^{(1)} - \sum_{j=1}^p\dot{u}_{i,j}^{(1)\prime}O_j^{(1)}O_j^{(1)\prime}\tilde{v}_{t,j}^{(1)}X_{j,it}\right)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(Y_{it}-u_{i,0}^{0\prime}O_0^{(1)\prime}\tilde{v}_{t,0}^{(1)}-\sum_{j=1}^{p}u_{i,j}^{0\prime}O_j^{(1)\prime}\tilde{v}_{t,j}^{(1)}X_{j,it}\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\left(\tau-\mathbf{1}\left\{w_{1,it}\le 0\right\}\right)\right]\\
&+\frac{1}{T}\sum_{t=1}^{T}\int_0^{\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}}\left(\mathbf{1}\left\{w_{1,it}\le s\right\}-\mathbf{1}\left\{w_{1,it}\le 0\right\}\right)ds\\
&=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\tilde{\Phi}_{it}^{(1)\prime}\left(\tau-\mathbf{1}\left\{w_{1,it}\le 0\right\}\right)\middle|\mathscr{D}_{e_i}^{I_1}\right]\dot{\Delta}_{i,u}\\
&+\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\Phi}_{it}^{(1)\prime}\left(\tau-\mathbf{1}\left\{w_{1,it}\le 0\right\}\right)-\mathbb{E}\left(\tilde{\Phi}_{it}^{(1)\prime}\left(\tau-\mathbf{1}\left\{w_{1,it}\le 0\right\}\right)\middle|\mathscr{D}_{e_i}^{I_1}\right)\right]\right\}\dot{\Delta}_{i,u}\\
&+\frac{1}{T}\sum_{t=1}^{T}\int_0^{\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}}\left(\mathbf{1}\left\{\varepsilon_{it}\le s\right\}-\mathbf{1}\left\{\varepsilon_{it}\le 0\right\}\right)-\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}\le s\right\}-\mathbf{1}\left\{\varepsilon_{it}\le 0\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)ds\\
&+\frac{1}{T}\sum_{t=1}^{T}\int_0^{\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}}\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}\le s\right\}-\mathbf{1}\left\{\varepsilon_{it}\le 0\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)ds\\
&+\frac{1}{T}\sum_{t=1}^{T}\int_0^{\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}}\left(\mathbf{1}\left\{w_{1,it}\le s\right\}-\mathbf{1}\left\{\varepsilon_{it}\le s\right\}\right)-\mathbb{E}\left(\mathbf{1}\left\{w_{1,it}\le s\right\}-\mathbf{1}\left\{\varepsilon_{it}\le s\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)ds\\
&+\frac{1}{T}\sum_{t=1}^{T}\int_0^{\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}}\mathbb{E}\left(\mathbf{1}\left\{w_{1,it}\le s\right\}-\mathbf{1}\left\{\varepsilon_{it}\le s\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)ds\\
&+\frac{1}{T}\sum_{t\in[T]}\left[\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\left(\mathbf{1}\left\{\varepsilon_{it}\le 0\right\}-\mathbf{1}\left\{w_{1,it}\le 0\right\}\right)\right]\\
&:=\sum_{m=1}^{7}A_{m,i}, \qquad\qquad \text{(A.5)}
\end{aligned}
$$

where the first inequality holds by the definition of the estimator and the second equality holds by Knight's identity in Knight (1998) which states that

$$
\rho_\tau(u-v)-\rho_\tau(u)=v(\tau-\mathbf{1}\{u\le 0\})+\int_0^v\left(\mathbf{1}\{u\le s\}-\mathbf{1}\{u\le 0\}\right)ds.
$$

After simple manipulation, we have

$$
\begin{aligned}
\left|A_{4,i}\right|=A_{4,i}&\le -A_{1,i}-A_{2,i}-A_{3,i}-A_{5,i}-A_{6,i}-A_{7,i}\\
&\le\left|A_{1,i}\right|+\left|A_{2,i}\right|+\left|A_{3,i}\right|+\left|A_{5,i}\right|+\left|A_{6,i}\right|+\left|A_{7,i}\right|.
\end{aligned}
$$

Define, for some constant M, an event set

$$
\mathscr{A}_3(M)=\left\{\max_{i\in I_2}\left(|A_{m,i}|/\left\|\dot{\Delta}_{i,u}\right\|_2\right)\le M\eta_N,\quad m=1,2,3,5,6,7\right\}
$$

and

$$q_i^I = \inf_{\Delta} \frac{\left[\frac{1}{T}\sum_{t\in[T]}\left(\tilde{\Phi}_{it}^{(1)\prime}\Delta\right)^2\right]^{\frac{3}{2}}}{\frac{1}{T}\sum_{t\in[T]}\left|\tilde{\Phi}_{it}^{(1)\prime}\Delta\right|^3}. \tag{A.6}$$

Then, under $\mathscr{A}_3(M)$, we have

$$\begin{aligned} &M\eta_N \left\|\dot{\Delta}_{i,u}\right\|_2 \geq |A_{4,i}| \\ &\geq \min\left(\frac{\left(3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'\right)c_\phi \max_{i\in I_2}\left\|\dot{\Delta}_{i,u}\right\|_2^2}{12}, \frac{\left(3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'\right)\sqrt{c_\phi}q_i^I \max_{i\in I_2}\left\|\dot{\Delta}_{i,u}\right\|_2}{6\sqrt{2}}\right), \end{aligned} \tag{A.7}$$

where $c_{11} < \min(\frac{3\underline{\mathfrak{f}}}{\bar{\mathfrak{f}}'}, 1)$ and the second inequality holds by Lemma A.15. In addition, note that

$$\begin{aligned} \max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2^3 &= \max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left(\left\|\tilde{v}_{t,0}^{(1)}\right\|_2^2 + \sum_{j\in[p]}\left\|\tilde{v}_{t,j}^{(1)}X_{j,it}\right\|_2^2\right)^{3/2} \\ &\leq \max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left[\left(\frac{2M}{c_\sigma}\right)^2\left(1+\sum_{j\in[p]}X_{j,it}^2\right)\right]^{3/2} \leq C_9, \quad \text{a.s.} \end{aligned}$$

with C_9 being a positive constant, where the first inequality holds by Lemma A.13(ii) and the second inequality is by Assumption 2.1(iv). Then we have by Lemma A.14

$$q_i^I \geq \inf_{\Delta}\frac{\|\Delta\|_2^3\left[\lambda_{\min}\left(\tilde{\Phi}_i^{(1)}\right)\right]^{2/3}}{\|\Delta\|_2^3\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2^3} \geq \frac{\min_{i\in I_2}\left[\lambda_{\min}\left(\tilde{\Phi}_i^{(1)}\right)\right]^{2/3}}{\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2^3} > \frac{\left(c_\phi/2\right)^{2/3}}{C_9},$$

which implies

$$\frac{\left(3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'\right)\sqrt{c_\phi}q_i^I \max_{i\in I_2}\left\|\dot{\Delta}_{i,u}\right\|_2}{6\sqrt{2}} > C_{10}\max_{i\in I_2}\left\|\dot{\Delta}_{i,u}\right\|_2\eta_N,$$

as C_{10} is defined to be the positive constant and $\eta_N = o(1)$. Combining this with (A.7), we have

$$\max_{i\in I_2}\left\|O_j^{(1)\prime}\dot{u}_{i,j}^{(1)} - u_{i,j}^0\right\|_2 \leq \left\|\dot{\Delta}_{i,u}\right\|_2 \leq M'\eta_N$$

for some constant M' which may depends on M. In addition, for an arbitrary constant $e > 0$, we can find a sufficiently large constant M such that $\mathbb{P}(\mathscr{A}_2^c(M)) \leq e$, which implies $\max_{i\in I_2}\left\|O_j^{(1)\prime}\dot{u}_{i,j}^{(1)} - u_{i,j}^0\right\|_2 = O_p(\eta_N)$. ■

Proof of Statement (ii)

Differently from the proof in the previous subsection, owing to the dependence of $\dot{u}_{i,j}^{(1)}$ and ε_{it}, we can not directly use conditional exponential inequality. In this subsection, we will show how to handle this dependence in detail. Recall that $\mathscr{D}_{e_{it}}$ is the σ-field generated by $\{e_{j,it}\}_{j\in[p]} \cup \left\{V_j^0\right\}_{j\in[p]\cup\{0\}} \cup \left\{W_j^0\right\}_{j\in[p]}$ and define

$$\dot{Q}_{\tau t,v}\left(\left\{v_{t,j}\right\}_{j\in[p]\cup\{0\}}\right) = \frac{1}{N_2}\sum_{i\in I_2}\rho_\tau\left(Y_{it} - v_{t,0}'\dot{u}_{i,0}^{(1)} - \sum_{j=1}^{p} v_{t,j}'\dot{u}_{i,j}^{(1)}X_{j,it}\right),$$

$$v_t^0 = \left(v_{t,0}^{0\prime},\cdots,v_{t,p}^{0\prime}\right)', \quad \dot{\Delta}_{t,j} = O_j^{(1)\prime}\dot{v}_{t,j}^{(1)} - v_{t,j}^0, \quad \dot{\Delta}_{t,v} = \left(\dot{\Delta}_{t,0}',\cdots,\dot{\Delta}_{t,p}'\right)',$$

$$\dot{\Psi}_{it}^{(1)} = \left[\left(O_0^{(1)\prime}\dot{u}_{i,0}^{(1)}\right)', \left(O_1^{(1)\prime}\dot{u}_{i,1}^{(1)}X_{1,it}\right)', ..., \left(O_p^{(1)\prime}\dot{u}_{i,p}^{(1)}X_{p,it}\right)'\right]',$$

$$\dot{\Psi}_t^{(1)} = \frac{1}{N_2}\sum_{i\in I_2}\dot{\Psi}_{it}^{(1)}\dot{\Psi}_{it}^{(1)\prime}.$$

As in (A.5), we have

$$\begin{aligned}
0 &\geq \dot{Q}_{\tau t,v}\left(\left\{\dot{v}_{t,j}^{(1)}\right\}_{j\in[p]\cup\{0\}}\right) - \dot{Q}_{\tau t,v}\left(\left\{O_j^{(1)}v_{t,j}^0\right\}_{j\in[p]\cup\{0\}}\right) \\
&= \frac{1}{N_2}\sum_{i\in I_2}\rho_\tau\left(Y_{it} - \dot{v}_{t,0}^{(1)\prime}O_0^{(1)}O_0^{(1)\prime}\dot{u}_{i,0}^{(1)} - \sum_{j=1}^{p}\dot{v}_{i,j}^{(1)\prime}O_j^{(1)}O_j^{(1)\prime}\dot{u}_{t,j}^{(1)}X_{j,it}\right) \\
&\quad - \frac{1}{N_2}\sum_{i\in I_2}\rho_\tau\left(Y_{it} - v_{t,0}^{0\prime}O_0^{(1)\prime}\dot{u}_{i,0}^{(1)} - \sum_{j=1}^{p}v_{t,j}^{0\prime}O_j^{(1)\prime}\dot{u}_{t,j}^{(1)}X_{j,it}\right) \\
&= \frac{1}{N_2}\sum_{i\in I_2}\left[\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}\left(\tau - \mathbf{1}\left\{w_{3,it}\leq 0\right\}\right)\right] + \frac{1}{N_2}\sum_{i\in I_2}\int_0^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\left(\mathbf{1}\left\{w_{3,it}\leq s\right\} - \mathbf{1}\left\{w_{3,it}\leq 0\right\}\right)ds \\
&= \frac{1}{N_2}\sum_{i\in I_2}\left[\Psi_{it}^{0\prime}\dot{\Delta}_{t,v}\left(\tau - \mathbf{1}\{\varepsilon_{it}\leq 0\}\right)\right] + \frac{1}{N_2}\sum_{i\in I_2}\left[\left(\dot{\Psi}_{it}^{(1)} - \Psi_{it}^0\right)'\dot{\Delta}_{t,v}\left(\tau - \mathbf{1}\{\varepsilon_{it}\leq 0\}\right)\right] \\
&\quad + \frac{1}{N_2}\sum_{i\in I_2}\left[\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}\left(\mathbf{1}\{\varepsilon_{it}\leq 0\} - \mathbf{1}\left\{w_{3,it}\leq 0\right\}\right)\right] \\
&\quad + \frac{1}{N_2}\sum_{i\in I_2}\int_0^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\mathbb{E}\left[\left(\mathbf{1}\{\varepsilon_{it}\leq s\} - \mathbf{1}\{\varepsilon_{it}\leq 0\}\right)ds\middle|\mathscr{D}_{e_{it}}\right] \\
&\quad + \frac{1}{N_2}\sum_{i\in I_2}\left\{\int_0^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\left[\left(\mathbf{1}\{\varepsilon_{it}\leq s\} - \mathbf{1}\{\varepsilon_{it}\leq 0\}\right) - \mathbb{E}\left(\left(\mathbf{1}\{\varepsilon_{it}\leq s\} - \mathbf{1}\{\varepsilon_{it}\leq 0\}\right)ds\middle|\mathscr{D}_{e_{it}}\right)\right]\right\} \\
&\quad + \frac{1}{N_2}\sum_{i\in I_2}\int_0^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\left(\mathbf{1}\left\{w_{3,it}\leq s\right\} - \mathbf{1}\{\varepsilon_{it}\leq s\}\right)ds \\
&\quad + \frac{1}{N_2}\sum_{i\in I_2}\int_0^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\left(\mathbf{1}\{\varepsilon_{it}\leq 0\} - \mathbf{1}\left\{w_{3,it}\leq 0\right\}\right)ds
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N_2}\sum_{i\in I_2}\left[\Psi_{it}^{0\prime}\dot{\Delta}_{t,v}\left(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)\right]+\frac{1}{N_2}\sum_{i\in I_2}\left[\left(\dot{\Psi}_{it}^{(1)}-\Psi_{it}^{0}\right)'\dot{\Delta}_{t,v}\left(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)\right]\\
&+\frac{2}{N_2}\sum_{i\in I_2}\left[\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}\left(\mathbf{1}\{\varepsilon_{it}\le 0\}-\mathbf{1}\left\{w_{3,it}\le 0\right\}\right)\right]\\
&+\frac{1}{N_2}\sum_{i\in I_2}\int_0^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\mathbb{E}\left[\left(\mathbf{1}\{\varepsilon_{it}\le s\}-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)\middle|\mathscr{D}_{e_{it}}\right]ds\\
&+\frac{1}{N_2}\sum_{i\in I_2}\left\{\int_0^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\left[\left(\mathbf{1}\{\varepsilon_{it}\le s\}-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)-\mathbb{E}\left(\left(\mathbf{1}\{\varepsilon_{it}\le s\}-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)\middle|\mathscr{D}_{e_{it}}\right)\right]ds\right\}\\
&+\frac{1}{N_2}\sum_{i\in I_2}\int_0^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\left(\mathbf{1}\left\{w_{3,it}\le s\right\}-\mathbf{1}\{\varepsilon_{it}\le s\}\right)ds\\
&:=\sum_{m=1}^{6}B_{m,t} \qquad\qquad (A.8)
\end{aligned}
$$

where

$$
w_{3,it}=Y_{it}-v_{t,0}^{0\prime}O_0^{(1)\prime}\dot{u}_{i,0}^{(1)}-\sum_{j=1}^{p}v_{t,j}^{0\prime}O_j^{(1)\prime}\dot{u}_{i,j}^{(1)}X_{j,it}=Y_{it}-v_t^{0\prime}\dot{\Psi}_{it}^{(1)}=\varepsilon_{it}-v_t^{0\prime}\left(\dot{\Psi}_{it}^{(1)}-\Psi_{it}^{0}\right),
$$

the last equality is by the fact that the third and the last terms after the second equality are identical. Then we obtain $\left|B_{4,t}\right|\le\sum_{m\neq 4}\left|B_{m,t}\right|$. By Lemma A.16 and similar arguments for Theorem 2.2(i), we obtain that

$$
\max_{t\in[T]}\left\|O_j^{(1)\prime}\dot{v}_{t,j}^{(1)}-v_{t,j}^{0}\right\|_2\le\left\|\dot{\Delta}_{t,v}\right\|_2=O_p(\eta_N).
$$

■

Proof of Statement (iii)

In this proof we derive the linear expansion of $\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}$ for each $i\in I_3$. Recall that $\forall i\in I_3$,

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

Define $\dot{\varpi}_{it}=\left(\dot{v}_{t,0}^{(1)\prime},X_{1,it}\dot{v}_{t,1}^{(1)\prime},\cdots,X_{p,it}\dot{v}_{t,p}^{(1)\prime}\right)'$ and

$$
\begin{aligned}
\dot{\mathbb{H}}_i\left(\{u_{i,j}\}_{j\in[p]\cup\{0\}}\right)&=\frac{1}{T}\sum_{t=1}^{T}\left[\tau-\mathbf{1}\left\{\varepsilon_{it}\le g_{it}(\{u_{i,j}\}_{j\in[p]\cup\{0\}})\right\}\right]\dot{\varpi}_{it},\\
\dot{\mathscr{H}}_i\left(\{u_{i,j}\}_{j\in[p]\cup\{0\}}\right)&=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\left[\tau-\mathbf{1}\left\{\varepsilon_{it}\le g_{it}(\{u_{i,j}\}_{j\in[p]\cup\{0\}})\right\}\right]\dot{\varpi}_{it}\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right\},\\
&=\frac{1}{T}\sum_{t=1}^{T}\left\{\left[\tau-\mathfrak{F}_{it}\left(g_{it}(\{u_{i,j}\}_{j\in[p]\cup\{0\}})\right)\right]\dot{\varpi}_{it}\right\},
\end{aligned}
$$

$$\dot{\mathbb{W}}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)=\dot{\mathbb{H}}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)-\dot{\mathbb{H}}_i\left(\{O_j^{(1)}u_{i,j}^{0}\}_{j\in[p]\cup\{0\}}\right)$$
$$-\left\{\dot{\mathscr{H}}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)-\dot{\mathscr{H}}_i\left(\{O_j^{(1)}u_{i,j}^{0}\}_{j\in[p]\cup\{0\}}\right)\right\},$$

where $\mathscr{D}_{e_i}^{I_1\cup I_2}$ being the σ-field generated by

$$\{\varepsilon_{i^*t},e_{i^*t}\}_{i^*\in I_1\cup I_2,t\in[T]}\bigcup\{e_{it}\}_{t\in[T]}\bigcup\{V_j^0\}_{j\in[p]\cup\{0\}}\bigcup\{W_j^0\}_{j\in[p]},$$

and

$$g_{it}(\{u_{i,j}\}_{j\in[p]\cup\{0\}})=u_{i,0}'\dot{v}_{t,0}^{(1)}+\sum_{j\in[p]}u_{i,j}'\dot{v}_{t,j}^{(1)}X_{j,it}-u_{i,0}^{0\prime}v_{t,0}^{0}-\sum_{j\in[p]}u_{i,j}^{0\prime}v_{t,j}^{0}X_{j,it}.$$

Then we have

$$\dot{\mathbb{H}}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)=\dot{\mathbb{H}}_i\left(\{O_j^{(1)}u_{i,j}^{0}\}_{j\in[p]\cup\{0\}}\right)+\dot{\mathscr{H}}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)$$
$$-\dot{\mathscr{H}}_i\left(\{O_j^{(1)}u_{i,j}^{0}\}_{j\in[p]\cup\{0\}}\right)+\dot{\mathbb{W}}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right).\quad\text{(A.9)}$$

By Assumptions 2.1(v) and 2.2 and Theorem 2.2(ii), we have $\max_{i\in I_3,t\in[T]}\|\dot{\boldsymbol{\varpi}}_{it}\|_2\leq C_{11}\xi_N$ a.s.. By the first order condition of the quantile regression, we have

$$\max_{i\in I_3}\left\|\dot{\mathbb{H}}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)\right\|_2=O_p\left(\frac{1}{T}\max_{i\in I_3,t\in[T]}\|\dot{\boldsymbol{\varpi}}_{it}\|_2\right)=O_p\left(\frac{\xi_N}{T}\right).\quad\text{(A.10)}$$

Next, we show that $\max_{i\in I_3}\left\|\dot{\mathbb{W}}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)\right\|_2=o_p\left((N\vee T)^{-1/2}\right)$. Notice that

$$\begin{aligned}
&\dot{\mathbb{W}}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\dot{\boldsymbol{\varpi}}_{it}\left(\mathbf{1}\left\{\varepsilon_{it}\leq g_{it}(\{O_j^{(1)}u_{i,j}^{0}\}_{j\in[p]\cup\{0\}})\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq g_{it}(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}})\right\}\right)\\
&-\frac{1}{T}\sum_{t\in[T]}\dot{\boldsymbol{\varpi}}_{it}\left[\mathfrak{F}_{it}\left(g(\{O_j^{(1)}u_{i,j}^{0}\}_{j\in[p]\cup\{0\}})\right)-\mathfrak{F}_{it}\left(g_{it}(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}})\right)\right]\\
&=\dot{\mathbb{W}}_i^{I}-\dot{\mathbb{W}}_i^{II}\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right),
\end{aligned}\quad\text{(A.11)}$$

where $\dot{\mathbb{W}}_i^{I}=\frac{1}{T}\sum_{t=1}^{T}\dot{\mathbb{W}}_{it}^{I}$ with

$$\dot{\mathbb{W}}_{it}^{I}=\dot{\boldsymbol{\varpi}}_{it}\left\{\left(\mathbf{1}\left\{\varepsilon_{it}\leq g_{it}(\{O_j^{(1)}u_{i,j}^{0}\}_{j\in[p]\cup\{0\}})\right\}-\mathbf{1}\{\varepsilon_{it}\leq 0\}\right)\right.$$
$$\left.-\left[\mathfrak{F}_{it}\left(g_{it}(\{O_j^{(1)}u_{i,j}^{0}\}_{j\in[p]\cup\{0\}})\right)-\mathfrak{F}_{it}(0)\right]\right\}$$

and

$$\dot{\mathbb{W}}_i^{II}\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)=\frac{1}{T}\sum_{t=1}^{T}\dot{\boldsymbol{\varpi}}_{it}\left\{\left(\mathbf{1}\left\{\varepsilon_{it}\leq g_{it}(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}})\right\}-\mathbf{1}\{\varepsilon_{it}\leq 0\}\right)\right.$$

$$-\left[\mathfrak{F}_{it}\left(g_{it}(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}})\right)-\mathfrak{F}_{it}(0)\right]\Bigg\}.$$

Noting that

$$g_{it}(\{O_j^{(1)}u_{i,j}^0\}_{j\in[p]\cup\{0\}})=\left(O_0^{(1)}u_{i,0}\right)'\left(\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^0\right)$$
$$+\sum_{j\in[p]}\left(O_j^{(1)}u_{i,j}\right)'\left(\dot{v}_{t,j}^{(1)}-O_j^{(1)}v_{t,j}^0\right)X_{j,it},$$

and $\max_{i\in I_2,t\in[T]}\left|g_{it}(\{O_j^{(1)}u_{i,j}^0\}_{j\in[p]\cup\{0\}})\right|=O_p(\xi_N\eta_N)$, we have, $\max_{i\in I_3,t\in[T]}\left\|\dot{\mathbb{W}}_{it}^I\right\|_2=O_p(\xi_N)$, $\mathbb{E}\left(\mathbb{W}_{it}^I\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)=0$, and for some positive constants C_{11},

$$\begin{aligned}
&\max_{i\in I_3,t\in[T]}\left\|Var\left(\dot{\mathbb{W}}_{it}^I\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)\right\|_F\\
&\leq\max_{i\in I_3,t\in[T]}\left\|\mathbb{E}\left\{\dot{\boldsymbol{\varpi}}_{it}\dot{\boldsymbol{\varpi}}_{it}'\left(\mathbf{1}\left\{\varepsilon_{it}\leq g(\{O_j^{(1)}u_{i,j}^0\}_{j\in[p]\cup\{0\}})\right\}-\mathbf{1}\{\varepsilon_{it}\leq 0\}\right)^2\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right\}\right\|_F\\
&\leq\max_{i\in I_3,t\in[T]}\|\dot{\boldsymbol{\varpi}}_{it}\|_2^2\,\mathbb{E}\left(\mathbf{1}\left\{0\leq|\varepsilon_{it}|\leq\left|g(\{O_j^{(1)}u_{i,j}^0\}_{j\in[p]\cup\{0\}})\right|\right\}\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)=O_p(\xi_N^3\eta_N),
\end{aligned}$$

and

$$\begin{aligned}
&\max_{i\in I_3,t\in[T]}\sum_{s=t+1}^T\left\|Cov\left(\dot{\mathbb{W}}_{it}^I,\dot{\mathbb{W}}_{is}^I\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)\right\|_F\\
&\lesssim\max_{i\in I_3,t\in[T]}\sum_{s=t+1}^T\left[\mathbb{E}\left(\|\dot{\mathbb{W}}_{it}^I\|_F^{2+\vartheta}\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)\right]^{\frac{1}{2+\vartheta}}\left[\mathbb{E}\left(\|\dot{\mathbb{W}}_{is}^I\|_F^{2+\vartheta}\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)\right]^{\frac{1}{2+\vartheta}}[\alpha(t-s)]^{1-\frac{2}{2+\vartheta}}\\
&\leq\max_{i\in I_3,t\in[T]}\left[\mathbb{E}\left(\|\dot{\mathbb{W}}_{it}^I\|_F^{2+\vartheta}\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)\right]^{\frac{2}{2+\vartheta}}\max_{t\in[T]}\sum_{s=t+1}^T[\alpha(t-s)]^{1-\frac{2}{2+\vartheta}}\text{ a.s.}\\
&=O_p\left(\xi_N^{\frac{6+2\vartheta}{2+\vartheta}}\eta_N^{\frac{2}{2+\vartheta}}\right),
\end{aligned}$$

for any $\vartheta>0$, where the first inequality holds by Davydov's inequality for conditional strong mixing processes, and the last equality holds by the fact that

$$\mathbb{E}\left(\left\|\dot{\mathbb{W}}_{it}^I\right\|^{2+\vartheta}\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)=O_p\left(\xi_N^{3+\vartheta}\eta_N\right)$$

following the similar argument as in (A.10). Combining (A.10) and (A.11) yields

$$\max_{i\in I_3,t\in[T]}\left\{\left\|Var\left(\dot{\mathbb{W}}_{it}^I\right)\right\|_F+2\sum_{s>t}\left\|Cov\left(\dot{\mathbb{W}}_{it}^I,\dot{\mathbb{W}}_{is}^I\right)\right\|_F\right\}=O_p\left(\xi_N^{\frac{6+2\vartheta}{2+\vartheta}}\eta_N^{\frac{2}{2+\vartheta}}\right).$$

For any constant $C_{12}>0$, define

$$\mathscr{A}_{5,i}(C_{12})=\left\{\max_{t\in[T]}\left\{\left\|Var\left(\dot{\mathbb{W}}_{it}^I\right)\right\|_F+2\sum_{s>t}\left\|Cov\left(\dot{\mathbb{W}}_{it}^I,\dot{\mathbb{W}}_{is}^I\right)\right\|_F\right\}\leq C_{12}\xi_N^{\frac{6+2\vartheta}{2+\vartheta}}\eta_N^{\frac{2}{2+\vartheta}}\right\},$$

$$\mathscr{A}_5(C_{12}) = \left\{ \max_{i\in I_3, t\in[T]} \left\{ \left\|Var\left(\dot{\mathbb{W}}_{it}^{I}\right)\right\|_F + 2\sum_{s>t} \left\|Cov\left(\dot{\mathbb{W}}_{it}^{I}, \dot{\mathbb{W}}_{is}^{I}\right)\right\|_F \right\} \le C_{12}\xi_N^{\frac{6+2\vartheta}{2+\vartheta}} \eta_N^{\frac{2}{2+\vartheta}} \right\}.$$

For any $e > 0$, we can find a sufficiently large constants C_{12} such that $\mathbb{P}(\mathscr{A}_5^c(C_{12})) \le e$. Therefore, we have

$$\begin{aligned}
&\mathbb{P}\left\{ \max_{i\in I_3} \left\| \frac{1}{T}\sum_{t=1}^{T} \dot{\mathbb{W}}_{it}^{I} \right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}} \left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}} \sqrt{\frac{\log(N\vee T)}{T}} \right\} \\
&\le \mathbb{P}\left\{ \max_{i\in I_3} \left\| \frac{1}{T}\sum_{t=1}^{T} \dot{\mathbb{W}}_{it}^{I} \right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}} \left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}} \sqrt{\frac{\log(N\vee T)}{T}}, \mathscr{A}_5(C_{12}) \right\} + e \\
&\le \sum_{i\in I_3} \mathbb{P}\left\{ \left\| \frac{1}{T}\sum_{t=1}^{T} \dot{\mathbb{W}}_{it}^{I} \right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}} \left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}} \sqrt{\frac{\log(N\vee T)}{T}}, \mathscr{A}_5(C_{12}) \right\} + e \\
&\le \sum_{i\in I_3} \mathbb{P}\left\{ \left\| \frac{1}{T}\sum_{t=1}^{T} \dot{\mathbb{W}}_{it}^{I} \right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}} \left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}} \sqrt{\frac{\log(N\vee T)}{T}}, \mathscr{A}_{5,i}(C_{12}) \right\} + e \\
&= \sum_{i\in I_3} \mathbb{E}\mathbb{P}\left\{ \left\| \frac{1}{T}\sum_{t=1}^{T} \dot{\mathbb{W}}_{it}^{I} \right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}} \left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}} \sqrt{\frac{\log(N\vee T)}{T}} \middle| \mathscr{D}_{e_i}^{I_1\cup I_2} \right\} 1\{\mathscr{A}_{5,i}(C_{12})\} \\
&+ e, \qquad \text{(A.12)}
\end{aligned}$$

where the second inequality is by the union bound, the third inequality is by $\mathscr{A}_5(C_{12}) \subset \mathscr{A}_{5,i}(C_{12})$, and the last equality is owing to the fact that $\mathscr{A}_{5,i}(C_{12})$ is $\mathscr{D}_{e_i}^{I_1\cup I_2}$ measurable. Given $\mathscr{D}_{e_i}^{I_1\cup I_2}$, the randomness in $\dot{\mathbb{W}}_i^I$ only comes from $\{\varepsilon_{it}\}_{t\in[T]}$, which are strong mixing given $\mathscr{D}_{e_i}^{I_1\cup I_2}$. Therefore, on $\mathscr{A}_{5,i}(C_{12})$, Lemma A.12(ii) implies

$$\begin{aligned}
&\mathbb{P}\left\{ \left\| \frac{1}{T}\sum_{t=1}^{T} \dot{\mathbb{W}}_{it}^{I} \right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}} \left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}} \sqrt{\frac{\log(N\vee T)}{T}} \middle| \mathscr{D}_{e_i}^{I_1\cup I_2} \right\} \\
&= \mathbb{P}\left\{ \left\| \sum_{t=1}^{T} \dot{\mathbb{W}}_{it}^{I} \right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}} \left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}} \sqrt{T\log(N\vee T)} \middle| \mathscr{D}_{e_i}^{I_1\cup I_2} \right\} \\
&\le \exp\left\{ -\frac{c_{12}C_{13}^2 T \left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{2+\vartheta}} \xi_N^{\frac{10+2\vartheta}{2+\vartheta}} \log(N\vee T)}{\mathscr{D}_1 + \mathscr{D}_2 + \mathscr{D}_3} \right\}, \qquad \text{(A.13)}
\end{aligned}$$

with

$$\mathscr{D}_1 = C_{12}T\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{2+\vartheta}} \xi_N^{\frac{10+2\vartheta}{2+\vartheta}}, \ \mathscr{D}_2 = C_{11}^2\xi_N^2, \text{ and}$$

$$\mathscr{D}_3 = C_{11}C_{13}\sqrt{T\log(N\vee T)}\xi_N^{\frac{7+2\vartheta}{2+\vartheta}} \left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}} (\log T)^2,$$

which further implies

$$\mathbb{P}\left\{\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\dot{\mathbb{W}}_{it}^{I}\right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right\} = o(1)+e.$$

As e is arbitrary, by Assumption 2.1(ix), we obtain that

$$\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\dot{\mathbb{W}}_{it}^{I}\right\|_2 = O_p\left(C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right)$$
$$= o_p\left((N\vee T)^{-1/2}\right).$$

For $\dot{\mathbb{W}}_{i}^{II}\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)$, we observe that

$$g_{it}(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}) = \dot{u}_{i,0}^{(1)\prime}\dot{v}_{t,0}^{(1)} + \sum_{j\in[p]}\dot{u}_{i,j}^{(1)\prime}\dot{v}_{t,j}^{(1)}X_{j,it} - u_{i,0}^{0\prime}v_{t,0}^{0} - \sum_{j\in[p]}u_{i,j}^{0\prime}v_{t,j}^{0}X_{j,it}$$
$$= \left(\dot{u}_{i,0}^{(1)} - O_0^{(1)}u_{i,0}^{0}\right)'\dot{v}_{t,0}^{(1)} + \left(O_0^{(1)}u_{i,0}^{0}\right)'\left(\dot{v}_{t,0}^{(1)} - O_0^{(1)}v_{t,0}^{0}\right)$$
$$+ \sum_{j\in[p]}\left(\dot{u}_{i,j}^{(1)} - O_j^{(1)}u_{i,j}^{0}\right)'\dot{v}_{t,j}^{(1)}X_{j,it} + \sum_{j\in[p]}\left(O_j^{(1)}u_{i,j}^{0}\right)'\left(\dot{v}_{t,j}^{(1)} - O_j^{(1)}v_{t,j}^{0}\right)X_{j,it}$$
$$= \dot{\Delta}_{i,u}'\dot{\Phi}_{it}^{(1)} + \dot{\Delta}_{t,v}'\Psi_{it}^{0},$$

where $\dot{\Delta}_{i,u} = \left(\dot{\Delta}_{i,0}',\cdots,\dot{\Delta}_{i,p}'\right)'$ with $\dot{\Delta}_{i,j} = O_j^{(1)\prime}\dot{u}_{i,j}^{(1)} - u_{i,j}^{0}$, $\dot{\Delta}_{t,v} = \left(\dot{\Delta}_{t,0}',\cdots,\dot{\Delta}_{t,p}'\right)'$ with $\dot{\Delta}_{t,j} = O_j^{(1)\prime}\dot{v}_{t,j}^{(1)} - v_{t,j}^{0}$,

$$\dot{\Phi}_{it}^{(1)} = \left[\left(O_0^{(1)\prime}\dot{v}_{t,0}^{(1)}\right)', \left(O_1^{(1)\prime}\dot{v}_{t,1}^{(1)}X_{1,it}\right)', \cdots, \left(O_p^{(1)\prime}\dot{v}_{t,p}^{(1)}X_{p,it}\right)'\right]', \quad \text{and}$$
$$\Psi_{it}^{0} = (u_{i,0}^{0\prime}, u_{i,1}^{0\prime}X_{1,it}, ..., u_{i,p}^{0\prime}X_{p,it})'.$$

Unlike the analysis for $\dot{\mathbb{W}}_{it}^{I}$, to handle the dependence between ε_{it} and $\dot{\Delta}_{i,u}$, for any constant $C_{14} > 0$, we first define an event set $\mathscr{A}_6(C_{14}) = \left\{\max_{i\in I_3}\left\|\dot{\Delta}_{i,u}\right\|_2 \leq C_{14}\eta_N\right\}$ with $\mathbb{P}(\mathscr{A}_6^c(C_{14})) \leq e$ for any $e > 0$ by Theorem 2.2(i), then we have

$$\mathbb{P}\left(\max_{i\in I_3}\left\|\dot{\mathbb{W}}_{i}^{II}\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)\right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right)$$
$$\leq \mathbb{P}\left(\max_{i\in I_3}\left\|\dot{\mathbb{W}}_{i}^{II}\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)\right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}, \mathscr{A}_6(C_{14})\right)$$
$$+ e$$
$$\leq \mathbb{P}\left(\sup_{s\in\mathbb{S}}\max_{i\in I_3}\left\|\overline{\mathbb{W}}_{i}^{II}(s)\right\|_2 > C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right) + e \tag{A.14}$$

with $\mathbb{S} = \left\{ s \in \mathbb{R}^{\left(\sum_{j\in[p]\cup\{0\}} K_j\right)\times 1} : \|s\|_2 \le C_{14}\eta_N \right\}$ and

$$\overline{\mathbb{W}}_i^{II}(s) = \frac{1}{T}\sum_{t=1}^{T} \dot{\varpi}_{it} \Bigg\{ \left(\mathbf{1}\left\{ \varepsilon_{it} \le s'\dot{\Phi}_{it}^{(1)} + \dot{\Delta}_{t,v}'\Psi_{it}^0 \right\} - \mathbf{1}\left\{\varepsilon_{it} \le 0\right\} \right) - \left[\mathfrak{F}_{it}\left(s'\dot{\Phi}_{it}^{(1)} + \dot{\Delta}_{t,v}'\Psi_{it}^0 \right) - \mathfrak{F}_{it}(0) \right] \Bigg\}.$$

Similarly as in (A.95), we sketch the proof. Divide $\mathbb{S}$ into $\mathbb{S}_m$ with center s_m for $m = 1, \cdots, n_{\mathbb{S}}$ if $s \in \mathbb{S}_m$, then $\|s - s_m\|_2 < \frac{\varepsilon}{T}$ and $n_{\mathbb{S}} \lesssim T^{\sum_{j\in[p]\cup\{0\}} K_j}$. Then, $\forall s \in \mathbb{S}_m$, we have

$$\left\| \overline{\mathbb{W}}_i^{II}(s) \right\|_2 \le \left\| \overline{\mathbb{W}}_i^{II}(s_m) \right\|_2 + \left\| \overline{\mathbb{W}}_i^{II}(s) - \overline{\mathbb{W}}_i^{II}(s_m) \right\|_2, \tag{A.15}$$

with

$$\begin{aligned}
&\max_{i\in I_3, m\in[n_{\mathbb{S}}]} \sup_{s\in\mathbb{S}_m} \left\| \overline{\mathbb{W}}_i^{II}(s) - \overline{\mathbb{W}}_i^{II}(s_m) \right\|_2 \\
&\le \max_{i\in I_3, m\in[n_{\mathbb{S}}]} \sup_{s\in\mathbb{S}_m} \left\| \frac{1}{T}\sum_{t=1}^{T} \dot{\varpi}_{it} \left(\mathbf{1}\left\{ \varepsilon_{it} \le s'\dot{\Phi}_{it}^{(1)} + \dot{\Delta}_{t,v}'\Psi_{it}^0 \right\} - \mathbf{1}\left\{ \varepsilon_{it} \le s_m'\dot{\Phi}_{it}^{(1)} + \dot{\Delta}_{t,v}'\Psi_{it}^0 \right\} \right) \right\|_2 \\
&+ \max_{i\in I_3, m\in[n_{\mathbb{S}}]} \sup_{s\in\mathbb{S}_m} \left\| \frac{1}{T}\sum_{t=1}^{T} \dot{\varpi}_{it} \left[\mathfrak{F}_{it}\left(s'\dot{\Phi}_{it}^{(1)} + \dot{\Delta}_{t,v}'\Psi_{it}^0 \right) - \mathfrak{F}_{it}\left(s_m'\dot{\Phi}_{it}^{(1)} + \dot{\Delta}_{t,v}'\Psi_{it}^0 \right) \right] \right\|_2 \\
&\le \max_{i\in I_3, m\in[n_{\mathbb{S}}]} \frac{1}{T}\sum_{t\in[T]} \|\dot{\varpi}_{it}\|_2 \mathbb{E}\left(\mathbf{1}\left\{ \left| \varepsilon_{it} - \dot{\Delta}_{t,v}'\Psi_{it}^0 - s_m'\dot{\Phi}_{it}^{(1)} \right| \le \frac{\varepsilon \|\dot{\Phi}_{it}^{(1)}\|_2}{T} \right\} \middle| \mathscr{D}_{e_i}^{I_1\cup I_2} \right) \\
&+ \max_{i\in I_2, m\in[n_{\mathbb{S}}]} \left| \overline{\mathbb{W}}_i^{III}(m) \right| + \max_{i\in I_3, m\in[n_{\mathbb{S}}]} \sup_{s\in\mathbb{S}_m} \frac{1}{T}\sum_{t\in[T]} \|\dot{\varpi}_{it}\|_2 \|\dot{\Phi}_{it}^{(1)}\|_2 \|s - s_m\|_2
\end{aligned} \tag{A.16}$$

such that $\overline{\mathbb{W}}_i^{III}(m) := \frac{1}{T}\sum_{t\in[T]} \overline{\mathbb{W}}_{it}^{III}(m)$ and

$$\overline{\mathbb{W}}_{it}^{III}(m) := \|\dot{\varpi}_{it}\|_2 \Bigg[\left(\mathbf{1}\left\{ \left| \varepsilon_{it} - \dot{\Delta}_{t,v}'\Psi_{it}^0 - s_m'\dot{\Phi}_{it}^{(1)} \right| \le \frac{\varepsilon \|\dot{\Phi}_{it}^{(1)}\|_2}{T} \right\} \right) - \mathbb{E}\left(\mathbf{1}\left\{ \left| \varepsilon_{it} - \dot{\Delta}_{t,v}'\Psi_{it}^0 - s_m'\dot{\Phi}_{it}^{(1)} \right| \le \frac{\varepsilon \|\dot{\Phi}_{it}^{(1)}\|_2}{T} \right\} \middle| \mathscr{D}_{e_i}^{I_1\cup I_2} \right) \Bigg].$$

Like (A.97), we can show that

$$\max_{i\in I_3, m\in[n_{\mathbb{S}}]} \frac{1}{T}\sum_{t\in[T]} \|\dot{\varpi}_{it}\|_2 \mathbb{E}\left(\mathbf{1}\left\{ \left| \varepsilon_{it} - \dot{\Delta}_{t,v}'\Psi_{it}^0 - s_m'\dot{\Phi}_{it} \right| \le \frac{\varepsilon \|\dot{\Phi}_{it}^{(1)}\|_2}{T} \right\} \middle| \mathscr{D}_{e_i}^{I_1\cup I_2} \right) = O_p(\frac{\varepsilon}{T})$$

because $\dot{\Delta}_{t,v}'\Psi_{it}^0 + s_m'\dot{\Phi}_{it}^{(1)}$ and $\|\dot{\Phi}_{it}\|$ are measurable in $\mathscr{D}_{e_i}^{I_1\cup I_2}$ and the conditional density of ε_{it} given $\mathscr{D}_{e_i}^{I_1\cup I_2}$ is bounded. Also we have

$$\max_{i\in I_3, m\in[n_{\mathbb{S}}]} \sup_{s\in\mathbb{S}_m} \frac{1}{T}\sum_{t\in[T]} \|\dot{\varpi}_{it}\|_2 \|\dot{\Phi}_{it}^{(1)}\|_2 \|s - s_m\|_2 = \frac{\varepsilon}{T} \max_{i\in I_3} \frac{1}{T}\sum_{t\in[T]} \|\dot{\varpi}_{it}\|_2 \|\dot{\Phi}_{it}^{(1)}\|_2 = O_p\left(\frac{\varepsilon}{T}\right).$$

In addition, we note that

$$\max_{i\in I_3,m\in[n_{\mathbb{S}}],t\in[T]}\|\dot{\varpi}_{it}\|_2=O_p(\xi_N),\qquad \max_{i\in I_3,m\in[n_{\mathbb{S}}],t\in[T]}Var\left(\overline{\mathbb{W}}_{it}^{III}(m)\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)=O_p\left(\frac{\xi_N^3\varepsilon}{T}\right),$$

$$\max_{i\in I_3,m\in[n_{\mathbb{S}}],t\in[T]}\sum_{s=t+1}^{T}\left|Cov\left(\overline{\mathbb{W}}_{it}^{III}(m),\overline{\mathbb{W}}_{is}^{III}(m)\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)\right|=O_p\left(\xi_N^{8/3}\left(\frac{\varepsilon}{T}\right)^{2/3}\right),$$

and for any positive constant C_{15} and C_{16}, define event set

$$\mathscr{A}_{8,N}=\left(\begin{array}{c}\max_{i\in I_3,m\in[n_{\mathbb{S}}],t\in[T]}\left\{Var\left(\overline{\mathbb{W}}_{it}^{III}(m)\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)+2\sum_{s=t+1}^{T}\left|Cov\left(\overline{\mathbb{W}}_{it}^{III}(m),\overline{\mathbb{W}}_{is}^{III}(m)\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)\right|\right\}\leq C_{15}\xi_N^{8/3}\left(\frac{\varepsilon}{T}\right)^{2/3}\\ \text{and}\quad \max_{i\in I_3,t\in[T]}\|\dot{\varpi}_{it}\|_2\leq C_{16}\xi_N\end{array}\right),$$

$$\mathscr{A}_{8,N,i}=\left(\begin{array}{c}\max_{m\in[n_{\mathbb{S}}],t\in[T]}\left\{Var\left(\overline{\mathbb{W}}_{it}^{III}(m)\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)+2\sum_{s=t+1}^{T}\left|Cov\left(\overline{\mathbb{W}}_{it}^{III}(m),\overline{\mathbb{W}}_{is}^{III}(m)\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)\right|\right\}\leq C_{15}\xi_N^{8/3}\left(\frac{\varepsilon}{T}\right)^{2/3}\\ \text{and}\quad \max_{t\in[T]}\|\dot{\varpi}_{it}\|_2\leq C_{16}\xi_N\end{array}\right),$$

with $\mathbb{P}\left(\mathscr{A}_{8,N}^c\right)\leq e$ for any positive e. Then we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{i\in I_3,m\in[n_{\mathbb{S}}]}\left|\overline{\mathbb{W}}_{i}^{III}(m)\right|>C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right)\\
&\leq\mathbb{P}\left(\max_{i\in I_3,m\in[n_{\mathbb{S}}]}\left|\overline{\mathbb{W}}_{i}^{III}(m)\right|>C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}},\mathscr{A}_{8,N}\right)+e\\
&\leq\sum_{i\in[I_3],m\in[n_{\mathbb{S}}]}\mathbb{P}\left(\left|\overline{\mathbb{W}}_{i}^{III}(m)\right|>C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}},\mathscr{A}_{8,N,i}\right)+e\\
&=\sum_{i\in[I_3],m\in[n_{\mathbb{S}}]}\mathbb{E}\mathbb{P}\left(\left|\overline{\mathbb{W}}_{i}^{III}(m)\right|>C_{13}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right)\mathbf{1}\{\mathscr{A}_{8,N,i}\}\\
&+e\\
&=o(1)+e,
\end{aligned}\tag{A.17}$$

where the last line is by Bernstein's inequality similar to (A.13). As e is arbitrary, we have

$$\max_{i\in I_3,m\in[n_{\mathbb{S}}]}\left\|\overline{\mathbb{W}}_{i}^{III}(m)\right\|_2=O_p\left(\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right),$$

which implies that

$$\max_{i\in I_3,m\in[n_{\mathbb{S}}]}\sup_{s\in\mathbb{S}_m}\left\|\overline{\mathbb{W}}_{i}^{II}(s)-\overline{\mathbb{W}}_{i}^{II}(s_m)\right\|_2=O_p\left(\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right).$$

Following the same argument in (A.17), we can show that

$$\max_{i\in I_3, m\in[n_{\mathbb{S}}]}\left\|\overline{\mathbb{W}}_i^{II}(s_m)\right\|_2 = O_p\left(\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right),$$

which, combined with (A.14) and (A.15), implies that

$$\begin{aligned}
&\max_{i\in I_3}\left\|\dot{\mathbb{W}}_i^{II}\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)\right\|_2 \\
&= O_p\left(\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right) = o_p\left((N\vee T)^{-1/2}\right), \quad \text{and thus,} \\
&\max_{i\in I_3}\left\|\dot{\mathbb{W}}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j\in[p]\cup\{0\}}\right)\right\|_2 = o_p\left((N\vee T)^{-1/2}\right). \qquad \text{(A.18)}
\end{aligned}$$

Next, we observe that

$$\begin{aligned}
&\mathscr{H}_i\left(\{O_j^{(1)}u_{i,j}^0\}_{j=0}^p\right) - \mathscr{H}_i\left(\{\dot{u}_{i,j}^{(1)}\}_{j=0}^p\right) \\
&= \frac{1}{T}\sum_{t=1}^T \mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)\dot{\varpi}_{it}\dot{\varpi}_{it}'\dot{\Delta}_{i,u} + O_p\left(\max_{i\in i_3}\left\|\dot{\Delta}_{i,u}\right\|_2^2\right) \\
&= \frac{1}{T}\sum_{t=1}^T \mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)\dot{\varpi}_{it}\dot{\varpi}_{it}'\dot{\Delta}_{i,u} + o_p\left((N\vee T)^{-1/2}\right) \quad \text{uniformly over } i\in I_3,
\end{aligned} \tag{A.19}$$

where the first equality holds by Taylor expansion and Lemma A.17 and the second equality is by Assumption 2.1(ix). Combining (A.9), (A.10), (A.18) and (A.19), we have shown that

$$\dot{\Delta}_{i,u} = \left[\dot{D}_i^I\right]^{-1}\dot{D}_i^{II} + o_p\left((N\vee T)^{-1/2}\right) \quad \text{uniformly over } i\in I_3, \tag{A.20}$$

where $\dot{D}_i^I := \frac{1}{T}\sum_{t=1}^T \mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)\dot{\varpi}_{it}\dot{\varpi}_{it}'$ and $\dot{D}_i^{II} := \frac{1}{T}\sum_{t=1}^T\left[\tau - \mathbf{1}\left\{\varepsilon_{it}\leq\dot{\Delta}_{t,v}'\Psi_{it}^0\right\}\right]\dot{\varpi}_{it}$. Recall that $\varpi_{it}^0 = \left(\left(O_0^{(1)}v_{t,0}^0\right)', \left(O_1^{(1)}v_{t,1}^0\right)'X_{1,it}, \cdots, \left(O_1^{(1)}v_{t,p}^0\right)'X_{p,it}\right)'$, $D_i^I = \frac{1}{T}\sum_{t=1}^T \mathfrak{f}_{it}(0)\varpi_{it}^0\varpi_{it}^{0\prime}$, and $D_i^{II} = \frac{1}{T}\sum_{t=1}^T\left[\tau - \mathbf{1}\{\varepsilon_{it}\leq 0\}\right]\varpi_{it}^0$. Noting that

$$\begin{aligned}
\left[\dot{D}_i^I\right]^{-1}\dot{D}_i^{II} &= \left[D_i^I\right]^{-1}D_i^{II} + \left[D_i^I\right]^{-1}\left(\dot{D}_i^{II} - D_i^{II}\right) + \left[\left(\dot{D}_i^I\right)^{-1} - \left(D_i^I\right)^{-1}\right]D_i^{II} \\
&\quad + \left[\left(\dot{D}_i^I\right)^{-1} - \left(D_i^I\right)^{-1}\right]\left(\dot{D}_i^{II} - D_i^{II}\right),
\end{aligned}$$

we have by (A.20) and Lemma A.18, uniformly

$$\begin{aligned}
\dot{\Delta}_{i,u} &= \left[D_i^I\right]^{-1}D_i^{II} + \left[D_i^I\right]^{-1}\frac{1}{T}\sum_{t=1}^T\left[\mathbf{1}\{\varepsilon_{it}\leq 0\} - \mathbf{1}\left\{\varepsilon_{it}\leq\dot{\Delta}_{t,v}'\Psi_{it}^0\right\}\right]\varpi_{it}^0 \\
&\quad + o_p\left((N\vee T)^{-1/2}\right)
\end{aligned}$$

uniformly over $i\in I_3$, where we use the fact that $\sqrt{\frac{\log(N\vee T)}{T}}\xi_N\eta_N = o\left((N\vee T)^{-1/2}\right)$

by Assumption 2.1(ix). ■

A.1.3 Proof of Theorem 2.3

In this section, we extend the distribution theory of the least squares framework in Chernozhukov et al. (2019) to the quantile regression framework and obtain the uniform error bound. We assume the model has only one regressor in this section for notation simplicity.

Proof of Statement (i)

For $\forall i \in I_3$, recall from (2.7) that

$$\left\{\hat{u}_{i,j}^{(3,1)}\right\}_{j\in[p]} = \underset{\left\{u_{i,j}\right\}_{j\in[p]}}{\arg\min} \frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(\tilde{Y}_{it} - u_{i,0}'\dot{v}_{t,0}^{(1)} - u_{i,1}'\dot{v}_{t,1}^{(1)}\hat{e}_{1,it}\right). \tag{A.21}$$

where $\tilde{Y}_{it} = Y_{it} - \hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}$. Let $\hat{\Delta}_{i,u} = \begin{bmatrix}\hat{u}_{i,0}^{(3,1)} - O_0^{(1)}u_{i,0}^0 \\ \hat{u}_{i,1}^{(3,1)} - O_1^{(1)}u_{i,1}^0\end{bmatrix}$ and $\dot{\omega}_{it} = \begin{bmatrix}\dot{v}_{t,0}^{(1)} \\ \dot{v}_{t,1}^{(1)}\hat{e}_{1,it}\end{bmatrix}$.

With generic $(u_{i,0}, u_{i,1}, \mathrm{u}_{i,1})$, define

$$\hat{\mathbb{H}}_i(u_{i,0}, u_{i,1}, \mathrm{u}_{i,1}) = \frac{1}{T}\sum_{t=1}^{T}\left[\tau - \mathbf{1}\left\{\varepsilon_{it} \le \iota_{it}(u_{i,0}, u_{i,1}, \mathrm{u}_{i,1}),\right\}\right]\dot{\omega}_{it} \text{ and}$$

$$\iota_{it}(u_{i,0}, u_{i,1}, \mathrm{u}_{i,1}) = u_{i,0}'\dot{v}_{t,0}^{(1)} - u_{i,0}^{0\prime}v_{t,0}^0 + \hat{\mu}_{1,it}\mathrm{u}_{i,1}'\dot{v}_{t,1}^{(1)} - \mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^0 + \hat{e}_{1,it}u_{i,1}'\dot{v}_{t,1}^{(1)} - e_{1,it}u_{i,1}^{0\prime}v_{t,1}^0. \tag{A.22}$$

We can see that

$$\hat{\mathbb{H}}_i\left(u_{i,0}, u_{i,1}, \dot{u}_{i,1}^{(1)}\right) = \frac{1}{T}\sum_{t=1}^{T}\left[\tau - \mathbf{1}\left\{\tilde{Y}_{it} - u_{i,0}'\dot{v}_{t,0}^{(1)} - u_{i,1}'\dot{v}_{t,1}^{(1)}\hat{e}_{1,it} \le 0\right\}\right]\dot{\omega}_{it}$$

is the first order subgradient of (A.21). In addition, we define

$$\begin{aligned}\hat{\mathscr{H}}_i(u_{i,0}, u_{i,1}, \mathrm{u}_{i,1}) &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\left[\tau - \mathbf{1}\left\{\varepsilon_{it} \le \left(\iota_{it}(u_{i,0}, u_{i,1}, \mathrm{u}_{i,1})\right)\right\}\right]\dot{\omega}_{it}\middle|\mathscr{D}_e^{I_1\cup I_2}\right\} \\ &= \frac{1}{T}\sum_{t=1}^{T}\left[\tau - F_{it}\left(\iota_{it}(u_{i,0}, u_{i,1}, \mathrm{u}_{i,1})\right)\right]\dot{\omega}_{it},\end{aligned}$$

and

$$\begin{aligned}\hat{\mathbb{W}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right) &= \hat{\mathbb{H}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right) - \hat{\mathbb{H}}_i\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right) \\ &\quad - \left\{\hat{\mathscr{H}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right) - \hat{\mathscr{H}}_i\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right\},\end{aligned}$$

where $\mathscr{D}_e^{I_1 \cup I_2}$ is the σ-field generated by

$$\{\varepsilon_{it}\}_{i \in I_1 \cup I_2, t \in [T]} \bigcup \{e_{it}\}_{i \in [N], t \in [T]} \bigcup \left\{V_j^0\right\}_{j \in [p] \cup \{0\}} \bigcup \left\{W_j^0\right\}_{j \in [p]}.$$

Then it is clear that

$$\begin{aligned}\hat{\mathbb{H}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right) &= \hat{\mathbb{H}}_i\left(O_0^{(1)} u_{i,0}^0, O_1^{(1)} u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)) + \hat{\mathscr{H}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right) \\ &\quad - \hat{\mathscr{H}}_i\left(O_0^{(1)} u_{i,0}^0, O_1^{(1)} u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right) + \hat{\mathbb{W}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right).\end{aligned} \tag{A.23}$$

For specific $u_{i,0}$ and $u_{i,1}$, let $u_i = \left(u_{i,0}', u_{i,1}'\right)'$. Following similar arguments as used in the proof of Lemma A.17, the second order partial derivative of the function $\hat{\mathscr{H}}_i(\cdot)$ with respect to u_i at the true value can be shown to be bounded in probability. By Taylor expansion, it yields

$$\begin{aligned}&\hat{\mathscr{H}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right) - \hat{\mathscr{H}}_i\left(O_0^{(1)} u_{i,0}^0, O_1^{(1)} u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right) \\ &= \frac{\partial \hat{\mathscr{H}}_i\left(O_0^{(1)} u_{i,0}^0, O_1^{(1)} u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)}{\partial u_i'} \hat{\Delta}_{i,u} + R_i,\end{aligned} \tag{A.24}$$

where $\max_{i \in I_3} |R_i| \lesssim \max_{i \in I_3} \left\|\hat{\Delta}_{i,u}\right\|_2^2$ and

$$\begin{aligned}&\frac{\partial \hat{\mathscr{H}}_i\left(O_0^{(1)} u_{i,0}^0, O_1^{(1)} u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)}{\partial u_i'} \\ &= -\frac{1}{T}\sum_{t=1}^{T} f_{it}\left[\iota_{it}\left(O_0^{(1)} u_{i,0}^0, O_1^{(1)} u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right]\begin{bmatrix} \dot{v}_{t,0}^{(1)} \dot{v}_{t,0}^{(1)\prime} & \hat{e}_{1,it} \dot{v}_{t,0}^{(1)} \dot{v}_{t,1}^{(1)\prime} \\ \hat{e}_{1,it} \dot{v}_{t,1}^{(1)} \dot{v}_{t,0}^{(1)\prime} & \hat{e}_{1,it}^2 \dot{v}_{t,1}^{(1)} \dot{v}_{t,1}^{(1)\prime} \end{bmatrix} = -\dot{D}_i^F.\end{aligned}$$

Combing (A.23) and (A.24), we have

$$\begin{aligned}\hat{\Delta}_{i,u} &= \left(\dot{D}_i^F\right)^{-1}\left\{\hat{\mathbb{H}}_i\left(O_0^{(1)} u_{i,0}^0, O_1^{(1)} u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right) + \hat{\mathbb{W}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right)\right. \\ &\quad \left. - \hat{\mathbb{H}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right) + R_i\right\} \\ &= \left(\dot{D}_i^F\right)^{-1}\left\{\hat{\mathbb{H}}_i\left(O_0^{(1)} u_{i,0}^0, O_1^{(1)} u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right) + \hat{\mathbb{W}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right)\right. \\ &\quad \left. + o_p\left((N \vee T)^{-1/2}\right)\right\},\end{aligned} \tag{A.25}$$

uniformly over $i \in I_3$, where the second line is due to the fact that

$$\max_{i \in I_3}\left\|\hat{\mathbb{H}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right)\right\|_{\max} = O_p\left(\frac{\xi_N}{T}\right) \quad \text{and}$$

$$\max_{i\in I_3}\left\|\hat{\Delta}_{i,u}\right\|_2^2 = O_p\left(\eta_N^2\right) = o_p\left((N\vee T)^{-1/2}\right),$$

following similar arguments as in (A.9) and the proof of Theorem 2.2.

Next, we analyze the term $\hat{\mathbb{H}}_i\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)$ in (A.24), which can be written as

$$\begin{aligned}
&\hat{\mathbb{H}}_i\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left(\tau-\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right\}\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left(\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right)\\
&+\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left(\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right\}\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left[\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right)+\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left\{F_{it}(0)-F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right]\right\}\\
&+\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left(\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-F_{it}(0)\right]\\
&-\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left\{\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right\}-F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right]\right\}.
\end{aligned}\tag{A.26}$$

For the second term after the last equality, we notice that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left\{F_{it}(0)-F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right]\right\}\\
&=-\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}f_{it}\left(\tilde{\iota}_{it}\right)\left[(\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^0)' \quad (\hat{e}_{1,it}\dot{v}_{t,1}^{(1)}-e_{1,it}O_1^{(1)}v_{t,1}^0)'\right]\begin{bmatrix}O_0^{(1)}u_{i,0}^0\\ O_1^{(1)}u_{i,1}^0\end{bmatrix}\\
&-\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}f_{it}\left(\tilde{\iota}_{it}\right)\left(\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}-\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^0\right)\\
&:=\dot{D}_i^J\begin{bmatrix}O_0^{(1)}u_{i,0}^0\\ O_1^{(1)}u_{i,1}^0\end{bmatrix}+\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}f_{it}\left(\tilde{\iota}_{it}\right)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^0-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right),
\end{aligned}\tag{A.27}$$

where $|\tilde{\iota}_{it}|$ lies between 0 and $\left|\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right|$ and

$$\dot{D}_i^J=\frac{1}{T}\sum_{t=1}^{T}f_{it}(\tilde{\iota}_{it})\begin{bmatrix}\dot{v}_{t,0}^{(1)}\left(O_0^{(1)}v_{t,0}^0-\dot{v}_{t,0}^{(1)}\right)' & \dot{v}_{t,0}^{(1)}\left(e_{1,it}O_1^{(1)}v_{t,1}^0-\hat{e}_{it}\dot{v}_{t,1}^{(1)}\right)'\\ \hat{e}_{1,it}\dot{v}_{t,1}^{(1)}\left(O_0^{(1)}v_{t,0}^0-\dot{v}_{t,0}^{(1)}\right)' & \hat{e}_{1,it}\dot{v}_{t,1}^{(1)}\left(e_{1,it}O_1^{(1)}v_{t,1}^0-\hat{e}_{1,it}\dot{v}_{t,1}^{(1)}\right)'\end{bmatrix}.$$

The first equality above is due to the mean-value theorem and the definition for

$\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)$ in (A.22). Inserting (A.27) into (A.26), we obtain that

$$\begin{aligned}
&\hat{\mathbb{H}}_i\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right) \\
&= \dot{D}_i^J \begin{bmatrix} O_0^{(1)}u_{i,0}^0 \\ O_1^{(1)}u_{i,1}^0 \end{bmatrix} + \frac{1}{T}\sum_{t=1}^{T} \dot{\omega}_{it}\left[\tau - \mathbf{1}\{\varepsilon_{it} \le 0\}\right) \\
&+ \frac{1}{T}\sum_{t=1}^{T} \dot{\omega}_{it} f_{it}(\tilde{\iota}_{it})\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^0 - \hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right) \\
&+ \frac{1}{T}\sum_{t=1}^{T} \dot{\omega}_{it}\Bigg\{\left[\mathbf{1}\{\varepsilon_{it} \le 0\} - \mathbf{1}\left\{\varepsilon_{it} \le \iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right\}\right] \\
&- \left(F_{it}(0) - F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right]\right)\Bigg\} \\
&:= \dot{D}_i^J \begin{bmatrix} O_0^{(1)}u_{i,0}^0 \\ O_1^{(1)}u_{i,1}^0 \end{bmatrix} + \hat{\mathbb{I}}_{1,i} + \hat{\mathbb{I}}_{2,i} + \hat{\mathbb{I}}_{3,i}. \qquad (A.28)
\end{aligned}$$

Combining (A.25) and (A.28), we obtain that

$$\begin{aligned}
\hat{\Delta}_{i,u} = \left(\dot{D}_i^F\right)^{-1}\Bigg\{&\dot{D}_i^J \begin{bmatrix} O_0^{(1)}u_{i,0}^0 \\ O_1^{(1)}u_{i,1}^0 \end{bmatrix} + \hat{\mathbb{I}}_{1,i} + \hat{\mathbb{I}}_{2,i} + \hat{\mathbb{I}}_{3,i} + \hat{\mathbb{W}}_i\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)}\right) \\
&+ o_p\left((N \vee T)^{-1/2}\right)\Bigg\}, \qquad (A.29)
\end{aligned}$$

where the $o_p\left((N \vee T)^{-1/2}\right)$ term holds uniformly over $i \in I_3$. To prove Theorem 2.3(i), we analyze each term in (A.29) step by step.

Step 1: Uniform Convergence for $\dot{D}_i^F$ and $\dot{D}_i^J$.

Define

$$\begin{aligned}
D_i^F &= \frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\begin{bmatrix} O_0^{(1)}v_{t,0}^0 v_{t,0}^{0\prime}O_0^{(1)\prime} & 0 \\ 0 & e_{1,it}^2 O_1^{(1)}v_{t,1}^0 v_{t,1}^{0\prime}O_1^{(1)\prime} \end{bmatrix}, \\
D_i^J &= \frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\begin{bmatrix} \dot{v}_{t,0}^{(1)}\left(O_0^{(1)}v_{t,0}^0 - \dot{v}_{t,0}^{(1)}\right)' & 0 \\ 0 & e_{1,it}^2 O_1^{(1)}v_{t,1}^0\left(O_1^{(1)}v_{t,1}^0 - \dot{v}_{t,1}^{(1)}\right)' \end{bmatrix}.
\end{aligned}$$

Lemmas A.22 and A.23 show that

$$\max_{i\in I_3}||\dot{D}_i^F - D_i^F||_F = O_p(\eta_N) \quad \text{and} \quad \max_{i\in I_3}||\dot{D}_i^J - D_i^J||_F = \left\|\begin{bmatrix} O_p(\eta_N^2) & O_p(\eta_N) \\ O_p(\eta_N^2) & O_p(\eta_N^2) \end{bmatrix}\right\|_F,$$

with $\eta_N = \frac{\sqrt{\log(N\vee T)}\xi_N^2}{\sqrt{N\wedge T}}$.

Step 2: Uniform Convergence for $\hat{\mathbb{I}}_{1,i}$.

Let $\omega_{it}^{0}=\begin{bmatrix} O_0^{(1)}v_{t,0}^{0} \\ O_1^{(1)}v_{t,1}^{0}e_{1,it}\end{bmatrix}$. Then we can see that

$$\hat{\mathbb{I}}_{1,i}=\frac{1}{T}\sum_{t=1}^{T}\omega_{it}^{0}\left(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)+\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^{0}\right)\left(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\}\right).$$

Noting that

$$\begin{aligned}
&\dot{\omega}_{it}-\omega_{it}^{0}\\
&=\begin{bmatrix}\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^{0}\\ \dot{v}_{t,1}^{(1)}\hat{e}_{1,it}-O_1^{(1)}v_{t,1}^{0}e_{1,it}^{0}\end{bmatrix}\\
&=\begin{bmatrix}\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^{0}\\ \left(\dot{v}_{t,1}^{(1)}-O_0^{(1)}v_{t,1}^{0}\right)(\hat{e}_{1,it}-e_{1,it})+e_{1,it}\left(\dot{v}_{t,1}^{(1)}-O_0^{(1)}v_{t,1}^{0}\right)+O_1^{(1)}v_{t,1}^{(1)}(\hat{e}_{1,it}-e_{1,it})\end{bmatrix},
\end{aligned}\tag{A.30}$$

we have

$$\max_{i\in I_3}\frac{1}{T}\sum_{t\in[T]}\left\|\dot{\omega}_{it}-\omega_{it}^{0}\right\|_2=O_p(\eta_N).\tag{A.31}$$

In addition, $\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^{0}\right)\left(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)\right\|_2=O_p\left(\eta_N^2\right)$ by Lemma A.24. It follows that

$$\hat{\mathbb{I}}_{1,i}=\frac{1}{T}\sum_{t=1}^{T}\omega_{it}^{0}\left(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)+O_p\left(\eta_N^2\right),\tag{A.32}$$

uniformly over $i\in I_3$.

Step 3: Uniform Convergence for $\hat{\mathbb{I}}_{2,i}$.

Note that

$$\begin{aligned}
\hat{\mathbb{I}}_{2,i}&=\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}f_{it}(\tilde{\iota}_{it})\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\omega_{it}^{0}f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\\
&+\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^{0}\right)\left[f_{it}(\tilde{\iota}_{it})-f_{it}(0)\right]\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\\
&+\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^{0}\right)f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\\
&+\frac{1}{T}\sum_{t=1}^{T}\omega_{it}^{0}\left[f_{it}(\tilde{\iota}_{it})-f_{it}(0)\right]\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)
\end{aligned}\tag{A.33}$$

where

$$\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^{0}\right)\left[f_{it}\left(\tilde{\iota}_{it}\right)-f_{it}(0)\right]\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\right\|_2=O_p(\eta_N^3),$$

$$\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^{0}\right)f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\right\|_2=O_p(\eta_N^2),$$

$$\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\omega_{it}^{0}\left[f_{it}\left(\tilde{\iota}_{it}\right)-f_{it}(0)\right]\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\right\|_2=O_p(\eta_N^2).$$

To see why these three equalities hold, we focus on the third one. By Cauchy's inequality, Theorem 2.2, and Lemma A.21, we have

$$\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\omega_{it}^{0}\left[f_{it}\left(\tilde{\iota}_{it}\right)-f_{it}(0)\right]\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\right\|_2$$
$$\lesssim\sqrt{\max_{i\in I_3}\frac{1}{T}\sum_{t\in[T]}\left\|\omega_{it}^{0}\right\|_2^2\left|\tilde{\iota}_{it}\right|^2}\sqrt{\max_{i\in I_3}\frac{1}{T}\sum_{t\in[T]}\left|\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right|^2}=O_p\left(\eta_N^2\right).$$

For the first term on the right hand side (RHS) of the second equality of (A.33), we have by Lemma A.25

$$\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}O_0^{(1)}v_{t,0}^{0}f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\right\|_2=O_p(\eta_N),$$

$$\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}e_{1,it}O_1^{(1)}v_{t,1}^{0}f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\right\|_2=O_p\left(\eta_N^2\right),$$

and thus $\max_{i\in I_3}\left\|\hat{\mathbb{I}}_{2,i}\right\|_2=\left\|\begin{bmatrix}O_p(\eta_N)\\O_p(\eta_N^2)\end{bmatrix}\right\|_2$.

Step 4: Uniform Convergence for $\hat{\mathbb{I}}_{3,i}$.

Note that

$$\begin{aligned}\hat{\mathbb{I}}_{3,i}=&\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left\{\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(O_0^{(1)}u_{i,0}^{0},O_1^{(1)}u_{i,1}^{0},\dot{u}_{i,1}^{(1)}\right)\right\}\right]\right.\\&\left.-\left(F_{it}(0)-F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^{0},O_1^{(1)}u_{i,1}^{0},\dot{u}_{i,1}^{(1)}\right)\right]\right)\right\}\\=&\frac{1}{T}\sum_{t=1}^{T}\omega_{it}^{0}\left\{\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(O_0^{(1)}u_{i,0}^{0},O_1^{(1)}u_{i,1}^{0},\dot{u}_{i,1}^{(1)}\right)\right\}\right]\right.\\&\left.-\left(F_{it}(0)-F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^{0},O_1^{(1)}u_{i,1}^{0},\dot{u}_{i,1}^{(1)}\right)\right]\right)\right\}\\&+\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^{0}\right)\left\{\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(O_0^{(1)}u_{i,0}^{0},O_1^{(1)}u_{i,1}^{0},\dot{u}_{i,1}^{(1)}\right)\right\}\right]\right.\end{aligned}$$

$$-\left(F_{it}(0)-F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]\right)\Bigg\}. \tag{A.34}$$

By (A.132), we can show that

$$\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\le R_{\iota,it}^1\left(\left|\mu_{1,it}\right|+\left|e_{1,it}\right|\right)+R_{\iota,it}^2 \quad \text{with}$$
$$\max_{i\in I_3,t\in[T]}\left|R_{\iota,it}^1\right|=O_p(\eta_N),\ \max_{i\in I_3,t\in[T]}\left|R_{\iota,it}^2\right|=O_p(\eta_N). \tag{A.35}$$

For the second term on the RHS of the second equality in (A.34), we notice that

$$\max_{i\in I_3}\Bigg\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^0\right)\Bigg\{\left[\mathbf{1}\left\{\varepsilon_{it}\le 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right\}\right]$$
$$-\left(F_{it}(0)-F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]\right)\Bigg\}\Bigg\|_2$$
$$\le\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|\dot{\omega}_{it}-\omega_{it}^0\right\|_2\mathbf{1}\left\{0\le|\varepsilon_{it}|\le\left|\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right|\right\}$$
$$+\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|\dot{\omega}_{it}-\omega_{it}^0\right\|_2\left|F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]-F_{it}(0)\right|$$
$$\le\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|\dot{\omega}_{it}-\omega_{it}^0\right\|_2\mathbf{1}\left\{0\le|\varepsilon_{it}|\le\left|\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right|\right\}$$
$$+O_p(\eta_N^2), \tag{A.36}$$

where the last line is by (A.30), (A.35) and Assumption 2.1(iv).

Define the event

$$\mathscr{A}_9(M):=\left\{\max_{i\in I_3,t\in[T]}\left|\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right|\le M\eta_N\left(\left|\mu_{1,it}\right|+\left|e_{1,it}\right|+1\right)\right\}$$

with $\mathbb{P}\left\{\mathscr{A}_9^c(M)\right\}\le e$ for any $e>0$. Then for a large enough constant C_{17}, we have

$$\mathbb{P}\left(\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|\dot{\omega}_{it}-\omega_{it}^0\right\|_2\mathbf{1}\left\{0\le|\varepsilon_{it}|\le\left|\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right|\right\}>C_{17}\eta_N^2\right)$$
$$\le\mathbb{P}\left(\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|\dot{\omega}_{it}-\omega_{it}^0\right\|_2\mathbf{1}\left\{0\le|\varepsilon_{it}|\le\left|\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right|\right\}>C_{17}\eta_N^2\middle|\mathscr{A}_9(M)\right)+e$$
$$\le\mathbb{P}\left(\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|\dot{\omega}_{it}-\omega_{it}^0\right\|_2\mathbf{1}\left\{0\le|\varepsilon_{it}|\le M\eta_N\left(|\mu_{1,it}|+|e_{1,it}|+1\right)\right\}>C_{17}\eta_N^2\right)+e$$
$$\le\mathbb{E}\left\{\mathbb{P}\left(\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|\dot{\omega}_{it}-\omega_{it}^0\right\|_2\left[\mathbf{1}_{it}-\bar{\mathbf{1}}_{it}\right]>\frac{C_{17}\eta_N^2}{2}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)\right\}$$
$$+\mathbb{E}\left\{\mathbb{P}\left(\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\left\|\dot{\omega}_{it}-\omega_{it}^0\right\|_2\mathbf{1}\left\{0\le|\varepsilon_{it}|\le M\eta_N\left(|\mu_{1,it}|+|e_{1,it}|+1\right)\right\}\middle|\mathscr{D}_e^{I_1\cup I_2}\right\}>\frac{C_{17}\eta_N^2}{2}\right)\right\}+e$$
$$=o(1)+e \tag{A.37}$$

where

$$\mathbf{1}_{it}-\bar{\mathbf{1}}_{it}:=\mathbf{1}\left\{0\le|\varepsilon_{it}|\le M\eta_N\left(\left|\mu_{1,it}\right|+\left|e_{1,it}\right|+1\right)\right\}$$
$$-\mathbb{E}\left(\mathbf{1}\left\{0\le|\varepsilon_{it}|\le M\eta_N\left(\left|\mu_{1,it}\right|+\left|e_{1,it}\right|+1\right)\right\}\middle|\mathscr{D}_e^{I_1\cup I_2}\right),$$

the last line holds by the fact that

$$\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\left\|\dot{\omega}_{it}-\omega_{it}^0\right\|_2\mathbf{1}\left\{0\le|\varepsilon_{it}|\le M\eta_N\left(\left|\mu_{1,it}\right|+\left|e_{1,it}\right|+1\right)\right\}\middle|\mathscr{D}_e^{I_1\cup I_2}\right\}=O_p(\eta_N^2)$$

and

$$\mathbb{P}\left(\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|\dot{\omega}_{it}-\omega_{it}^0\right\|_2\left[\mathbf{1}_{it}-\bar{\mathbf{1}}_{it}\right]>C_{17}\eta_N^2\middle|\mathscr{D}_e^{I_1\cup I_2}\right)$$
$$\lesssim\exp\left(-\frac{T^2\eta_N^4}{T\xi_N^2\eta_N^2+T\eta_N^3\xi_N\log T\log\log T}\right)=o(1)$$

by Bernstein's inequality in Lemma A.12(i). Combining (A.36) and (A.37), we have shown the second term on the RHS of the second equality in (A.34) is $O_p(\eta_N^2)$ uniformly over $i\in I_3$. This result, in conjunction with Lemma A.25, (A.34) and Assumption 2.1(ix), implies that $\max_{i\in I_3}\left\|\hat{\mathbb{I}}_{3,i}\right\|=o_p\left((N\vee T)^{-\frac{1}{2}}\right)$.

Step 5: Uniform Convergence for $\hat{\mathbb{W}}_i\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)$.

Note that

$$\begin{aligned}
&\hat{\mathbb{W}}_i\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left(\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right\}-\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right\}\right)\\
&-\frac{1}{T}\sum_{t=1}^{T}\dot{\omega}_{it}\left(F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]-F_{it}\left[\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right]\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^0\right)\left(\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right\}-\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right\}\right)\\
&-\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^0\right)\left(F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]-F_{it}\left[\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right]\right)\\
&+\frac{1}{T}\sum_{t=1}^{T}\omega_{it}^0\left\{\left(\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right\}-\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right\}\right)\right.\\
&\left.-\left(F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]-F_{it}\left[\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right]\right)\right\}\\
&:=\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^0\right)\left(\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right\}-\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right\}\right)
\end{aligned}$$

$$-\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^{0}\right)\left(F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]-F_{it}\left[\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right]\right)$$
$$+\hat{\mathbb{W}}_i^{I}-\hat{\mathbb{W}}_i^{II},$$

where we define

$$\hat{\mathbb{W}}_i^{I}=\frac{1}{T}\sum_{t=1}^{T}\omega_{it}^{0}\Bigg\{\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right\}-\mathbf{1}\{\varepsilon_{it}\leq 0\}$$
$$-\left\{F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]-F_{it}(0)\right\}\Bigg\}\quad\text{and}$$
$$\hat{\mathbb{W}}_i^{II}=\frac{1}{T}\sum_{t=1}^{T}\omega_{it}^{0}\Bigg\{\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right\}$$
$$-\left\{F_{it}(0)-F_{it}\left[\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)},\right)\right]\right\}\Bigg\}.$$

We first observe that

$$\begin{aligned}&\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)-\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\\&=\left(\hat{u}_{i,0}^{(3,1)}-O_0^{(1)}u_{i,0}^0\right)'\dot{v}_{t,0}^{(1)}+\left(\hat{u}_{i,1}^{(3,1)}-O_1^{(1)}u_{i,1}^0\right)'\dot{v}_{t,1}^{(1)}\hat{e}_{1,it}\\&=\left(O_0^{(1)\prime}\hat{u}_{i,0}^{(3,1)}-u_{i,0}^0\right)'v_{t,0}^0+\left(O_1^{(1)\prime}\hat{u}_{i,1}^{(3,1)}-u_{i,1}^0\right)'v_{t,1}^{(1)}e_{1,it}+O_p(\eta_N^2)\\&=R_{\iota,it}^{3}e_{1,it}+R_{\iota,it}^{4}\end{aligned}\tag{A.38}$$

such that $\max_{i\in I_3,t\in[T]}\left|R_{\iota,it}^{3}\right|=O_p(\eta_N)$ and $\max_{i\in I_3,t\in[T]}\left|R_{\iota,it}^{4}\right|=O_p(\eta_N)$. As in Step 4, we can show that

$$\max_{i\in I_3}\Bigg\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^{0}\right)\Bigg(\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right\}$$
$$-\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right\}\Bigg)\Bigg\|=O_p(\eta_N^2)$$

and

$$\max_{i\in I_3}\Bigg\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{\omega}_{it}-\omega_{it}^{0}\right)\Bigg(F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]$$
$$-F_{it}\left[\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right]\Bigg)\Bigg\|=O_p(\eta_N^2).$$

Then by Lemma A.26, Lemma A.27 and Assumption 2.1(ix), we obtain that

$$\max_{i\in I_3}\left\|\hat{\mathbb{W}}_i\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right\|_2=o_p\left((N\vee T)^{-\frac{1}{2}}\right).$$

Step 6: Distribution Theory for $\hat{\Delta}_{i,u}$

Combining the above results, we have that uniformly over $i \in I_3$,

$$\begin{aligned}
\begin{bmatrix} \hat{u}_{i,0}^{(3,1)} \\ \hat{u}_{i,1}^{(3,1)} \end{bmatrix} &= \begin{bmatrix} O_0^{(1)} u_{i,0}^0 \\ O_1^{(1)} u_{i,1}^0 \end{bmatrix} + \left(\dot{D}_i^F\right)^{-1} \left\{ \dot{D}_i^J \begin{bmatrix} O_0^{(1)} u_{i,0}^0 \\ O_1^{(1)} u_{i,1}^0 \end{bmatrix} + \hat{\mathbb{I}}_{1,i} + \hat{\mathbb{I}}_{2,i} + \mathbb{I}_{3,i} \right. \\
&\quad \left. + \hat{\mathbb{W}}_i \left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \dot{u}_{i,1}^{(1)} \right) + o_p\left((N \vee T)^{-\frac{1}{2}} \right) \right\} \\
&= \left[I_{K_0+K_1} + \left(D_i^F\right)^{-1} D_i^J \right] \begin{bmatrix} O_0^{(1)} u_{i,0}^0 \\ O_1^{(1)} u_{i,1}^0 \end{bmatrix} + \left[\left(\dot{D}_i^F\right)^{-1} \dot{D}_i^J - \left(D_i^F\right)^{-1} D_i^J \right] \begin{bmatrix} O_0^{(1)} u_{i,0}^0 \\ O_1^{(1)} u_{i,1}^0 \end{bmatrix} \\
&\quad + \left(D_i^F\right)^{-1} \frac{1}{T} \sum_{t=1}^{T} \omega_{it}^0 \left(\tau - \mathbf{1}\{\varepsilon_{it} \le 0\} \right) \\
&\quad + \left[\left(\dot{D}_i^F\right)^{-1} - \left(D_i^F\right)^{-1} \right] \frac{1}{T} \sum_{t=1}^{T} \omega_{it}^0 \left(\tau - \mathbf{1}\{\varepsilon_{it} \le 0\} \right) \\
&\quad + \left(D_i^F\right)^{-1} \hat{\mathbb{I}}_{2,i} + o_p\left((N \vee T)^{-\frac{1}{2}} \right). \qquad \text{(A.39)}
\end{aligned}$$

Owing to the fact that

$$\max_{i \in I_3} \left\| \frac{1}{T} \sum_{t=1}^{T} \omega_{it}^0 \left(\tau - \mathbf{1}\{\varepsilon_{it} \le 0\} \right) \right\|_2 = O_p\left(\sqrt{\frac{\log(N \vee T)}{T}} \right) \qquad \text{(A.40)}$$

by similar arguments as (A.116) using Bernstein's inequality in Lemma A.12(ii) and Lemma A.22, we notice that

$$\max_{i \in I_3} \left\| \left[\left(\dot{D}_i^F\right)^{-1} - \left(D_i^F\right)^{-1} \right] \frac{1}{T} \sum_{t=1}^{T} \omega_{it}^0 \left(\tau - \mathbf{1}\{\varepsilon_{it} \le 0\} \right) \right\|_2 = o_p\left((N \vee T)^{-\frac{1}{2}} \right).$$

Next, we define

$$D^F = O^{(1)} \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[f_{it}(0) \middle| \mathscr{D} \right] v_{t,0}^0 v_{t,0}^{0\prime} & 0 \\ 0 & \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[e_{1,it}^2 f_{it}(0) \middle| \mathscr{D} \right] v_{t,1}^0 v_{t,1}^{0\prime} \end{bmatrix} O^{(1)\prime},$$

$$\begin{aligned}
D^J = diag\Bigg(&\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[f_{it}(0) \middle| \mathscr{D} \right] \dot{v}_{t,0}^{(1)} \left(O_0^{(1)} v_{t,0}^0 - \dot{v}_{t,0}^{(1)} \right)', \\
&O_1^{(1)} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[e_{1,it}^2 f_{it}(0) \middle| \mathscr{D} \right] v_{t,1}^0 \left(O_1^{(1)} v_{t,1}^0 - \dot{v}_{t,1}^{(1)} \right)' \Bigg),
\end{aligned}$$

where $O^{(1)} = diag\left(O_0^{(1)}, O_1^{(1)} \right)$. Here D^F and D^J do not depend on i owing to stationary assumption of sequence $\left\{ f_{it}, f_{it}(0) e_{j,it} \right\}_{j \in [p]}$ conditional on all factors in Assumption 2.9(iii). By Bernstein's inequality conditional on all factors similarly as in (A.116), we can show that $\max_{i \in I_3} \left\| D_i^F - D^F \right\|_F = O_p\left(\sqrt{\frac{\log(N \vee T)}{T}} \xi_N \right)$. Analogously, by Bernstein's inequality conditional on $\mathscr{D}^{I_1 \cup I_2}$, we can show that

$\max_{i\in I_3}\left\|D_i^J - D_j\right\|_F = O_p\left(\sqrt{\frac{\log(N\vee T)}{T}}\eta_N\xi_N\right)$. Then it follows that

$$\max_{i\in I_3}\left\|\left(D_i^F\right)^{-1}D_i^J - \left(D^F\right)^{-1}D^J\right\|_F = O_p(\eta_N^2).$$

In addition, uniformly over $i\in I_3$,

$$\begin{aligned}
&\left(\dot{D}_i^F\right)^{-1}\dot{D}_i^J - \left(D_i^F\right)^{-1}D_i^J \\
&= \left[\left(\dot{D}_i^F\right)^{-1} - \left(D_i^F\right)^{-1}\right]\left[\dot{D}_i^J - D_i^J\right] + D_i^J\left[\left(\dot{D}_i^F\right)^{-1} - \left(D_i^F\right)^{-1}\right] + \left(D_i^F\right)^{-1}\left[\dot{D}_i^J - D_i^J\right] \\
&= \left(D_i^F\right)^{-1}\left[\dot{D}_i^J - D_i^J\right] + O_p(\eta_N^2) \\
&= \begin{bmatrix} O_p\left(\eta_N^2\right) & O_p(\eta_N) \\ O_p\left(\eta_N^2\right) & O_p\left(\eta_N^2\right)\end{bmatrix},
\end{aligned}$$

where the upper right block is dominated by $\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)O_0^{(1)}v_{t,0}^0 v_{t,1}^{0\prime}O_1^{(1)\prime}(e_{1,it}-\hat{e}_{1,it})$ in the analysis of $J_{1,i}^2$ in (A.121).

Let $I_{K_0+K_1} + \left(D^F\right)^{-1}D^J = O_{u,1}^{(1)} := \begin{pmatrix}\bar{O}_{u,0} & 0 \\ 0 & \bar{O}_{u,1}\end{pmatrix}$, $O_{u,0}^{(1)} = \bar{O}_{u,0}O_0^{(1)}$, and $O_{u,1}^{(1)} = \bar{O}_{u,1}O_1^{(1)}$. Combining the above arguments, we obtain that

$$\hat{u}_{i,1}^{(3,1)} - O_{u,1}^{(1)}u_{i,1}^0 = O_1^{(1)}\hat{V}_{u_1}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_{1,it}v_{t,1}^0\left(\tau - \mathbf{1}\{\varepsilon_{it}\le 0\}\right) + \mathscr{R}_{i,u}^1, \tag{A.41}$$

$$\begin{aligned}
\hat{u}_{i,0}^{(3,1)} - O_{u,0}^{(1)}u_{i,0}^0 = {} & O_0^{(1)}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t=1}^{T} v_{t,0}^0\left(\tau - \mathbf{1}\{\varepsilon_{it}\le 0\}\right) \\
& + O_0^{(1)}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)v_{t,0}^0 v_{t,1}^{0\prime}u_{i,1}^0\left(e_{1,it} - \hat{e}_{1,it}\right) \\
& + O_0^{(1)}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)v_{t,0}^0\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^0 - \hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right) + \mathscr{R}_{i,u}^0
\end{aligned} \tag{A.42}$$

such that

$$\max_{i\in I_3}\left|\mathscr{R}_{i,u}^0\right| = o_p\left((N\vee T)^{-\frac{1}{2}}\right), \max_{i\in I_3}\left|\mathscr{R}_{i,u}^1\right| = o_p\left(\frac{1}{\sqrt{T}}\right),$$

$$\hat{V}_{u_0} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[f_{it}(0)\middle|\mathscr{D}\right]v_{t,0}^0 v_{t,0}^{0\prime} \text{ and } \hat{V}_{u_1} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[f_{it}(0)e_{1,it}^2\middle|\mathscr{D}\right]v_{t,1}^0 v_{t,1}^{0\prime}.$$

From (A.41), owing to the fact that $\hat{V}_{u_1}$ is bounded a.s. and

$$\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T} e_{1,it}v_{t,1}^0\left(\tau - \mathbf{1}\{\varepsilon_{it}\le 0\}\right)\right\|_2 = O_p\left(\sqrt{\frac{\log(N\vee T)}{T}}\right)$$

by Bernstein's inequality in Lemma A.12(ii), we obtain that

$$\max_{i\in I_3}\left\|\hat{u}_{i,1}^{(3,1)}-O_{u,1}^{(1)}u_{i,1}^{0}\right\|_2=O_p\left(\sqrt{\frac{\log(N\vee T)}{T}}\right)\quad\text{and}$$
$$\max_{i\in I_3}\left\|\hat{u}_{i,0}^{(3,1)}-O_{u,0}^{(1)}u_{i,0}^{0}\right\|_2=O_p(\eta_N). \tag{A.43}$$

Last, noting that $O_1^{(1)}$ is a rotation matrix and the normal distribution is invariant to rotation, for each $i\in I_3$, we have

$$\sqrt{T}\left(\hat{u}_{i,1}^{(3,1)}-O_{u,1}^{(1)}u_{i,1}^{0}\right)\rightsquigarrow\mathbb{N}(0,\Sigma_{u_1,i}),$$

where $\Sigma_{u_1,i}=O_1^{(1)}V_{u_1}^{-1}\Omega_{u_1}V_{u_1}^{-1}O_1^{(1)\prime}$, $V_{u_1,i}=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(f_{it}(0)e_{1,it}^2v_{t,1}^{0}v_{t,1}^{0\prime}\right)$, and

$$\Omega_{u_1,i}=Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_{1,it}v_{t,1}^{0}\left(\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right)\right].\quad\blacksquare$$

Proof of Statement (ii)

Steps for the proof for statement (ii) are the same as those in the proof of statement (i). Hence, we only sketch the proof. Recall from (2.8) that $\forall t\in[T]$,

$$\{\hat{v}_{t,j}^{(3,1)}\}_{j\in[p]}=\underset{\{v_{t,j}\}_{j\in[p]}}{\arg\min}\frac{1}{N_3}\sum_{i\in I_3}\rho_\tau\left(\hat{Y}_{it}-v_{t,0}'\hat{u}_{i,0}^{(3,1)}-v_{t,1}'\hat{u}_{i,1}^{(3,1)}\hat{e}_{1,it}\right),$$

where $\hat{Y}_{it}=Y_{it}-\hat{\mu}_{1,it}\hat{u}_{i,1}^{(3,1)\prime}\dot{v}_{t,1}^{(1)}$. Let

$$\hat{\Delta}_{t,v}=\begin{bmatrix}\hat{v}_{t,0}^{(3,1)}-\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^{0}\\ \hat{v}_{t,1}^{(3,1)}-\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^{0}\end{bmatrix}\text{ and }\hat{\omega}_{it}=\begin{bmatrix}\hat{u}_{i,0}^{(3,1)}\\ \hat{u}_{i,1}^{(3,1)}\hat{e}_{1,it}\end{bmatrix}.$$

For generic $(u_{i,0},u_{i,1},v_{i,0},v_{i,1})$, define

$$\hat{\mathbb{S}}_t\left(u_{i,0},u_{i,1},v_{i,0},v_{i,1}\right)=\frac{1}{N_3}\sum_{i\in I_3}\left[\tau-\mathbf{1}\left\{\varepsilon_{it}\leq\rho_{it}\left(u_{i,0},u_{i,1},v_{i,0},v_{i,1}\right)\right\}\right]\hat{\omega}_{it},$$

with

$$\rho_{it}\left(u_{i,0},u_{i,1},v_{i,0},v_{i,1}\right)=u_{i,0}'v_{t,0}-u_{i,0}^{0\prime}v_{t,0}^{0}+\hat{\mu}_{1,it}u_{i,1}'\dot{v}_{t,1}^{(1)}-\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}+\hat{e}_{1,it}u_{i,1}'v_{t,1}-e_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}.$$

We also define

$$\hat{\mathscr{S}}_t\left(u_{i,0},u_{i,1},v_{i,0},v_{i,1}\right)=\frac{1}{N_3}\sum_{i\in I_3}\mathbb{E}\left\{\left[\tau-\mathbf{1}\left\{\varepsilon_{it}\leq\left(\rho_{it}\left(u_{i,0},u_{i,1},v_{i,0},v_{i,1}\right)\right)\right\}\right]\dot{\omega}_{it}\middle|\mathscr{D}_e^{I_1\cup I_2}\right\}$$
$$=\frac{1}{N_3}\sum_{i\in I_3}\left[\tau-F_{it}\left(\rho_{it}\left(u_{i,0},u_{i,1},v_{i,0},v_{i,1}\right)\right)\right]\dot{\omega}_{it}$$

and

$$\hat{\mathbb{M}}_t\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\hat{v}_{t,0}^{(3,1)},\hat{v}_{t,1}^{(3,1)}\right)$$
$$=\hat{\mathbb{S}}_t\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\hat{v}_{t,0}^{(3,1)},\hat{v}_{t,1}^{(3,1)}\right)-\hat{\mathbb{S}}_t\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^{0},\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^{0}\right)$$
$$-\left\{\hat{\mathscr{S}}_t\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\hat{v}_{t,0}^{(3,1)},\hat{v}_{t,1}^{(3,1)}\right)-\hat{\mathscr{S}}_t\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^{0},\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^{0}\right)\right\}.$$

Then a similar result to that in (A.29) holds:

$$\hat{\Delta}_{t,v}=\left(\hat{D}_t^F\right)^{-1}\left\{\hat{D}_t^J\begin{bmatrix}\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^{0}\\\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^{0}\end{bmatrix}+\hat{\mathbb{I}}_{4,t}+\hat{\mathbb{I}}_{5,t}+\hat{\mathbb{I}}_{6,t}\right.$$
$$\left.+\hat{\mathbb{M}}_t\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\hat{v}_{t,0}^{(1)},\hat{v}_{t,1}^{(3,1)}\right)+o_p\left((N\vee T)^{-\frac{1}{2}}\right)\right\},\qquad\text{(A.44)}$$

where

$$\hat{D}_t^F=\frac{1}{N_3}\sum_{i\in I_3}f_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^{0},\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^{0}\right)\right]\cdot$$
$$\begin{bmatrix}\hat{u}_{i,0}^{(3,1)}\hat{u}_{i,0}^{(3,1)\prime}&\hat{e}_{1,it}\hat{u}_{i,0}^{(3,1)}\hat{u}_{i,1}^{(3,1)\prime}\\\hat{e}_{1,it}\hat{u}_{i,1}^{(3,1)}\hat{u}_{i,0}^{(3,1)\prime}&\hat{e}_{1,it}^{2}\hat{u}_{i,1}^{(3,1)}\hat{u}_{i,1}^{(3,1)\prime}\end{bmatrix},$$

$$\hat{D}_t^J=\frac{1}{N_3}\sum_{i\in I_3}f_{it}(\tilde{\rho}_{it})\begin{bmatrix}\hat{u}_{i,0}^{(3,1)}\left(O_{u,0}^{(1)}u_{i,0}^{0}-\hat{u}_{i,0}^{(3,1)}\right)'&\hat{u}_{i,0}^{(3,1)}\left(e_{1,it}O_{u,1}^{(1)}u_{i,1}^{0}-\hat{e}_{it}\hat{u}_{i,1}^{(3,1)}\right)'\\\hat{e}_{1,it}\hat{u}_{i,1}^{(3,1)}\left(O_{u,0}^{(1)}u_{i,0}^{0}-\hat{u}_{i,0}^{(3,1)}\right)'&\hat{e}_{1,it}\hat{u}_{i,1}^{(3,1)}\left(e_{1,it}O_{u,1}^{(1)}u_{i,1}^{0}-\hat{e}_{1,it}\hat{u}_{i,1}^{(3,1)}\right)'\end{bmatrix},$$

$$\hat{\mathbb{I}}_{4,t}=\frac{1}{N_3}\sum_{i\in I_3}\hat{\omega}_{it}\left[\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\}\right),$$

$$\hat{\mathbb{I}}_{5,t}=\frac{1}{N_3}\sum_{i\in I_3}\hat{\omega}_{it}f_{it}\left(\tilde{\rho}_{it}\right)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\hat{u}_{i,1}^{(3,1)\prime}\dot{v}_{t,1}^{(1)}\right),$$

$$\hat{\mathbb{I}}_{6,t}=\frac{1}{N_3}\sum_{i\in I_3}\hat{\omega}_{it}\left\{\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{\varepsilon_{it}\leq\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^{0},\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^{0}\right)\right\}\right]\right.$$
$$\left.-\left(f_{it}(0)-f_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^{0},\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^{0}\right)\right]\right)\right\},$$

and $|\tilde{\rho}_{it}|$ lies between 0 and $\left|\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^{0},\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^{0}\right)\right|$.

Then, we derive the linear expansion for $\hat{\Delta}_{t,v}$ by analyzing each term in (A.44).

Define

$$D_t^F=\frac{1}{N_3}\sum_{i\in I_3}f_{it}(0)\begin{bmatrix}O_{u,0}^{(1)}u_{i,0}^{0}u_{i,0}^{0\prime}O_{u,0}^{(1)\prime}&0\\0&e_{1,it}^{2}O_{u,1}^{(1)}u_{i,1}^{0}u_{i,1}^{0\prime}O_{u,1}^{(1)\prime}\end{bmatrix},$$

$$D_t^J=\frac{1}{N_3}\sum_{i\in I_3}f_{it}(0)\begin{bmatrix}\hat{u}_{i,0}^{(3,1)}\left(O_{u,0}^{(1)}u_{i,0}^0-\hat{u}_{i,0}^{(3,1)}\right)' & 0\\ 0 & e_{1,it}^2O_{u,1}^{(1)}u_{i,1}^0\left(O_{u,1}^{(1)}u_{i,1}^0-\hat{u}_{i,1}^{(3,1)}\right)'\end{bmatrix},$$

such that

$$\max_{t\in[T]}\left\|\hat{D}_t^F-D_t^F\right\|_F=O_p(\eta_N)\text{ and }\max_{t\in[T]}\left\|\hat{D}_t^J-D_t^J\right\|_F=\left\|\begin{bmatrix}O_p\left(\eta_N^2\right) & O_p(\eta_N)\\ O_p\left(\eta_N^2\right) & O_p\left(\eta_N^2\right)\end{bmatrix}\right\|_F$$

by Lemma A.28. Let $\omega_{it}^*=\begin{bmatrix}O_{u,0}^{(1)}u_{i,0}^0\\ e_{1,it}O_{u,1}^{(1)}u_{i,1}^0\end{bmatrix}$. We can show that

$$\hat{\mathbb{I}}_{4,t}=\frac{1}{N_3}\sum_{i\in I_3}\omega_{it}^*\left[\tau-\mathbf{1}\left\{\varepsilon_{it}\le 0\right\}\right]+O_p\left(\eta_N^2\right)$$

uniformly over $t\in[T]$ by analogous analysis as in Step 2 in the previous subsection. With Lemma A.29 and by similar arguments as in Step 3 in the previous subsection, we obtain

$$\max_{t\in[T]}\left\|\hat{\mathbb{I}}_{5,t}\right\|_2=\left\|\begin{bmatrix}O_p(\eta_N)\\ o_p\left((N\vee T)^{-\frac{1}{2}}\right)\end{bmatrix}\right\|_2$$

uniformly over $t\in[T]$.

Next, for $\hat{\mathbb{I}}_{6,t}$, From (A.146), we first note that

$$\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^0\right)\le R_{\rho,it}^1\left(\left|\mu_{1,it}\right|+\left|e_{1,it}\right|\right)+R_{\rho,it}^2\quad\text{with}$$

$$\max_{i\in I_3,t\in[T]}\left|R_{\rho,it}^1\right|=O_p(\eta_N),\ \max_{i\in I_3,t\in[T]}\left|R_{\rho,it}^2\right|=O_p(\eta_N).\tag{A.45}$$

We then observe that, uniformly over $t\in[T]$,

$$\begin{aligned}\hat{\mathbb{I}}_{6,t}=&\frac{1}{N_3}\sum_{i\in I_3}\hat{\omega}_{it}\Bigg\{\left[\mathbf{1}\left\{\varepsilon_{it}\le 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\le\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^0\right)\right\}\right]\\&-\left(F_{it}(0)-F_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^0\right)\right]\right)\Bigg\}\\=&\frac{1}{N_3}\sum_{i\in I_3}\omega_{it}^*\Bigg\{\left[\mathbf{1}\left\{\varepsilon_{it}\le 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\le\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^0\right)\right\}\right]\\&-\left(F_{it}(0)-F_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^0\right)\right]\right)\Bigg\}\\&+\frac{1}{N_3}\sum_{i\in I_3}\left(\hat{\omega}_{it}-\omega_{it}^*\right)\Bigg\{\left[\mathbf{1}\left\{\varepsilon_{it}\le 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\le\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^0\right)\right\}\right]\\&-\left(F_{it}(0)-F_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^0\right)\right]\right)\Bigg\}\end{aligned}$$

$$= \frac{1}{N_3} \sum_{i \in I_3} \omega_{it}^* \Bigg\{ \left[\mathbf{1}\{\varepsilon_{it} \le 0\} - \mathbf{1}\left\{ \varepsilon_{it} \le \rho_{it}\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \left(O_{u,0}^{(1)}\right)^{\prime -1} v_{t,0}^0, \left(O_{u,1}^{(1)}\right)^{\prime -1} v_{t,1}^0 \right) \right\} \right]$$

$$- \left(F_{it}(0) - F_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \left(O_{u,0}^{(1)}\right)^{\prime -1} v_{t,0}^0, \left(O_{u,1}^{(1)}\right)^{\prime -1} v_{t,1}^0 \right) \right] \right) \Bigg\} + o_p\left((N \vee T)^{-\frac{1}{2}} \right)$$

$$= \begin{bmatrix} O_p(\eta_N) \\ o_p\left((N \vee T)^{-\frac{1}{2}} \right) \end{bmatrix},$$

where the third equality holds by (A.45) and similar arguments as used in (A.36) and (A.37), and the last equality holds by Lemma A.30.

Similarly, combining Lemma A.30 and Lemma A.31, it yields

$$\max_{t \in [T]} \left\| \hat{\mathbb{M}}_t \left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \hat{v}_{t,0}^{(3,1)}, \hat{v}_{t,1}^{(3,1)} \right) \right\|_2 = \left\| \begin{bmatrix} O_p(\eta_N) \\ o_p\left((N \vee T)^{-\frac{1}{2}} \right) \end{bmatrix} \right\|_2 .$$

Let $\hat{V}_{v_1,t}^3 = \frac{1}{N_3} \sum_{i \in I_3} f_{it}(0) e_{1,it}^2 u_{i,1}^0 u_{i,1}^{0\prime}$, and

$$O_{v_1,t}^{(1)} = \left\{ I_{K_1} + \left(O_{u,1}^{(1)\prime} \right)^{-1} \left[\hat{V}_{v_1,t}^3 \right]^{-1} \left[\frac{1}{N_3} \sum_{i \in I_3} f_{it}(0) e_{1,it}^2 u_{i,1}^0 \left(O_{u,1}^{(1)} u_{i,1}^0 - \hat{u}_{i,1}^{(3,1)} \right)' \right] \right\} \left(O_{u,1}^{(1)\prime} \right)^{-1}.$$

Combining arguments above from (A.44), we have

$$\hat{v}_{t,1}^{(3,1)} - O_{v_1,t}^{(1)} v_{t,1}^0 = \left(O_{u,1}^{(1)\prime} \right)^{-1} \left(\hat{V}_{v_1,t}^3 \right)^{-1} \frac{1}{N_3} \sum_{i \in I_3} e_{1,it} u_{i,1}^0 \left(\tau - \mathbf{1}\{\varepsilon_{it} \le 0\} \right) + \mathscr{R}_{t,v}^1,$$

such that $\max_{t \in [T]} \left| \mathscr{R}_{t,v}^1 \right| = o_p\left((N \vee T)^{-\frac{1}{2}} \right)$. This, in conjunction with the result in Lemma A.32, i.e., $\max_{t \in [T]} \left\| O_{v_1,t}^{(1)} - \left(O_{u,1}^{(1)\prime} \right)^{-1} \right\|_F = o_p\left((N \vee T)^{-\frac{1}{2}} \right)$, implies that

$$\begin{aligned} \hat{v}_{t,1}^{(3,1)} - \left(O_{u,1}^{(1)} \right)^{\prime -1} v_{t,1}^0 &= \left(O_{u,1}^{(1)\prime} \right)^{-1} \left(\hat{V}_{v_1,t}^3 \right)^{-1} \frac{1}{N_3} \sum_{i \in I_3} e_{1,it} u_{i,1}^0 \left(\tau - \mathbf{1}\{\varepsilon_{it} \le 0\} \right) + \mathscr{R}_{t,v}^1, \\ &= O_1^{(1)} \left(\hat{V}_{v_1,t}^3 \right)^{-1} \frac{1}{N_3} \sum_{i \in I_3} e_{1,it} u_{i,1}^0 \left(\tau - \mathbf{1}\{\varepsilon_{it} \le 0\} \right) + \mathscr{R}_{t,v}^1. \end{aligned} \tag{A.46}$$

where the second line holds by the fact that $\left\| O_{u,1}^{(1)} - O_1^{(1)} \right\|_F = O_p(\eta_N)$, $\hat{V}_{v_1,t}^I$ is uniformly bounded and that

$$\max_{t \in [T]} \left\| \frac{1}{N_3} \sum_{i \in I_3} e_{1,it} u_{i,1}^0 \left(\tau - \mathbf{1}\{\varepsilon_{it} \le 0\} \right) \right\|_2 = O_p\left(\sqrt{\frac{\log(N \vee T)}{N}} \right)$$

by similar arguments as in (A.40). Further define

$$\Sigma_{v_1} = O_1^{(1)} V_{v_1}^{-1} \Omega_{u_1} V_{u_1}^{-1} O_1^{(1)\prime}, \quad V_{v_1} = \frac{1}{N_3} \sum_{i \in I_3} \mathbb{E}\left(f_{it}(0) e_{1,it}^2 u_{i,1}^0 u_{i,1}^{0\prime} \right), \text{ and}$$

$$\Omega_{v_1} = \tau(1-\tau)\frac{1}{N_3}\sum_{i\in I_3}\mathbb{E}\left(e_{1,it}^2 u_{i,1}^0 u_{i,1}^{0\prime}\right).$$

Then we have

$$\begin{aligned}
&\sqrt{N_3}\left(\hat{v}_{t,1}^{(3,1)} - \left(O_{u,1}^{(1)}\right)^{\prime -1} v_{t,1}^0\right) \rightsquigarrow \mathbb{N}(0,\Sigma_{v_1}),\\
&\max_{t\in[T]}\left\|\hat{v}_{t,1}^{(3,1)} - O_{u,1}^{(1)} v_{t,1}^0\right\|_2 = O_p\left(\sqrt{\frac{\log(N\vee T)}{N}}\right), \quad \text{and}\\
&\max_{t\in[T]}\left\|\hat{v}_{t,0}^{(3,1)} - O_{u,0}^{(1)} v_{t,0}^0\right\|_2 = O_p(\eta_N),
\end{aligned} \tag{A.47}$$

where the second line holds by Bernstein's inequality with independent data and is similar to (A.43). ■

A.1.4 Proof of Proposition 2.1

Proof of Statement (i)

Focusing on the slope estimators for $i \in I_3$, we notice that

$$\hat{\Theta}_{j,it} = \frac{1}{2}\left\{\hat{u}_{t,j}^{(3,1)\prime}\hat{v}_{t,j}^{(3,1)} + \hat{u}_{t,j}^{(3,2)\prime}\hat{v}_{t,j}^{(3,2)}\right\}.$$

It follows that

$$\begin{aligned}
&\hat{\Theta}_{j,it} - \Theta_{j,it}^0\\
&= \frac{1}{2}\Big\{\left(\hat{u}_{i,j}^{(3,1)} - O_{u,j}^{(1)} u_{i,j}^0\right)^\prime\left(\hat{v}_{t,j}^{(3,1)} - O_{u,j}^{(1)} v_{t,j}^0\right) + \left(O_{u,j}^{(1)} u_{i,j}^0\right)^\prime\left(\hat{v}_{t,j}^{(3,1)} - O_{u,j}^{(1)} v_{t,j}^0\right)\\
&+ \left(\hat{u}_{i,j}^{(3,1)} - O_{u,j}^{(1)} u_{i,j}^0\right)^\prime O_{u,j}^{(1)} v_{t,j}^0\Big\} + \frac{1}{2}\Big\{\left(\hat{u}_{i,j}^{(3,2)} - O_{u,j}^{(2)} u_{i,j}^0\right)^\prime\left(\hat{v}_{t,j}^{(3,2)} - O_{u,j}^{(2)} v_{t,j}^0\right)\\
&+ \left(O_{u,j}^{(2)} u_{i,j}^0\right)^\prime\left(\hat{v}_{t,j}^{(3,2)} - O_{u,j}^{(2)} v_{t,j}^0\right) + \left(\hat{u}_{i,j}^{(3,2)} - O_{u,j}^{(2)} u_{i,j}^0\right)^\prime O_{u,j}^{(1)} v_{t,j}^0\Big\}\\
&= u_{i,j}^{0\prime}\left(\hat{V}_{v_j,t}^{(3)}\right)^{-1}\frac{1}{N_3}\sum_{i\in I_3} e_{j,it} u_{i,j}^0\left(\tau - \mathbf{1}\{\varepsilon_{it}\le 0\}\right)\\
&+ v_{t,j}^{0\prime}\hat{V}_{u_j}^{-1}\frac{1}{T}\sum_{t=1}^T e_{j,it} v_{t,j}^0\left(\tau - \mathbf{1}\{\varepsilon_{it}\le 0\}\right) + \mathscr{R}_{it}^j\\
&= u_{i,j}^{0\prime}\left(\hat{V}_{v_j,t}^{(3)}\right)^{-1}\frac{1}{N_3}\sum_{i\in I_3}\xi_{j,it}^0 + v_{t,j}^{0\prime}\hat{V}_{u_j}^{-1}\frac{1}{T}\sum_{t=1}^T b_{j,it}^0 + \mathscr{R}_{it}^j,
\end{aligned}$$

such that $\max_{i\in I_3, t\in[T]}\left|\mathscr{R}_{it}^j\right| = o_p\left((N\vee T)^{-1/2}\right)$ by Theorem 2.3 and the second equality above combines Theorem 2.3 and the fact that $\left\|O_{u,1}^{(1)} - O_1^{(1)}\right\|_F = O_p(\eta_N)$. With similar results hold for slope estimators for subsamples I_1 and I_2, and then we obtain the statement (i).

Proof of Statement (ii)

Combining (A.43), (A.47) and Lemma A.13(i), it's clear that

$$\max_{i\in I_3, t\in[T]} \left|\hat{\Theta}_{j,it} - \Theta^0_{j,it}\right| = O_p\left(\sqrt{\frac{\log(N\vee T)}{N\wedge T}}\right) \ \forall j\in[p] \quad \text{and}$$
$$\max_{i\in I_3, t\in[T]} \left|\hat{\Theta}_{0,it} - \Theta^0_{0,it}\right| = O_p(\eta_N). \tag{A.48}$$

Proof of Statement (iii)

For $i \in I_a$ and $a \in [3]$, with the distribution theory defined in Theorem 2.3, we notice that

$$\left(\frac{1}{T}v^{0\prime}_{t,j}\Xi_{u_j,i}v^0_{t,j} + \frac{1}{N_a}u^{0\prime}_{i,j}\Xi^a_{v_j}u^0_{i,j}\right)^{-1/2}\left(\hat{\Theta}_{j,it} - \Theta^0_{j,it}\right) \rightsquigarrow \mathbb{N}(0,1),$$

which leads to the proof. ■

A.1.5 Proof of Theorem 2.4

Proof of Statement (i)

The proof is analogous to that in Castagnetti et al. (2015) and Lu and Su (2023). Recall that $S_{u_j} = \max(S^{(1,2)}_{u_j}, S^{(2,3)}_{u_j}, S^{(3,1)}_{u_j})$. For $a \in [3]$ and $b \in [3]\setminus\{a\}$, Theorem 2.3 shows that

$$\hat{u}^{(a,b)}_{i,j} - O^{(b)}_{u,j}u^0_{i,j} = O^{(b)}_j\hat{V}^{-1}_{u_j}\frac{1}{T}\sum_{t=1}^T b^0_{j,it} + \mathscr{R}^j_{i,u} \ \forall i\in I_a,$$

where $\max_{i\in I_a}\left|\mathscr{R}^j_{i,u}\right| = o_p\left((N\vee T)^{-1/2}\right)$. Recall that $\hat{\bar{u}}^{(a)}_j = \frac{1}{N_a}\sum_{i\in I_a}\hat{u}^{(a,b)}_{i,j}$. Under $H^I_0: u^0_{i,j} = u_j, \forall i\in[N]$, we have

$$\hat{\bar{u}}^{(a,b)}_j - O^{(b)}_{u,j}u_j = O^{(b)}_j\hat{V}^{-1}_{u_j}\frac{1}{N_aT}\sum_{i\in I_a}\sum_{t=1}^T b^0_{j,it} + \frac{1}{N_a}\sum_{i\in I_a}\mathscr{R}^j_{i,u} = o_p\left((N\vee T)^{-1/2}\right), \tag{A.49}$$

where the last equality holds by a simple application of Bernstein's inequality. Note that

$$\begin{aligned}
&T\left(\hat{u}^{(a,b)}_{i,j} - \hat{\bar{u}}^{(a,b)}_j\right)'\left(\hat{\Sigma}_{u_j}\right)^{-1}\left(\hat{u}^{(a,b)}_{i,j} - \hat{\bar{u}}^{(a,b)}_j\right)\\
&= T\left(\hat{u}^{(a,b)}_{i,j} - O^{(b)}_{u,j}u_j\right)'\left(\hat{\Sigma}_{u_j}\right)^{-1}\left(\hat{u}^{(a,b)}_{i,j} - O^{(b)}_{u,j}u_j\right)\\
&+ T\left(\hat{\bar{u}}^{(a,b)}_j - O^{(b)}_{u,j}u_j\right)'\left(\hat{\Sigma}_{u_j}\right)^{-1}\left(\hat{\bar{u}}^{(a,b)}_j - O^{(b)}_{u,j}u_j\right)
\end{aligned}$$

$$-2T\left(\hat{u}_{i,j}^{(a,b)}-O_{u,j}^{(b)}u_j\right)'\left(\hat{\Sigma}_{u_j}\right)^{-1}\left(\hat{\bar{u}}_{j}^{(a,b)}-O_{u,j}^{(b)}u_j\right)$$
$$:=I_{1ij}+I_{2ij}-2I_{3ij}. \tag{A.50}$$

For I_{2ij}, we have

$$\max_{i\in I_a}\left|I_{2ij}\right|\leq\max_{i\in I_a}T\left\|\hat{\bar{u}}_{j}^{(a,b)}-O_{u,j}^{(b)}u_j\right\|_2^2\left\{\lambda_{\min}(\Sigma_{u_j})+o_p(1)\right\}^{-1}=o_p(1) \tag{A.51}$$

by Lemma A.33, Assumption 2.11 and (A.49). For I_{3ij}, we have

$$\max_{i\in I_a}\left|I_{3ij}\right|\leq T\left(\max_{i\in I_a}\left\|\hat{u}_{i,j}^{(a,b)}-O_{u,j}^{(b)}u_j\right\|_2\right)\left(\left\|\hat{\bar{u}}_{j}^{(a,b)}-O_{u,j}^{(b)}u_j\right\|_2\right)\left[\lambda_{\min}(\Sigma_{u_j})+o_p(1)\right]^{-1}$$
$$=TO_p\left(\sqrt{\frac{\log(N\vee T)}{T}}\right)o_p\left((N\vee T)^{-1/2}\right)=o_p(1) \tag{A.52}$$

by (A.43) and (A.49). It suffices to study I_{1ij} below.

Now, let $\mathbb{Z}_{\mathfrak{B}}^{(1)}=\left(\mathbb{Z}_{\mathfrak{B},1}^{(1)\prime},\cdots,\mathbb{Z}_{\mathfrak{B},N}^{(1)\prime}\right)'$, where $\mathbb{Z}_{\mathfrak{B},i}^{(1)}\sim\mathbb{N}\left(0,O_j^{(1)\prime}\Sigma_{u_j}O_j^{(1)}\right)$ for $i\in I_3$, $\mathbb{Z}_{\mathfrak{B},i}^{(1)}\sim\mathbb{N}\left(0,O_j^{(3)\prime}\Sigma_{u_j}O_j^{(3)}\right)$ for $i\in I_2$ and $\mathbb{Z}_{\mathfrak{B},i}^{(1)}\sim\mathbb{N}\left(0,O_j^{(2)\prime}\Sigma_{u_j}O_j^{(2)}\right)$ for $i\in I_1$. Note that

$$I_{1ij}=\left(O_j^{(b)}\mathbb{Z}_{\mathfrak{B},i}^{(1)}+o_{p(1)}+\mathscr{R}_{i,u}^{j}\right)'\Sigma_{u_j}^{-1}\left(O_j^{(b)}\mathbb{Z}_{\mathfrak{B},i}^{(1)}+o_p(1)+\mathscr{R}_{i,u}^{j}\right)$$
$$+T\left(\hat{u}_{i,j}^{(a,b)}-O_{u,j}^{(b)}u_j\right)'\left[\hat{\Sigma}_{u_j}^{-1}-\Sigma_{u_j}^{-1}\right]\left(\hat{u}_{i,j}^{(a,b)}-O_{u,j}^{(b)}u_j\right)$$
$$=\mathbb{Z}_{\mathfrak{B},i}^{(1)\prime}\left(O_j^{(b)\prime}\Sigma_{u_j}O_j^{(b)}\right)^{-1}\mathbb{Z}_{\mathfrak{B},i}^{(1)}+o_p(1)\quad\text{uniformly over } i\in[N]. \tag{A.53}$$

It follows that

$$S_{u_j}=\max_{i\in[N]}\begin{cases}\mathbb{Z}_{\mathfrak{B},i}^{(1)\prime}\left(O_j^{(2)\prime}\Sigma_{u_j}O_j^{(2)}\right)^{-1}\mathbb{Z}_{\mathfrak{B},i}^{(1)}+o_p(1), & \forall i\in I_1,\\ \mathbb{Z}_{\mathfrak{B},i}^{(1)\prime}\left(O_j^{(3)\prime}\Sigma_{u_j}O_j^{(3)}\right)^{-1}\mathbb{Z}_{\mathfrak{B},i}^{(1)}+o_p(1), & \forall i\in I_2,\\ \mathbb{Z}_{\mathfrak{B},i}^{(1)\prime}\left(O_j^{(1)\prime}\Sigma_{u_j}O_j^{(1)}\right)^{-1}\mathbb{Z}_{\mathfrak{B},i}^{(1)}+o_p(1), & \forall i\in I_3,\end{cases} \tag{A.54}$$

with $\mathbb{Z}_{\mathfrak{B},i}^{(1)\prime}\left(O_j^{(2)\prime}\Sigma_{u_j}O_j^{(2)}\right)^{-1}\mathbb{Z}_{\mathfrak{B},i}^{(1)}\to\chi^2(1)$ for each $i\in I_1$, $\mathbb{Z}_{\mathfrak{B},i}^{(1)\prime}\left(O_j^{(3)\prime}\Sigma_{u_j}O_j^{(3)}\right)^{-1}\mathbb{Z}_{\mathfrak{B},i}^{(1)}\to\chi^2(1)$ for each $i\in I_2$, and $\mathbb{Z}_{\mathfrak{B},i}^{(1)\prime}\left(O_j^{(1)\prime}\Sigma_{u_j}O_j^{(1)}\right)^{-1}\mathbb{Z}_{\mathfrak{B},i}^{(1)}\to\chi^2(1)$ for each $i\in I_3$. As in Castagnetti et al. (2015), we can conclude that

$$\mathbb{P}\left(\frac{1}{2}S_{u_j}\leq x+\mathsf{b}(N)\right)\to e^{-e^{-x}}\text{ as }(N,T)\to\infty. \tag{A.55}$$

For the test statistic for H_0^{II}, the proof is similar. Let $\mathbb{Z}_{\mathfrak{B}}^{(2)}=\left(\mathbb{Z}_{\mathfrak{B},1}^{(2)\prime},\cdots,\mathbb{Z}_{\mathfrak{B},T}^{(2)\prime}\right)'$, where $\mathbb{Z}_{\mathfrak{B},t}^{(2)}\sim N\left(0,O_j^{(1)\prime}\Sigma_{v_j}O_j^{(1)}\right)$. As in (A.54), we can show that

$$S_{v_j}^{(3,1)}=\max_{t\in[T]}\mathbb{Z}_{\mathfrak{B},t}^{(2)\prime}\left(O_j^{(1)\prime}\Sigma_{v_j}O_j^{(1)}\right)^{-1}\mathbb{Z}_{\mathfrak{B},t}^{(2)}+o_p(1).$$

By the strong mixing condition in Assumption 2.1(iii), we have

$$\max_{j\in[p],i\in[N]}\left\|Cov\left(\mathfrak{b}_{j,it}^{(2)},\mathfrak{b}_{j,is}^{(2)}\right)\right\|_{\infty}\log(t-s)=o(1)$$

as $t-s\to\infty$ by Davydov's inequality. Then by Theorem 3.5.1 in Leadbetter and Rootzen (1988), we have that

$$\mathbb{P}\left(\frac{1}{2}\left(\max_{t\in[T]}\mathbb{Z}_{\mathfrak{B},t}^{(2)\prime}\left(O_j^{(1)\prime}\Sigma_{v_j}O_j^{(1)}\right)^{-1}\mathbb{Z}_{\mathfrak{B},t}^{(2)}\right)\le x+\mathsf{b}(T)\right)\to e^{-e^{-x}},$$

which implies that

$$\mathbb{P}\left(\frac{1}{2}S_{v_j}^{(3,1)}\le x+\mathsf{b}(T)\right)\to e^{-e^{-x}}\text{ as }(N,T)\to\infty.$$

Recall that $\tilde{S}_{v_j}^{(a,b)}=\frac{1}{2}S_{v_j}^{(a,b)}-\mathsf{b}(T)$ and $S_{v_j}=\max(\tilde{S}_{v_j}^{(1,2)},\tilde{S}_{v_j}^{(2,3)},\tilde{S}_{v_j}^{(3,1)})$. Noting that S_{v_j} is asymptotically distributed as the maximum of three independent Gumbel random variables under H_0^{II}, we have $\mathbb{P}(S_{v_j}\le x)\to e^{-3e^{-x}}$ as $(N,T)\to\infty$.

Proof of Statement (ii)

Under H_1^I, we have that $\forall i\in I_a$,

$$\begin{aligned}\hat{u}_{i,j}^{(a,b)}-\hat{\bar{u}}_j^{(a,b)}&=\left(\hat{u}_{i,j}^{(a,b)}-O_{u,j}^{(b)}u_{i,j}^0\right)+O_{u,j}^{(b)}\left(u_{i,j}^0-u_j\right)-\left(\hat{\bar{u}}_j^{(a,b)}-O_{u,j}^{(b)}u_j\right)\\&=\left(\hat{u}_{i,j}^{(a,b)}-O_{u,j}^{(b)}u_{i,j}^0\right)+O_{u,j}^{(b)}c_{i,j}^u-\left(\hat{\bar{u}}_j^{(a,b)}-O_{u,j}^{(b)}u_j\right).\end{aligned}$$

Then

$$\begin{aligned}&T\left(\hat{u}_{i,j}^{(a,b)}-\hat{\bar{u}}_j^{(a,b)}\right)'\left(\hat{\Sigma}_{u_j}\right)^{-1}\left(\hat{u}_{i,j}^{(a,b)}-\hat{\bar{u}}_j^{(a,b)}\right)\\&=T\left(\hat{u}_{i,j}^{(a,b)}-O_{u,j}^{(b)}u_j\right)'\hat{\Sigma}_{u_j}^{-1}\left(\hat{u}_{i,j}^{(1)}-O_{u,j}^{(b)}u_j\right)+T\left(\hat{\bar{u}}_j^{(1)}-O_{u,j}^{(b)}u_j\right)'\hat{\Sigma}_{u_j}^{-1}\left(\hat{\bar{u}}_j^{(a,b)}-O_{u,j}^{(b)}u_j\right)\\&-2T\left(\hat{u}_{i,j}^{(a,b)}-O_{u,j}^{(b)}u_j\right)'\hat{\Sigma}_{u_j}^{-1}\left(\hat{\bar{u}}_j^{(1)}-O_{u,j}^{(b)}u_j\right)+2T\left(O_{u,j}^{(b)}c_{i,j}^u\right)'\hat{\Sigma}_{u_j}^{-1}\left(\hat{u}_{i,j}^{(a,b)}-O_{u,j}^{(b)}u_j\right)\\&+T\left(O_{u,j}^{(b)}c_{i,j}^u\right)'\hat{\Sigma}_{u_j}^{-1}O_{u,j}^{(b)}c_{i,j}^u-2T\left(O_{u,j}^{(b)}c_{i,j}^u\right)'\hat{\Sigma}_{u_j}^{-1}\left(\hat{\bar{u}}_j^{(a,b)}-O_{u,j}^{(b)}u_j\right)\\&:=S_{u_j,i,1}^{(b)}+S_{u_j,2}^{(b)}+S_{u_j,i,3}^{(b)}+S_{u_j,i,4}^{(b)}+S_{u_j,i,5}^{(b)}+S_{u_j,i,6}^{(b)},\end{aligned}$$

where $\max_{i\in I_a}\left|S_{u_j,i,1}^{(b)}\right|=O_p(\log N)$ by (A.53) and (A.55), $\left|S_{u_j,2}^{(b)}\right|=o_p(1)$ by (A.51), $\max_{i\in I_3}\left|S_{u_j,i,3}^{(b)}\right|=o_p(1)$ by (A.52). Next,

$$\begin{aligned}\max_{i\in I_a}\left|S_{u_j,i,5}^{(b)}\right|&=\max_{i\in I_a}T\left(O_{u,j}^{(b)}c_{i,j}^u\right)'\Sigma_{u_j}^{-1}O_{u,j}^{(b)}c_{i,j}^u\left\{1+o_p(1)\right\}\\&\gtrsim\left[\max_{i\in I_a}\lambda_{\max}\left(\Sigma_{u,j}\right)\right]^{-1}T\max_{i\in I_a}\left\|c_{i,j}^u\right\|_2^2\\&\asymp T\max_{i\in I_a}\left\|c_{i,j}^u\right\|_2^2,\end{aligned}$$

which diverges to infinity at the rate faster than $\log N$ by condition in statement (ii). By Cauchy-Schwarz inequality, we obtain that

$$\max_{i\in I_a}\left|S^{(b)}_{u_j,i,4}\right| \lesssim \max_{i\in I_a}\left(S^{(b)}_{u_j,i,1}\right)^{1/2}\left(S^{(b)}_{u_j,i,5}\right)^{1/2} = o_p\left(\max_{i\in I_a}\left|S^{(b)}_{u_j,i,5}\right|\right), \text{ and}$$

$$\max_{i\in I_a}\left|S^{(b)}_{u_j,i,6}\right| \lesssim \max_{i\in I_a}\left(S^{a}_{u_j,i,2}\right)^{1/2}\left(S^{(b)}_{u_j,i,5}\right)^{1/2} = o_p\left(\max_{i\in I_a}\left|S^{(b)}_{u_j,i,5}\right|\right).$$

It follows that $\mathbb{P}\left\{S_{u_j} > c_{\alpha,N}\right\} \to 1$ as $c_{\alpha,N} \asymp \log N$, and the final result follows.

The power of the test statistic S_{v_j} can be analyzed analogously. ■

A.1.6 Proof of Theorem 2.5

We first derive the linear expansion of $\hat{\Theta}^*_{j,it} - \Theta^*_{j,it}$ for $i \in I_3$, and similar results hold for $i \in I_1 \cup I_2$. Let $\bar{v}^0_j := \frac{1}{T}\sum_{t\in[T]} v^0_{t,j}$ and $\bar{u}^{0,I_a}_j := \frac{1}{N_a}\sum_{i\in I_a} u^0_{i,j}$. By Proposition 2.1, we obtain that uniformly in $i \in I_a$,

$$\begin{aligned}
\bar{\hat{\Theta}}_{j,i\cdot} - \bar{\Theta}_{j,i\cdot} &= \frac{1}{T}\sum_{t\in[T]}\left(\hat{\Theta}_{j,it} - \Theta^0_{j,it}\right) \\
&= \frac{1}{T}\sum_{t\in[T]} u^{0\prime}_{i,j}\left(\hat{V}^{(a)}_{v_j,t}\right)^{-1}\frac{1}{N_a}\sum_{i^*\in I_a}\xi^0_{j,i^*t} + \frac{1}{T}\sum_{t\in[T]} v^{0\prime}_{t,j}\hat{V}^{-1}_{u_j}\frac{1}{T}\sum_{t=1}^{T} b^0_{j,it} + \frac{1}{T}\sum_{t\in[T]}\mathscr{R}^j_{it} \\
&= \bar{v}^{0\prime}_j\hat{V}^{-1}_{u_j}\frac{1}{T}\sum_{t=1}^{T} b^0_{j,it} + o_p\left((N\vee T)^{-1/2}\right), \qquad \text{(A.56)}
\end{aligned}$$

where the third equality holds by the fact that

$$\begin{aligned}
&\max_{i\in I_a}\left|\frac{1}{T}\sum_{t\in[T]} u^{0\prime}_{i,j}\left(\hat{V}^{(a)}_{v_j,t}\right)^{-1}\frac{1}{N_a}\sum_{i^*\in I_a}\xi^0_{j,i^*t}\right| = \max_{i\in I_a}\left|\frac{1}{N_aT}\sum_{i^*\in I_a}\sum_{t\in[T]} u^{0\prime}_{i,j}\left(\hat{V}^{(a)}_{v_j,t}\right)^{-1}\xi^0_{j,i^*t}\right| \\
&= O_p\left(\sqrt{\frac{\log N}{NT}}\xi_N\right) = o_p\left((N\vee T)^{-1/2}\right),
\end{aligned}$$

by conditional Bernstein's inequality given $\mathscr{D}_e$ in Lemma A.12(i), Assumption 2.1(i)-(ii) and Assumption 2.1(ix). Similarly, uniformly in $t \in [T]$, we have

$$\begin{aligned}
\bar{\hat{\Theta}}^{I_a}_{j,\cdot t} - \bar{\Theta}^{I_a}_{j,\cdot t} &= \frac{1}{N_a}\sum_{i\in I_3}\left(\hat{\Theta}_{j,it} - \Theta^0_{j,it}\right) \\
&= \frac{1}{N_a}\sum_{i\in I_a} u^{0\prime}_{i,j}\left(\hat{V}^{(a)}_{v_j,t}\right)^{-1}\frac{1}{N_a}\sum_{i\in I_a}\xi^0_{j,it} + \frac{1}{N_a}\sum_{i\in I_a} v^{0\prime}_{t,j}\hat{V}^{-1}_{u_j}\frac{1}{T}\sum_{t^*=1}^{T} b^0_{j,it^*} + \frac{1}{N_a}\sum_{i\in I_a}\mathscr{R}^j_{it} \\
&= \bar{u}^{0,I_a\prime}_j\left(\hat{V}^{(a)}_{v_j,t}\right)^{-1}\frac{1}{N_a}\sum_{i\in I_a}\xi^0_{j,it} + o_p\left((N\vee T)^{-1/2}\right), \qquad \text{(A.57)}
\end{aligned}$$

and

$$\bar{\hat{\Theta}}^{I_a}_j - \bar{\Theta}^{I_a}_j = \frac{1}{N_aT}\sum_{i\in I_a}\sum_{t\in[T]}\left(\hat{\Theta}_{j,it} - \Theta^0_{j,it}\right)$$

$$= \frac{1}{N_a T}\sum_{i\in I_a}\sum_{t\in[T]} u_{i,j}^{0\prime}\left(\hat{V}_{v_j,t}^{(a)}\right)^{-1}\frac{1}{N_a}\sum_{i\in I_a}\xi_{j,it}^{0} + \frac{1}{N_a T}\sum_{i\in I_a}\sum_{t\in[T]} v_{t,j}^{0\prime}\hat{V}_{u_j}^{-1}\frac{1}{T}\sum_{t=1}^{T} b_{j,it}^{0}$$
$$+ \frac{1}{N_a T}\sum_{i\in I_a}\sum_{t\in[T]}\mathscr{R}_{it}^{j}$$
$$= o_p\left((N\vee T)^{-1/2}\right). \tag{A.58}$$

Combining (A.56)-(A.58), we obtain that $\forall i \in I_a$ and $t \in [T]$,

$$\hat{\Theta}_{j,it}^{*} - \Theta_{j,it}^{*} = \hat{\Theta}_{j,it} - \Theta_{j,it}^{0} - \left(\hat{\bar{\Theta}}_{j,i\cdot} - \bar{\Theta}_{j,i\cdot}\right) - \left(\hat{\bar{\Theta}}_{j,\cdot t}^{I_a} - \bar{\Theta}_{j,\cdot t}^{I_a}\right) + \left(\hat{\bar{\Theta}}_{j}^{I_a} - \bar{\Theta}_{j}^{I_a}\right)$$
$$= \left(u_{i,j}^{0} - \bar{u}_{j}^{0,I_a}\right)'\left(\hat{V}_{v_j,t}^{(a)}\right)^{-1}\frac{1}{N_a}\sum_{i\in I_a}\xi_{j,it}^{0} + \left(v_{t,j}^{0} - \bar{v}_{j}^{0}\right)'\hat{V}_{u_j}^{-1}\frac{1}{T}\sum_{t=1}^{T} b_{j,it}^{0} + \bar{\mathscr{R}}_{it}^{j},$$

such that $\max_{i\in I_3,t\in[T]}\left|\bar{\mathscr{R}}_{it}^{j}\right| = o_p\left((N\vee T)^{-1/2}\right)$. It follows that $\hat{\Theta}_{j,it}^{*} - \Theta_{j,it}^{*} \rightsquigarrow \mathbb{N}(0,\Sigma_{j,it}^{*})$ with

$$\Sigma_{j,it}^{*} = \sum_{a\in[3]}\frac{1}{N_a}\left(O_j^{(b)}u_{i,j}^{0} - O_j^{(b)}\bar{u}_j^{0}\right)'\Sigma_{v_j}\left(O_j^{(b)}u_{i,j}^{0} - O_j^{(b)}\bar{u}_j^{0}\right)\mathbf{1}_{ia}$$
$$+ \frac{1}{T}\left(O_j^{(b)}v_{t,j}^{0} - O_j^{(b)}\bar{v}_j^{0}\right)'\Sigma_{u_j}\left(O_j^{(b)}v_{t,j}^{0} - O_j^{(b)}\bar{v}_j^{0}\right),$$

and Σ_{u_j} and Σ_{v_j} are as defined in Theorem 2.3. The reason why $\Sigma_{j,it}^{*}$ is not indexed with b is owing to the fact that $O_j^{(b)}$ shown in the right side of the equality can be absorbed by $O_j^{(b)}$ not shown in Σ_{u_j} and Σ_{v_j}.

Define $\hat{\Sigma}_{j,it}^{*} = \frac{1}{2}\sum_b \hat{\Sigma}_{j,it}^{(b)*}$ with

$$\hat{\Sigma}_{j,it}^{(b)*} = \sum_{a\in[3]}\left[\frac{1}{N_a}\left(\hat{u}_{i,j}^{(a,b)} - \hat{\bar{u}}_j^{(a,b)}\right)'\hat{\Sigma}_{v_j}\left(\hat{u}_{i,j}^{(a,b)} - \bar{u}_j^{(a,b)}\right)\mathbf{1}_{ia}\right.$$
$$\left. + \frac{1}{T}\left(\hat{v}_{t,j}^{(a,b)} - \hat{\bar{v}}_j^{(a,b)}\right)'\hat{\Sigma}_{u_j}\left(\hat{v}_{t,j} - \hat{\bar{v}}_j^{(a,b)}\right)\right],$$

where $\hat{\bar{u}}_j^{(a,b)} = \frac{1}{N_a}\sum_{i\in I_a}\hat{u}_{i,j}^{(a,b)}$ and $\hat{\bar{v}}_j^{(a,b)} = \frac{1}{T}\sum_{t\in[T]}\hat{v}_{t,j}^{(a,b)}$. By Theorem 2.3 and Lemma A.33, we have

$$\max_{j\in[p],i\in[N],t\in[T]}\left|\hat{\Sigma}_{j,it}^{*} - \Sigma_{j,it}^{*}\right| = o_p(1).$$

By arguments as used in the proof of Theorem 2.3, we have that as $(N,T)\to\infty$, $\mathbb{P}\left(\frac{1}{2}S_{NT} \le x + \mathsf{b}(NT)\right) \to e^{-e^{-x}}$ under H_0^{III}, and $\mathbb{P}\left(S_{NT} > c_{\alpha,3\cdot NT}\right) \to 1$ under H_1^{III} provided $\frac{N\wedge T}{\log NT}\max_{i\in[N],t\in[T]}\left|\Theta_{j,it}^{*}\right|^2 \to \infty$. ■

A.2 Some Technical Lemmas

In this section we state and prove some technical lemmas that are used in the proofs of the main results in the paper.

A.2.1 Lemmas for the Proof of Theorem 2.1

Lemma A.1. *Consider a matrix sequence $\{A_i, i=1,\cdots,N\}$ whose values are symmetric matrices with dimension d,*

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

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

(ii) suppose $\{A_i, i=1,\cdots,N\}$ is sequence of martingale difference matrices with $\mathbb{E}_{i-1}(A_i)=0$ and $\|A_i\|_{op} \leq M$ 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. Further let $\left\|\sum_{i\in[N]} \mathbb{E}_{i-1}\left(A_i^2\right)\right\|_{op} \leq \sigma^2$. Then for all $t>0$, we have

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

Proof Lemma A.1(i) and (ii) are Matrix Bernstein inequality and Matrix Freedman inequality, which are respectively stated in Theorem 1.3 and Corollary 4.2 in Tropp (2011). ■

Lemma A.2. *Consider a specific matrix $A \in \mathbb{R}^{N\times T}$ whose rows (denoted as A_i' where $A_i \in \mathbb{R}^T$) are independent random vectors in $\mathbb{R}^T$ with $\mathbb{E}A_i = 0$ and $\Sigma_i = \mathbb{E}(A_i A_i')$. Suppose $\max_{i\in[N]} \|A_i\|_2 \leq \sqrt{m}$ a.s. and $\max_{i\in[N]} \|\Sigma_i\|_{op} \leq M$ for some positive constant M. Then for every $t>0$, with probability $1-2T\exp\left(-c_1 t^2\right)$, we have*

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

where c_1 is an absolute constant.

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

$$\|Z_i\|_{op} \leq \frac{1}{N}\left(\|A_iA_i'\|_{op} + \|\Sigma_i\|_{op}\right) \leq \frac{1}{N}\left(\|A_i\|_2^2 + \|\Sigma_i\|_{op}\right) \leq \frac{m+M}{N}, \quad \text{a.s.} \tag{A.59}$$

Next, noting that $\mathbb{E}\left[(A_iA_i')^2\right] = \mathbb{E}\left[\|A_i\|_2 A_iA_i'\right] \leq m\Sigma_i$ and $Z_i^2 = \frac{1}{N^2}[(A_iA_i')^2 - A_iA_i'\Sigma_i - \Sigma_iA_iA_i' + \Sigma_i^2]$, we have

$$\begin{aligned}\|\mathbb{E}(Z_i^2)\|_{op} &= \frac{1}{N^2}\left\|\mathbb{E}\left[(A_iA_i')^2 - \Sigma_i^2\right]\right\|_{op} \leq \frac{1}{N^2}\left\{\left\|\mathbb{E}\left[(A_iA_i')^2\right]\right\|_{op} + \|\Sigma_i\|_{op}^2\right\} \\ &\leq \frac{1}{N^2}\left(m\|\Sigma_i\|_{op} + \|\Sigma_i\|_{op}^2\right) \leq \frac{mM+M^2}{N^2} \text{ uniformly in } i.\end{aligned}$$

It follows that

$$\left\|\sum_{i\in[N]}\mathbb{E}(Z_i^2)\right\|_{op} \leq N\max_{i\in[N]}\|\mathbb{E}(Z_i^2)\|_{op} \leq \frac{mM+M^2}{N}. \tag{A.60}$$

Let $\varepsilon = \max\left(\sqrt{M}\delta, \delta^2\right)$ with $\delta = t\sqrt{\frac{m+M}{N}}$. By (A.59)-(A.60) and the matrix Bernstein inequality in Lemma A.1(i), we have

$$\begin{aligned}\mathbb{P}\left\{\left\|\frac{1}{N}\left(A'A - \sum_{i\in[N]}\Sigma_i\right)\right\|_{op} \geq \varepsilon\right\} &= \mathbb{P}\left(\left\|\sum_{i\in[N]}Z_i\right\|_{op} \geq \varepsilon\right) \\ &\leq 2T\exp\left\{-c_1\min\left(\frac{\varepsilon^2}{\frac{mM+M^2}{N}}, \frac{\varepsilon}{\frac{m+M}{N}}\right)\right\} \\ &\leq 2T\exp\left\{-c_1\min\left(\frac{\varepsilon^2}{M}, \varepsilon\right)\frac{N}{m+M}\right\} \\ &\leq 2T\exp\left\{-\frac{c_1\delta^2 N}{m+M}\right\} = 2T\exp\left\{-c_1t^2\right\},\end{aligned}$$

for some positive constant c_1, where the third inequality is due to the fact that

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

That is,

$$\left\| \frac{1}{N}A'A - \frac{1}{N}\sum_{i\in[N]}\Sigma_i \right\|_{op} \leq \max\left(\sqrt{M}\delta, \delta^2\right) \tag{A.61}$$

with probability $1-\exp\left(-c_1 t^2\right)$. Combining the fact that $\|\Sigma_i\| \leq M$ uniformly over i and (A.61), we show that

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

It follows that $\|A\|_{op} \leq \sqrt{NM} + t\sqrt{m+M}$. ■

Lemma A.3. *Recall* $a_{it} = \tau - \mathbf{1}\{\varepsilon_{it} \leq 0\}$ *and* $a = \{a_{it}\} \in \mathbb{R}^{N\times T}$. *Under Assumption 2.1, we have* $\left\|X_j \odot a\right\|_{op} = O_p\left(\sqrt{N} + \sqrt{T\log T}\right)$ $\forall j \in [p]$ *and* $\|a\|_{op} = O_p\left(\sqrt{N} + \sqrt{T\log T}\right)$.

Proof We focus on $\left\|X_j \odot a\right\|_{op}$ as the result for $||a||_{op}$ can be derived in the same manner. We first note that, conditional on $\mathscr{D}$, the i-th row of $X_j \odot a$ only depends on $\{e_{it}, \varepsilon_{it}\}_{t\in[T]}$, which are independent across i. Therefore $X_j \odot a$ has independent rows, denoted as $A_i = X_{j,i} \odot a_i$, given $\mathscr{D}$, where $X_{j,i}$ and a_i being the i-th row of matrix X_j and a, respectively. In addition, for the t-th element of A_i, we have

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

where the second equality holds by Assumption 2.1(ii) and the fact that given $\mathscr{D}_e$, $X_{j,it}$ is known. Therefore, In order to apply Lemma A.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 2.1(iv), we have $\frac{1}{T}\sum_{t\in[T]}\left(X_{j,it}a_{it}\right)^2 \leq \frac{1}{T}\sum_{t\in[T]}X_{j,it}^2 \leq c_2$ a.s. for some positive constant c_2, which implies

$$||A_i||_2 = \left\|X_{j,i}\odot a_i\right\|_2 \leq c_2\sqrt{T} \quad \text{a.s.} \tag{A.62}$$

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

and $\infty-$norms, i.e.,

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

By Davydov's inequality for conditional strong mixing sequence, we can show that

$$\begin{aligned}
&\max_{s\in[T]} \sum_{t\in[T]} \left|\mathbb{E}\left(X_{j,it}X_{j,is}a_{it}a_{is}\middle|\mathscr{D}\right)\right| \\
&= \max_{s\in[T]} \sum_{t\in[T]} \left|Cov\left(X_{j,it}a_{it}, X_{j,is}a_{is}\middle|\mathscr{D}\right)\right| \\
&\leq \max_{s\in[T]} \sum_{t\in[T]} \left\{\mathbb{E}\left[\left|X_{j,it}a_{it}\right|^q \middle|\mathscr{D}\right]\right\}^{1/q} \left\{\mathbb{E}\left[\left|X_{j,is}a_{is}\}\right|^q \middle|\mathscr{D}\right]\right\}^{1/q} \times \alpha(t-s)^{(q-2)/q} \\
&\leq \max_{i\in[N], t\in[T]} \left\{\mathbb{E}\left[\left|X_{j,it}^q\right| \middle|\mathscr{D}\right]\right\}^{2/q} \max_{s\in[T]} \sum_{t\in[T]} \alpha(t-s)^{(q-2)/q} \\
&\leq c_3 \; a.s.,
\end{aligned}$$

where c_3 is a positive constant which does not depend on i and the last line is by Assumption 2.1(iii) and 2.1(iv). Similarly, we have $\max_{t\in[T]} \sum_{s\in[T]} \left|\mathbb{E}\left(X_{j,it}X_{j,is}a_{it}a_{is}\middle|\mathscr{D}\right)\right| \leq c_3$ a.s. Then by Corollary 2.3.2 in Golub and Van Loan (1996), we have

$$\max_{i\in[N]} \|\Sigma_i\|_{op} \leq \sqrt{\|\Sigma_i\|_1 \|\Sigma_i\|_\infty} \leq c_3 \quad \text{a.s.} \tag{A.63}$$

Combining (A.59), (A.60), and Lemma A.1 with $t = \sqrt{\log T}$, we obtain the desired result. ∎

Recall that $\Delta_{\Theta_j} = \Theta_j - \Theta_j^0$ for any Θ_j and define

$$\mathscr{R}(C_1) := \left\{ \{\Delta_{\Theta_j}\}_{j=0}^p : \sum_{j=0}^p \left\|\mathscr{P}_j^\perp(\Delta_{\Theta_j})\right\|_* \leq C_1 \sum_{j=0}^p \left\|\mathscr{P}_j(\Delta_{\Theta_j})\right\|_* \right\}.$$

Lemma A.4. *Suppose Assumptions 2.1-2.3 hold. Then* $\{\tilde{\Delta}_{\Theta_j}\}_{j=0}^p \in \mathscr{R}(3)$ *w.p.a.1.*

Proof Define $\mathbb{Q}_\tau\left(\{\Theta_j\}_{j=0}^p\right) = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \rho_\tau\left(Y_{it} - \Theta_{0,it} - \sum_{j=1}^p X_{j,it}\Theta_{j,it}\right)$ for generic $\{\Theta_j\}_{j=0}^p$. By the definition of the nuclear norm estimator in (2.9), we have

$$\mathbb{Q}_\tau\left(\{\Theta_j^0\}_{j=0}^p\right) - \mathbb{Q}_\tau\left(\{\tilde{\Theta}_j\}_{j=0}^p\right) + \sum_{j=0}^p \nu_j \left[\|\Theta_j^0(\tau)\|_* - \|\tilde{\Theta}_j(\tau)\|_*\right] \geq 0. \tag{A.64}$$

In addition, we have

$$\mathbb{Q}_\tau\left(\{\Theta_j^0\}_{j=0}^p\right) - \mathbb{Q}_\tau\left(\{\tilde{\Theta}_j\}_{j=0}^p\right)$$

$$
\begin{aligned}
&= -\left\{ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{it} \left(\tilde{\Delta}_{\Theta_{0,it}} + \sum_{j=1}^{p} X_{j,it} \tilde{\Delta}_{\Theta_{j,it}} \right) \right. \\
&\left. + \int_0^{\tilde{\Delta}_{\Theta_{0,it}} + \sum_{j=1}^{p} X_{j,it} \tilde{\Delta}_{\Theta_{j,it}}} \mathbf{1}\{\varepsilon_{it} \leq s\} - \mathbf{1}\{\varepsilon_{it} \leq 0\} \, ds \right\} \\
&\leq \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{it} \left(\tilde{\Delta}_{\Theta_{0,it}} + \sum_{j=1}^{p} X_{j,it} \tilde{\Delta}_{\Theta_{j,it}} \right) \right| \\
&\leq \sum_{j=1}^{p} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{it} X_{j,it} \tilde{\Delta}_{\Theta_{j,it}} \right| + \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} a_{it} \tilde{\Delta}_{\Theta_{0,it}} \right| \\
&= \sum_{j=1}^{p} \frac{1}{NT} \left| tr\left[\tilde{\Delta}'_{\Theta_j} \left(X_j \odot a \right) \right] \right| + \frac{1}{NT} \left| tr\left(\tilde{\Delta}'_{\Theta_0} a \right) \right| \\
&\leq \sum_{j=1}^{p} \frac{1}{NT} \left\| \tilde{\Delta}_{\Theta_j} \right\|_* \left\| X_j \odot a \right\|_{op} + \frac{1}{NT} \left\| \tilde{\Delta}_{\Theta_0} \right\|_* \|a\|_{op} \\
&\leq c_4 \sum_{j=0}^{p} \frac{(\sqrt{N} \vee \sqrt{T \log T})}{NT} \left\| \tilde{\Delta}_{\Theta_j} \right\|_*, \quad \text{w.p.a.1}, \qquad \text{(A.65)}
\end{aligned}
$$

where the first equality holds by Knight's identity in Knight (1998) which states that

$$
\rho_\tau(u-v) - \rho_\tau(u) = v(\tau - \mathbf{1}\{u \leq 0\}) + \int_0^v (\mathbf{1}\{u \leq s\} - \mathbf{1}\{u \leq 0\}) \, ds, \qquad \text{(A.66)}
$$

the first inequality is due to the fact that the second term in the bracket of the second line is non-negative, the second inequality holds by triangle inequality, the third inequality is by the fact that $tr(AB) \leq ||A||_{op}||B||_*$, and the last inequality holds by Lemma A.3.

Combining (A.64) and (A.65), w.p.a.1, we have

$$
0 \leq c_4 \sum_{j=0}^{p} \left\{ \frac{(\sqrt{N} \vee \sqrt{T \log T})}{NT} ||\tilde{\Delta}_{\Theta_j}||_* + \nu_j \left(\|\Theta_j^0\|_* - \|\tilde{\Theta}_j\|_* \right) \right\}. \qquad \text{(A.67)}
$$

Besides, we can show that

$$
\begin{aligned}
\|\tilde{\Theta}_j\|_* &= \|\tilde{\Delta}_{\Theta_j} + \Theta_j^0\|_* = \left\| \Theta_j^0 + \mathscr{P}_j^\perp(\tilde{\Delta}_{\Theta_j}) + \mathscr{P}_j(\tilde{\Delta}_{\Theta_j}) \right\|_* \\
&\geq \left\| \Theta_j^0 + \mathscr{P}_j^\perp(\tilde{\Delta}_{\Theta_j}) \right\|_* - \|\mathscr{P}_j(\tilde{\Delta}_{\Theta_j})\|_* = \|\Theta_j^0\|_* + \left\| \mathscr{P}_j^\perp(\tilde{\Delta}_{\Theta_j}) \right\|_* - \|\mathscr{P}_j(\tilde{\Delta}_{\Theta_j})\|_*
\end{aligned} \qquad \text{(A.68)}
$$

where the second equality holds by Lemma D.2(i) in Chernozhukov et al. (2019), the first inequality holds by triangle inequality, and the last equality is by the construction of the linear space $\mathscr{P}_j^\perp$ and $\mathscr{P}_j$. Then combining (A.67) and (A.68), we obtain

$$\sum_{j=0}^{p} \nu_j \left\{ \left\| \mathscr{P}_j^{\perp}\left(\tilde{\Delta}_{\Theta_j}\right) \right\|_* - \left\| \mathscr{P}_j\left(\tilde{\Delta}_{\Theta_j}\right) \right\|_* \right\} \leq c_4 \sum_{j=0}^{p} \frac{(\sqrt{N} \vee \sqrt{T \log T})}{NT} \|\tilde{\Delta}_{\Theta_j}\|_*$$

$$= c_4 \sum_{j=0}^{p} \frac{(\sqrt{N} \vee \sqrt{T \log T})}{NT} \left\{ \left\| \mathscr{P}_j\left(\tilde{\Delta}_{\Theta_j}\right) \right\|_* + \left\| \mathscr{P}_j^{\perp}\left(\tilde{\Delta}_{\Theta_j}\right) \right\|_* \right\}, \text{ w.p.a.1.}$$

By setting $\nu_j = \frac{2c_4(\sqrt{N} \vee \sqrt{T \log T})}{NT}$, we obtain $\sum_{j=0}^{p} \left\| \mathscr{P}_j^{\perp}\left(\tilde{\Delta}_{\Theta_j}\right) \right\|_* \leq 3 \sum_{j=0}^{p} \left\| \mathscr{P}_j\left(\tilde{\Delta}_{\Theta_j}\right) \right\|_*$ w.p.a.1. ■

Recall $\mathscr{G}_{i,t-1}$ is the σ-field generated by $\{V_j^0\}_{j \in [p] \cup \{0\}}$, $\{W_j^0\}_{j \in [p]}$, $\{e_{ls}\}_{l \leq i-1, s \in [T]}$, $\{e_{is}\}_{s \leq t}$, $\{\varepsilon_{ls}\}_{l \leq i-1, s \in [T]}$, and $\{\varepsilon_{is}\}_{s \leq t-1}$ and $\mathsf{F}_{it}(\cdot)$ and $\mathsf{f}_{it}(\cdot)$ are the conditional CDF and PDF of ε_{it} given $\mathscr{G}_{i,t-1}$, respectively. Specifically, we note that $(\{X_{j,it}\}_{j \in [p]}, \{\Theta_{j,it}^0\}_{j \in [p] \cup \{0\}})$ are measurable w.r.t. $\mathscr{G}_{i,t-1}$.

Lemma A.5. *For all $u_1, u_2 \in \mathbb{R}$ and all $c_5 \in (0,1]$, we have*

(i) $\int_0^{u_2} (\mathbf{1}\{u_1 \leq z\} - \mathbf{1}\{u_1 \leq 0\})\, dz \geq \int_0^{c_5 u_2} (\mathbf{1}\{u_1 \leq z\} - \mathbf{1}\{u_1 \leq 0\})\, dz \geq 0,$

(ii) $\int_0^{u_2} \{\mathsf{F}_{it}(u_1 + z) - \mathsf{F}_{it}(u_1)\}\, dz \geq \int_0^{c_5 u_2} \{\mathsf{F}_{it}(u_1 + z) - \mathsf{F}_{it}(u_1)\}\, dz \geq 0.$

Proof Statement (i) is just Feng (2019, Lemma A2). To prove statement (ii), notice that if $u_2 \geq 0$, then $z \geq 0$ and $\mathsf{F}_{it}(u_1 + z) - \mathsf{F}_{it}(u_1) \geq 0$ for all $z \in [0, u_2]$, which leads to the existence of the second inequality naturally:

$$\int_0^{u_2} \{\mathsf{F}_{it}(u_1 + z) - \mathsf{F}_{it}(u_1)\}\, dz - \int_0^{c_5 u_2} \{\mathsf{F}_{it}(u_1 + z) - \mathsf{F}_{it}(u_1)\}\, dz$$
$$= \int_{c_5 u_2}^{u_2} \{\mathsf{F}_{it}(u_1 + z) - \mathsf{F}_{it}(u_1)\}\, dz \geq 0.$$

On the other hand, if $u_2 < 0$, we have

$$\int_0^{u_2} \{\mathsf{F}_{it}(u_1 + z) - \mathsf{F}_{it}(u_1)\}\, dz - \int_0^{c_5 u_2} \{\mathsf{F}_{it}(u_1 + z) - \mathsf{F}_{it}(u_1)\}\, dz$$
$$= \int_{u_2}^{0} \{\mathsf{F}_{it}(u_1) - \mathsf{F}_{it}(u_1 + z)\}\, dz - \int_{c_5 u_2}^{0} \{\mathsf{F}_{it}(u_1) - \mathsf{F}_{it}(u_1 + z)\}\, dz$$
$$= \int_{u_2}^{c_5 u_2} \{\mathsf{F}_{it}(u_1) 5\mathsf{F}_{it}(u_1 + z)\}\, dz \geq 0,$$

where the last inequality holds as the same reason for $u_2 \geq 0$ case. ■

Lemma A.6. *Under Assumptions 2.1-2.4, for any $\{\Delta_{\Theta_j}\}_{j=0}^{p} \in \mathscr{R}(3, C_2)$ such that $||\Delta_{\Theta_j}||_{\max} \leq M$ for some constant $M > 0$, we have*

$$Q_\tau\left(\{\Theta_j^0 + \Delta_{\Theta_j}\}_{j=0}^{p}\right) - Q_\tau\left(\{\Theta_j^0\}_{j=0}^{p}\right) \geq \frac{c_7 C_3}{NT\xi_N^2} \sum_{j=0}^{p} \|\Delta_{\Theta_j}\|_F^2 - \frac{c_7 C_4}{NT\xi_N^2}(N+T) \text{ w.p.a.1},$$

where $Q_\tau(\{\Theta_j\}_{j=0}^p) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left[\rho_\tau\left(Y_{it} - \Theta_{0,it} - \sum_{j=1}^p X_{j,it}\Theta_{j,it}\right)\middle|\mathscr{G}_{i,t-1}\right]$, $c_7 = \frac{\underline{\mathsf{f}}c_6^2}{4}$ *with* c_6 *being a positive constant between 0 and 1.*

Proof We can choose a sufficiently large constant M such that $c_6 := \frac{3\underline{\mathsf{f}}}{2\bar{\mathsf{f}}' M(1+p)} \in (0,1]$. Then we have

$$
\begin{aligned}
& Q_\tau\left(\{\Theta_j^0 + \Delta_{\Theta_j}\}_{j=0}^p\right) - Q_\tau\left(\{\Theta_j^0\}_{j=0}^p\right) \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left\{\rho_\tau\left(Y_{it} - \Theta_{0,it} - \sum_{j=1}^p X_{j,it}\Theta_{j,it}\right)\right. \\
&\left. - \rho_\tau\left(Y_{it} - \Theta_{0,it}^0 - \sum_{j=1}^p X_{j,it}\Theta_{j,it}^0\right)\middle|\mathscr{G}_{i,t-1}\right\} \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left\{\left(\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}\right)(\tau - \mathbf{1}\{\varepsilon_{it} \le 0\})\middle|\mathscr{G}_{i,t-1}\right\} \\
&+ \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left\{\int_0^{\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}} (\mathbf{1}\{\varepsilon_{it} \le s\} - \mathbf{1}\{\varepsilon_{it} \le 0\})\, ds\middle|\mathscr{G}_{i,t-1}\right\} \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left\{\int_0^{\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}} (\mathbf{1}_{\varepsilon_{it} \le s} - \mathbf{1}_{\varepsilon_{it} \le 0})\, ds\middle|\mathscr{G}_{i,t-1}\right\} \\
&\ge \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left\{\int_0^{c_6\xi_N^{-1}\left(\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}\right)} (\mathbf{1}_{\varepsilon_{it} \le s} - \mathbf{1}_{\varepsilon_{it} \le 0})\, ds\middle|\mathscr{G}_{i,t-1}\right\} \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \int_0^{c_6\xi_N^{-1}\left(\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}\right)} [\mathsf{F}_{it}(s) - \mathsf{F}_{it}(0)]\, ds \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \int_0^{c_6\xi_N^{-1}\left(\Delta_{\Theta_{0,it}} + \sum_{j=1}^0 X_{j,it}\Delta_{\Theta_{j,it}}\right)} \left[s\mathsf{f}_{it}(0) + \frac{s^2}{2}\mathsf{f}'_{it}(\tilde{s})\right] ds \\
&\ge \frac{\underline{\mathsf{f}}c_6^2}{2NT\xi_N^2}\sum_{i=1}^N\sum_{t=1}^T \left(\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}\right)^2 - \frac{\bar{\mathsf{f}}' c_6^3}{6NT\xi_N^3}\sum_{i=1}^N\sum_{t=1}^T \left|\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}\right|^3 \\
&= \frac{\underline{\mathsf{f}}c_6^2}{4NT\xi_N^2}\sum_{i=1}^N\sum_{t=1}^T \left(\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}\right)^2 \\
&+ \frac{1}{NT\xi_N^2}\sum_{i=1}^N\sum_{t=1}^T \left\{\frac{\underline{\mathsf{f}}c_6^2}{4}\left(\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}\right)^2\left(1 - \frac{2c_6\bar{\mathsf{f}}'}{3\underline{\mathsf{f}}\xi_N}\left|\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}\right|\right)\right\} \\
&\ge \frac{\underline{\mathsf{f}}c_6^2}{4NT\xi_N^2}\sum_{i=1}^N\sum_{t=1}^T \left\{\Delta_{\Theta_{0,it}} + \sum_{j=1}^p X_{j,it}\Delta_{\Theta_{j,it}}\right\}^2 \\
&= \frac{\underline{\mathsf{f}}c_6^2}{4NT\xi_N^2}\left\|\Delta_{\Theta_0} + \sum_{j=1}^p X_j \odot \Delta_{\Theta_j}\right\|_F^2
\end{aligned}
$$

$$\geq \frac{\underline{f}c_6^2}{4NT\xi_N^2}\left\{C_3\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F^2 - C_4\left(N+T\right)\right\} \quad \text{w.p.a.1}$$

$$= \frac{c_7C_3}{NT\xi_N^2}\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F^2 - \frac{c_7C_4}{NT\xi_N^2}\left(N+T\right) \tag{A.69}$$

where the second equality is by (A.66), the first inequality is by Lemma A.5 and the fact that $c_6/\xi_N \leq 1$, the fifth equality is by the mean-value theorem, the third inequality is by the fact that

$$1-\frac{2c_6\bar{f}'}{3\underline{f}\xi_N}\left|\Delta_{\Theta_{0,it}}+\sum_{j=1}^{p}X_{j,it}\Delta_{\Theta_{j,it}}\right| \geq 1-\frac{2c_6\bar{f}'}{3\underline{f}\xi_N}M(1+p)\xi_N \geq 0,$$

and the fourth inequality holds under Assumption 2.4. This concludes the proof. ■

Lemma A.7. *Under Assumptions 2.1–2.4, for any* $\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p} \in \mathscr{R}(3,C_2)$*, we have* $\left\|\Delta_{\Theta_j}\right\|_* \leq c_8\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F$ $\forall j \in \{0,\cdots,p\}$ *where* $c_8 = 4\sqrt{2\bar{K}}$*.*

Proof For $\forall j \in \{0,\cdots,p\}$, we obtain that

$$\begin{aligned}
\left\|\Delta_{\Theta_j}\right\|_* &= \left\|\mathscr{P}_j(\Delta_{\Theta_j})\right\|_* + \left\|\mathscr{P}_j^{\perp}(\Delta_{\Theta_j})\right\|_* \leq \left\|\mathscr{P}_j(\Delta_{\Theta_j})\right\|_* + 3\sum_{j=0}^{p}\left\|\mathscr{P}_j(\Delta_{\Theta_j})\right\|_* \\
&\leq 4\sum_{j=0}^{p}\left\|\mathscr{P}_j(\Delta_{\Theta_j})\right\|_* \leq 4\sum_{j=0}^{p}\sqrt{2K_j}\left\|\mathscr{P}_j\left(\Delta_{\Theta_j}\right)\right\|_F \\
&\leq 4\sqrt{2\bar{K}}\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F := c_8\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F,
\end{aligned}$$

where the first equality is by Chernozhukov et al. (2019, Lemma D.2(i)), the first inequality is by the definition of $\mathscr{R}(C_1,C_2)$, and the last two inequalities follow the facts that $\|A\|_* \leq \sqrt{\text{rank}(A)}\,\|A\|_F$ for any matrix $A \in \mathbb{R}^{N\times T}$ and $\text{rank}\left(\mathscr{P}_j\left(\Delta_{\Theta_j}\right)\right) \leq 2K_j$, which hold by Chernozhukov et al. (2019, Lemma D.2.(iii)). ■

Let $\mathscr{Z}$ be a separable metric space, $\{Z_1,\cdots,Z_n\}$ be a sequence of random variables in $\mathscr{Z}$ adapted to the filtration $\{\mathscr{F}_t\}_{t\in[n]}$, $\mathscr{F} = \{f : \mathscr{Z} \to \mathbb{R}\}$ be a set of bounded real valued functions on $\mathbb{R}$, and $u_1,\cdots,u_n$ be i.i.d. Rademacher random variables. Here we allow the dependence of sequence $\{Z_1,\cdots,Z_n\}$. Similarly as in Rakhlin et al. (2015), we define a $\mathscr{Z}$-valued tree $\mathbf{z}$ of depth n with the sequence $(\mathbf{z}_1,\cdots,\mathbf{z}_n)$ such that $\mathbf{z}_t : \{u_1,\cdots,u_{t-1}\} \to \mathscr{Z}$. For simplicity, we denote as $\mathbf{z}_t\left(u\right) = \mathbf{z}_t\left(u_1,\cdots,u_{t-1}\right)$ and $\mathbb{E}_t\left[\cdot\right] = \mathbb{E}\left[\cdot\middle|\mathscr{F}_t\right]$ for short. Also, denote $u_{1:t} := (u_1,\cdots,u_t)$ and similarly for $Z_{1:t}$.

Lemma A.8. *Let $\mathscr{F}$ be a class of functions. For any $\alpha > 0$, it holds that*

$$\beta_n \mathbb{P}\left\{\sup_{f\in\mathscr{F}}\left|\frac{1}{n}\sum_{t=1}^{n}\left(f\left(Z_t\right)-\mathbb{E}_{t-1}\left[f\left(Z_t\right)\right]\right)\right|>\alpha\right\} \leq 2\sup_{\mathbf{z}}\mathbb{P}\left\{\sup_{f\in\mathscr{F}}\left|\frac{1}{n}\sum_{t=1}^{n}u_t f\left(\mathbf{z}_t\left(u\right)\right)\right|>\frac{\alpha}{4}\right\}.$$

where $\beta_n \geq 1-\sup_{f\in\mathscr{F}}\frac{4\sum_{t=1}^{n} Var\left(f(Z_t)\middle|\mathscr{F}_{t-1}\right)}{n^2\alpha^2}$ and the outer supremum is taken over all $\mathscr{Z}$-valued tree of depth n.[1]

Proof Let $Z'_{1:n}$ be a decoupled sequence tangent to $Z_{1:n}$. For the sequence of random variables $\{Z_t : t \in [n]\}$ adapted to the filtration $\{\mathscr{F}_t : t \in [n]\}$, the sequence $Z'_{1:n} = \{Z_t^*, t \in [n]\}$ is said to be a decoupled sequence tangent to $\{Z_t : t \in [n]\}$ if for each $t \in [n]$, Z_t^* is generated from the conditional distribution of Z_t given $\mathscr{F}_{t-1}$ and independent of everything else. This means the sequence $Z'_{1:n}$ is conditionally independent given $\mathscr{F}_n$ and for any measurable function f of Z_t^*,

$$\mathbb{E}(f(Z_t^*)|\mathscr{F}_n) = \mathbb{E}(f(Z_t^*)|\mathscr{F}_{t-1})\ a.s. \tag{A.70}$$

For $\forall f \in \mathscr{F}$, with Chebyshev's inequality, we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{n}\left|\sum_{t=1}^{n} f\left(Z_t^*\right)-\mathbb{E}_{t-1}\left[f\left(Z_t^*\right)\right]\right|>\alpha/2\middle|\mathscr{F}_n\right) \\
&\leq \frac{4\mathbb{E}\left\{\left(\sum_{t=1}^{n} f\left(Z_t^*\right)-\mathbb{E}_{t-1}\left[f\left(Z_t^*\right)\right]\right)^2\middle|\mathscr{F}_n\right\}}{n^2\alpha^2} \\
&= \frac{4\sum_{t=1}^{n}\mathbb{E}\left\{\left(f\left(Z_t^*\right)-\mathbb{E}_{t-1}\left[f\left(Z_t^*\right)\right]\right)^2\middle|\mathscr{F}_n\right\}}{n^2\alpha^2} \\
&= \frac{4\sum_{t=1}^{n} Var\left(f\left(Z_t\right)\middle|\mathscr{F}_{t-1}\right)}{n^2\alpha^2},
\end{aligned}$$

where the first equality holds by the fact that, given $\mathscr{F}_n$, $\{Z_t^*\}_{t\in[n]}$ are independent and the last equality holds by (A.70). This implies

$$\begin{aligned}
\beta_n :&= \inf_{f\in\mathscr{F}}\mathbb{P}\left(\frac{1}{n}\left|\sum_{t=1}^{n} f\left(Z_t^*\right)-\mathbb{E}_{t-1}\left[f\left(Z_t^*\right)\right]\right|\leq\alpha/2\middle|\mathscr{F}_n\right) \\
&= 1-\sup_{f\in\mathscr{F}}\mathbb{P}\left(\frac{1}{n}\left|\sum_{t=1}^{n} f\left(Z_t^*\right)-\mathbb{E}_{t-1}\left[f\left(Z_t^*\right)\right]\right|>\alpha/2\middle|\mathscr{F}_n\right) \\
&\geq 1-\sup_{f\in\mathscr{F}}\frac{4\sum_{t=1}^{n} Var\left(f\left(Z_t\right)\middle|\mathscr{F}_{t-1}\right)}{n^2\alpha^2}.
\end{aligned}$$

Let function f^* be the function that maximizes $\frac{1}{n}\left|\sum_{t=1}^{n}\left[f\left(Z_t^*\right)-\mathbb{E}_{t-1}\left(f\left(Z_t^*\right)\right)\right]\right|$ con-

[1]This means the supremum is taken over all $\mathbf{z} = \{\mathbf{z}_t(\cdot)\}_{t\in[n]}$.

dition on $\mathscr{F}_n$, and define the event $A_1 = \left\{\sup_{f\in\mathscr{F}} \frac{1}{n} \left|\sum_{t=1}^{n} \left[f(Z_t) - \mathbb{E}_{t-1}(f(Z_t))\right]\right| > \alpha\right\}$. Then we obtain that

$$\beta_n \leq \mathbb{P}\left(\frac{1}{n}\left|\sum_{t=1}^{n} f^*(Z_t^*) - \mathbb{E}_{t-1}\left[f^*(Z_t)\right]\right| \leq \alpha/2 \middle| \mathscr{F}_n\right),$$

where the inequality follows by the definition of β_n and the fact that $\mathbb{E}_{t-1}[f^*(Z_t^*)] = \mathbb{E}_{t-1}[f^*(Z_t)]$. As $A_1 \in \mathscr{F}_n$, we have

$$\beta_n \leq \mathbb{P}\left(\frac{1}{n}\left|\sum_{t=1}^{n} f^*(Z_t^*) - \mathbb{E}_{t-1}\left[f^*(Z_t^*)\right]\right| \leq \alpha/2 \middle| A_1\right).$$

It follows that

$$\begin{aligned}
&\beta_n \mathbb{P}\left\{\sup_{f\in\mathscr{F}}\left|\frac{1}{n}\sum_{t=1}^{n}\left(f(Z_t) - \mathbb{E}_{t-1}\left[f(Z_t)\right]\right)\right| > \alpha\right\} \\
&\leq \mathbb{P}\left(\frac{1}{n}\left|\sum_{t=1}^{n} f^*(Z_t^*) - \mathbb{E}_{t-1}\left[f^*(Z_t^*)\right]\right| \leq \alpha/2 \middle| A_1\right)\mathbb{P}(A_1) \\
&= \mathbb{P}\left(\left\{\frac{1}{n}\left|\sum_{t=1}^{n} f^*(Z_t^*) - \mathbb{E}_{t-1}\left[f^*(Z_t^*)\right]\right| \leq \alpha/2\right\}\bigcup A_1\right) \\
&\leq \mathbb{P}\left(\frac{1}{n}\left|\sum_{t=1}^{n} f^*(Z_t) - f^*(Z_t^*)\right| \geq \alpha/2\right) \\
&\leq \mathbb{P}\left(\frac{1}{n}\sup_{f\in\mathscr{F}}\left|\sum_{t=1}^{n} f(Z_t) - f(Z_t^*)\right| \geq \alpha/2\right), \qquad \text{(A.71)}
\end{aligned}$$

where the second inequality holds by the implication rule. Let $\phi(\cdot) = \mathbf{1}\{\cdot > n\alpha/2\}$. By Lemma 18 in Rakhlin et al. (2015), we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{n}\sup_{f\in\mathscr{F}}\left|\sum_{t=1}^{n} f(Z_t) - f(Z_t^*)\right| \geq \alpha/2\right) \\
&= \mathbb{E}\mathbf{1}\left\{\frac{1}{n}\sup_{f\in\mathscr{F}}\left|\sum_{t=1}^{n} f(Z_t) - f(Z_t^*)\right| \geq \alpha/2\right\} \\
&\leq \sup_{z_1,z_1'}\mathbb{E}_{u_1}\cdots\sup_{z_n,z_n'}\mathbb{E}_{u_n}\mathbf{1}\left\{\sup_{f\in\mathscr{F}}\left|\sum_{t=1}^{n} u_t\left[f(z_t) - f(Z_t^*)\right]\right| \geq n\alpha/2\right\} \\
&\leq \sup_{z_1,z_1'}\mathbb{E}_{u_1}\cdots\sup_{z_n,z_n'}\mathbb{E}_{u_n}\mathbf{1}\left\{\sup_{f\in\mathscr{F}}\left|\sum_{t=1}^{n} u_t f(z_t)\right| \geq n\alpha/4\right\} \\
&+ \sup_{z_1,z_1'}\mathbb{E}_{u_1}\cdots\sup_{z_n,z_n'}\mathbb{E}_{u_n}\mathbf{1}\left\{\sup_{f\in\mathscr{F}}\left|\sum_{t=1}^{n} u_t f(Z_t^*)\right| \geq n\alpha/4\right\} \\
&= 2\sup_{z_1,z_1'}\mathbb{E}_{u_1}\cdots\sup_{z_n,z_n'}\mathbb{E}_{u_n}\mathbf{1}\left\{\sup_{f\in\mathscr{F}}\left|\sum_{t=1}^{n} u_t f(z_t)\right| \geq n\alpha/4\right\}
\end{aligned}$$

$$= 2 \sup_{\mathbf{z}} \mathbb{P}\left\{ \sup_{f\in\mathscr{F}} \left| \frac{1}{n}\sum_{t=1}^{n} u_t f(\mathbf{z}_t(u)) \right| > \frac{\alpha}{4} \right\}, \tag{A.72}$$

where the standard text z_t and Z_t^* are $\mathscr{Z}$-valued, the bold text $\mathbf{z}_t = \mathbf{z}_t(u)$ is the t-th root of a tree $\mathbf{z}(\cdot)$ (i.e., a function of $u_{1:t-1}$), and the outer supremum in the last line is taken over all $\mathscr{Z}$-valued tree of depth n. Combining (A.71) and (A.72), we can conclude that

$$\beta_n \mathbb{P}\left\{ \sup_{f\in\mathscr{F}} \left| \frac{1}{n}\sum_{t=1}^{n} (f(Z_t) - \mathbb{E}_{t-1}[f(Z_t)]) \right| > \alpha \right\} \le 2\sup_{\mathbf{z}} \mathbb{P}\left\{ \sup_{f\in\mathscr{F}} \left| \frac{1}{n}\sum_{t=1}^{n} u_t f(\mathbf{z}_t(u)) \right| > \frac{\alpha}{4} \right\}.$$

∎

The next lemma is an extension of the contraction principle, i.e., Ledoux and Talagrand (1991, Theorem 4.12), to the case with sequential symmetrization.

Lemma A.9. *Let function $F : \mathbb{R}_+ \to \mathbb{R}_+$ be convex and increasing and $\phi_t : \mathbb{R} \to \mathbb{R}$ be contractions such that $\phi_t(0) = 0$, $\mathbf{z}_t$ is the t-th root of a tree ($\mathbf{z}$) which depends on $\{u_1, \cdots, u_{t-1}\}$. Then we have*

$$\mathbb{E}F\left\{ \frac{1}{2} \sup_{f\in\mathscr{F}} \left| \sum_{t=1}^{n} u_t \phi_t(f(\mathbf{z}_t(u))) \right| \right\} \le \mathbb{E}F\left\{ \sup_{f\in\mathscr{F}} \left| \sum_{t=1}^{n} u_t f(\mathbf{z}_t(u)) \right| \right\}.$$

Proof We first consider the statement without the absolute value. Let function $G : \mathbb{R} \to \mathbb{R}$ be convex and increasing. We observe that

$$\mathbb{E}G\left\{ \sup_{f\in\mathscr{F}} \sum_{t=1}^{n} u_t \phi_t(f(\mathbf{z}_t(u))) \right\} = \mathbb{E}\left\{ \mathbb{E}\left[G\left\{ \sup_{f\in\mathscr{F}} \sum_{t=1}^{n} u_t \phi_t(f(\mathbf{z}_t(u))) \right\} \middle| u_{1:n-1} \right] \right\}$$

and

$$\begin{aligned}
&\mathbb{E}\left\{ G\left(\sup_{f\in\mathscr{F}} \sum_{t=1}^{n} u_t \phi_t(f(\mathbf{z}_t(u))) \middle| u_{1:n-1} \right) \right\} \\
&= \mathbb{E}\left\{ G\left(\sup_{f\in\mathscr{F}} \sum_{t=1}^{n-1} u_t \phi_t(f(\mathbf{z}_t(u))) + u_n \phi_n(f(\mathbf{z}_n(u))) \middle| u_{1:n-1} \right) \right\} \\
&= \mathbb{E}\left\{ G\left(\sup_{k_1,k_2\in\mathscr{K}} (k_1 + u_n \phi_n(k_2)) \right) \middle| u_{1:n-1} \right\},
\end{aligned}$$

where $k_1 = \sum_{t=1}^{n-1} u_t \phi_t(f(\mathbf{z}_t(u)))$, $k_2 = f(\mathbf{z}_n(u))$, and $\mathscr{K} = \{(k_1, k_2) : f \in \mathscr{F}\} \subset \mathbb{R}^2$. We also note that k_1 and k_2 only depend on $u_{1:n-1}$ and is independent of u_n. The proof in Ledoux and Talagrand (1989, Theorem 4.12) shows

$$\mathbb{E}G\left\{ \sup_{k_1,k_2\in\mathscr{K}} k_1 + u_2\phi_2(k_2) \right\} \le \mathbb{E}G\left\{ \sup_{k_1,k_2\in\mathscr{K}} k_1 + u_2 k_2 \right\},$$

which implies

$$\begin{aligned}&\mathbb{E}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n}u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)\Big|u_{1:n-1}\right)\right\}\\&\leq\mathbb{E}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n-1}u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)+u_nf\left(\mathbf{z}_n(u)\right)\Big|u_{1:n-1}\right)\right\}.\end{aligned}$$

Taking expectation on both sides, we have

$$\mathbb{E}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n}u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)\right)\right\}\leq\mathbb{E}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n-1}u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)+u_nf\left(\mathbf{z}_n(u)\right)\right)\right\}.$$

Next, let $k_1=\sum_{t=1}^{n-2}u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)$, $k_2=f(\mathbf{z}_{n-1}(u))$ and $k_3(u_{n-1})=u_nf(\mathbf{z}_n(u))$. We emphasize that k_1 and k_2 only depend on $u_{1:n-2}$ and u_n while k_3 also depends on u_{n-1}. The dependence of $(k_1,k_2,k_3(\cdot))$ on f is made implicit for notation simplicity. Furthermore, given the fact that u_{n-1} only takes values {-1,1}, we have

$$k_3(u_{n-1})=\frac{u_{n-1}+1}{2}k_3(1)+\frac{1-u_{n-1}}{2}k_3(-1)=\frac{k_3(1)+k_3(-1)}{2}+u_{n-1}\frac{k_3(1)-k_3(-1)}{2}.$$

Given these notation and conditioning on $(u_{1:n-2},u_n)$, we have

$$\begin{aligned}&\mathbb{E}_{u_{n-1}}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n-1}u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)+u_nf\left(\mathbf{z}_n(u)\right)\right)\right\}\\&=\mathbb{E}_{u_{n-1}}\left[G\left(\sup_{f\in\mathscr{F}}k_1+u_{n-1}\phi_{n-1}(k_2)+k_3(u_{n-1})\right)\right]\\&=\mathbb{E}_{u_{n-1}}\left[G\left(\sup_{f\in\mathscr{F}}\left(k_1+\frac{k_3(1)-k_3(-1)}{2}\right)+u_{n-1}\left(\phi_{n-1}(k_2)+\frac{k_3(1)-k_3(-1)}{2}\right)\right)\right]\\&=\mathbb{E}_{u_{n-1}}\left[G\left(\sup_{(h_1,h_2,h_3)\in\mathscr{H}}h_1+u_{n-1}\left(\phi_{n-1}(h_2)+h_3\right)\right)\right],\end{aligned}$$

where $\mathbb{E}_{u_{n-1}}$ means the expectation is taken conditionally on $(u_{1:n-2},u_n)$, $h_1=\sum_{t=1}^{n-2}u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)+\frac{u_n(f(\mathbf{z}_n(u_{1:n-2},1))+f(\mathbf{z}_n(u_{1:n-2},-1)))}{2}$, $h_2=k_2$, $h_3=\frac{u_n(f(\mathbf{z}_n(u_{1:n-2},1))-f(\mathbf{z}_n(u_{1:n-2},-1)))}{2}$, and $\mathscr{H}=((h_1,h_2,h_3):f\in\mathscr{F})\in\mathbb{R}^3$. Suppose $(h_1^*,h_2^*,h_3^*)\in\mathscr{H}$ and $(h_1^\dagger,h_2^\dagger,h_3^\dagger)\in\mathscr{H}$ achieve the supremum of

$$h_1+\left(\phi_{n-1}(h_2)+h_3\right)\quad\text{and}\quad h_1-\left(\phi_{n-1}(h_2)+h_3\right),\quad\text{respectively.}$$

Then, we have

$$\begin{aligned}&\mathbb{E}_{u_{n-1}}\left[G\left(\sup_{(h_1,h_2,h_3)\in\mathscr{H}}h_1+u_{n-1}\left(\phi_{n-1}(h_2)+h_3\right)\right)\right]\\&=\frac{1}{2}G((h_1^*+h_3^*)+\phi_{n-1}(h_2^*))+\frac{1}{2}G((h_1^\dagger-h_3^\dagger)-\phi_{n-1}(h_2^\dagger))\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{2}G((h_1^* + h_3^*) + h_2^*) + \frac{1}{2}G((h_1^\dagger - h_3^\dagger) - h_2^\dagger) \\
&\leq \mathbb{E}_{u_{n-1}}\left[G\left(\sup_{(h_1,h_2,h_3)\in\mathscr{H}} h_1 + u_{n-1}(h_2+h_3)\right)\right],
\end{aligned}
$$

where the first inequality is by the fact proved in the proof of Ledoux and Talagrand (1991, Theorem 4.12) that for any t_1, s_1, t_2, s_2,

$$
\frac{1}{2}G(t_1 + \phi_{n-1}(t_2)) + \frac{1}{2}G(s_1 - \phi_{n-1}(s_2)) \leq \frac{1}{2}G(t_1+t_2) + \frac{1}{2}G(s_1 - s_2).
$$

Plugging back the definition of h_1, h_2, h_3, we have

$$
\begin{aligned}
&\mathbb{E}_{u_{n-1}}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n-1} u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right) + u_n f\left(\mathbf{z}_n(u)\right)\right)\right\} \\
&\leq \mathbb{E}_{u_{n-1}}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n-2} u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right) + \sum_{t=n-1}^{n} u_t f\left(\mathbf{z}_t(u)\right)\right)\right\}.
\end{aligned}
$$

Taking expectation on both sides, we have

$$
\begin{aligned}
&\mathbb{E}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n-1} u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right) + u_n f\left(\mathbf{z}_n(u)\right)\right)\right\} \\
&\leq \mathbb{E}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n-2} u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right) + \sum_{t=n-1}^{n} u_t f\left(\mathbf{z}_t(u)\right)\right)\right\}.
\end{aligned}
$$

We can repeat a similar argument by taking conditional expectations given $(u_{1:t-1}, u_{t+1:n})$ and removing ϕ_t for all $t = n-2, \cdots, 2$. This leads to the result that

$$
\mathbb{E}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n} u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)\right)\right\} \leq \mathbb{E}\left\{G\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n} u_t f\left(\mathbf{z}_t(u)\right)\right)\right\}. \quad \text{(A.73)}
$$

Next, we come back to the case with the absolute value. Note that

$$
\begin{aligned}
&\mathbb{E}F\left[\frac{1}{2}\sup_{f\in\mathscr{F}}\left|\sum_{t=1}^{n} u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)\right|\right] \\
&\leq \frac{1}{2}\left\{\mathbb{E}F\left[\sup_{f\in\mathscr{F}}\left(\sum_{t=1}^{n} u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)\right)^+\right] + \mathbb{E}F\left[\sup_{f\in\mathscr{F}}\left(\sum_{t=1}^{n} u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)\right)^-\right]\right\} \\
&= \frac{1}{2}\left\{\mathbb{E}F\left[\sup_{f\in\mathscr{F}}\left(\sum_{t=1}^{n} u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)\right)^+\right] + \mathbb{E}F\left[\sup_{f\in\mathscr{F}}\left(\sum_{t=1}^{n} u_t^*\phi_t\left(f\left(\mathbf{z}_t^*(u^*)\right)\right)\right)^+\right]\right\} \\
&= \frac{1}{2}\left\{\mathbb{E}F\left[\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n} u_t\phi_t\left(f\left(\mathbf{z}_t(u)\right)\right)\right)^+\right] + \mathbb{E}F\left[\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n} u_t^*\phi_t\left(f\left(\mathbf{z}_t^*(u^*)\right)\right)\right)^+\right]\right\} \\
&\leq \frac{1}{2}\left\{\mathbb{E}F\left[\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n} u_t\left(f\left(\mathbf{z}_t(u)\right)\right)\right)^+\right] + \mathbb{E}F\left[\left(\sup_{f\in\mathscr{F}}\sum_{t=1}^{n} u_t^* f\left(\mathbf{z}_t^*(u^*)\right)\right)^+\right]\right\}
\end{aligned}
$$

$$= \frac{1}{2}\left\{\mathbb{E}F\left[\sup_{f\in\mathscr{F}}\left(\sum_{t=1}^{n} u_t\left(f\left(\mathbf{z}_t(u)\right)\right)\right)^{+}\right] + \mathbb{E}F\left[\sup_{f\in\mathscr{F}}\left(\sum_{t=1}^{n} u_t^{*} f\left(\mathbf{z}_t^{*}(u^{*})\right)\right)^{+}\right]\right\}$$
$$= \frac{1}{2}\left\{\mathbb{E}F\left[\sup_{f\in\mathscr{F}}\left(\sum_{t=1}^{n} u_t\left(f\left(\mathbf{z}_t(u)\right)\right)\right)^{+}\right] + \mathbb{E}F\left[\sup_{f\in\mathscr{F}}\left(\sum_{t=1}^{n} u_t f\left(\mathbf{z}_t(u)\right)\right)^{-}\right]\right\}$$
$$\leq \mathbb{E}F\left[\sup_{f\in\mathscr{F}}\left|\sum_{t=1}^{n} u_t f\left(\mathbf{z}_t(u)\right)\right|\right], \tag{A.74}$$

where $u_t^* = -u_t$, $\mathbf{z}_t^*(u) = \mathbf{z}_t(-u)$, the first inequality is by the convexity of F, the first and second equalities are by the fact that $(v)^- = (-v)^+$ for any v, and the second inequality is by (A.73) and the fact that $F((\cdot)^+)$ is convex and increasing. This leads to the desired result. ∎

Define Rademacher sequence $u = (u_{11}, \cdots, u_{1T}, \cdots, u_{N1}, \cdots, u_{NT}) = \left(u_{(1)}, \cdots, u_{(NT)}\right) \in \mathbb{R}^{NT\times 1}$. In the matrix notation, let $U = \{u_{it}\} \in \mathbb{R}^{N\times T}$. By vectorization, for a sequence of independent variables $X_{j,it}$, we define

$$\left(X_{j,11}, \cdots, X_{j,1T}, \cdots, X_{j,N1}, \cdots, X_{j,NT}\right) = \left(X_{j,(1)}, \cdots, X_{j,(NT)}\right),$$
$$\left(\varepsilon_{11}, \cdots, \varepsilon_{1T}, \cdots, \varepsilon_{N1}, \cdots, \varepsilon_{NT}\right) = \left(\varepsilon_{(1)}, \cdots, \varepsilon_{(NT)}\right),$$

$\forall j \in [p]$. Using the binary tree representation, let $x_{j,(l)}^*$ be $(l)^{th}$ root of the tree which takes values in the support of $X_{j,(l)}$, i.e., $x_{j,(l)}^* : u^{l-1} \mapsto [-\xi_N, \xi_N]$ for $u^{l-1} := \left(u_{(1)}, \cdots, u_{(l-1)}\right)$ such that $\max_{i\in[N]} \sum_{t\in[T]} \left(x_{j,it}^*\right)^2 \leq MT$ and $\max_{t\in[T]} \sum_{i\in[N]} \left(x_{j,it}^*\right)^{2\ell} \leq MN$ for some fixed constant $M < \infty$ and $\ell = 1, 2$. Similar notation follows for $\varepsilon_{(l)}^* = \varepsilon_{(l)}^*\left(u^{l-1}\right)$. In the matrix notation, let $x_j^* = \left\{x_{j,it}^*\right\} \in \mathbb{R}^{N\times T}$ such that $x_{j,it}^* = x_{j,(l)}$ with $i = \lceil \frac{l}{T} \rceil$ and $t = l - (i-1)T$.

Lemma A.10. *Under Assumption 2.1, for $j \in [p]$, there exists an absolute constant C that is independent of the trees (x^*, ε^*) such that when $\log(N \vee T) \geq 2$,*

$$\mathbb{E}\exp\left(\frac{\|U\|_{op}}{\sqrt{(N\vee T)\log(N\vee T)}}\right) \leq C \quad \textit{and} \quad \mathbb{E}\exp\left(\frac{\left\|U \odot x_j^*\right\|_{op}}{\sqrt{(N\vee T)\log(N\vee T)}}\right) \leq C.$$

Proof The proof here follows similarly as Lemma A.2 except that we have martingale difference matrices rather than independent matrices. For a specific j, let $A = U \odot x_j^* = (A_1, \cdots, A_N)' \in \mathbb{R}^{N\times T}$, $\mathscr{F}_i$ be the σ-field generated by $\{u_{i^*t}\}_{i^*\leq i, t\in[T]}$, $\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot|\mathscr{F}_i)$, $\Sigma_i = \mathbb{E}_{i-1}\left(A_i A_i'\right)$ and $Z_i = \frac{1}{N}\left(A_i A_i' - \Sigma_i\right)$, with $\frac{1}{N}\left(A'A - \sum_{i\in[N]}\Sigma_i\right) =$

$\sum_{i\in[N]} Z_i$. Note that $\mathbb{E}_{i-1}(Z_i) = 0$,

$$\max_{i\in[N]} \|A_i\|_2 = \max_{i\in[N]} \sqrt{\sum_{t\in[T]} \left(x^*_{j,it}u_{it}\right)^2} = \max_{i\in[N]} \sqrt{\sum_{t\in[T]} \left(x^*_{j,it}\right)^2} \leq \sqrt{MT} \ a.s., \quad (A.75)$$

$$\max_{i\in[N]} \|\Sigma_i\|_{op} = \max_{i\in[N]} \left\|\operatorname{diag}\left(\left(x^*_{j,i1}\right)^2, \cdots, \left(x^*_{j,iT}\right)^2\right)\right\|_{op} \leq \xi_N^2 \ a.s., \quad (A.76)$$

and for $\ell = 1, 2$,

$$\left\|\sum_{i\in[N]} \Sigma_i^\ell\right\|_{op} = \left\|\operatorname{diag}\left(\sum_{i\in[N]} \left(x^*_{j,i1}\right)^{2\ell}, \cdots, \sum_{i\in[N]} \left(x^*_{j,iT}\right)^{2\ell}\right)\right\|_{op} \leq MN \ a.s.. \quad (A.77)$$

Combining (A.75) and (A.76) yields

$$\begin{aligned}
\max_{i\in[N]} \|Z_i\|_{op} &\leq \max_{i\in[N]} \frac{1}{N}\left(\|A_iA_i'\|_{op} + \|\Sigma_i\|_{op}\right) \\
&\leq \frac{1}{N}\left(\max_{i\in[N]} \|A_i\|_2^2 + \max_{i\in[N]} \|\Sigma_i\|_{op}\right) \leq \frac{MT + \xi_N^2}{N} \ a.s. \quad (A.78)
\end{aligned}$$

In addition,

$$\begin{aligned}
\left\|\sum_{i\in[N]} \mathbb{E}_{i-1}\left(Z_i^2\right)\right\|_{op} &= \left\|\sum_{i\in[N]} \mathbb{E}_{i-1}\left\{\frac{1}{N^2}\left[\left(A_iA_i'\right)^2 - \Sigma_i^2\right]\right\}\right\|_{op} \\
&\leq \frac{1}{N^2}\left[\sum_{i\in[N]} \left\|\mathbb{E}_{i-1}\left(\|A_i\|_2^2 A_iA_i'\right)\right\|_{op} + \left\|\sum_{i\in[N]} \Sigma_i^2\right\|_{op}\right] \\
&\leq \frac{1}{N^2}\left[MT\left\|\sum_{i\in[N]} \Sigma_i\right\|_{op} + \left\|\sum_{i\in[N]} \Sigma_i^2\right\|_{op}\right] \\
&\leq \frac{(MT+1)M}{N}, \quad (A.79)
\end{aligned}$$

where the last inequality holds by (A.75) and (A.77).

Combining (A.78) and (A.79), by matrix Bernstein's inequality in Lemma A.1(ii), for some sufficiently large constant $\bar{c}$ that depends on M, we have

$$\mathbb{P}\left(\left\|\sum_{i\in[N]} Z_i\right\|_{op} > \bar{c}\frac{(N\vee T)}{N}\log(N\vee T)\right)$$

$$\leq T\exp\left\{-\frac{\frac{1}{2}\bar{c}^2\left(\frac{(N\vee T)}{N}\right)^2 (\log(N\vee T))^2}{\frac{(MT+1)M}{N} + \frac{\frac{MT+\xi_N^2}{N}\bar{c}\frac{(N\vee T)}{N}\log(N\vee T)}{3}}\right\} = \exp\left(-\left(\frac{\bar{c}}{2} - 1\right)\log((N\vee T))\right),$$

which implies that with probability greater than $1 - \exp(-(\bar{c}/2 - 1)\log(N\vee T))$,

$$\left\|\frac{1}{N}\left(A'A - \sum_{i\in[N]} \Sigma_i\right)\right\|_{op} \leq \bar{c}\left(\frac{(N\vee T)}{N}\log(N\vee T)\right),$$

$$\left\|\frac{1}{N}A'A\right\|_{op} \le \left\|\frac{1}{N}\sum_{i\in[N]}\Sigma_i\right\|_{op} + \left\|\frac{1}{N}\left(A'A - \sum_{i\in[N]}\Sigma_i\right)\right\|_{op} \le M + \bar{c}\left(\frac{(N\vee T)}{N}\log(N\vee T)\right),$$

$$\|A\|_{op} = \left\|U\odot x_j^*\right\|_{op} \le \sqrt{(1+\bar{c})(N\vee T)\log(N\vee T)}.$$

Consequently, when $\log(N\vee T)\ge 2$, we have

$$\begin{aligned}
&\mathbb{E}\exp\left(\frac{\left\|U\odot x_j^*\right\|_{op}}{\sqrt{(N\vee T)\log(N\vee T)}}\right)\\
&=\int_0^\infty \exp(u)\mathbb{P}\left(\frac{\left\|U\odot x_j^*\right\|_{op}}{\sqrt{(N\vee T)\log(N\vee T)}}\ge u\right)du\\
&=\left(\int_0^2+\int_2^\infty\right)\exp(u)\mathbb{P}\left(\left\|U\odot x_j^*\right\|_{op}\ge u\sqrt{(N\vee T)\log(N\vee T)}\right)du\\
&\le \int_0^2\exp(u)du+\int_2^\infty \exp(u)\exp\left(-\frac{(u^2-3)}{2}\log(N\vee T)\right)du\\
&\le \int_0^2\exp(u)du+\int_2^\infty \exp\left(-\left(u-\frac{1}{2}\right)^2+\frac{13}{4}\right)du\\
&\le \exp(2)-1+\sqrt{2\pi}\exp\left(\frac{13}{4}\right):=C,
\end{aligned}$$

where the first inequality is by the fact that

$$\mathbb{P}\left(\frac{\left\|U\odot x_j^*\right\|_{op}}{\sqrt{(N\vee T)\log(N\vee T)}}\ge\sqrt{1+\bar{c}}\right)\le \exp\left(-\left(\frac{\bar{c}}{2}-1\right)\log((N\vee T))\right)$$

and by letting $u=\sqrt{1+\bar{c}}\ge 2$ for large $\bar{c}$. Similarly, we can show that when $\log(N\vee T)\ge 2$, $\mathbb{E}\exp\left(\frac{\|U\|_{op}}{\sqrt{(N\vee T)\log(N\vee T)}}\right)\le C$ for some absolute constant C. ■

Recall that $\tilde{\rho}_{it}\left(\left\{\Delta_{\Theta_j,it},X_{j,it}\right\}_{j=0}^p,\varepsilon_{it}\right)$ is defined in (A.4) and $\mathscr{G}_{i,t}$ is defined in Assumption 2.1.

Lemma A.11. *If Assumptions 2.1-2.4 hold, then we have*

$$\sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^p\in\mathscr{R}(3,C_2)}\frac{\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\tilde{\rho}_{it}\left(\left\{\Delta_{\Theta_j,it},X_{j,it}\right\}_{j=0}^p,\varepsilon_{it}\right)\right|}{\sum_{j=0}^p\left\|\Delta_{\Theta_j}\right\|_F}=O_p\left(a_{NT}\right),$$

where $a_{NT}=\frac{\sqrt{(N\vee T)\log(N\vee T)}}{NT}$.

Proof Let $n=NT$ and for $l\in[n]$, $Z_l=(\{X_{j,it}\}_{j\in[p]},\varepsilon_{it})$ with $i=\lceil\frac{l}{T}\rceil$ and $t=l-(i-1)T$, and $\mathscr{F}_l=\mathscr{G}_{it}$. Then, Z_l is adapted to the filtration $\{\mathscr{F}_l\}_{l\in[n]}$. Lemma

A.8 implies

$$\beta_{NT}\mathbb{P}\left\{\sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p}\in\mathscr{R}(3,C_2)}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\tilde{\rho}_{it}\left(\left\{\Delta_{\Theta_j,it},X_{j,it}\right\}_{j=0}^{p},\varepsilon_{it}\right)}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F}\right|>C_5a_{NT}\right\}$$

$$\leq 2\sup_{x^*,\varepsilon^*}\mathbb{P}\left\{\sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p}\in\mathscr{R}(3,C_2)}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{u_{it}\rho_{it}\left(\left\{\Delta_{\Theta_j,it},x^*_{j,it}\right\}_{j=0}^{p},\varepsilon^*_{it}\right)}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F}\right|>\frac{C_5a_{NT}}{4}\right\}, \tag{A.80}$$

for some positive constant C_5, where the outer supremum on the RHS of the above display is taken over $[-\xi_N,\xi_N]^p\times\mathbb{R}$-valued trees with depth n and

$$\beta_{NT}=1-\sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p}\in\mathscr{R}(3,C_2)}\frac{4}{C_5^2(NT)^2a_{NT}^2}\sum_{(i,t)\in[N]\times[T]}Var\left(\frac{\tilde{\rho}_{it}\left(\left\{\Delta_{\Theta_j,it},X_{j,it}\right\}_{j=0}^{p},\varepsilon_{it}\right)}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F}\middle|\mathscr{G}_{i,t-1}\right).$$

We first note

$$\begin{aligned}
&\sum_{(i,t)\in[N]\times[T]}Var\left(\frac{\tilde{\rho}_{it}\left(\left\{\Delta_{\Theta_j,it},X_{j,it}\right\}_{j=0}^{p},\varepsilon_{it}\right)}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F}\middle|\mathscr{G}_{i,t-1}\right)\\
&\leq\sum_{(i,t)\in[N]\times[T]}\mathbb{E}\left[\left(\frac{\rho_{it}\left(\left\{\Delta_{\Theta_j,it},X_{j,it}\right\}_{j=0}^{p},\varepsilon_{it}\right)}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F}\right)^2\middle|\mathscr{G}_{i,t-1}\right]\\
&\leq 2\sum_{(i,t)\in[N]\times[T]}\left(\frac{\Delta_{\Theta_0,it}+\sum_{j=1}^{p}X_{j,it}\Theta_{j,it}}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F}\right)^2\\
&\lesssim\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\sum_{j=1}^{p}\left(X_{j,it}\Delta_{\Theta_j,it}\right)^2+\Delta_{\Theta_0,it}^2}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F^2}\leq c_{10}\xi_N^2
\end{aligned}$$

with some positive constant c_{10}, where the first inequality holds by Jensen inequality, the second inequality is by $|\rho_\tau(u)-\rho_\tau(v)|\leq 2|u-v|$, and the last line holds by Assumption 2.1(iv) and the fact that $\left(\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F\right)^2=O\left(\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F^2\right)$. Therefore, we have

$$\begin{aligned}
\beta_{NT}&\geq 1-\sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p}\in\mathscr{R}(3,C_2)}\frac{4}{C_5^2(NT)^2a_{NT}^2}\sum_{(i,t)\in[N]\times[T]}Var\left(\frac{\tilde{\rho}_{it}\left(\left\{\Delta_{\Theta_j,it},X_{j,it}\right\}_{j=0}^{p},\varepsilon_{it}\right)}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F}\right)\\
&\geq 1-\sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p}\in\mathscr{R}(3,C_2)}\frac{4c_{10}\xi_N^2}{C_5^2(N\vee T)\log(N\vee T)}\\
&\geq 1-O\left(\frac{\xi_N^2}{(N\vee T)\log(N\vee T)}\right)\to 1,
\end{aligned}\tag{A.81}$$

where the last line is by Assumption 2.1(ix).

Define

$$\mathscr{A}_0 = \sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p} \in \mathscr{R}(3,C_2)} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{u_{it} \Delta_{\Theta_0,it}}{\sum_{j=0}^{p} \left\|\Delta_{\Theta_j}\right\|_F} \right|,$$

$$\mathscr{A}_j = \sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p} \in \mathscr{R}(3,C_2)} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{u_{it} x_{j,it}^* \Delta_{\Theta_j,it}}{\sum_{j=0}^{p} \left\|\Delta_{\Theta_j}\right\|_F} \right|, \quad \forall j \in [p],$$

$$\mathscr{A}_{p+1} = \sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p} \in \mathscr{R}(3,C_2)} \left| \frac{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} u_{it} \phi_{it} \left(\Delta_{\Theta_0,it} + \sum_{j=1}^{p} x_{j,it}^* \Delta_{\Theta_j,it}\right)}{\sum_{j=0}^{p} \left\|\Delta_{\Theta_j}\right\|_F} \right|,$$

where $\phi_{it}(u) = (\varepsilon_{it}^* - u)^- - (\varepsilon_{it}^*)^-$. Notice that

$$\rho_{it}\left(\left\{\Delta_{\Theta_j,it}, x_{j,it}^*\right\}_{j=0}^{p}, \varepsilon_{it}^*\right) = \tau\left(\Delta_{\Theta_0,it} + \sum_{j=1}^{p} x_{j,it}^* \Delta_{\Theta_j,it}\right) + \phi_{it}\left(\Delta_{\Theta_0,it} + \sum_{j=1}^{p} x_{j,it}^* \Delta_{\Theta_j,it}\right),$$

we obtain that

$$\sup_{x^*,\varepsilon^*} \mathbb{P}\left\{ \sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p} \in \mathscr{R}(3,C_2)} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{u_{it} \rho_{it}\left(\left\{\Delta_{\Theta_j,it}, x_{j,it}^*\right\}_{j=0}^{p}, \varepsilon_{it}^*\right)}{\sum_{j=0}^{p} \left\|\Delta_{\Theta_j}\right\|_F} \right| > \frac{C_5 a_{NT}}{4} \right\}$$

$$\leq \sum_{j=0}^{p} \sup_{x^*} \mathbb{P}\left\{\tau \mathscr{A}_j > \frac{C_5 a_{NT}}{4(p+2)}\right\} + \sup_{x^*,\varepsilon^*} \mathbb{P}\left\{\mathscr{A}_{p+1} > \frac{C_5 a_{NT}}{4(p+2)}\right\}. \tag{A.82}$$

We first bound $\mathscr{A}_j$ for $j \in [p]$. We have

$$\begin{aligned}
\mathscr{A}_j &= \frac{1}{NT} \sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p} \in \mathscr{R}(3,C_2)} \left| \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{u_{it} x_{j,it}^* \Delta_{\Theta_j,it}}{\sum_{j=0}^{p} \left\|\Delta_{\Theta_j}\right\|_F} \right| \\
&\leq \frac{1}{NT} \sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p} \in \mathscr{R}(3,C_2)} \frac{\left| tr\left[\Delta_{\Theta_j}'\left(U \odot x_j^*\right)\right] \right|}{\sum_{j=0}^{p} \left\|\Delta_{\Theta_j}\right\|_F} \\
&\leq \frac{1}{NT} \left\|U \odot x_j^*\right\|_{op} \sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p} \in \mathscr{R}(3,C_2)} \frac{\left\|\Delta_{\Theta_j}\right\|_*}{\sum_{j=0}^{p} \left\|\Delta_{\Theta_j}\right\|_F} \\
&\leq \frac{c_8}{NT} \left\|U \odot x_j^*\right\|_{op},
\end{aligned} \tag{A.83}$$

where the first inequality holds by $tr(AB) \leq \|A\|_{op} \|B\|_*$ and the second inequality holds by Lemma A.7. Then

$$\sup_{x^*,\varepsilon^*} \mathbb{P}\left\{\tau \mathscr{A}_j > \frac{C_5 a_{NT}}{4(p+2)}\right\}$$

$$
\begin{aligned}
&\le \sup_{x^*} \mathbb{P}\left\{ \frac{\left\| U \odot x_j^* \right\|}{\sqrt{(N\vee T)\log(N\vee T)}} > \frac{C_5}{4c_8\tau(p+2)} \right\} \\
&\le \sup_{x^*} \left\{ \exp\left(-\frac{C_5}{4c_8\tau(p+2)} \right) \mathbb{E}\left[\exp\left(\frac{\left\| U \odot x_j^* \right\|}{\sqrt{(N\vee T)\log(N\vee T)}} \right) \right] \right\} \\
&\le C\exp\left(-\frac{C_5}{4c_8\tau(p+2)} \right) \qquad \text{(A.84)}
\end{aligned}
$$

for some absolute constant C that independent of (x^*, ε^*), where the last inequality holds by Lemma A.10. Similarly, we can establish

$$
\sup_{x^*,\varepsilon^*} \mathbb{P}\left\{ \tau\mathscr{A}_0 > \frac{C_5 a_{NT}}{4(p+2)} \right\} \le C\exp\left(-\frac{C_5}{4c_8\tau(p+2)} \right). \qquad \text{(A.85)}
$$

Next, we turn to $\mathscr{A}_{p+1}$. We have

$$
\begin{aligned}
&\sup_{x^*,\varepsilon^*} \mathbb{P}\left\{ \mathscr{A}_{p+1} > \frac{C_5 a_{NT}}{4(p+2)} \right\} \\
&\le \exp\left(-\frac{C_5}{4c_8(p+1)(p+2)} \right) \sup_{x^*,\varepsilon^*} \mathbb{E}\left\{ \exp\left(\frac{NT\mathscr{A}_{p+1}}{c_8(p+1)\sqrt{(N\vee T)\log(N\vee T)}} \right) \right\}.
\end{aligned} \qquad \text{(A.86)}
$$

Because $\phi_{it}(\cdot)$ is a contraction, Lemma A.9 implies

$$
\begin{aligned}
&\mathbb{E}\left\{ \exp\left(\frac{NT\mathscr{A}_{p+1}}{c_8(p+1)\sqrt{(N\vee T)(\log(N\vee T))}} \right) \right\} \\
&= \mathbb{E}\left\{ \exp\left[\frac{1}{c_8(p+1)\sqrt{(N\vee T)(\log(N\vee T))}} \sup_{\{\Delta_{\Theta_j}\}_{j=0}^p \in \mathscr{R}(3,C_2)} \left| \frac{\sum_{i=1}^N \sum_{t=1}^T u_{it}\phi_{it}\left(\Delta_{\Theta_0,it} + \sum_{j=1}^p x_{j,it}^* \Delta_{\Theta_j,it} \right)}{\sum_{j=0}^p \left\| \Delta_{\Theta_j} \right\|_F} \right| \right] \right\} \\
&\le \mathbb{E}\left\{ \exp\left[\frac{1}{c_8(p+1)\sqrt{(N\vee T)(\log(N\vee T))}} \sup_{\{\Delta_{\Theta_j}\}_{j=0}^p \in \mathscr{R}(3,C_2)} \left| \sum_{i=1}^N \sum_{t=1}^T \frac{u_{it}(\Delta_{\Theta_0,it} + \sum_{j=1}^p x_{j,it}^* \Delta_{\Theta_j,it})}{\sum_{j=0}^p \left\| \Delta_{\Theta_j} \right\|_F} \right| \right] \right\} \\
&\le \mathbb{E}\left\{ \exp\left[\frac{\left(\|U\|_{op} + \sum_{j\in[p]} \left\| U \odot x_j^* \right\|_{op} \right)}{(p+1)\sqrt{(N\vee T)(\log(N\vee T))}} \right] \right\} \\
&\le \mathbb{E}\left\{ \exp\left[\frac{\left(\|U\|_{op} \right)}{\sqrt{(N\vee T)(\log(N\vee T))}} \right] \right\}^{1/(1+p)} \Pi_{j\in[p]} \left[\mathbb{E}\left\{ \exp\left[\frac{\left(\left\| U \odot x_j^* \right\|_{op} \right)}{\sqrt{(N\vee T)(\log(N\vee T))}} \right] \right\}^{1/(1+p)} \right] \le C
\end{aligned} \qquad \text{(A.87)}
$$

for some absolute constant C, where the first inequality is by Lemma A.9, the second inequality is by (A.83), the third inequality is due to the fact that, for random variables $\{A_i\}_{i\in[p+1]}$,

$$\mathbb{E}(\Pi_{i\in[p+1]}|A_i|) \le \Pi_{i\in[p+1]}[\mathbb{E}|A_i^{1+p}|]^{1/(1+p)},$$

and the final inequality is by Lemma A.10 with an absolute constant C that is independent of (x^*, ε^*).

Combining (A.86) and (A.87), we have

$$\sup_{x^*,\varepsilon^*} \mathbb{P}\left\{\mathscr{A}_{p+1} > \frac{C_5 a_{NT}}{4(p+2)}\right\} \le C\exp\left(-\frac{C_5}{4c_8(p+1)(P+2)}\right),$$

which, combined with (A.80), (A.82), (A.84), and (A.85), further implies that

$$\mathbb{P}\left\{\sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p}\in\mathscr{R}(3,C_2)} \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\tilde{\rho}_{it}\left(\left\{\Delta_{\Theta_j,it}, X_{j,it}\right\}_{j=0}^{p}, \varepsilon_{it}\right)}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F}\right| > C_5 a_{NT}\right\}$$

$$\le 2\beta_{NT}^{-1}\sup_{x^*,\varepsilon^*}\mathbb{P}\left\{\sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p}\in\mathscr{R}(3,C_2)} \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{u_{it}\rho_{it}\left(\left\{\Delta_{\Theta_j,it}, x^*_{j,it}\right\}_{j=0}^{p}, \varepsilon^*_{it}\right)}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F}\right| > \frac{C_5 a_{NT}}{4}\right\}$$

$$\le C(1+o(1))\left[(p+1)\exp\left(-\frac{C_5}{4c_8\tau(p+2)}\right) + \exp\left(-\frac{C_5}{4c_8(p+1)(p+2)}\right)\right].$$

The RHS of the last inequality converges to zero as $C_5 \to \infty$, which implies that

$$\sup_{\left\{\Delta_{\Theta_j}\right\}_{j=0}^{p}\in\mathscr{R}(3,C_2)} \frac{\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\rho}_{it}\left(\left\{\Delta_{\Theta_j,it}, X_{j,it}\right\}_{j=0}^{p}, \varepsilon_{it}\right)\right|}{\sum_{j=0}^{p}\left\|\Delta_{\Theta_j}\right\|_F} = O_p(a_{NT}).$$

■

A.2.2 Lemmas for the Proof of Theorem 2.2

Lemma A.12. *Let $\{\Upsilon_t, t = 1,\cdots,T\}$ be a zero-mean strong mixing process, not necessarily stationary, with the mixing coefficients satisfying $\alpha(z) \le c_\alpha \rho^z$ for some $c_\alpha > 0$ and $\rho \in (0,1)$. If $\sup_{1\le t\le T}|\Upsilon_t| \le M_T$, then there exist a constant c_9 depending on c_α and ρ such that for any $T \ge 2$ and $d > 0$,*

(i) $\mathbb{P}\left\{\left|\sum_{t=1}^{T}\Upsilon_t\right| > d\right\} \le \exp\left\{-\frac{c_9 d^2}{M_T^2 T + d M_T(\log T)(\log\log T)}\right\}$.

(ii) $\mathbb{P}\left\{\left|\sum_{t=1}^{T}\Upsilon_t\right| > d\right\} \le \exp\left\{-\frac{c_9 d^2}{\upsilon_0^2 T + M_T^2 + d M_T(\log T)^2}\right\}$,

with $\upsilon_0^2 = \sup_{t\in[T]}\left[Var(\Upsilon_t) + 2\sum_{s>t}|Cov(\Upsilon_t,\Upsilon_s)|\right]$.

Proof The proof is the same as that of Theorems 1 and 2 in Merlevède et al. (2009)) with the condition assumed $\alpha(a) \leq \exp\{-2ca\}$ for some $c > 0$ changed with $c_\alpha = 1$ and $\rho = \exp\{-2c\}$ in our lemma instead. ∎

Lemma A.13. *Suppose Assumptions 2.1-2.4 hold. Then, for $j \in \{0, \cdots, p\}$, we have*

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

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

(iii) $\max_{i\in I_2} \frac{1}{T}\sum_{t\in[T]} \left\|\tilde{\phi}^{(1)}_{it}\right\|_2^2 \leq \frac{4M^2}{c_\sigma} + \frac{4M^2pC}{c_\sigma}$ *w.p.a.1.*

Proof (i) Recall that $\frac{1}{\sqrt{NT}}\Theta^0_j = \mathscr{U}^0_j\Sigma^0_j\mathscr{V}^{0\prime}_j$, $U^0_j = \sqrt{N}\mathscr{U}^0_j\Sigma^0_j$ and $V_j = \sqrt{T}\mathscr{V}_j$. Then we have

$$\frac{1}{\sqrt{T}}\Theta^0_j\mathscr{V}^0_j = \sqrt{N}\mathscr{U}^0_j\Sigma^0_j = U^0_j \quad \text{and} \quad \frac{1}{\sqrt{N}}\mathscr{U}^{0\prime}_j\Theta^0_j = \sqrt{T}\Sigma^0_j\mathscr{V}^{0\prime}_j = \Sigma^0_jV^{0\prime}_j. \tag{A.88}$$

Hence, it's natural to see that

$$\|u^0_{i,j}\|_2 = \frac{1}{\sqrt{T}}\left\|\left[\Theta^0_j\mathscr{V}^0_j\right]_{i.}\right\|_2 \leq \frac{1}{\sqrt{T}}\left\|\left[\Theta_j\right]_{i.}\right\|_2 \leq M,$$

where the first inequality is due to the fact that $\mathscr{V}_j$ is the unitary matrix and the last inequality holds by Assumption 2.2. Since the upper bound M is not dependent on i, this result holds uniformly in i. Analogously,

$$\|v^0_{t,j}\|_2 \leq \frac{1}{\sqrt{N}}c_\sigma^{-1}\left\|\left[\mathscr{U}^{0\prime}_j\Theta^0_j\right]_{.t}\right\|_2 \leq \frac{1}{\sqrt{N}}c_\sigma^{-1}\left\|\left[\Theta^0_j\right]_{.t}\right\|_2 \leq \frac{M}{c_\sigma}.$$

(ii) As in (A.88), we have

$$\frac{1}{\sqrt{N}}\tilde{\mathscr{U}}'_j\tilde{\Theta}_j = \sqrt{T}\tilde{\Sigma}_j\tilde{\mathscr{V}}'_j = \tilde{\Sigma}_j\tilde{V}'_j.$$

It follows that

$$\|O'_j\tilde{v}_{t,j}\|_2 \leq \frac{1}{\sqrt{N}}\tilde{\sigma}^{-1}_{K_j,j}\left\|\left[\tilde{\mathscr{U}}'_j\tilde{\Theta}_j\right]_{.t}\right\|_2 \leq \frac{1}{\sqrt{N}}\tilde{\sigma}^{-1}_{K_j,j}\left\|\left[\tilde{\Theta}_j\right]_{.t}\right\|_2 \leq \frac{2M}{c_\sigma},$$

where the last inequality holds due to the fact that $\max_{k\in[K_j]}\left|\tilde{\sigma}^{-1}_{k,j} - \Sigma^{-1}_{k,j}\right| \leq \Sigma^{-1}_{K_j,j}$ w.p.a.1. and the bounded parameter space where $\tilde{\Theta}_j$ lies in by Assumption 2.2 and ADMM algorithm proposed in the last section. The upper bound of $\max_{t\in[T]}\left\|O^{(1)\prime}_j\tilde{v}^{(1)}_{t,j}\right\|_2$ follows the same argument as above.

(iii) We observe that

$$\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2^2\leq\frac{1}{T}\sum_{t\in[T]}\left\|O_0'\tilde{v}_{t,0}^{(1)}\right\|_2^2+\max_{i\in I_2}\sum_{j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left\|O_j'\tilde{v}_{t,j}^{(1)}\right\|_2^2\left|X_{j,it}\right|^2$$
$$\leq\frac{4M^2}{c_\sigma^2}+\frac{4M^2pC}{c_\sigma^2}\quad\text{w.p.a.1},$$

where the last inequality holds by Lemma A.13(ii) and Assumption 2.1(iv). ■

Lemma A.14. *Under Assumptions 2.1–2.5, we have*

(i) $\min_{i\in I_2}\lambda_{\min}\left(\tilde{\Phi}_i^{(1)}\right)\geq\frac{c_\phi}{2}$, $\max_{i\in I_2}\lambda_{\max}\left(\tilde{\Phi}_i^{(1)}\right)\leq 2C_\phi$ *w.p.a.1,*

(ii) For $\forall j\in[p]$, $\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left[X_{j,it}^2-\mathbb{E}\left(X_{j,it}^2\middle|\mathscr{D}_{\{e_{is}\}_{s<t}}^{I_1}\right)\right]\left\|\tilde{O}_j^{(1)\prime}\tilde{v}_{t,j}^{(1)}-v_{t,j}^0\right\|_2^2=O_p(\eta_N^2)$,

(iii) $\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}-\Phi_{it}^0\right\|_2^2=O_p(\eta_N^2)$.

Proof (i) Recall that

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

Uniformly over $i\in I_2$, it is clear that

$$\begin{aligned}&\left\|\tilde{\Phi}_i^{(1)}-\Phi_i\right\|_F\\&\lesssim\frac{4M}{c_\sigma T}\sum_{t=1}^T\left\|O_0^{(1)\prime}\tilde{v}_{t,0}^{(1)}-v_{t,0}^0\right\|_2+\frac{4M}{c_\sigma T}\sum_{j=1}^p\sum_{t=1}^T\left\|O_j^{(1)\prime}\tilde{v}_{t,j}^{(1)}-v_{t,j}^0\right\|_2\left|X_{j,it}\right|\\&\leq\frac{4M}{c_\sigma}\frac{1}{\sqrt{T}}\left\|O_0^{(1)\prime}\tilde{V}_0^{(1)}-V_0^0\right\|_F+\frac{4M^2}{c_\sigma}\sum_{j=1}^p\frac{1}{\sqrt{T}}\left\|O_j^{(1)\prime}\tilde{V}_j^{(1)}-V_j^0\right\|_F\left(\frac{1}{T}\sum_{t\in[T]}\left|X_{j,it}\right|^2\right)^{1/2}\\&=O_p\left(\eta_N\right),\end{aligned}$$

where the third line holds by Lemma A.13(i) and Assumption 2.1(iv). It follows that

$$\min_{i\in I_2}\lambda_{\min}\left[\tilde{\Phi}_i^{(1)}\right]\geq\min_{i\in I_2}\lambda_{\min}\left[\Phi_i\right]-O\left(\eta_N\right)\geq\frac{c_\phi}{2},\quad\text{w.p.a.1}$$

and

$$\max_{i\in I_2}\lambda_{\max}\left[\tilde{\Phi}_i^{(1)}\right]\leq\max_{i\in I_2}\lambda_{\max}\left[\Phi_i\right]+O\left(\eta_N\right)\leq 2C_\phi,\quad\text{w.p.a.1}.$$

(ii) Let $I_{j,i} := \frac{1}{T}\sum_{t\in[T]} I_{j,it}$ such that

$$I_{j,it} = \left[X_{j,it}^2 - \mathbb{E}\left(X_{j,it}^2 \middle| \mathscr{D}^{I_1}_{\{e_{is}\}_{s<t}}\right)\right] \left\|\tilde{O}_j^{(1)\prime}\tilde{v}_{t,j}^{(1)} - v_{t,j}^0\right\|_2^2.$$

Then for a constant c, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{i\in I_2}\left|\sum_{t\in[T]} I_{j,it}\right| > cT\eta_N^2\right) \leq \sum_{i\in I_2}\mathbb{P}\left(\left|\sum_{t\in[T]} I_{j,it}\right| > cT\eta_N^2\right) \\
&= \sum_{i\in I_2}\mathbb{E}\mathbb{P}\left(\left|\sum_{t\in[T]} I_{j,it}\right| > cT\eta_N^2 \middle| \mathscr{D}^{I_1}_{\{e_{is}\}_{s<T}}\right) \\
&\leq 2\sum_{i\in I_2}\exp\left\{-\frac{2(cT\eta_N^2)^2}{\sum_{t\in[T]}\left[2\xi_N^2\left\|\tilde{O}_j^{(1)\prime}\tilde{v}_{t,j}^{(1)} - v_{t,j}^0\right\|_2^2\right]^2}\right\} \\
&\leq 2\exp\left\{-\frac{2(cT\eta_N^2)^2}{4\xi_N^4\left[\frac{M^2}{c_\sigma^2}+\frac{4M^2}{c_\sigma^2}\right]\sum_{t\in[T]}\left\|\tilde{O}_j^{(1)\prime}\tilde{v}_{t,j}^{(1)} - v_{t,j}^0\right\|_2^2} + \log N\right\} = o(1),
\end{aligned}$$

where the first inequality combines the fact that $I_{j,it}$ is the martingale difference sequence, Assumption 2.1(v) and the Azuma-Hoeffding inequality in Wainwright (2019, Corollary 2.20). The last inequality is by Lemma A.13(i) and A.13(ii), and the finial result is by the definition of η_N.

(iii) Note that

$$\begin{aligned}
&\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)} - \Phi_{it}^0\right\|_2^2 \\
&\leq \frac{1}{T}\sum_{t\in[T]}\left\|\tilde{O}_0^{(1)\prime}\tilde{v}_{t,0}^{(1)} - v_{t,0}^0\right\|_2^2 + p\max_{i\in I_2, j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left|X_{j,it}\right|^2\left\|\tilde{O}_j^{(1)\prime}\tilde{v}_{t,j}^{(1)} - v_{t,j}^0\right\|_2^2 \\
&= \frac{1}{T}\sum_{t\in[T]}\left\|\tilde{O}_0^{(1)\prime}\tilde{v}_{t,0}^{(1)} - v_{t,0}^0\right\|_2^2 \\
&+ p\max_{i\in I_2, j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left[X_{j,it}^2 - \mathbb{E}\left(X_{j,it}^2 \middle| \mathscr{D}^{I_1}_{\{e_{is}\}_{s<t}}\right)\right]\left\|\tilde{O}_j^{(1)\prime}\tilde{v}_{t,j}^{(1)} - v_{t,j}^0\right\|_2^2 \\
&+ \max_{i\in I_2, j\in[p]}\frac{1}{T}\sum_{t\in[T]}\mathbb{E}\left(X_{j,it}^2 \middle| \mathscr{D}^{I_1}_{\{e_{is}\}_{s<t}}\right)\left\|\tilde{O}_j^{(1)\prime}\tilde{v}_{t,j}^{(1)} - v_{t,j}^0\right\|_2^2 \\
&\leq \frac{1}{T}\sum_{t\in[T]}\left\|\tilde{O}_0^{(1)\prime}\tilde{v}_{t,0}^{(1)} - v_{t,0}^0\right\|_2^2 + O_p(\eta_N^2) + pM\max_{j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{O}_j^{(1)\prime}\tilde{v}_{t,j}^{(1)} - v_{t,j}^0\right\|_2^2 \\
&= \frac{1}{T}\left\|O_0^{(1)}\tilde{V}_0^{(1)} - V_0^0\right\|_F^2 + p\max_{j\in[p]}\frac{1}{T}\left\|O_j^{(1)}\tilde{V}_j^{(1)} - V_j^0\right\|_F^2 + O_p(\eta_N^2) \\
&= O_p(\eta_N^2),
\end{aligned}$$

where the the second inequality holds by Assumption 2.1(iv) and Lemma A.14(ii), and the last equality holds by the Theorem 2.1(ii). ■

Lemma A.15. *Recall* $\{A_{1,i},\cdots,A_{7,i}\}_{i\in I_2}$ *and* q_i^I *defined in* (A.5) *and* (A.6), *respectively. Suppose Assumptions 2.1–2.5 hold. Then for any constant* $c_{11} < \min(\frac{3\underline{\mathfrak{f}}}{\bar{\mathfrak{f}}'},1)$, *we have*

$$\max_{i\in I_2}(|A_{m,i}|/\left\|\dot{\Delta}_{i,u}\right\|_2) = O_p(\eta_N), \forall m\in\{1,2,3,5,6,7\} \quad and$$
$$|A_{4,i}| \geq \min\left(\frac{\left(3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'\right)c_\phi\left\|\dot{\Delta}_{i,u}\right\|_2^2}{12}, \frac{\left(3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'\right)\sqrt{c_\phi}q_i^I\left\|\dot{\Delta}_{i,u}\right\|_2}{6\sqrt{2}}\right), \forall i\in I_2, \ w.p.a.1.$$

Proof Recall that $w_{1,it} = \varepsilon_{it} - u_i^{0\prime}\left(\tilde{\Phi}_{it}^{(1)} - \Phi_{it}^0\right)$. Let $w_{2,it} = \tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}$. For some positive constant $c_{11}\in(0,1]$, we observe that

$$\begin{aligned}
A_{4,i} &= \frac{1}{T}\sum_{t=1}^{T}\int_0^{w_{2,it}}\mathbb{E}\left(\mathbf{1}\{\varepsilon_{it}\leq s\} - \mathbf{1}\{\varepsilon_{it}\leq 0\}\middle|\mathscr{D}_{e_i}^{I_1}\right)ds \\
&= \frac{1}{T}\sum_{t=1}^{T}\int_0^{w_{2,it}}\left[\mathfrak{F}_{it}(s) - \mathfrak{F}_{it}(0)\right]ds \geq \frac{1}{T}\sum_{t=1}^{T}\int_0^{c_{11}w_{2,it}}\left[\mathfrak{F}_{it}(s) - \mathfrak{F}_{it}(0)\right]ds \\
&= \frac{1}{T}\sum_{t=1}^{T}\int_0^{c_{11}w_{2,it}}\left[s\mathfrak{f}_{it}(0) + \frac{s^2}{2}\mathfrak{f}_{it}'(\tilde{s})\right]ds \\
&\geq \frac{1}{T}\sum_{t=1}^{T}\left[\frac{c_{11}^2\underline{\mathfrak{f}}\left(\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\right)^2}{2} - \frac{c_{11}^3\bar{\mathfrak{f}}'\left|\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\right|^3}{6}\right]
\end{aligned} \tag{A.89}$$

where $\tilde{s}\in(0,s)$. Here, due to Assumption 2.1(i), the conditional CDF of ε_{it} given $\mathscr{D}_{e_i}^{I_1}$ is the same as that given the σ-field generated by

$$\left\{\{e_{it}\}_{t\in[T]}\bigcup\{V_j^0\}_{j\in[p]\cup\{0\}}\bigcup\{W_j^0\}_{j\in[p]}\right\}$$

, which leads to the second equality of the above display by Assumption 2.1(vii); the first inequality holds by Lemma A.5(ii) for any $c_{11}\in(0,1]$. Here, we choose c_{11} such that $c_{11} < \frac{3\underline{\mathfrak{f}}}{\bar{\mathfrak{f}}'}$; and the last inequality holds by Assumption 2.1(vii).

Let $q_i^{II} = \left[\frac{1}{T}\sum_{t\in[T]}\left(\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\right)^2\right]^{\frac{1}{2}}$ and recall that $q_i^I = \inf_{\Delta}\frac{\left[\frac{1}{T}\sum_{t\in[T]}\left(\tilde{\Phi}_{it}^{(1)\prime}\Delta\right)^2\right]^{\frac{3}{2}}}{\frac{1}{T}\sum_{t\in[T]}\left|\tilde{\Phi}_{it}^{(1)\prime}\Delta\right|^3}$. If $q_i^I\geq q_i^{II}$, we notice that $\frac{1}{T}\sum_{t\in[T]}\left|\tilde{\Phi}_{it}^{(1)\prime}\Delta\right|^3 < \left(q_i^{II}\right)^2$ and $A_{4,i}\geq\frac{c_{11}^2\underline{\mathfrak{f}}\left(q_i^{II}\right)^2}{2} - \frac{c_{11}^3\bar{\mathfrak{f}}'\left(q_i^{II}\right)^2}{6} = \frac{3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'}{6}\left(q_i^{II}\right)^2$. If $q_i^I < q_i^{II}$, we have $\left[\frac{1}{T}\sum_{t\in[T]}\left(\tilde{\Phi}_{it}^{(1)\prime}\Delta_{i,u}^*\right)^2\right]^{\frac{1}{2}} = q_i^I$ with $\Delta_{i,u}^* =$

$\frac{q_i^{II}\dot{\Delta}_{i,u}}{q_i^{I}}$. Define the function

$$F(\Delta) = \frac{1}{T}\sum_{t=1}^{T}\int_0^{\tilde{\Phi}_{it}^{(1)\prime}\Delta}\left[\mathfrak{F}_{it}(s) - \mathfrak{F}_{it}(0)\right]ds.$$

Note that the second-order derivative of function $F(\Delta)$ is no less than zero, which implies $F(\Delta)$ is convex. Therefore, we have

$$\begin{aligned}
F\left(\dot{\Delta}_{i,u}\right) = F\left(\frac{q_i^{I}\Delta_{i,u}^{*}}{q_i^{II}}\right) &\geq \frac{q_i^{I}}{q_i^{II}}F\left(\Delta_{i,u}^{*}\right) \geq \frac{q_i^{II}}{q_i^{I}}\frac{3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'}{6}\frac{1}{T}\sum_{t\in[T]}\left(\tilde{\Phi}_{it}^{(1)\prime}\Delta_{i,u}^{*}\right)^2 \\
&= \frac{\left(3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'\right)q_i^{I}q_i^{II}}{6}.
\end{aligned}$$

Combining these two cases, we have

$$\begin{aligned}
A_{4,i} &\geq \min\left(\frac{3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'}{6}\left(q_i^{II}\right)^2, \frac{\left(3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'\right)q_i^{I}q_i^{II}}{6}\right) \\
&\geq \min\left(\frac{\left(3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'\right)c_\phi\left\|\dot{\Delta}_{i,u}\right\|_2^2}{12}, \frac{\left(3c_{11}^2\underline{\mathfrak{f}} - c_{11}^3\bar{\mathfrak{f}}'\right)\sqrt{c_\phi}q_i^{I}\left\|\dot{\Delta}_{i,u}\right\|_2}{6\sqrt{2}}\right), \quad \text{(A.90)}
\end{aligned}$$

where the second inequality holds by Lemma A.14(i).

As for $\left|A_{1,i}\right|$, we notice that

$$\begin{aligned}
&\max_{i\in I_2}\left(\left|A_{1,i}\right| / \left\|\dot{\Delta}_{i,u}\right\|_2\right) \\
&= \max_{i\in I_2}\frac{\left|\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\tilde{\Phi}_{it}^{(1)\prime}\left(\tau - \mathbf{1}\left\{\varepsilon_{it} \leq u_i^{0\prime}\left(\tilde{\Phi}_{it}^{(1)} - \Phi_{it}^{0}\right)\right\}\right)\middle|\mathscr{D}_{e_i}^{I_1}\right\}\dot{\Delta}_{i,u}\right|}{\left\|\dot{\Delta}_{i,u}\right\|_2} \\
&= \max_{i\in I_2}\frac{\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\left(\tau - \mathfrak{F}_{it}\left[u_i^{0\prime}\left(\tilde{\Phi}_{it}^{(1)} - \Phi_{it}^{0}\right)\right]\right)\right|}{\left\|\dot{\Delta}_{i,u}\right\|_2} \\
&= \max_{i\in I_2}\frac{\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\left[\mathfrak{f}_{it}(s_{it})u_i^{0\prime}\left(\tilde{\Phi}_{it}^{(1)} - \Phi_{it}^{0}\right)\right]\right|}{\left\|\dot{\Delta}_{i,u}\right\|_2} \\
&\leq \max_{i\in I_2}\frac{\bar{\mathfrak{f}}}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\left\|\tilde{\Phi}_{it}^{(1)} - \Phi_{it}^{0}\right\|_2\left\|u_i^{0}\right\|_2 \\
&\leq \max_{i\in I_2}\sqrt{\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|^2}\sqrt{\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)} - \Phi_{it}^{0}\right\|^2}\left\|u_i^{0}\right\|_2 \\
&\leq O_p(\eta_N), \qquad \text{(A.91)}
\end{aligned}$$

where the second and third equalities hold by Assumption 2.1(vii) and mean-value theorem with some $|s_{it}| \in \left(0, \left|u_i^{0\prime}(\tilde{\Phi}_{it}^{(1)} - \Phi_{it}^{0})\right|\right)$, the second inequality holds by Cauchy-Schwarz inequality, and the third inequality holds by Lemmas A.13(i), A.13(iii) and A.14(ii).

For $A_{3,i}$, note that

$$
\begin{aligned}
A_{3,i} &= \frac{1}{T}\sum_{t\in[T]}\int_0^{\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}} \left(\mathbf{1}\{\varepsilon_{it}\le s\}-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)-\left\{\mathbb{E}\left(\mathbf{1}\{\varepsilon_{it}\le s\}-\mathbf{1}\{\varepsilon_{it}\le 0\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right\}ds \\
&= \int_0^1\left\{\frac{1}{T}\sum_{t\in[T]}\tilde{\Phi}_{it}^{(1)}\left(\mathbf{1}\left\{\varepsilon_{it}\le\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}s^*\right\}-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)\right. \\
&\quad \left.-\left[\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}\le\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}s^*\right\}-\mathbf{1}\{\varepsilon_{it}\le 0\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right]\right\}'\dot{\Delta}_{i,u}ds^* \\
&\le \sup_{s\in\mathbb{R}}\left\|A_{3,i}^{I}(s)\right\|_2\left\|\dot{\Delta}_{i,u}\right\|_2 \qquad\qquad (A.92)
\end{aligned}
$$

by change of variables with $s^*=\frac{s}{\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}}$, $A_{3,i}^{I}(s)=\frac{1}{T}\sum_{t\in[T]}A_{3,it}^{I}(s)$ and

$$
A_{3,it}^{I}(s)=\tilde{\Phi}_{it}^{(1)}\left[\left(\mathbf{1}\{\varepsilon_{it}\le s\}-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)-\mathbb{E}\left(\mathbf{1}\{\varepsilon_{it}\le s\}-\mathbf{1}\{\varepsilon_{it}\le 0\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right].
$$

Below, we aim to show $\sup_{s\in(-\infty,+\infty)}\max_{i\in I_2}\left\|A_{3,i}^{I}(s)\right\|_2=O_p(\eta_N)$.

When $|s|>T^{1/4}$, we notice that

$$
\begin{aligned}
&\sup_{|s|>T^{1/4}}\max_{i\in I_2}\left\|A_{3,i}^{I}(s)\right\|_2 \\
&\le \sup_{|s|>T^{1/4}}\max_{i\in I_2}\left\|\frac{1}{T}\sum_{t\in[T]}\tilde{\Phi}_{it}^{(1)}\left[\mathbf{1}\{\varepsilon_{it}\le s\}-\mathbb{E}\left(\mathbf{1}\{\varepsilon_{it}\le s\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right]\right\|_2 \\
&+\max_{i\in I_2}\left\|\frac{1}{T}\sum_{t\in[T]}\tilde{\Phi}_{it}^{(1)}\left[\mathbf{1}\{\varepsilon_{it}\le 0\}-\mathbb{E}\left(\mathbf{1}\{\varepsilon_{it}\le 0\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right]\right\|_2 \\
&\le \max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}+\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right) \\
&+\max_{i\in I_2}\left\|\frac{1}{T}\sum_{t\in[T]}\tilde{\Phi}_{it}^{(1)}\left[\mathbf{1}\{\varepsilon_{it}\le 0\}-\mathbb{E}\left(\mathbf{1}\{\varepsilon_{it}\le 0\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right]\right\|_2 \\
&\le 2\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right) \\
&+\max_{i\in I_2}\left|\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\left[\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}-\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right]\right| \\
&+\max_{i\in I_2}\left|\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\left[\mathbf{1}\{\varepsilon_{it}\le 0\}-\mathbb{E}\left(\mathbf{1}\{\varepsilon_{it}\le 0\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right]\right|,
\end{aligned}
$$

where the first and third inequalities hold by triangle inequality. Besides, we observe that

$$
\begin{aligned}
&\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\\
&=\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\mathbb{P}\left(\varepsilon_{it}>T^{1/4}\middle|\mathscr{D}_{e_i}^{I_1}\right)\leq\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\frac{\mathbb{E}\left(\varepsilon_{it}^2\middle|\mathscr{D}_{e_i}^{I_1}\right)}{\sqrt{T}}\\
&\leq T^{-1/2}\max_{i\in I_2}\sqrt{\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2^2}\max_{i\in I_2}\sqrt{\frac{1}{T}\sum_{t\in[T]}\left[\mathbb{E}\left(\varepsilon_{it}^2\middle|\mathscr{D}_{e_i}\right)\right]^2}=O_p\left(T^{-1/2}\right),
\end{aligned}
\tag{A.93}
$$

where the last equality holds by Lemma A.13(iii) and Assumption 2.1(iv). For a positive constant c_{12},

$$
\begin{aligned}
&\mathbb{P}\left(\max_{i\in I_2}\left|\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\left[\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}-\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right]\right|>c_{12}\frac{\xi_N\sqrt{\log(N\vee T)}}{\sqrt{T}}\right)\\
&\leq\sum_{i\in I_2}\mathbb{E}\mathbb{P}\left(\left|\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\left[\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}-\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right]\right|>c_{12}\frac{\xi_N\sqrt{\log(N\vee T)}}{\sqrt{T}}\middle|\mathscr{D}_{e_i}^{I_1}\right)\\
&\leq\sum_{i\in I_2}\exp\left\{-\frac{c_9c_{12}^2T\xi_N^2\log(N\vee T)}{\frac{16TM^2}{c_\sigma^2}\left(1+p\xi_N^2\right)+\frac{4Mc_{12}}{c_\sigma}\sqrt{1+p\xi_N}\sqrt{T}\xi_N\sqrt{\log(N\vee T)}\log T\left(\log\log T\right)}\right\}=o(1),
\end{aligned}
$$

where the first inequality holds by the union bound and Assumption 2.1(i), and the second inequality holds by Assumption 2.1(iii) and the conditional Bernstein's inequality in Lemma A.12(i) with the fact that $\max_{i\in I_2,t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\leq\frac{2M}{c_\sigma}\sqrt{1+p\xi_N^2}$ w.p.a.1. Here, we can apply the conditional Bernstein's inequality because $\tilde{\Phi}_{it}^{(1)}$ and $\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}>T^{1/4}\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)$ are deterministic given $\left\{\mathscr{D}_{e_i}^{I_1}\right\}_{i\in I_2,t\in[T]}$ so that the only randomness comes from $\{\varepsilon_{it}\}_{t\in[T]}$. Furthermore, the joint distribution of $\{\varepsilon_{it}\}_{t\in[T]}$ given $\mathscr{D}_{e_i}^{I_1}$ is the same as that given the σ-field generated by $\mathscr{D}_{e_i}$ due to the independence structure assumed in Assumption 2.1(i). Last, given the σ-field generated by $\mathscr{D}_{e_i}$, $\{\varepsilon_{it}\}_{t\in[T]}$ is strong mixing with mixing coefficient $\alpha_i(\cdot)$ as assumed by Assumption 2.1(iii). Similarly, we obtain that

$$
\max_{i\in I_2}\left|\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right]\right|=O_p\left(\frac{\sqrt{\log(N\vee T)}\xi_N}{\sqrt{T}}\right),
$$

which implies

$$
\sup_{|s|>T^{1/4}}\max_{i\in I_2}\left|A_{3,i}^I(s)\right|=O_p(\eta_N). \tag{A.94}
$$

For $|s| \leq T^{1/4}$, let $\mathbb{S} = [-T^{1/4}, T^{1/4}]$ and divide $\mathbb{S}$ into $\mathbb{S}_m$ for $m = 1, \cdots, n_{\mathbb{S}}$ such that $|s - \bar{s}| < \frac{\varepsilon}{T}$ for s and $\bar{s} \in \mathbb{S}_m$ and $n_{\mathbb{S}} \asymp T^{5/4}$. Let $s_m \in \mathbb{S}_m$. For any $s \in \mathbb{S}_m$, we have

$$\left\| \frac{1}{T} \sum_{t \in [T]} A_{3,it}^{I}(s) \right\|_2 \leq \left\| \frac{1}{T} \sum_{t \in [T]} A_{3,it}^{I}(s_m) \right\|_2 + \left\| \frac{1}{T} \sum_{t \in [T]} \left[A_{3,it}^{I}(s) - A_{3,it}^{I}(s_m) \right] \right\|_2, \tag{A.95}$$

such that

$$\begin{aligned}
&\max_{i \in I_2, m \in [n_{\mathbb{S}}]} \sup_{s \in \mathbb{S}_m} \left\| \frac{1}{T} \sum_{t \in [T]} \left[A_{3,it}^{I}(s) - A_{3,it}^{I}(s_m) \right] \right\|_2 \\
&= \max_{i \in I_2, m \in [n_{\mathbb{S}}]} \sup_{s \in \mathbb{S}_m} \Bigg\| \frac{1}{T} \sum_{t \in [T]} \tilde{\Phi}_{it}^{(1)} \left(\mathbf{1}\{\varepsilon_{it} \leq s\} - \mathbf{1}\{\varepsilon_{it} \leq s_m\} \right) \\
&- \tilde{\Phi}_{it}^{(1)} \left[\mathbb{E}\left(\mathbf{1}\{\varepsilon_{it} \leq s\} - \mathbf{1}\{\varepsilon_{it} \leq s_m\} \middle| \mathscr{D}_{e_i}^{I_1} \right) \right] \Bigg\|_2 \\
&\leq \max_{i \in I_2, m \in [n_{\mathbb{S}}]} \frac{2}{T} \sum_{t \in [T]} \left\| \tilde{\Phi}_{it}^{(1)} \right\|_2 \mathbb{E}\left(\mathbf{1}\left\{ \varepsilon_{it} \in \left[s_m - \frac{\varepsilon}{T}, s_m + \frac{\varepsilon}{T} \right] \right\} \middle| \mathscr{D}_{e_i}^{I_1} \right) \\
&+ \max_{i \in I_2, m \in [n_{\mathbb{S}}]} \Bigg| \frac{1}{T} \sum_{t \in [T]} \left\| \tilde{\Phi}_{it}^{(1)} \right\|_2 \Bigg[\left(\mathbf{1}\left\{ \varepsilon_{it} \in \left[s_m - \frac{\varepsilon}{T}, s_m + \frac{\varepsilon}{T} \right] \right\} \right) \\
&- \mathbb{E}\left(\mathbf{1}\left\{ \varepsilon_{it} \in \left[s_m - \frac{\varepsilon}{T}, s_m + \frac{\varepsilon}{T} \right] \right\} \middle| \mathscr{D}_{e_i}^{I_1} \right) \Bigg] \Bigg| \\
&:= \max_{i \in I_2, m \in [n_{\mathbb{S}}]} \frac{2}{T} \sum_{t \in [T]} \left\| \tilde{\Phi}_{it}^{(1)} \right\|_2 \mathbb{E}\left(\mathbf{1}\left\{ \varepsilon_{it} \in \left[s_m - \frac{\varepsilon}{T}, s_m + \frac{\varepsilon}{T} \right] \right\} \middle| \mathscr{D}_{e_i}^{I_1} \right) \\
&+ \max_{i \in I_2, m \in [n_{\mathbb{S}}]} \left| A_{3,i}^{II}(m) \right|.
\end{aligned} \tag{A.96}$$

For the first term in (A.96), we notice that

$$\begin{aligned}
&\max_{i \in I_2, m \in [n_{\mathbb{S}}]} \frac{2}{T} \sum_{t \in [T]} \left\| \tilde{\Phi}_{it}^{(1)} \right\|_2 \mathbb{E}\left(\mathbf{1}\left\{ \varepsilon_{it} \in \left[s_m - \frac{\varepsilon}{T}, s_m + \frac{\varepsilon}{T} \right] \right\} \middle| \mathscr{D}_{e_i}^{I_1} \right) \\
&= \max_{i \in I_2, m \in [n_{\mathbb{S}}]} \frac{2}{T} \sum_{t \in [T]} \left\| \tilde{\Phi}_{it}^{(1)} \right\|_2 \left(\mathfrak{F}_{it}(s_m + \frac{\varepsilon}{T}) - \mathfrak{F}_{it}(s_m - \frac{\varepsilon}{T}) \right) \\
&\leq \max_{i \in I_2} \frac{2}{T} \sum_{t \in [T]} \left\| \tilde{\Phi}_{it}^{(1)} \right\|_2 \frac{2\varepsilon}{T} \\
&\leq \frac{6M\varepsilon}{c_\sigma T} (1 + \sqrt{p}C) \quad \text{w.p.a.1},
\end{aligned} \tag{A.97}$$

where the first inequality is by mean-value theorem, and the second inequality is due to the fact that

$$\begin{aligned}\max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2 &\le \max_{i\in I_2}\frac{1}{T}\sum_{t\in[T]}\sqrt{\frac{4M^2}{c_\sigma}\left(1+p\max_{j\in[p]}\left|X_{j,it}\right|^2\right)}\\ &\le \max_{i\in I_2}\frac{2M}{c_\sigma}\frac{1}{T}\sum_{t\in[T]}\left(1+\sqrt{p}\max_{j\in[p]}\left|X_{j,it}\right|\right)\\ &\le \frac{2M\left(1+\sqrt{p}C\right)}{c_\sigma}\quad \text{w.p.a.1.}\end{aligned}$$

For $A_{3,i}^{II}(m)$, let $A_{3,i}^{II}(m)=\frac{1}{T}\sum_{t\in[T]}A_{3,it}^{II}(m)$ with

$$A_{3,it}^{II}(m)=\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\left[\left(\mathbf{1}\left\{\varepsilon_{it}\in\left[s_m-\frac{\varepsilon}{T},s_m+\frac{\varepsilon}{T}\right]\right\}\right)-\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}\in\left[s_m-\frac{\varepsilon}{T},s_m+\frac{\varepsilon}{T}\right]\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\right].$$

We first observe that

$$\begin{aligned}\max_{i\in I_2,m\in[n_{\mathbb{S}}],t\in[T]}Var\left(A_{3,it}^{II}(m)\middle|\mathscr{D}_{e_i}^{I_1}\right) &\le \max_{i\in I_2,t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2^2\mathbb{E}\left(\mathbf{1}\left\{\varepsilon_{it}\in\left[s_m-\frac{\varepsilon}{T},s_m+\frac{\varepsilon}{T}\right]\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)\\ &\le \max_{i\in I_2,t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2^2\frac{2\varepsilon}{T}\lesssim\frac{\xi_N^2\varepsilon}{T},\quad w.p.a.1,\end{aligned}$$

where the first inequality is by $Var(x)\le\mathbb{E}(x^2)$ for any random variable x and the second inequality is by the mean-value theorem and Assumption 2.1(v). Similarly, we have

$$\begin{aligned}&\max_{i\in I_2,m\in[n_{\mathbb{S}}],t\in[T]}\sum_{s=t+1}^{T}\left|Cov\left(A_{3,it}^{II}(m),A_{3,is}^{II}(m)\middle|\mathscr{D}_{e_i}^{I_1}\right)\right|\\ &\le \max_{i\in I_2,m\in[n_{\mathbb{S}}],t\in[T]}\sum_{s=t+1}^{T}\mathbb{E}\left(\left|A_{3,it}^{II}(m)\right|^3\middle|\mathscr{D}_{e_i}^{I_1}\right)^{1/3}\mathbb{E}\left(\left|A_{3,is}^{II}(m)\right|^3\middle|\mathscr{D}_{e_i}^{I_1}\right)^{1/3}[\alpha(t-s)]^{1/3}\\ &\lesssim \max_{i\in I_2,m\in[n_{\mathbb{S}}],t\in[T]}\mathbb{E}\left(\left|A_{3,it}^{II}(m)\right|^3\middle|\mathscr{D}_{e_i}^{I_1}\right)^{2/3}\lesssim\xi_N^2\left(\frac{\varepsilon}{T}\right)^{2/3},\quad w.p.a.1.\end{aligned}$$

It follows that

$$\begin{aligned}&\max_{i\in I_2,m\in[n_{\mathbb{S}}],t\in[T]}\left\{Var\left(A_{3,it}^{II}(m)\middle|\mathscr{D}_{e_i}^{I_1}\right)+2\sum_{s=t+1}^{T}\left|Cov\left(A_{3,it}^{II}(m),A_{3,is}^{II}(m)\middle|\mathscr{D}_{e_i}^{I_1}\right)\right|\right\}\\ &\le c_{13}\xi_N^2\left(\frac{\varepsilon}{T}\right)^{2/3},\quad w.p.a.1.\end{aligned}$$

and $\max_{i\in I_2,m\in[n_{\mathbb{S}}],t\in[T]}\left|A_{3,it}^{II}(m)\right|\le C_{13}\xi_N$ w.p.a.1 for positive constants c_{13} and c_{14}. Denote the events

$$\mathscr{A}_{3,N}=\left(\begin{array}{c}\max_{i\in I_2,m\in[n_{\mathbb{S}}],t\in[T]}\left\{Var\left(A_{3,it}^{II}(m)\middle|\mathscr{D}_{e_i}^{I_1}\right)+2\sum_{s=t+1}^{T}\left|Cov\left(A_{3,it}^{II}(m),A_{3,is}^{II}(m)\middle|\mathscr{D}_{e_i}^{I_1}\right)\right|\right\}\le c_{13}\xi_N^2\left(\frac{\varepsilon}{T}\right)^{2/3}\\ \text{and}\quad \max_{i\in I_2,t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|_2\le c_{14}\xi_N\end{array}\right)$$

$$\mathscr{A}_{3,N,i} = \left(\begin{array}{c} \max_{m\in[n_{\mathbb{S}}],t\in[T]} \left\{ Var\left(A_{3,it}^{II}(m) \middle| \mathscr{D}_{e_i}^{I_1} \right) + 2\sum_{s=t+1}^{T} \left| Cov\left(A_{3,it}^{II}(m), A_{3,is}^{II}(m) \middle| \mathscr{D}_{e_i}^{I_1} \right) \right| \right\} \leq c_{13}\xi_N^2 \left(\frac{\varepsilon}{T}\right)^{2/3} \\ \text{and} \quad \max_{t\in[T]} \left\| \tilde{\Phi}_{it}^{(1)} \right\|_2 \leq c_{14}\xi_N \end{array} \right).$$

Then, we have $\mathbb{P}(\mathscr{A}_{3,N}^c) \to 0$ and

$$\begin{aligned}
&\mathbb{P}\left(\max_{i\in I_2, m\in[n_{\mathbb{S}}]} \left| A_{3,i}^{II}(m) \right| > c_{12}\eta_N \right) \\
&\leq \mathbb{P}\left(\max_{i\in I_2, m\in[n_{\mathbb{S}}]} \left| A_{3,i}^{II}(m) \right| > c_{12}\eta_N, \mathscr{A}_{3,N} \right) + \mathbb{P}(\mathscr{A}_{3,N}^c) \\
&\leq \sum_{i\in[I_2]} \mathbb{P}\left(\max_{m\in[n_{\mathbb{S}}]} \left| A_{3,i}^{II}(m) \right| > c_{12}\eta_N, \mathscr{A}_{3,N} \right) + \mathbb{P}(\mathscr{A}_{3,N}^c) \\
&\leq \sum_{i\in[I_2]} \mathbb{P}\left(\max_{m\in[n_{\mathbb{S}}]} \left| A_{3,i}^{II}(m) \right| > c_{12}\eta_N, \mathscr{A}_{3,N,i} \right) + \mathbb{P}(\mathscr{A}_{3,N}^c) \\
&= \sum_{i\in[I_2]} \mathbb{E}\mathbb{P}\left(\max_{m\in[n_{\mathbb{S}}]} \left| A_{3,i}^{II}(m) \right| > c_{12}\eta_N \middle| \mathscr{D}_{e_i}^{I_1} \right) \mathbf{1}\{\mathscr{A}_{3,N,i}\} + \mathbb{P}(\mathscr{A}_{3,N}^c) \\
&\leq \sum_{i\in I_2, m\in[n_{\mathbb{S}}]} \exp\left(-\frac{c_9 c_{12}^2 T^2 \eta_N^2}{c_{13} T \xi_N^2 \left(\frac{\varepsilon}{T}\right)^{2/3} + c_{14}^2 \xi_N^2 + c_{12}c_{14} T \eta_N \xi_{N,1} (\log T)^2} \right) + o(1) \\
&= o(1),
\end{aligned}$$

where second inequality is by the union bound, the first inequality is by $\mathscr{A}_{3,N} \subset \mathscr{A}_{3,N,i}$, the equality is by the fact that $\mathscr{A}_{3,N,i}$ is $\mathscr{D}_{e_i}^{I_1}$ measurable, and the last inequality is by Lemma A.12(ii), the definition of $\mathscr{A}_{3,N,i}$, and the fact that $\{\varepsilon_{it}\}_{t\in[T]}$ is strong mixing given $\mathscr{D}_{e_i}^{I_1}$. This implies

$$\max_{i\in I_2, m\in[n_{\mathbb{S}}]} \sup_{s\in\mathbb{S}_m} \left\| \frac{1}{T} \sum_{t\in[T]} \left[A_{3,it}^{I}(s) - A_{3,it}^{I}(s_m) \right] \right\|_2 = O_p(\eta_N).$$

Last, we turn to the first term of (A.95). Denote $A_{3,i}^{I,k}(s_m)$ as the k^{th} element of $A_{3,i}^{I,k}(s_m)$ and the event set $\mathscr{A}_{4,N,i} = \{\max_{t\in[T]} \left\| \tilde{\Phi}_{it}^{(1)} \right\|_2 \leq c_{14}\xi_N\}$. Similarly, we have $\mathbb{P}(\cap_{i\in I_2} \mathscr{A}_{4,N,i}^c) = \mathbb{P}(\max_{i\in I_2, t\in[T]} \left\| \tilde{\Phi}_{it}^{(1)} \right\|_2 > c_{14}\xi_N) = o(1)$. Following the same argument as above, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{i\in I_2, m\in[n_{\mathbb{S}}]} \left| A_{3,i}^{I,k}(s_m) \right| > c_{12}\eta_N \right) \\
&\leq \sum_{m\in[n_{\mathbb{S}}], i\in I_2} \mathbb{E}\mathbb{P}\left(\left| A_{3,i}^{I,k}(s_m) \right| > \frac{c_{12}}{2}\eta_N \middle| \mathscr{D}_{e_i}^{I_1} \right) 1\{\mathscr{A}_{4,N,i}\} + o(1) \\
&\leq \sum_{m\in[n_{\mathbb{S}}], i\in I_2} \exp\left(-\frac{c_9 c_{12}^2 T^2 \eta_N^2 / 4}{c_{14}^2 T \xi_N^2 + c_{12}c_{14} T \eta_N \xi_N \log T (\log\log T)/2} \right) + o(1) = o(1),
\end{aligned} \tag{A.98}$$

where the second inequality combines Lemma A.12(i). Combining (A.92), (A.94) and (A.98), we have

$$\max_{i\in I_2}\frac{|A_{3,i}|}{\|\dot{\Delta}_{i,u}\|_2}=O_p(\eta_N).$$

We now turn to $A_{2,i}$. Let

$$A_{2,it}^{I}=\left(\tau-\mathbf{1}\left\{w_{1,it}\leq 0\right\}\right)\tilde{\Phi}_{it}^{(1)}-\mathbb{E}\left(\left(\tau-\mathbf{1}\left\{w_{1,it}\leq 0\right\}\right)\tilde{\Phi}_{it}^{(1)}\middle|\mathscr{D}_{e_i}^{I_1}\right).$$

Then $A_{2,i}=\frac{1}{T}\sum_{t=1}^{T}\left(A_{2,it}^{I}\right)'\dot{\Delta}_{i,u}$. By conditional Bernstein's inequality and similarly as (A.98), we can show that

$$\mathbb{P}\left\{\max_{i\in I_2}\left\|\frac{1}{T}\sum_{t=1}^{T}A_{2,it}^{I}\right\|_2\geq c_{12}\eta_N\right\}=o(1),$$

which implies $\max_{i\in I_2}\frac{|A_{2,i}|}{\|\dot{\Delta}_{i,u}\|_2}=O_p\left(\eta_N\right)$. By similar arguments for $A_{3,i}$, we can also show that $\max_{i\in I_2}\frac{|A_{5,i}|}{\|\dot{\Delta}_{i,u}\|_2}=O_p\left(\eta_N\right)$. For $A_{6,i}$, we note that

$$\begin{aligned}
\max_{i\in I_2}\frac{|A_{6,i}|}{\|\dot{\Delta}_{i,u}\|_2}&=\max_{i\in I_2}\frac{\left|\frac{1}{T}\sum_{t=1}^{T}\int_0^{\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}}\mathbb{E}\left(\mathbf{1}\left\{w_{1,it}\leq s\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq s\right\}\middle|\mathscr{D}_{e_i}^{I_1}\right)ds\right|}{\|\dot{\Delta}_{i,u}\|_2}\\
&=\max_{i\in I_2}\frac{\left|\frac{1}{T}\sum_{t=1}^{T}\int_0^{\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}}\left[\mathfrak{F}_{it}\left(u_i^{0\prime}\left(\tilde{\Phi}_{it}^{(1)}-\Phi_{it}^{0}\right)+s\right)-\mathfrak{F}_{it}(s)\right]ds\right|}{\|\dot{\Delta}_{i,u}\|_2}\\
&\leq\max_{i\in I_2}\frac{\left|\frac{1}{T}\sum_{t=1}^{T}\int_0^{\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}}u_i^{0\prime}\left(\tilde{\Phi}_{it}^{(1)}-\Phi_{it}^{0}\right)\mathfrak{f}_{it}(s)ds\right|}{\|\dot{\Delta}_{i,u}\|_2}\\
&\leq\max_{i\in I_2}\frac{\frac{\bar{\mathfrak{f}}}{T}\sum_{t=1}^{T}\left|\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\right|\left|u_i^{0\prime}\left(\tilde{\Phi}_{it}^{(1)}-\Phi_{it}^{0}\right)\right|}{\|\dot{\Delta}_{i,u}\|_2}\\
&\leq\bar{\mathfrak{f}}\max_{i\in I_2}\sqrt{\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}\right\|^2}\sqrt{\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\Phi}_{it}^{(1)}-\Phi_{it}^{0}\right\|^2}\|u_i^0\|_2\\
&=O_p\left(\eta_N\right),
\end{aligned}$$

where the first inequality is by mean-value theorem and the other inequalities holds by similar reasons as those in (A.91).

Last, for $A_{7,i}$, we have

$$\begin{aligned}
&\frac{1}{T}\sum_{t\in[T]}\left[\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\left(\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{w_{1,it}\leq 0\right\}\right)\right]\\
&=\frac{1}{T}\sum_{t\in[T]}\mathbb{E}\left[\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\left(\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{w_{1,it}\leq 0\right\}\right)\middle|\mathscr{D}_{e_i}^{I_1}\right]
\end{aligned}$$

$$+\frac{1}{T}\sum_{t\in[T]}\left\{\left[\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\left(\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{w_{1,it}\leq 0\right\}\right)\right]\right.$$
$$\left.-\mathbb{E}\left[\tilde{\Phi}_{it}^{(1)\prime}\dot{\Delta}_{i,u}\left(\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{w_{1,it}\leq 0\right\}\right)\middle|\mathscr{D}_{e_i}^{I_1}\right]\right\}$$
$$:=A_{7,i}^{I}+A_{7,i}^{II},$$

Then, following the same analyses of $A_{1,i}$ and $A_{3,i}$, we have

$$\max_{i\in I_2}\frac{\left|A_{7,i}^{I}\right|}{\left\|\dot{\Delta}_{i,u}\right\|_2}=O_p(\eta_N)\quad\text{and}\quad\max_{i\in I_2}\frac{\left|A_{7,i}^{II}\right|}{\left\|\dot{\Delta}_{i,u}\right\|_2}=O_p(\eta_N),$$

which implies $\max_{i\in I_2}\frac{|A_{7,i}|}{\left\|\dot{\Delta}_{i,u}\right\|_2}=O_p(\eta_N)$. ■

Recall that event $\mathscr{A}_4=\left\{\max_{i\in I_2}\left\|O_j^{(1)\prime}\dot{u}_{i,j}^{(1)}-u_{i,j}^{0}\right\|_2=O(\eta_N),\forall j\in[p]\cup\{0\}\right\}$ where we define $q_i^{III}=\inf_{\Delta}\frac{\left[\frac{1}{N_2}\sum_{i\in I_2}\left(\tilde{\Psi}_{it}^{(1)\prime}\Delta\right)^2\right]^{\frac{3}{2}}}{\frac{1}{N_2}\sum_{i\in I_2}\left|\tilde{\Psi}_{it}^{(1)\prime}\Delta\right|^3}$.

Lemma A.16. . *Suppose Assumptions 2.1–2.5 hold. Then for* $\{B_{1,t},\cdots,B_{6,t}\}_{t\in[T]}$ *defined in* (A.8)*, for any constant* $0<c_{11}<\min(\frac{3\underline{\mathfrak{f}}}{\bar{\mathfrak{f}}'},1)$*, we have*

$$\max_{t\in[T]}\frac{|B_{m,t}|}{\left\|\dot{\Delta}_{t,v}\right\|_2}=O_p(\eta_N)\quad\forall m\in\{1,2,3,5,6\},$$
$$|B_{4,t}|\geq\min\left(\frac{\left(3c_{11}^2\underline{\mathfrak{f}}-c_{11}^3\bar{\mathfrak{f}}'\right)c_{\psi}\left\|\dot{\Delta}_{t,v}\right\|_2^2}{12},\frac{\left(3c_{11}^2\underline{\mathfrak{f}}-c_{11}^3\bar{\mathfrak{f}}'\right)\sqrt{c_{\psi}}q_i^{III}\left\|\dot{\Delta}_{t,v}\right\|_2}{6\sqrt{2}}\right),\quad\forall t\in[T].$$

Proof We first deal with $B_{4,t}$. Let $w_{4,it}=\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}$. Following the same arguments as used to derive the lower bound for $A_{4,i}$ in the proof of Lemma A.15 by replacing $w_{2,it}$ with $w_{4,it}$, we can show that, for $t\in[T]$ and any constants $c_{11}<\min(\frac{3\underline{\mathfrak{f}}}{\bar{\mathfrak{f}}'},1)$,

$$\left|B_{4,t}\right|\geq\min\left(\frac{\left(3c_{11}^2\underline{\mathfrak{f}}-c_{11}^3\bar{\mathfrak{f}}'\right)c_{\psi}\left\|\dot{\Delta}_{t,v}\right\|_2^2}{12},\frac{\left(3c_{11}^2\underline{\mathfrak{f}}-c_{11}^3\bar{\mathfrak{f}}'\right)\sqrt{c_{\psi}}q_i^{III}\left\|\dot{\Delta}_{t,v}\right\|_2}{6\sqrt{2}}\right).$$

For $B_{1,t}$, we have

$$B_{1,t}=\left(\frac{1}{N_2}\sum_{i\in I_2}\Psi_{it}^{0}\left(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\}\right)\right)^{\prime}\dot{\Delta}_{t,v}:=\left(\frac{1}{N_2}\sum_{i\in I_2}B_{1,it}^{I}\right)^{\prime}\dot{\Delta}_{t,v}.$$

Conditional on the fixed effects $\left\{V_j^0\right\}_{j\in[p]\cup\{0\}}$ and $\left\{W_j^0\right\}_{j\in[p]}$, the randomness in $B_{1,t}$ is from $\{\varepsilon_{it}\}_{i\in I_2,t\in[T]}$ and $\left\{e_{j,it}\right\}_{j\in[p],i\in I_2,t\in[T]}$, which are independent across i.

Owing to this, by conditional Hoeffding's inequality, we can show that

$$\mathbb{P}\left\{\max_{t\in[T]}\frac{1}{N_2}\left\|\sum_{i\in I_2}B_{1,it}^{I,k}\right\|_2 > c_{15}\eta_N\middle|\mathscr{D}\right\} \leq \sum_{t\in[T]}\mathbb{P}\left\{\left\|\sum_{i\in I_2}B_{1,it}^{I,k}\right\|_2 > c_{15}N_2\eta_N\middle|\mathscr{D}\right\}$$
$$\leq 2\sum_{t\in[T]}\exp\left(-\frac{c_{15}^2N_2^2\eta_N^2}{4M^2\xi_N^2N_2}\right) = o(1), \tag{A.99}$$

with $B_{1,it}^{I,k}$ being the k^{th} element in $B_{1,it}^{I}$, c_{15} is a positive constant, where the second inequality is by Hoeffding's inequality with the fact that $\max_{i\in I_2,t\in[T]}\left|B_{1,it}^{I,k}\right| \leq M\xi_N$ a.s. by Assumption 2.1(v) and Lemma A.13(i). It follows that $\max_{t\in[T]}\frac{|B_{1,t}|}{\|\dot{\Delta}_{t,v}\|_2} = O_p(\eta_N)$. If we use the conditional Bernstein's inequality for the independent sequence rather than the Hoeffding's inequality above, we can show that $\max_{t\in[T]}\frac{1}{N_2}\left\|\sum_{i\in I_2}B_{1,it}^{I,k}\right\|_2 = O_p\left(\sqrt{\frac{\log N\vee T}{N}}\right)$, but here we only need to show the uniform convergence rate to be η_N.

Let $X_{0,it} = 1$. As for $B_{2,t}$, note that

$$\max_{t\in[T]}\frac{|B_{2,t}|}{\|\dot{\Delta}_{t,v}\|_2}$$
$$= \max_{t\in[T]}\frac{\left|\left[\frac{1}{N_2}\sum_{i\in I_2}\left(\dot{\Psi}_{it}^{(1)}-\Psi_{it}^{0}\right)(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right]'\dot{\Delta}_{t,v}\right|}{\|\dot{\Delta}_{t,v}\|_2}$$
$$\leq \max_{t\in[T]}\left\|\frac{1}{N_2}\sum_{i\in I_2}\left(\dot{\psi}_{it}^{(1)}-\Psi_{it}^{0}\right)(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right\|_2$$
$$\leq \max_{t\in[T]}\frac{1}{N_2}\sum_{i\in I_2}\left\|\dot{\Psi}_{it}^{(1)}-\Psi_{it}^{0}\right\|_2 = \max_{t\in[T]}\frac{1}{N_2}\sum_{i\in I_2}\sqrt{\sum_{j=0}^{p}\left\|O_0^{(1)\prime}\dot{u}_{i,j}^{(1)}-u_{i,j}^{0}\right\|_2^2X_{j,it}^2}$$
$$\leq \max_{i\in I_2}\max_{j\in[p]\cup\{0\}}\left(||O_0^{(1)\prime}\dot{u}_{i,j}^{(1)}-u_{i,j}^{0}||_2\right)\left[\frac{1}{N_2}\sum_{i\in I_2}\sum_{j=0}^{p}X_{j,it}^2\right]^{1/2} = O_p(\eta_N), \tag{A.100}$$

where the first inequality is by Cauchy's inequality, the second inequality is by Jensen's inequality, and the last equality is by Theorem 2.2(i) and Assumption 2.1(iv).

Next, we deal with $B_{3,t}$. Following similar arguments as used in (A.100), we obtain that

$$\max_{t\in[T]}\frac{\left|\frac{1}{N_2}\sum_{i\in I_2}\left(\dot{\Psi}_{it}^{(1)}-\Psi_{it}^{0}\right)'\dot{\Delta}_{t,v}\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\{w_{3,it}\leq 0\}\right]\right|}{\|\dot{\Delta}_{t,v}\|_2} = O_p(\eta_N),$$

which implies

$$\max_{t\in[T]}\frac{\left|B_{3,t}\right|}{2\left\|\dot{\Delta}_{t,v}\right\|_2}$$

$$=\max_{t\in[T]}\frac{\left|\left\{\frac{1}{N_2}\sum_{i\in I_2}\Psi_{it}^0\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{w_{3,it}\leq 0\right\}\right]\right\}'\dot{\Delta}_{t,v}\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2}+O_p\left(\eta_N\right)$$

$$\leq\max_{t\in[T]}\left\|\frac{1}{N_2}\sum_{i\in I_2}\Psi_{it}^0\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{w_{3,it}\leq 0\right\}\right]\right\|_2+O_p\left(\eta_N\right)$$

$$\leq\max_{t\in[T]}\frac{1}{N_2}\sum_{i\in I_2}\left\|\Psi_{it}^0\right\|_2\mathbf{1}\left\{\left|\varepsilon_{it}\right|\leq\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\left\|\dot{u}_{i,j}^{(1)}-O_j^{(1)}u_{i,j}^0\right\|_2\right\}+O_p\left(\eta_N\right)$$

$$\equiv\Xi_{NT}+O_p\left(\eta_N\right).$$

Now define the event set $\mathscr{B}_{N,1}(M)=\{\max_{i\in I_2,j\in[p]\cup\{0\}}||\dot{u}_{i,j}^{(1)}-O_j^{(1)}u_{i,j}^0||\leq M\eta_N\}$. Then, Theorem 2.2(i) implies, for any $e>0$, there is a sufficiently large M such that $\mathbb{P}(\mathscr{B}_{N,1}^c(M))\leq e$. Recall that $\mathscr{D}_{e_{it}}$ is the σ-field generated by $e_{it}\cup\left\{V_j^0\right\}_{j\in[p]\cup\{0\}}\cup\left\{W_j^0\right\}_{j\in[p]}$. Then, we have

$$\begin{aligned}\mathbb{P}\left(\Xi_{NT}\geq C\eta_N\right)&\leq\mathbb{P}\left(\Xi_{NT}\geq C\eta_N,\mathscr{B}_{N,1}\right)+e\\&\leq\mathbb{P}\left(\frac{1}{N_2}\sum_{i\in I_2}\left\|\Psi_{it}^0\right\|_2\mathbf{1}\left\{\left|\varepsilon_{it}\right|\leq M\eta_N\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\right\}\geq C\eta_N\right)+e\\&\leq\mathbb{P}\left(\max_{t\in[T]}B_{3,t}^I\geq C\eta_N+\max_{t\in T}B_{3,t}^{II}\right)+e,\end{aligned}\tag{A.101}$$

where

$$B_{3,t}^I=\frac{1}{N_2}\sum_{i\in I_2}\left[\left\|\Psi_{it}^0\right\|_2\mathbf{1}\left\{\left|\varepsilon_{it}\right|\leq M\eta_N\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\right\}-B_{3,t}^{II}\right]\quad\text{and}$$

$$\begin{aligned}B_{3,t}^{II}&=\frac{1}{N_2}\sum_{i\in I_2}\mathbb{E}\left(\left\|\Psi_{it}^0\right\|_2\mathbf{1}\left\{\left|\varepsilon_{it}\right|\leq M\eta_N\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\right\}\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right)\\&=\frac{1}{N_2}\sum_{i\in I_2}\mathbb{E}\left(\left\|\Psi_{it}^0\right\|_2\mathbf{1}\left\{\left|\varepsilon_{it}\right|\leq M\eta_N\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\right\}\middle|\mathscr{D}_{e_{it}}\right),\end{aligned}$$

and the second equality for $B_{3,t}^{II}$ holds by Assumption 2.1(i). Following this, we show that

$$\max_{t\in[T]}B_{3,t}^{II}$$

$$=\max_{t\in[T]}\frac{1}{N_2}\sum_{i\in I_2}\left\|\Psi_{it}^0\right\|_2\left[\mathfrak{F}_{it}\left(M\eta_N\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\right)-\mathfrak{F}_{it}\left(-M\eta_N\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\right)\right]$$

$$
\begin{aligned}
&\leq \max_{t\in[T]}\frac{1}{N_2}\sum_{i\in I_2}\sqrt{1+p\max_{j\in[p]}\left|X_{j,it}\right|^2}\frac{2(1+p)M^2}{c_\sigma}\eta_N \\
&\leq \frac{2(1+p)^2M^2C}{c_\sigma}\eta_N \quad \text{a.s.,} \qquad\qquad (A.102)
\end{aligned}
$$

where $\mathfrak{F}_{it}(\cdot)$ is the conditional CDF of ε_{it} given $\mathscr{D}_{e_i}$ who also has bounded PDF by Assumption 2.1(vii), the first inequality is by mean-value theorem and facts that $\left\|\Psi_{it}^0\right\|_2^2 \leq M^2\left(1+p\max_{j\in[p]}\left|X_{j,it}\right|^2\right)$ together with Lemma A.13(i), and the second inequality is due to Assumption 2.1(iv). In addition, given $\{\mathscr{D}_{e_{it}}\}_{i\in I_2}$, $\{\varepsilon_{it}\}_{i\in I_2}$ are still independent across i. Therefore, by Hoeffding's inequality and similar arguments for term $B_{1,t}$ in (A.99), we can show that

$$
\begin{aligned}
\mathbb{P}\left(\max_{t\in[T]}\left\|B_{3,t}^I\right\|_2 > c_{12}\eta_N\right) &\leq \sum_{t\in[T]}\mathbb{P}\left(\left\|B_{3,t}^I\right\|_2 > c_{12}\eta_N\right) \\
&= \sum_{t\in[T]}\mathbb{E}\mathbb{P}\left(\left\|B_{3,t}^I\right\|_2 > c_{12}\eta_N\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right) = o(1). \qquad (A.103)
\end{aligned}
$$

Combining (A.101)-(A.103), we obtain that $\max_{t\in[T]}\frac{|B_{3,t}|}{\left\|\dot{\Delta}_{t,v}\right\|_2} = O_p(\eta_N)$.

For $B_{5,t}$, we observe that

$$
\begin{aligned}
&B_{5,t} \\
&= \frac{1}{N_2}\sum_{i\in I_2}\left\{\int_0^{\Psi_{it}^{0\prime}\dot{\Delta}_{t,v}}\left[\left(\mathbf{1}\{\varepsilon_{it}\leq s\}-\mathbf{1}\{\varepsilon_{it}\leq 0\}\right)-\mathbb{E}\left(\left(\mathbf{1}\{\varepsilon_{it}\leq s\}-\mathbf{1}\{\varepsilon_{it}\leq 0\}\right)\middle|\mathscr{D}_{e_{it}}\right)\right]ds\right\} \\
&+ \frac{1}{N_2}\sum_{i\in I_2}\left\{\int_{\Psi_{it}^{0\prime}\dot{\Delta}_{t,v}}^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\left[\left(\mathbf{1}\{\varepsilon_{it}\leq s\}-\mathbf{1}\{\varepsilon_{it}\leq 0\}\right)-\mathbb{E}\left(\left(\mathbf{1}\{\varepsilon_{it}\leq s\}-\mathbf{1}\{\varepsilon_{it}\leq 0\}\right)\middle|\mathscr{D}_{e_{it}}\right)\right]ds\right\} \\
&:= B_{5,t}^I + B_{5,t}^{II}.
\end{aligned}
$$

Following similar arguments for $A_{3,i}$ with Bernstein's inequality replacing by Hoeffding's inequality, we can show that $\max_{t\in[T]}\frac{\left|B_{5,t}^I\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2} = O_p(\eta_N)$. Besides, we obtain that

$$
\max_{t\in[T]}\frac{\left|B_{5,t}^{II}\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2} \leq \max_{t\in[T]}\frac{4}{N_2}\sum_{i\in I_2}\left\|\dot{\Psi}_{it}^{(1)}-\Psi_{it}^0\right\|_2 = O_p(\eta_N),
$$

which implies $\max_{t\in[T]}\frac{|B_{5,t}|}{\left\|\dot{\Delta}_{t,v}\right\|_2} = O_p(\eta_N)$.

For $B_{6,t}$, we first note that

$$
\max_{t\in[T]}\frac{\left|\frac{1}{N_2}\sum_{i\in I_2}\int_{\Psi_{it}^{0\prime}\dot{\Delta}_{t,v}}^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\left[\mathbf{1}\{\varepsilon_{it}\leq s\}-\mathbf{1}\{w_{3,it}\leq s\}\right]\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2}
$$

$$\leq \max_{t\in[T]} \frac{2}{N_2}\sum_{i\in I_2}\left\|\dot{\Psi}_{it}^{(1)}-\Psi_{it}^{0}\right\|_2 = O_p(\eta_N),$$

and this implies

$$\begin{aligned}
\max_{t\in[T]}\frac{\left|B_{6,t}\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2} &= \max_{t\in[T]}\frac{\left|\frac{1}{N_2}\sum_{i\in I_2}\int_0^{\dot{\Psi}_{it}^{(1)\prime}\dot{\Delta}_{t,v}}\left(\mathbf{1}\left\{w_{3,it}\leq s\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq s\right\}\right)ds\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2}\\
&= \max_{t\in[T]}\frac{\left|\frac{1}{N_2}\sum_{i\in I_2}\int_0^{\Psi_{it}^{0\prime}\dot{\Delta}_{t,v}}\left(\mathbf{1}\left\{w_{3,it}\leq s\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq s\right\}\right)ds\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2}+O_p\left(\eta_N\right).
\end{aligned}$$

In addition, for the first term on the RHS of the last equality, we have

$$\begin{aligned}
&\max_{t\in[T]}\frac{\left|\frac{1}{N_2}\sum_{i\in I_2}\int_0^{\Psi_{it}^{0\prime}\dot{\Delta}_{t,v}}\left(\mathbf{1}\left\{w_{3,it}\leq s\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq s\right\}\right)ds\right|}{\left\|\dot{\Delta}_{t,v}\right\|_2}\\
&\leq \max_{t\in[T]}\frac{\frac{1}{N_2}\sum_{i\in I_2}\int_0^{\left\|\Psi_{it}^{0}\right\|_2\left\|\dot{\Delta}_{t,v}\right\|_2}\left(\mathbf{1}\left\{\left|\varepsilon_{it}-s\right|\leq\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^{0}\right\|_2\left\|\dot{u}_{i,j}^{(1)}-O_j^{(1)}u_{i,j}^{0}\right\|_2\right\}\right)ds}{\left\|\dot{\Delta}_{t,v}\right\|_2}\\
&= \max_{t\in[T]}\frac{1}{N_2}\sum_{i\in I_2}\int_0^1\left\|\Psi_{it}^{0}\right\|_2\mathbf{1}\left\{\left|\varepsilon_{it}-\left\|\Psi_{it}^{0}\right\|_2\left\|\dot{\Delta}_{t,v}\right\|_2 s\right|\leq\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^{0}\right\|_2\left\|\dot{u}_{i,j}^{(1)}-O_j^{(1)}u_{i,j}^{0}\right\|_2\right\}ds\\
&\leq \sup_{s\geq0,t\in[T]}\frac{1}{N_2}\sum_{i\in I_2}\left\|\Psi_{it}^{0}\right\|_2\mathbf{1}\left\{\left|\varepsilon_{it}-s\right|\leq\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^{0}\right\|_2\left\|\dot{u}_{i,j}^{(1)}-O_j^{(1)}u_{i,j}^{0}\right\|_2\right\}.
\end{aligned}$$

Following the same argument for $B_{3,t}$, we only need to upper bound the RHS of the last display by $\sup_{s\geq0,t\in[T]}B_{6,t}^{I}(s)$ on $\mathscr{B}_{N,1}(M)$ for some sufficiently large but fixed constant M, where

$$B_{6,t}^{I}(s)=\frac{1}{N_2}\sum_{i\in I_2}\left\|\Psi_{it}^{0}\right\|_2\mathbf{1}\left\{\left|\varepsilon_{it}-s\right|\leq M\eta_N\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^{0}\right\|_2\right\}.$$

Let

$$\begin{aligned}
B_{6,t}^{II}(s) &= \mathbb{E}(B_{6,t}^{I}(s)|\{\mathscr{D}_{e_{it}}\}_{i\in I_2})\\
&= \frac{1}{N_2}\sum_{i\in I_2}\mathbb{E}\left(\left\|\Psi_{it}^{0}\right\|_2\mathbf{1}\left\{\left|\varepsilon_{it}-s\right|\leq M\eta_N\sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^{0}\right\|_2\right\}\middle|\mathscr{D}_{e_{it}}\right).
\end{aligned}$$

We note that $\sup_{s\geq0,t\in[T]}B_{6,t}^{II}(s)=O_p(\eta_N)$. Similar to the arguments in (A.102), to show $\max_{t\in[T]}\frac{|B_{6,t}|}{\|\dot{\Delta}_{t,v}\|_2}=O_p(\eta_N)$, it suffices to show

$$\mathbb{P}\left(\sup_{s\geq0,t\in[T]}\left|B_{6,t}^{I}(s)-B_{6,t}^{II}(s)\right|>c_{12}\eta_N\right)=o(1). \tag{A.104}$$

Further denote

$$B_{6,t}^{III}(s) = \frac{1}{N_2}\sum_{i\in I_2}\left\|\Psi_{it}^0\right\|_2 \mathbf{1}\left\{\varepsilon_{it} > s - M\eta_N \sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\right\},$$

$$B_{6,t}^{IV}(s) = \frac{1}{N_2}\sum_{i\in I_2}\left\|\Psi_{it}^0\right\|_2 \mathbf{1}\left\{\varepsilon_{it} \geq s + M\eta_N \sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\right\}.$$

Then, we have $B_{6,t}^{I}(s) = B_{6,t}^{III}(s) - B_{6,t}^{IV}(s)$ and thus,

$$\begin{aligned}
&\sup_{s>T^{1/4},t\in[T]}\left|B_{6,t}^{I}(s) - B_{6,t}^{II}(s)\right| \\
&\leq \sup_{s>T^{1/4},t\in[T]}\left|B_{6,t}^{III}(s) - \mathbb{E}\left(B_{6,t}^{III}(s)\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right)\right| \\
&+ \sup_{s>T^{1/4},t\in[T]}\left|B_{6,t}^{IV}(s) - \mathbb{E}\left(B_{6,t}^{IV}(s)\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right)\right| \\
&\leq \max_{t\in[T]} B_{6,t}^{III}(T^{1/4}) + \max_{t\in[T]}\mathbb{E}\left(B_{6,t}^{III}(T^{1/4})\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right) \\
&+ \max_{t\in[T]} B_{6,t}^{IV}(T^{1/4}) + \max_{t\in[T]}\mathbb{E}\left(B_{6,t}^{IV}(T^{1/4})\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right) \\
&\leq \max_{t\in[T]}\left|B_{6,t}^{III}(T^{1/4}) - \mathbb{E}\left(B_{6,t}^{III}(T^{1/4})\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right)\right| + 2\max_{t\in[T]}\mathbb{E}\left(B_{6,t}^{III}(T^{1/4})\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right) \\
&+ \max_{t\in[T]}\left|B_{6,t}^{IV}(T^{1/4}) - \mathbb{E}\left(B_{6,t}^{IV}(T^{1/4})\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right)\right| + 2\max_{t\in[T]}\mathbb{E}\left(B_{6,t}^{IV}(T^{1/4})\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right),
\end{aligned}$$

where the second inequality holds because both $B_{6,t}^{III}(s)$ and $B_{6,t}^{IV}(s)$ are non-decreasing in s. Further note that

$$\left\|\Psi_{it}^0\right\|_2 \mathbf{1}\left\{\varepsilon_{it} > s - M\eta_N \sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\right\} \quad \text{and} \quad \left\|\Psi_{it}^0\right\|_2 \mathbf{1}\left\{\varepsilon_{it} > s + M\eta_N \sum_{j\in[p]\cup\{0\}}\left\|v_{t,j}^0\right\|_2\right\}$$

are independent across $i \in I_2$ given $\{\mathscr{D}_{e_{it}}\}_{i\in I_2}$. Therefore, following the same argument in the analysis of $A_{6,t}$ with the Bernstein's inequality replaced by the Hoeffding's inequality, we have

$$\begin{aligned}
&\max_{t\in[T]}\left|B_{6,t}^{III}(T^{1/4}) - \mathbb{E}\left(B_{6,t}^{III}(T^{1/4})\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right)\right| = O_p(\eta_N), \\
&\max_{t\in[T]}\mathbb{E}\left(B_{6,t}^{III}(T^{1/4})\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right) = O_p(\eta_N) \\
&\max_{t\in[T]}\left|B_{6,t}^{IV}(T^{1/4}) - \mathbb{E}\left(B_{6,t}^{IV}(T^{1/4})\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right)\right| = O_p(\eta_N), \\
&\max_{t\in[T]}\mathbb{E}\left(B_{6,t}^{IV}(T^{1/4})\middle|\{\mathscr{D}_{e_{it}}\}_{i\in I_2}\right) = O_p(\eta_N),
\end{aligned}$$

which implies

$$\sup_{s>T^{1/4},t\in[T]}\left|B_{6,t}^{I}(s) - B_{6,t}^{II}(s)\right| = O_p(\eta_N).$$

In addition, following the same analysis in (A.95) and (A.96), we have

$$\sup_{s\in[0,T^{1/4}],t\in[T]} \left|B_{6,t}^{I}(s)-B_{6,t}^{II}(s)\right| = O_p(\eta_N),$$

which leads to the desired result that $\max_{t\in[T]} \frac{|B_{6,t}|}{\|\dot{\Delta}_{t,v}\|_2} = O_p(\eta_N)$. ■

Recall for $i\in I_3$, $\dot{\mathscr{H}}_i\left(\{u_{i,j}\}_{j\in[p]\cup\{0\}}\right) = \frac{1}{T}\sum_{t=1}^{T}\left\{\left[\tau-\mathfrak{F}_{it}\left(g_{it}(\{u_{i,j}\}_{j\in[p]\cup\{0\}})\right)\right]\dot{\boldsymbol{\varpi}}_{it}\right\}$, where

$$g_{it}(\{u_{i,j}\}_{j\in[p]\cup\{0\}}) = u_{i,0}'\dot{v}_{t,0}^{(1)} + \sum_{j\in[p]} u_{i,j}'\dot{v}_{t,j}^{(1)}X_{j,it} - u_{i,0}^{0\prime}v_{t,0}^{0} - \sum_{j\in[p]} u_{i,j}^{0\prime}v_{t,j}^{0}X_{j,it},$$

and $\dot{\boldsymbol{\varpi}}_{it} = \left(\dot{v}_{t,0}^{(1)\prime}, X_{1,it}\dot{v}_{t,1}^{(1)\prime}, \cdots, X_{p,it}\dot{v}_{t,p}^{(1)\prime}\right)'$.

Lemma A.17. *Under Assumptions 2.1–2.5, the second-order derivative of* $\dot{\mathscr{H}}_i\left(\{u_{i,j}\}_{j=0}^{p}\right)$ *is bounded in probability.*

Proof Noted that

$$\frac{\dot{\mathscr{H}}_i\left(\{u_{i,j}\}_{j=0}^{p}\right)}{\partial u_i'} = \frac{1}{T}\sum_{t=1}^{T} -\mathfrak{f}_{it}\left(u_{i,0}'\dot{v}_{t,0}^{(1)} - u_{i,0}^{0\prime}v_{t,0}^{0} + X_{1,it}u_{i,1}'\dot{v}_{t,1}^{(1)} - X_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}\right)\dot{\boldsymbol{\varpi}}_{it}\dot{\boldsymbol{\varpi}}_{it}'.$$

For notation simplicity, we focus on the case with $p=1$ and denote $u_i = \left(u_{i,0}', u_{i,1}'\right)'$. Further denote $\dot{\mathscr{H}}_{i,k}\left(\{u_{i,j}\}_{j=0}^{p}\right)$ as the k^{th} element in $\dot{\mathscr{H}}_i\left(\{u_{i,j}\}_{j=0}^{p}\right)$ and $\dot{v}_{t,0,k}^{(1)}$ as the k^{th} element in $\dot{v}_{t,0}^{(1)}$. For $k\in[K_0]$, we have

$$\frac{\partial\dot{\mathscr{H}}_{i,k}\left(\{u_{i,j}\}_{j=0}^{p}\right)}{\partial u_i} = \frac{1}{T}\sum_{t=1}^{T} -\mathfrak{f}_{it}\left(u_{i,0}'\dot{v}_{t,0}^{(1)} - u_{i,0}^{0\prime}v_{t,0}^{0} + X_{1,it}u_{i,1}'\dot{v}_{t,1}^{(1)} - X_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}\right)\dot{v}_{t,0,k}^{(1)}\dot{\boldsymbol{\varpi}}_{it} \quad \text{and}$$

$$\frac{\partial^2\dot{\mathscr{H}}_{i,k}\left(\{u_{i,j}\}_{j=0}^{p}\right)}{\partial u_i\partial u_i'} = \frac{1}{T}\sum_{t=1}^{T} -\mathfrak{f}_{it}'\left(u_{i,0}'\dot{v}_{t,0}^{(1)} - u_{i,0}^{0\prime}v_{t,0}^{0} + X_{1,it}u_{i,1}'\dot{v}_{t,1}^{(1)} - X_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}\right)\dot{v}_{t,0,k}^{(1)}\dot{\boldsymbol{\varpi}}_{it}\dot{\boldsymbol{\varpi}}_{it}'.$$

Therefore, we have

$$\begin{aligned}
&\left\|\frac{\partial^2\dot{\mathscr{H}}_{i,k}\left(\{u_{i,j}\}_{j=0}^{p}\right)}{\partial u_i\partial u_i'}\right\|_F \\
&\leq \frac{\bar{\mathfrak{f}}'}{T}\sum_{t=1}^{T}\left\|\dot{v}_{t,0,k}^{(1)}\dot{\boldsymbol{\varpi}}_{it}\dot{\boldsymbol{\varpi}}_{it}'\right\|_F \\
&\leq c\bar{\mathfrak{f}}'\left[\max_{t\in[T]}\left\|\dot{v}_{t,0}^{(1)}\right\|_2\right]\frac{1}{T}\sum_{t=1}^{T}\|\dot{\boldsymbol{\varpi}}_{it}\|_2^2 \\
&\leq c\bar{\mathfrak{f}}'\left[\max_{t\in[T]}\left\|\dot{v}_{t,0}^{(1)}\right\|_2\right]\left[\max_{t\in[T]}(\left\|\dot{v}_{t,0}^{(1)}\right\|_2^2 + \left\|\dot{v}_{t,1}^{(1)}\right\|_2^2)\right]\left(1+\frac{1}{T}\sum_{t\in[T]}X_{1,it}^2\right) = O_p(1),
\end{aligned}$$

where we use the fact that $\max_{t\in[T]}\left\|\dot{v}_{t,0}^{(1)}\right\|_2 = O_p(1)$ by Theorem 2.2(ii) and Lemma A.13(i).

For $k \in [K_0+1,\cdots,K_0+K_1]$, we have

$$\begin{aligned}
&\frac{\partial \dot{\mathscr{H}}_{i,k}\left(\{u_{i,j}\}_{j=0}^{p}\right)}{\partial u_i} \\
&= \frac{1}{T}\sum_{t=1}^{T} -\mathfrak{f}_{it}\left(u_{i,0}'\dot{v}_{t,0}^{(1)} - u_{i,0}^{0\prime}v_{t,0}^{0} + X_{1,it}u_{i,1}'\dot{v}_{t,1}^{(1)} - X_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}\right)\dot{v}_{t,1,k-K_0}^{(1)}X_{1,it}\dot{\boldsymbol{\varpi}}_{it} \quad \text{and} \\
&\frac{\partial^2 \dot{\mathscr{H}}_{i,k}\left(\{u_{i,j}\}_{j=0}^{p}\right)}{\partial u_i \partial u_i'} \\
&= \frac{1}{T}\sum_{t=1}^{T} -\mathfrak{f}_{it}'\left(u_{i,0}'\dot{v}_{t,0}^{(1)} - u_{i,0}^{0\prime}v_{t,0}^{0} + X_{1,it}u_{i,1}'\dot{v}_{t,1}^{(1)} - X_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}\right)\dot{v}_{t,1,k-K_0}^{(1)}X_{1,it}\dot{\boldsymbol{\varpi}}_{it}\dot{\boldsymbol{\varpi}}_{it}'.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\left\|\frac{\partial^2 \dot{\mathscr{H}}_{i,k}\left(\{u_{i,j}\}_{j=0}^{p}\right)}{\partial u_i \partial u_i'}\right\|_F \\
&\leq c\bar{\mathfrak{f}}'\left[\max_{t\in[T]}\left\|\dot{v}_{t,0}^{(1)}\right\|_2\right]\frac{1}{T}\sum_{t=1}^{T}|X_{1,it}|\,\|\dot{\boldsymbol{\varpi}}_{it}\|_2^2 \\
&\leq c\bar{\mathfrak{f}}'\left[\max_{t\in[T]}\left\|\dot{v}_{t,0}^{(1)}\right\|_2\right]\left[\max_{t\in[T]}\left(\left\|\dot{v}_{t,0}^{(1)}\right\|_2^2 + \left\|\dot{v}_{t,1}^{(1)}\right\|_2^2\right)\right]\left(1+\frac{1}{T}\sum_{t\in T}X_{1,it}^3\right) = O_p(1).
\end{aligned}$$

■

Recall that

$$\begin{aligned}
&\dot{\boldsymbol{\varpi}}_{it} = \left(\dot{v}_{t,0}^{(1)\prime}, \dot{v}_{t,1}^{(1)\prime}X_{1,it},\cdots,\dot{v}_{t,p}^{(1)\prime}X_{p,it}\right)', \\
&\boldsymbol{\varpi}_{it}^{0} = \left(\left(O_0^{(1)}v_{t,0}^{0}\right)', \left(O_1^{(1)}v_{t,1}^{0}\right)'X_{1,it},\cdots,\left(O_p^{(1)}v_{t,p}^{0}\right)'X_{p,it}\right)', \\
&\dot{D}_i^{I} := \frac{1}{T}\sum_{t=1}^{T}\mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^{0}\right)\dot{\boldsymbol{\varpi}}_{it}\dot{\boldsymbol{\varpi}}_{it}', \quad \dot{D}_i^{II} := \frac{1}{T}\sum_{t=1}^{T}\left[\tau - \mathbf{1}\left\{\varepsilon_{it}\leq \dot{\Delta}_{t,v}'\Psi_{it}^{0}\right\}\right]\dot{\boldsymbol{\varpi}}_{it}, \\
&D_i^{I} = \frac{1}{T}\sum_{t=1}^{T}\mathfrak{f}_{it}(0)\boldsymbol{\varpi}_{it}^{0}\boldsymbol{\varpi}_{it}^{0\prime}, \quad D_i^{II} = \frac{1}{T}\sum_{t=1}^{T}\left[\tau - \mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right]\boldsymbol{\varpi}_{it}^{0}.
\end{aligned}$$

Lemma A.18. *Under Assumptions 2.1–2.5, we have*

(i) $\max_{i\in I_3}\left\|D_i^{II}\right\|_F = O_p\left(\sqrt{\frac{\log(N\vee T)}{T}}\xi_N\right)$,

(ii) $\max_{i\in\in I_3}\left\|\dot{D}_i^{I} - D_i^{I}\right\|_F = O_p(\eta_N)$,

(iii) $\max_{i\in I_3}\left\|\dot{D}_i^{II} - D_i^{II} - \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\} - \mathbf{1}\left\{\varepsilon_{it}\leq \dot{\Delta}_{t,v}'\Psi_{it}^{0}\right\}\right]\boldsymbol{\varpi}_{it}^{0}\right\|_2 = o_p\left((N\vee T)^{-\frac{1}{2}}\right)$.

Proof Throughout the proof, we assume there is only one regressor $p=1$ for notation simplicity.

(i) We notice that $\mathbb{E}\left(D_i^{II}\middle|\mathscr{D}\right)=0$. By conditional Bernstein's inequality, for a positive constant c_{16}, we have

$$\mathbb{P}\left(\max_{i\in I_3}\left\|\sum_{t\in[T]}\left[\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right]O_1^{(1)}v_{t,1}^0X_{1,it}\right\|_2>c_{16}\sqrt{T\log(N\vee T)}\xi_N\middle|\mathscr{D}\right)$$
$$\leq\sum_{i\in I_3}\exp\left(-\frac{c_9c_{16}^2T\xi_N^2\log(N\vee T)}{\frac{4M^2}{c_\sigma^2}T\xi_N^2+\frac{2Mc_{16}}{c_\sigma}\xi_N^2\sqrt{T\log(N\vee T)}\log T\log\log T}\right)=o(1),$$

where the inequality follows from Lemma A.12(i), Assumption 2.1(ii), Assumption 2.1(v), and the fact that $\max_{i\in I_3,t\in[T]}\left\|\left[\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right]O_1^{(1)}v_{t,1}^0X_{1,it}\right\|_2\leq\frac{2M}{c_\sigma}\xi_N$ a.s. Similar arguments hold for the upper block of D_i^{II}. This concludes the proof of (i).

(ii) Notice that

$$\dot{D}_i^I-D_i^I=\frac{1}{T}\sum_{t\in[T]}\begin{bmatrix}\mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)\dot{v}_{t,0}^{(1)}\dot{v}_{t,0}^{(1)}-\mathfrak{f}_{it}(0)v_{t,0}^0v_{t,0}^{0\prime} & \mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)\dot{v}_{t,0}^{(1)}\dot{v}_{t,1}^{(1)}X_{1,it}-\mathfrak{f}_{it}(0)v_{t,0}^0v_{t,1}^{0\prime}X_{1,it}\\ \mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)\dot{v}_{t,1}^{(1)}\dot{v}_{t,0}^{(1)}X_{1,it}-\mathfrak{f}_{it}(0)v_{t,1}^0v_{t,0}^{0\prime}X_{1,it} & \mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)}X_{1,it}^2-\mathfrak{f}_{it}(0)v_{t,1}^0v_{t,1}^{0\prime}X_{1,it}^2\end{bmatrix}.$$

To show the upper bound of $\dot{D}_i^I-D_i^I$, we take the lower block for instance and all other three blocks follow the same pattern. Noted that

$$\begin{aligned}
&\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t\in[T]}\left[\mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}X_{1,it}^2-\mathfrak{f}_{it}(0)v_{t,1}^0v_{t,1}^{0\prime}X_{1,it}^2\right]\right\|_F\\
&\leq\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t\in[T]}\left[\mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)-\mathfrak{f}_{it}(0)\right]\left[\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}-v_{t,1}^0v_{t,1}^{0\prime}\right]X_{1,it}^2\right\|_F\\
&+\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t\in[T]}\left[\mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)-\mathfrak{f}_{it}(0)\right]v_{t,1}^0v_{t,1}^{0\prime}X_{1,it}^2\right\|_F\\
&+\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t\in[T]}\mathfrak{f}_{it}(0)\left[\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}-v_{t,1}^0v_{t,1}^{0\prime}\right]X_{1,it}^2\right\|_F\\
&=O_p(\eta_N),
\end{aligned}\tag{A.105}$$

where the equality is by the fact that

$$\begin{aligned}
&\max_{i\in I_3,t\in[T]}\left|\mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)-\mathfrak{f}_{it}(0)\right|\lesssim\max_{i\in I_3,t\in[T]}\left\|\Psi_{it}^0\right\|_2^2\left\|\dot{\Delta}_{t,v}\right\|_2,\\
&\max_{i\in I_3}\left\|\dot{v}_{t,1}^{(1)}\right\|_2+\left\|v_{t,1}^0\right\|_2=O_p(1),\\
&\max_{i\in I_3}\frac{1}{T}\sum_{t\in[T]}X_{1,it}^2\leq C,
\end{aligned}$$

$$\max_{t\in[T]}\left\|\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}-v_{t,1}^{0}v_{t,1}^{0\prime}\right\|_F\leq\max_{i\in I_3}\left(\left\|\dot{v}_{t,1}^{(1)}\right\|_2+\left\|v_{t,1}^{0}\right\|_2\right)\left\|\dot{v}_{t,1}^{(1)}-v_{t,1}^{0}\right\|_2,$$

and

$$\begin{aligned}
&\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t\in[T]}\left[\mathfrak{f}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)-\mathfrak{f}_{it}(0)\right]\left[\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}-v_{t,1}^{0}v_{t,1}^{0\prime}\right]X_{1,it}^2\right\|_F\\
&\leq\max_{i\in I_3,t\in[T]}\bar{\mathfrak{f}}'\left\|\dot{\Delta}_{t,v}\right\|_2\left\|\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}-v_{t,1}^{0}v_{t,1}^{0\prime}\right\|_F\max_{i\in I_3}\frac{1}{T}\sum_{t\in[T]}\left\|\Psi_{it}^0\right\|_2\left|X_{1,it}\right|^2=O_p\left(\eta_N^2\right).
\end{aligned}$$

(iii) Note that

$$\begin{aligned}
&\dot{D}_i^{II}-D_i^{II}-\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq\dot{\Delta}_{t,v}'\Psi_{it}^0\right\}\right]\varpi_{it}^0\\
&=\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq\dot{\Delta}_{t,v}'\Psi_{it}^0\right\}\right]\left(\dot{\varpi}_{it}-\varpi_{it}^0\right)\\
&+\frac{1}{T}\sum_{t=1}^{T}\left[\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right]\left(\dot{\varpi}_{it}-\varpi_{it}^0\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq\dot{\Delta}_{t,v}'\Psi_{it}^0\right\}\right]\left(\dot{\varpi}_{it}-\varpi_{it}^0\right)\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right\}\\
&+\frac{1}{T}\sum_{t=1}^{T}\left\{\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq\dot{\Delta}_{t,v}'\Psi_{it}^0\right\}\right]\left(\dot{\varpi}_{it}-\varpi_{it}^0\right)\right.\\
&\left.-\mathbb{E}\left[\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq\dot{\Delta}_{t,v}'\Psi_{it}^0\right\}\right]\left(\dot{\varpi}_{it}-\varpi_{it}^0\right)\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right]\right\}\\
&+\frac{1}{T}\sum_{t=1}^{T}\left[\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right]\left(\dot{\varpi}_{it}-\varpi_{it}^0\right)\\
&:=S_{1,i}+S_{2,i}+S_{3,i},
\end{aligned}$$

where

$$\begin{aligned}
\max_{i\in I_3}\left\|S_{1,i}\right\|_2&\leq\max_{i\in I_3}\frac{1}{T}\sum_{t\in[T]}\left\|\dot{\varpi}_{it}-\varpi_{it}^0\right\|_2\left|\mathfrak{F}_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)-\mathfrak{F}_{it}(0)\right|\\
&\lesssim\max_{i\in I_3}\frac{1}{T}\sum_{t\in[T],j\in[p]}\left|X_{j,it}^2\right|\max_{t\in[T]}\left\|\dot{\Delta}_{t,v}\right\|_2^2=O_p\left(\eta_N^2\right).
\end{aligned}$$

As for $S_{2,i}$ and $S_{3,i}$, we first recall, for any $e>0$, there exists a sufficiently large constant M such that for

$$\mathscr{A}_7(M)=\left\{\max_{i\in I_3}\left\|O_j^{(1)\prime}\dot{u}_{i,j}^{(1)}-u_{i,j}^0\right\|_2\leq M\eta_N,\max_{t\in[T]}\left\|O_j^{(1)\prime}\dot{v}_{t,j}^{(1)}-v_{t,j}^0\right\|_2\leq M\eta_N,\forall j\in[p]\cup\{0\}\right\}$$

we have $\mathbb{P}(\mathscr{A}_7^c(M))\leq e$. In addition, let

$$\mathscr{A}_{7,i}(M)=\left\{\left\|O_j^{(1)\prime}\dot{u}_{i,j}^{(1)}-u_{i,j}^0\right\|_2\leq M\eta_N,\max_{t\in[T]}\left\|O_j^{(1)\prime}\dot{v}_{t,j}^{(1)}-v_{t,j}^0\right\|_2\leq M\eta_N,\forall j\in[p]\cup\{0\}\right\}.$$

Then, we have

$$\mathbb{P}\left(\max_{i\in I_3}\left\|S_{2,i}\right\|_2 > c_{17}\eta_N^2\right) \leq \mathbb{P}\left(\max_{i\in I_3}\left\|S_{2,i}\right\|_2 > c_{17}\eta_N^2, \mathscr{A}_7(M)\right) + e$$
$$\leq \sum_{i\in I_3}\mathbb{P}\left(\left\|S_{2,i}\right\|_2 > c_{17}\eta_N^2, \mathscr{A}_7(M)\right) + e$$
$$\leq \sum_{i\in I_3}\mathbb{P}\left(\left\|S_{2,i}\right\|_2 > c_{17}\eta_N^2, \mathscr{A}_{7,i}(M)\right) + e$$
$$= \sum_{i\in I_3}\mathbb{E}\mathbb{P}\left(\left\|S_{2,i}\right\|_2 > c_{17}\eta_N^2 \middle| \mathscr{D}_{e_i}^{I_1\cup I_2}\right) 1\{\mathscr{A}_{7,i}(m)\} + e$$
$$\leq \sum_{i\in I_2}\exp\left(-\frac{c_9 c_{17}^2 T^2\eta_N^4}{c_{18}T\eta_N^2\xi_N^2 + c_{17}c_{18}^{1/2}T\eta_N^3\xi_N\log T\log\log T}\right) + e = o(1) + e$$

with a positive constant c_{17} and the inequality above is by Lemma A.12(i) with the fact that, under $\mathscr{A}_{6,i}(M)$,

$$\max_{i\in I_3, t\in[T]}\Bigg\|\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\} - \mathbf{1}\left\{\varepsilon_{it}\leq \dot{\Delta}_{t,v}'\Psi_{it}^0\right\}\right]\left(\dot{\varpi}_{it} - \varpi_{it}^0\right)$$
$$-\mathbb{E}\left[\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\} - \mathbf{1}\left\{\varepsilon_{it}\leq \dot{\Delta}_{t,v}'\Psi_{it}^0\right\}\right]\left(\dot{\varpi}_{it} - \varpi_{it}^0\right)\middle|\mathscr{D}_{e_i}^{I_1\cup I_2}\right]\Bigg\|_2^2 \leq c_{18}\eta_N\xi_N.$$

As e is arbitrary, we have $\max_{i\in I_3}\left\|S_{2,i}\right\|_2 = O_p(\eta_N^2)$. Following a similar argument, we have $\max_{i\in I_3}\left\|S_{3,i}\right\|_2 = O_p(\eta_N^2)$. By Assumption 2.1(ix), we note that $O_p(\eta_N^2) = o_p\left((N\vee T)^{-1/2}\right)$. ■

A.2.3 Lemmas for the Proof Theorem 2.3

Lemma A.19. *Define* $H_{x,j}^l = \left(\frac{\hat{L}_j' L_j^0}{N}\right)^{-1}$ *and* $H_{x,j}^w = W_j^{0\prime}\hat{W}_j\left(\hat{W}_j'\hat{W}_j\right)^{-1}$. *Under Assumptions 2.6-2.8, we have*

(i) $H_{x,j}^l = H_{x,j}^w + O_p\left(\frac{1}{N\wedge T}\right)$,

(ii) $\frac{1}{N}\left\|\hat{L}_j - L_j^0 H_{x,j}^w\right\|_F^2 = O_p\left(\frac{1}{N\wedge T}\right)$,

(iii) $\frac{1}{T}\left\|\hat{W}_j - W_j\left(H_{x,j}^{l\prime}\right)^{-1}\right\|_F^2 = O_p\left(\frac{1}{N\wedge T}\right)$.

Proof The proof can be found in Bai and Ng (2020, Lemma 3 and Proposition 1). ■

Lemma A.20. *Under Assumptions 2.6-2.8, we have*

(i) $\frac{1}{N}L_j^{0\prime}\left(\hat{L}_j - L_j^0 H_{x,j}^w\right) = O_p\left(\frac{1}{N\wedge T}\right)$,

(ii) $\frac{1}{T}W_j^{0\prime}\left(\hat{W}_j - W_j^0\left(H_{x,j}^{l\prime}\right)^{-1}\right) = O_p\left(\frac{1}{N\wedge T}\right)$,

(iii) $\max_{t\in[T]}\frac{1}{N}\left(\hat{L}_j - L_j^0 H_{x,j}^w\right)^{\prime} e_{j,t} = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right)$,

(iv) $\max_{i\in[N]}\frac{1}{T}\left(\hat{W}_j - W_j^0\left(H_{x,j}^{l\prime}\right)^{-1}\right)^{\prime} e_{j,i} = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right)$.

Proof Statements (i) and (ii) are the same as Lemma 4(i) and (ii) in Bai and Ng (2020). Statements (iii) and (iv) are the uniform version of Lemma 4(iii) and (iv) in Bai and Ng (2020). Below we focus on part (iv) as the proof of part (iii) follows analogously.

Noting that $\hat{W}_j^{\prime} - \left(H_{x,j}^l\right)^{-1} W_j^{0\prime} = \frac{1}{N}\hat{L}_j^{\prime}X_j - \frac{\hat{L}_j^{\prime}L_j^0}{N}W_j^{0\prime} = \frac{1}{N}\hat{L}_j^{\prime}E_j$, we have

$$\frac{1}{T}\left(\hat{W}_j - W_j^0\left(H_{x,j}^{l\prime}\right)^{-1}\right)^{\prime} e_{j,i} = \frac{1}{NT}e_{j,i}^{\prime}E_j^{\prime}\hat{L}_j = \frac{1}{NT}e_{j,i}^{\prime}E_j^{\prime}L_j^0H_{x,j}^w + \frac{1}{NT}e_{j,i}^{\prime}E_j^{\prime}\left(\hat{L}_j - L_j^0H_{x,j}^w\right).$$

For the first term on the right side,

$$\max_{i\in[N]}\frac{1}{NT}e_{j,i}^{\prime}E_j^{\prime}L_j^0H_{x,j}^w \lesssim \max_{i\in[N]}\frac{1}{NT}\left\|e_{j,i}^{\prime}E_j^{\prime}L_j^0\right\|_2 = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right),$$

by Assumption 2.8(i). For the second term on the right side, we have

$$\begin{aligned}\max_{i\in[N]}\frac{1}{NT}\left\|e_{j,i}^{\prime}E_j^{\prime}\left(\hat{L}_j - L_j^0H_{x,j}^w\right)\right\|_2 &\leq \max_{i\in[N]}\frac{1}{\sqrt{NT}}\left\|e_{j,i}^{\prime}E_j^{\prime}\right\|_2\frac{\left\|\hat{L}_j - L_j^0H_{x,j}^w\right\|_F}{\sqrt{N}}\\ &= O_p\left(\frac{\log(N\vee T)}{\sqrt{N\wedge T}}\right)O_p\left(\frac{1}{\sqrt{N\wedge T}}\right),\end{aligned}$$

by Assumption 2.6(iv) and Lemma A.19(ii). Combining the above results completes the proof of part (iv). ■

Lemma A.21. *Under Assumption 2.1 and Assumptions 2.6-2.8, we have* $\forall j\in[p]$,

(i) $\hat{w}_{j,t} - \left(H_{x,j}^l\right)^{-1} w_{j,t}^0 = H_{x,j}^{l\prime}\frac{1}{N}\sum_{i\in[N]} l_{j,i}^0 e_{j,it} + \mathscr{R}_{w,t}$,

(ii) $\hat{l}_{j,i} - H_{x,j}^{w\prime}l_{j,i}^0 = \left(\frac{W_j^{0\prime}W_j^0}{T}\right)^{-1}\left(H_{x,j}^l\right)^{-1}\frac{1}{T}\sum_{t=1}^T w_{j,t}^0 e_{j,it} + \mathscr{R}_{l,i}$,

(iii) $\hat{\mu}_{j,it} - \mu_{j,it} = e_{j,it} - \hat{e}_{j,it} = w_{j,t}^{0\prime}\left(\frac{W_j^{0\prime}W_j^0}{T}\right)^{-1}\frac{1}{T}\sum_{t=1}^T w_{j,t}^0 e_{j,it} + l_{j,i}^{0\prime}\frac{1}{N}\sum_{i\in[N]} l_{j,i}^0 e_{j,it} + \mathscr{R}_{j,it}$,

such that

$$\max_{t\in[T]}\left|\mathscr{R}_{w,t}\right| = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right), \quad \max_{i\in[N]}\left|\mathscr{R}_{l,i}\right| = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right),$$

$$\max_{i\in[N],t\in[T],j\in[p]} \left|\mathscr{R}_{j,it}\right| = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right),$$

$$\max_{t\in[T]} \left\|\hat{w}_{j,t} - \left(H_{x,j}^{l}\right)^{-1} w_{j,t}^{0}\right\|_F = O_p(\eta_N), \quad \max_{i\in[N]} \left\|\hat{l}_{j,i} - H_{x,j}^{w\prime} l_{j,i}^{0}\right\|_F = O_p(\eta_N),$$

$$\max_{j\in[p],i\in[N],t\in[T]} \left|e_{j,it} - \hat{e}_{j,it}\right| = \max_{j\in[p],i\in[N],t\in[T]} \left|\hat{\mu}_{j,it} - \mu_{j,it}\right| = O_p(\eta_N),$$

with $\eta_N = \frac{\sqrt{\log(N\vee T)}}{\sqrt{N\wedge T}}\xi_N^2$.

Proof Recall that $X_{j,it} = \mu_{j,it} + e_{j,it} = l_{j,i}^{0\prime} w_{j,t}^{0} + e_{j,it}$ and $X_j = L_j^0 W_j^{0\prime} + E_j$ in matrix form, where $L_j^0 \in \mathbb{R}^{N\times r_j}$ is the factor loading and $W_j^0 \in \mathbb{R}^{T\times r_j}$ is the factor matrix. Following Bai and Ng (2002), Bai (2003) and Bai and Ng (2020), if we impose the normalization restrictions that

$$\frac{L_j^{0\prime} L_j^0}{N} = I_{r_j} \quad \text{and} \quad \frac{W_j^{0\prime} W_j^0}{T} \quad \text{is a diagonal matrix with descending diagonal elements,}$$

we have the principal components estimators:

$$\hat{L}_j = X_j \hat{W}_j \left(\hat{W}_j^{\prime} \hat{W}_j\right)^{-1} \text{ and } \hat{W}_j^{\prime} = \left(\hat{L}_j^{\prime} \hat{L}_j\right)^{-1} \hat{L}_j^{\prime} X_j = \frac{1}{N} \hat{L}_j^{\prime} X_j. \tag{A.106}$$

Let $H_{x,j}^{l} = \left(\frac{1}{N}\hat{L}_j^{\prime} L_j^0\right)^{-1}$. Premultiplying $\frac{1}{N}\hat{L}_j^{\prime}$ on both sides of $X_j = L_j^0 W_j^{0\prime} + E_j$ yields

$$\frac{1}{N}\hat{L}_j^{\prime} X_j = \frac{1}{N}\hat{L}_j^{\prime} L_j^0 W_j^{0\prime} + \frac{1}{N}\hat{L}_j^{\prime} E_j.$$

It follows that

$$\begin{aligned}
\hat{W}_j^{\prime} &= \left(H_{x,j}^{l}\right)^{-1} W_j^{0\prime} + \frac{1}{N}\hat{L}_j^{0\prime} E_j \\
&= \left(H_{x,j}^{l}\right)^{-1} W_j^{0\prime} + \frac{1}{N}\left(H_{x,j}^{l}\right)^{\prime} L_j^{0\prime} E_j + \frac{1}{N}\left[\hat{L} - L_j^0 H_{x,j}^{l}\right]^{\prime} E_j.
\end{aligned}$$

We then show the expansion for each factor, i.e.,

$$\hat{w}_{j,t} - \left(H_{x,j}^{l}\right)^{-1} w_{j,t}^{0} = H_{x,j}^{l\prime} \frac{1}{N} \sum_{i\in[N]} l_{j,i}^{0} e_{j,it} + \frac{1}{N}\left[\hat{L}_j - L_j^0 H_{x,j}^{l}\right]^{\prime} e_{j,t}. \tag{A.107}$$

For equation (A.107), we have the uniform bound for the second term, i.e.,

$$\max_{t\in[T]} \frac{1}{N}\left[\hat{L}_j - L_j^0 H_{x,j}^{l}\right]^{\prime} e_{j,t} = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right)$$

by Lemma A.20(iii). With c_{19} being a positive constant and $\max_{i\in[N],t\in[T]} \left\|l_{j,i}^{0} e_{j,it}\right\|_2 \leq c_{19}\xi_N$ a.s., we show that

$$\mathbb{P}\left(\max_{t\in[T]} \left\|\frac{1}{N}\sum_{i\in[N]} l_{j,i}^{0} e_{j,it}\right\|_2 > c_{20}\sqrt{\frac{\log(N\vee T)}{N}}\xi_N\right)$$

$$\leq \max_{t\in[T]} 2\exp\left(-\frac{c_{20}^2 N\xi_N^2\log(N\vee T)}{N4c_{19}^2\xi_N^2}\right) = o(1) \tag{A.108}$$

for a positive constant c_{20} by Hoeffding's inequality. It follows that $\hat{w}_{j,t} - \left(H_{x,j}^{l}\right)^{-1} w_{j,t}^{0} = H_{x,j}^{l\prime}\frac{1}{N}\sum_{i\in[N]} l_{j,i}^{0} e_{j,it} + \mathscr{R}_{w,t}$, such that

$$\max_{t\in[T]}\left\|\hat{w}_{j,t} - \left(H_{x,j}^{l}\right)^{-1} w_{j,t}^{0}\right\|_F = O_p\left(\sqrt{\frac{\log(N\vee T)}{T}}\xi_N\right), \tag{A.109}$$

and $\max_{t\in[T]}|\mathscr{R}_{w,t}| = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right)$.

Similarly, if we premultiply $\hat{W}_j\left(\hat{W}_j'\hat{W}_j\right)^{-1}$ to both sides of $X_j = L_j^0 W_j^{0\prime} + E_j$, it yields

$$X_j\hat{W}_j\left(\hat{W}_j'\hat{W}_j\right)^{-1} = L_j^0 W_j^{0\prime}\hat{W}_j\left(\hat{W}_j'\hat{W}_j\right)^{-1} + E_j\hat{W}_j\left(\hat{W}_j'\hat{W}_j\right)^{-1}.$$

It follows that

$$\begin{aligned}\hat{L}_j &= L_j^0 H_{x,j}^{w} + E_j\hat{W}_j\left(\hat{W}_j'\hat{W}_j\right)^{-1}\\ &= L_j^0 H_{x,j}^{w} + E_j W_j^0\left(H_{x,j}^{l\prime}\right)^{-1}\left(\hat{W}_j'\hat{W}_j\right)^{-1} + E_j\left(\hat{W}_j - W_j^0\left(H_{x,j}^{l\prime}\right)^{-1}\right)\left(\hat{W}_j'\hat{W}_j\right)^{-1},\end{aligned}$$

where the fist line is due to (A.106) and the definition that $H_{x,l} = W_j^{0\prime}\hat{W}_j\left(\hat{W}_j'\hat{W}_j\right)^{-1}$. Then we obtain the expansion for the factor loading

$$\begin{aligned}\hat{l}_{j,i} - H_{x,j}^{w\prime} l_{j,i}^{0} &= \left(\frac{\hat{W}_j'\hat{W}_j}{T}\right)^{-1}\left(H_{x,j}^{l}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} w_{j,t}^{0} e_{j,it}\\ &\quad + \left(\frac{\hat{W}_j'\hat{W}_j}{T}\right)^{-1}\frac{1}{T}\left[\left(\hat{W} - W\left(H_{x,j}^{l\prime}\right)^{-1}\right)' e_{j,i}\right].\end{aligned}$$

Note that

$$\begin{aligned}\frac{\hat{W}_j'\hat{W}_j}{T} &= \frac{\left(\hat{W}_j - W_j^0\left(H_{x,j}^{l\prime}\right)^{-1}\right)'\hat{W}_j}{T} + \frac{\left(H_{x,j}^{l}\right)^{-1} W_j^{0\prime}\hat{W}_j}{T}\\ &= \frac{\left(\hat{W}_j - W_j^0\left(H_{x,j}^{l\prime}\right)^{-1}\right)'\left(\hat{W}_j - W_j^0\left(H_{x,j}^{l\prime}\right)^{-1}\right)}{T}\\ &\quad + \frac{\left(\hat{W}_j - W_j^0\left(H_{x,j}^{l\prime}\right)^{-1}\right)' W_j^0\left(H_{x,j}^{l\prime}\right)^{-1}}{T}\\ &\quad + \frac{\left(H_{x,j}^{l}\right)^{-1} W_j^{0\prime}\left(\hat{W}_j - W_j\left(H_{x,j}^{l\prime}\right)^{-1}\right)}{T} + \frac{\left(H_{x,j}^{l}\right)' W_j^{0\prime} W_j^{0}\left(H_{x,j}^{l\prime}\right)^{-1}}{T}\end{aligned} \tag{A.110}$$

$$= \frac{\left(H_{x,j}^{l}\right)' W_j^{0\prime} W_j \left(H_{x,j}^{l\prime}\right)^{-1}}{T} + O_p\left(\frac{1}{N \wedge T}\right), \tag{A.111}$$

where the last equality holds by Lemma A.19(iii). Note that

$$\max_{i\in[N]} \frac{1}{T}\left(\hat{W} - W\left(H_{x,j}^{l\prime}\right)^{-1}\right)' e_i = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right)$$

by Lemma A.20(iv), and we can show that

$$\max_{i\in[N]} \left\| \frac{1}{T}\sum_{t=1}^{T} w_{j,t}^{0} e_{j,it} \right\|_2 = O_p\left(\sqrt{\frac{\log(N\vee T)}{T}}\xi_N\right) \tag{A.112}$$

as in (A.108) by conditional Bernstein's inequality in Lemma A.12(i) given $\left\{W_j^0\right\}_{j\in[p]}$.
It follows that $\hat{l}_{j,i} - H_{x,j}^{w\prime} l_{j,i}^{0} = \left(\frac{W_j^{0\prime}W_j^0}{T}\right)^{-1}\left(H_{x,j}^{l}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} w_{j,t}^{0} e_{j,it} + \mathscr{R}_{l,i}$ such that

$$\max_{i\in[N]} \left\|\hat{l}_{j,i} - H_{x,j}^{w\prime} l_{j,i}^{0}\right\|_F = O_p(\eta_N) \text{ and } \max_{i\in[N]} \left|\mathscr{R}_{l,i}\right| = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right).$$

Then, it's natural to obtain that

$$\begin{aligned}
&\hat{\mu}_{j,it} - \mu_{j,it} \\
&= \hat{l}_{j,i}' \hat{w}_t - l_{j,i}^{0\prime} w_{j,t}^{0} \\
&= \left(\hat{l}_{j,i} - \left(H_{x,j}^{l}\right)' l_{j,i}^{0}\right)'\left(\hat{w}_{j,t} - \left(H_{x,j}^{l}\right)^{-1} w_{j,t}^{0}\right) + \left(\hat{l}_{j,i} - \left(H_{x,j}^{l}\right)' l_{j,i}^{0}\right)'\left(H_{x,j}^{l}\right)^{-1} w_{j,t}^{0} \\
&+ l_{j,i}^{0\prime} H_{x,j}^{l}\left(\hat{w}_{j,t} - \left(H_{x,j}^{l}\right)^{-1} w_{j,t}^{0}\right) \\
&= \left(\hat{l}_{j,i} - \left(H_{x,j}^{l}\right)' l_{j,i}^{0}\right)'\left(H_{x,j}^{l}\right)^{-1} w_{j,t}^{0} + l_{j,i}^{0\prime} H_{x,j}^{l}\left(\hat{w}_{j,t} - \left(H_{x,j}^{l}\right)^{-1} w_{j,t}^{0}\right) + O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right) \\
&= w_{j,t}^{0}\left(\frac{W_j^{0\prime}W_j^0}{T}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} w_{j,t}^{0} e_{j,it} + l_{j,i}^{0\prime}\frac{1}{N}\sum_{i\in[N]} l_{j,i}^{0} e_{j,it} + \mathscr{R}_{j,it}
\end{aligned}$$

where the second equality holds by statements (i) and (ii), and the last equality holds with $\max_{i\in[N],t\in[T]} \left|\mathscr{R}_{j,it}\right| = O_p\left(\frac{\log(N\vee T)}{N\wedge T}\right)$. By Assumption 2.1(iv), $\max_{t\in[T]} \left\|w_{j,t}^{0}\right\|_2 \leq M$ a.s. and $\max_{i\in[N]} \left\|l_{j,i}^{0}\right\|_2 \leq M$ a.s., which leads to $\max_{j\in[p],i\in[N],t\in[T]} \left|\hat{\mu}_{j,it} - \mu_{j,it}\right| = O_p\left(\sqrt{\frac{\log(N\vee T)}{N\wedge T}}\xi_N\right) = O_p(\eta_N)$. ■

Lemma A.22. *Under Assumptions 2.1-2.9, for matrices $\dot{D}_i^F$ and D_i^F defined in the proof of Theorem 2.3, we have $\max_{i\in I_3} \left\|\dot{D}_i^F - D_i^F\right\|_F = O_p(\eta_N)$ with $\eta_N = \frac{\sqrt{\log(N\vee T)}\xi_N^2}{\sqrt{N\wedge T}}$.*

Proof Recall that

$$\dot{D}_i^F = \frac{1}{T}\sum_{t=1}^{T} f_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)\right]\begin{bmatrix} \dot{v}_{t,0}^{(1)}\dot{v}_{t,0}^{(1)\prime} & \hat{e}_{1,it}\dot{v}_{t,0}^{(1)}\dot{v}_{t,1}^{(1)\prime} \\ \hat{e}_{1,it}\dot{v}_{t,1}^{(1)}\dot{v}_{t,0}^{(1)\prime} & \hat{e}_{1,it}^2\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}\end{bmatrix} \text{ and}$$

$$D_i^F = \frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\begin{bmatrix} O_0^{(1)}v_{t,0}^0 v_{t,0}^{0\prime}O_0^{(1)\prime} & 0 \\ 0 & e_{1,it}^2 O_1^{(1)}v_{t,1}^0 v_{t,1}^{0\prime}O_1^{(1)\prime}\end{bmatrix},$$

with $\max_{i\in I_2}\left\|D_i^F\right\|_F = O(1)$ a.s..

Let $\dot{\iota}_{it}$ denote $\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right)$ for short. We have

$$\begin{aligned}
&\dot{D}_i^F - D_i^F \\
&= \frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\begin{bmatrix} \dot{v}_{t,0}^{(1)}\dot{v}_{t,0}^{(1)\prime} - O_0^{(1)}v_{t,0}^0 v_{t,0}^{0\prime}O_0^{(1)\prime} & \dot{v}_{t,0}^{(1)}\dot{v}_{t,1}^{(1)\prime}\hat{e}_{1,it} - O_0^{(1)}v_{t,0}^0 v_{t,1}^{0\prime}O_1^{(1)\prime}e_{1,it} \\ \dot{v}_{t,1}^{(1)}\dot{v}_{t,0}^{(1)\prime}\hat{e}_{1,it} - O_1^{(1)}v_{t,1}^0 v_{t,0}^{0\prime}O_0^{(1)\prime}e_{1,it} & \hat{e}_{1,it}^2\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime} - e_{1,it}^2 O_1^{(1)}v_{t,1}^0 v_{t,1}^{0\prime}O_1^{(1)\prime}\end{bmatrix} \\
&+ \frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\begin{bmatrix} 0 & O_0^{(1)}v_{t,0}^0 v_{t,1}^{0\prime}O_1^{(1)\prime}e_{1,it} \\ O_1^{(1)}v_{t,1}^0 v_{t,0}^{0\prime}O_0^{(1)\prime}e_{1,it} & 0\end{bmatrix} \\
&+ \frac{1}{T}\sum_{t=1}^{T}\left(f_{it}\left(\dot{\iota}_{it}\right) - f_{it}(0)\right)\begin{bmatrix} \dot{v}_{t,0}^{(1)}\dot{v}_{t,0}^{(1)\prime} & \dot{v}_{t,0}^{(1)}\dot{v}_{t,1}^{(1)\prime}\hat{e}_{1,it} \\ \dot{v}_{t,1}^{(1)}\dot{v}_{t,0}^{(1)\prime}\hat{e}_{1,it} & \hat{e}_{1,it}^2\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}\end{bmatrix} \\
&:= F_{1,i} + F_{2,i} + F_{3,i}.
\end{aligned}$$

We define $F_{1,i}^m$ for $m = \{1,2,3,4\}$ as fourth clockwise blocks in $F_{1,i} := \begin{bmatrix} F_{1,i}^1 & F_{1,i}^2 \\ F_{1,i}^4 & F_{1,i}^3\end{bmatrix}$. Define $F_{2,i}^m$ and $F_{3,i}^m$ similarly. We aim to show the uniform bound for each block.

First, we observe that

$$\begin{aligned}
F_{1,i}^1 &= \frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\left(\dot{v}_{t,0}^{(1)}\dot{v}_{t,0}^{(1)\prime} - O_0^{(1)}v_{t,0}^0 v_{t,0}^{0\prime}O_0^{(1)\prime}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\Big[\left(\dot{v}_{t,0}^{(1)} - O_0^{(1)}v_{t,0}^0\right)\left(\dot{v}_{t,0}^{(1)} - O_0^{(1)}v_{t,0}^0\right)' + O_0^{(1)}v_{t,0}^0\left(\dot{v}_{t,0}^{(1)} - O_0^{(1)}v_{t,0}^0\right)' \\
&+ \left(\dot{v}_{t,0}^{(1)} - O_0^{(1)}v_{t,0}^0\right)\left(O_0^{(1)}v_{t,0}^0\right)'\Big] \\
&= O_p\left(\eta_N^2\right) + O_p(\eta_N) = O_p(\eta_N)
\end{aligned} \tag{A.113}$$

uniformly over $i \in I_3$. For $F_{1,i}^2$, we have by Theorem 2.2 and Lemma A.21,

$$\begin{aligned}
F_{1,i}^2 &= \frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\left(\dot{v}_{t,0}^{(1)}\dot{v}_{t,1}^{(1)\prime}\hat{e}_{1,it} - O_0^{(1)}v_{t,0}^0 v_{t,1}^{0\prime}O_1^{(1)\prime}e_{1,it}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\left\{\dot{v}_{t,0}^{(1)}\dot{v}_{t,1}^{(1)\prime}\left(\hat{e}_{1,it} - e_{1,it}\right) + \left(\dot{v}_{t,0}^{(1)}\dot{v}_{t,1}^{(1)\prime} - O_0^{(1)}v_{t,0}^0 v_{t,1}^{0\prime}O_1^{(1)\prime}\right)e_{1,it}\right\}
\end{aligned}$$

$$= O_p(\eta_N) \text{ uniformly in } i \in I_3. \tag{A.114}$$

The same order holds for $\max_{i\in I_3}\left\|F^3_{1,i}\right\|_F$ as $F^3_{1,i} = F^{2\prime}_{1,i}$. Next, we study $F^4_{1,i}$. Noting that

$$\begin{aligned}|\hat{e}^2_{1,it} - e^2_{1,it}| &= \left|(\hat{e}_{1,it} + e_{1,it})\,(\hat{e}_{1,it} - e_{1,it})\right| \\ &\leq \left(2|e_{1,it}| + \max_{i\in I_3, t\in[T]}|\hat{e}_{1,it} - e_{1,it}|\right)\left(\max_{i\in I_3, t\in[T]}|\hat{e}_{1,it} - e_{1,it}|\right),\end{aligned}$$

we have

$$\begin{aligned}
&\max_{i\in I_3}||F^4_{1,i}||_F \\
&= \left\|\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\left(\dot{v}^{(1)}_{t,1}\dot{v}^{(1)\prime}_{t,1}\hat{e}^2_{1,it} - O^{(1)}_1 v^0_{t,1}v^{0\prime}_{t,1}O^{(1)\prime}_1 e^2_{1,it}\right)\right\|_F \\
&\leq \max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\left\|\dot{v}^{(1)}_{t,1}\right\|_2^2\left(|\hat{e}^2_{1,it} - e^2_{1,it}|\right) + \max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\left\|\dot{v}^{(1)}_{t,1}\dot{v}^{(1)\prime}_{t,1} - O^{(1)}_1 v^0_{t,1}v^{0\prime}_{t,1}O^{(1)\prime}_1\right\|_F e^2_{1,it} \\
&= \max_{t\in[T]}\left\|\dot{v}^{(1)}_{t,1}\right\|_2^2\left(\max_{i\in I_3, t\in[T]}|\hat{e}_{1,it} - e_{1,it}|\right)\max_{i\in I_3}\left(\frac{2}{T}\sum_{t=1}^{T} f_{it}(0)|e_{1,it}|\right) \\
&+ \max_{t\in[T]}\left\|\dot{v}^{(1)}_{t,1}\right\|_2^2\left(\max_{i\in I_3, t\in[T]}|\hat{e}_{1,it} - e_{1,it}|^2\right)\max_{i\in I_3}\left(\frac{2}{T}\sum_{t=1}^{T} f_{it}(0)\right) \\
&+ \max_{t\in[T]}\left\|\dot{v}^{(1)}_{t,1}\dot{v}^{(1)\prime}_{t,1} - O^{(1)}_1 v^0_{t,1}v^{0\prime}_{t,1}O^{(1)\prime}_1\right\|_F \max_{i\in I_3}\left(\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)e^2_{1,it}\right) = O_p(\eta_N).
\end{aligned} \tag{A.115}$$

Combining (A.113)-(A.115), we conclude $\max_{i\in I_3}\left\|F_{1,i}\right\|_2 = O_p(\eta_N)$.

For $F^2_{2,i}$, we note that $\mathbb{E}\left(f_{it}(0)v^0_{t,0}v^{0\prime}_{t,1}e_{1,it}\middle|\mathscr{D}\right) = 0$ by Assumption 2.9 and

$$\max_{i\in I_3, t\in[T]}\left\|f_{it}(0)v^0_{t,0}v^{0\prime}_{t,1}e_{1,it}\right\|_F \leq c_{22}\xi_N \; a.s.$$

by Assumption 2.1(iv) and Lemma A.13(i). Then, by conditional Bernstein's inequality in Lemma A.12(i) and Assumption 2.9(iii), we can show that, with positive constants c_{21} and c_{22},

$$\begin{aligned}
&\mathbb{P}\left(\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t\in[T]} f_{it}(0)v^0_{t,0}v^{0\prime}_{t,1}e_{1,it}\right\|_F > c_{21}\sqrt{\frac{\log(N\vee T)}{T}}\xi_N\middle|\mathscr{D}\right) \\
&\leq \sum_{i\in I_3}\exp\left(-\frac{c_{12}c^2_{21}T\xi^2_N\log(N\vee T)}{c^2_{22}T\xi^2_N + c_{21}c_{22}\sqrt{T\log(N\vee T)}\xi^2_N\log T\log\log T}\right) = o(1),
\end{aligned} \tag{A.116}$$

which yields $\max_{i\in I_3}\left\|F_{2,i}\right\|_F = O_p(\eta_N)$.

As for $F_{3,i}$, we show the bound for the first block and all other three blocks follow the same argument. Recall that

$$
\begin{aligned}
\mathfrak{i}_{it} &= u_{i,0}^{0\prime} O_0^{(1)} \dot{v}_{t,0}^{(1)} - u_{i,0}^{0\prime} v_{t,0}^0 + \hat{\mu}_{1,it} \dot{u}_{i,1}^{(1)\prime} \dot{v}_{t,1}^{(1)} - \mu_{1,it} u_{i,1}^{0\prime} v_{t,1}^0 + \hat{e}_{1,it} u_{i,1}^{0\prime} O_1^{(1)} \dot{v}_{t,1}^{(1)} - e_{1,it} u_{i,1}^{0\prime} v_{t,1}^0 \\
&= \left[\left(O_0^{(1)} u_{i,0}^0 \right)' \left(\dot{v}_{t,0}^{(1)} - O_0^{(1)} v_{t,0}^0 \right) \right] + \left(\hat{\mu}_{1,it} \dot{u}_{i,1}^{(1)\prime} \dot{v}_{t,1}^{(1)} - \mu_{1,it} u_{i,1}^{0\prime} v_{t,1}^0 \right) \\
&+ \left(\hat{e}_{1,it} u_{i,1}^{0\prime} O_1^{(1)} \dot{v}_{t,1}^{(1)} - e_{1,it} u_{i,1}^{0\prime} v_{t,1}^0 \right) \\
&:= \mathfrak{i}_{it}^I + \mathfrak{i}_{it}^{II} + \mathfrak{i}_{it}^{III},
\end{aligned}
\tag{A.117}
$$

with the fact that $\left|\mathfrak{i}_{it}^I\right| \leq \left\| O_0^{(1)} u_{i,0}^0 \right\|_2 \left\| \dot{v}_{t,0}^{(1)} - O_0^{(1)} v_{t,0}^0 \right\|_2 := R_{1,it}^I$ such that $\max_{i \in I_3, t \in [T]} R_{1,it}^I = O_p(\eta_N)$. In addition, we have

$$
\begin{aligned}
\max_{i \in I_3, t \in [T]} \left|\mathfrak{i}_{it}^{II}\right| &= \max_{i \in I_3, t \in [T]} \left| \left(\hat{\mu}_{1,it} - \mu_{1,it}\right) \dot{u}_{i,1}^{(1)\prime} \dot{v}_{t,1}^{(1)} + \mu_{1,it} \left(\dot{u}_{i,1}^{(1)\prime} \dot{v}_{t,1}^{(1)} - u_{i,1}^{0\prime} v_{t,1}^0 \right) \right| \\
&\leq R_{1,it}^{II} + R_{2,it}^{II} |\mu_{1,it}|,
\end{aligned}
\tag{A.118}
$$

where $\max_{i \in I_3, t \in [T]} |R_{1,it}^{II}| = O_p(\eta_N)$ and $\max_{i \in I_3, t \in [T]} |R_{2,it}^{II}| = O_p(\eta_N)$. Similarly, we have

$$
\begin{aligned}
\left|\mathfrak{i}_{it}^{III}\right| &= \left| \hat{e}_{1,it} u_{i,1}^{0\prime} O_1^{(1)} \dot{v}_{t,1}^{(1)} - e_{1,it} u_{i,1}^{0\prime} v_{t,1}^0 \right| \\
&\leq \left| \left(\hat{e}_{1,it} - e_{1,it}\right) \left(O_1^{(1)} u_{i,1}^0 \right)' \left(\dot{v}_{t,1}^{(1)} - O_1^{(1)} v_{t,1}^0 \right) + \left(\hat{e}_{1,it} - e_{1,it}\right) \left(O_1^{(1)} u_{i,1}^0 \right)' O_1^{(1)} v_{t,1}^0 \right| \\
&+ \left| e_{1,it} \left(O_1^{(1)} u_{i,1}^0 \right)' \left(\dot{v}_{t,1}^{(1)} - O_1^{(1)} v_{t,1}^0 \right) \right| \\
&= R_{1,it}^{III} + R_{2,it}^{III} |e_{1,it}|,
\end{aligned}
\tag{A.119}
$$

where $\max_{i \in I_3, t \in [T]} |R_{1,it}^{III}| = O_p(\eta_N)$ and $\max_{i \in I_3, t \in [T]} |R_{2,it}^{III}| = O_p(\eta_N)$. Therefore, we have

$$
\begin{aligned}
&\max_{i \in I_3} ||F_{3,i}^1||_F \\
&\leq \frac{1}{T} \sum_{t=1}^{T} \left| f_{it}\left(\mathfrak{i}_{it}\right) - f_{it}(0) \right| \left\| \dot{v}_{t,0}^{(1)} \right\|_2^2 \\
&\lesssim \max_{t \in [T]} \left\| \dot{v}_{t,0}^{(1)} \right\|_2 \max_{i \in I_3, t \in [T]} (R_{1,it}^I + R_{1,it}^{II} + R_{2,it}^{II} + R_{1,it}^{III} + R_{2,it}^{III}) \frac{1}{T} \sum_{t \in [T]} \left(1 + |e_{1,it}| + |\mu_{1,it}|\right) \\
&= O_p(\eta_N),
\end{aligned}
\tag{A.120}
$$

Combining all results above, we obtain that $\max_{i \in I_3} ||\hat{D}_i^F - D_i^F||_F = O_p(\eta_N)$. ■

Lemma A.23. *Under Assumptions 2.1-2.9, for matrices $\dot{D}_i^J$ and D_i^J defined in* *the proof of Theorem 2.3, we have*

$$\max_{i\in I_3}\left\|\dot{D}_i^J - D_i^J\right\|_F = \left\|\begin{bmatrix} O_p\left(\eta_N^2\right) & O_p(\eta_N) \\ O_p\left(\eta_N^2\right) & O_p\left(\eta_N^2\right)\end{bmatrix}\right\|_F .$$

Proof Recall that

$$D_i^J = \frac{1}{T}\sum_{t=1}^T f_{it}(0)\begin{bmatrix} \dot{v}_{t,0}^{(1)}\left(O_0^{(1)}v_{t,0}^0 - \dot{v}_{t,0}^{(1)}\right)' & 0 \\ 0 & e_{1,it}^2 O_1^{(1)}v_{t,1}^0\left(O_1^{(1)}v_{t,1}^0 - \dot{v}_{t,1}^{(1)}\right)'\end{bmatrix}.$$

We have

$$\begin{aligned}
&\hat{D}_i^J - D_i^J \\
&= \frac{1}{T}\sum_{t=1}^T f_{it}(0)\begin{bmatrix} 0 & \dot{v}_{t,0}^{(1)}v_{t,1}^{0\prime}O_1^{(1)\prime}\left(e_{1,it}-\hat{e}_{1,it}\right) + \dot{v}_{t,0}^{(1)}\left(O_1^{(1)}v_{t,1}^0 - \dot{v}_{t,1}^{(1)}\right)'\left(\hat{e}_{1,it}-e_{1,it}\right) \\ \left(\hat{e}_{1,it}-e_{1,it}\right)\dot{v}_{t,1}^{(1)}\left(O_0^{(1)}v_{t,0}^0 - \dot{v}_{t,0}^{(1)}\right)' & \dot{v}_{t,1}^{(1)}v_{t,1}^{0\prime}O_1^{(1)\prime}e_{1,it}\left(\hat{e}_{1,it}-e_{1,it}\right) - \dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}\left(\hat{e}_{1,it}^2 - e_{1,it}^2\right)\end{bmatrix} \\
&+ \frac{1}{T}\sum_{t=1}^T f_{it}(0)\begin{bmatrix} 0 & \dot{v}_{t,0}^{(1)}\left(O_1^{(1)}v_{t,1}^0 - \dot{v}_{t,1}^{(1)}\right)'e_{1,it} \\ \dot{v}_{t,1}^{(1)}\left(O_0^{(1)}v_{t,0}^0 - \dot{v}_{t,0}^{(1)}\right)'e_{1,it} & \left(\dot{v}_{t,1}^{(1)} - O_1^{(1)}v_{t,1}^0\right)\left(O_1^{(1)}v_{t,1}^0 - \dot{v}_{t,1}^{(1)}\right)'e_{1,it}^2\end{bmatrix} \\
&+ \frac{1}{T}\sum_{t=1}^T \left[f_{it}(\tilde{\iota}_{it}) - f_{it}(0)\right]\begin{bmatrix} \dot{v}_{t,0}^{(1)}\left(O_0^{(1)}v_{t,0}^0 - \dot{v}_{t,0}^{(1)}\right)' & \dot{v}_{t,0}^{(1)}\left(e_{1,it}O_1^{(1)}v_{t,1}^0 - \hat{e}_{1,it}\dot{v}_{t,1}^{(1)}\right)' \\ \hat{e}_{1,it}\dot{v}_{t,1}^{(1)}\left(O_0^{(1)}v_{t,0}^0 - \dot{v}_{t,0}^{(1)}\right)' & \hat{e}_{1,it}\dot{v}_{t,1}^{(1)}\left(e_{1,it}O_1^{(1)}v_{t,1}^0 - \hat{e}_{1,it}\dot{v}_{t,1}^{(1)}\right)'\end{bmatrix} \\
&:= J_{1,i} + J_{2,i} + J_{3,i}.
\end{aligned}$$

As in the proof of the last lemma, we define $J_{1,i}^m$, $J_{2,i}^m$ and $J_{3,i}^m$ for $m = 1,2,3,4$ as four clockwise blocks in $J_{1,i}$, $J_{2,i}$ and $J_{3,i}$, respectively.

First, we study $J_{1,i}$. For $J_{1,i}^2$, we notice that

$$\begin{aligned}
J_{1,i}^2 &= \frac{1}{T}\sum_{t=1}^T f_{it}(0)\dot{v}_{t,0}^{(1)}v_{t,1}^{0\prime}O_1^{(1)\prime}\left(e_{1,it}-\hat{e}_{1,it}\right) \\
&\quad + \frac{1}{T}\sum_{t=1}^T f_{it}(0)\dot{v}_{t,0}^{(1)}\left(O_1^{(1)}v_{t,1}^0 - \dot{v}_{t,1}^{(1)}\right)'\left(\hat{e}_{1,it}-e_{1,it}\right) \\
&= \frac{1}{T}\sum_{t=1}^T f_{it}(0)O_0^{(1)}v_{t,0}^0 v_{t,1}^{0\prime}O_1^{(1)\prime}\left(e_{1,it}-\hat{e}_{1,it}\right) \\
&\quad + \frac{1}{T}\sum_{t=1}^T f_{it}(0)\left(\dot{v}_{t,0}^{(1)} - O_0^{(1)}v_{t,0}^0\right)v_{t,1}^{0\prime}O_1^{(1)\prime}\left(e_{1,it}-\hat{e}_{1,it}\right) + O_p\left(\eta_N^2\right) \\
&= \frac{1}{T}\sum_{t=1}^T f_{it}(0)O_0^{(1)}v_{t,0}^0 v_{t,1}^{0\prime}O_1^{(1)\prime}\left(e_{1,it}-\hat{e}_{1,it}\right) + O_p\left(\eta_N^2\right) + O_p\left(\eta_N^2\right) \\
&= O_p(\eta_N) \quad \text{uniformly over } i \in I_3. \qquad \text{(A.121)}
\end{aligned}$$

Noted that the leading term in $J_{1,i}^2$ is $\frac{1}{T}\sum_{t=1}^T f_{it}(0)O_0^{(1)}v_{t,0}^0 v_{t,1}^{0\prime}O_1^{(1)\prime}\left(e_{1,it}-\hat{e}_{1,it}\right)$, which will remain as the bias term of $\hat{u}_{i,0}^{(3,1)}$.

Furthermore, it is clear that

$$J_{1,i}^{3}=\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)\left(\hat{e}_{1,it}-e_{1,it}\right)\dot{v}_{t,1}^{(1)}\left(O_{0}^{(1)}v_{t,0}^{0}-\dot{v}_{t,0}^{(1)}\right)'=O_{p}\left(\eta_{N}^{2}\right)\quad\text{uniformly over } i\in I_{3}.$$

Next, we obtain that

$$\begin{aligned}
J_{1,i}^{4}&=\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)\dot{v}_{t,1}^{(1)}v_{t,1}^{0\prime}O_{1}^{(1)\prime}e_{1,it}\left(\hat{e}_{1,it}-e_{1,it}\right)-\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}\left(\hat{e}_{1,it}^{2}-e_{1,it}^{2}\right)\\
&=O_{1}^{(1)}\left[\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}e_{1,it}\left(\hat{e}_{1,it}-e_{1,it}\right)\right]O_{1}^{(1)\prime}-\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}\left(\hat{e}_{1,it}-e_{1,it}\right)^{2}\\
&\quad-\frac{2}{T}\sum_{t=1}^{T}f_{it}(0)\dot{v}_{t,1}^{(1)}\dot{v}_{t,1}^{(1)\prime}e_{1,it}\left(\hat{e}_{1,it}-e_{1,it}\right)+O_{p}(\eta_{N}^{2})\\
&=-O_{1}^{(1)}\left[\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}e_{1,it}\left(\hat{e}_{1,it}-e_{1,it}\right)\right]O_{1}^{(1)\prime}+O_{p}(\eta_{N}^{2}),
\end{aligned}\tag{A.122}$$

where the first and second equalities hold by Theorem 2.2(ii) and Lemma A.21. We deal with the first term in the second equality of (A.122) by inserting the linear expansion of $\hat{e}_{1,it}-e_{1,it}$ in Lemma A.21(iii), i.e.,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}e_{1,it}\left(\hat{e}_{1,it}-e_{1,it}\right)\\
&=-\frac{1}{T}\sum_{t=1}^{T}\left\{f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}e_{1,it}\left[w_{1,t}^{0\prime}\left(\frac{W_{1}^{0\prime}W_{1}^{0}}{T}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}w_{1,t}^{0}e_{1,it}\right]\right\}\\
&\quad-\frac{1}{T}\sum_{t=1}^{T}\left\{f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}\left[e_{1,it}l_{1,i}^{0\prime}\frac{1}{N}\sum_{i^{*}\in[N]}l_{1,i^{*}}^{0}e_{1,i^{*}t}\right]\right\}+O_{p}\left(\frac{\log(N\vee T)}{N\wedge T}\right)
\end{aligned}\tag{A.123}$$

uniformly over $i\in I_{3}$. For the first term in the right side of (A.123), we notice that

$$\begin{aligned}
&\max_{i\in I_{3}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left\{f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}e_{1,it}\left[w_{1,t}^{0\prime}\left(\frac{W_{1}^{0\prime}W_{1}^{0}}{T}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}w_{1,t}^{0}e_{1,it}\right]\right\}\right\|_{F}\\
&=\max_{i\in I_{3}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left\{f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}e_{1,it}\left[\sum_{k\in[r_{1}]}w_{1,t,k}^{0}\left[\left(\frac{W_{1}^{0\prime}W_{1}^{0}}{T}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}w_{1,t}^{0}e_{1,it}\right]_{k}\right]\right\}\right\|_{F}\\
&\leq\sum_{k\in[r_{1}]}\max_{i\in I_{3}}\left\|\left[\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}w_{1,t,k}^{0}e_{1,it}\right]\right\|_{F}\max_{k\in[r_{1}],i\in I_{3}}\left|\left[\left(\frac{W_{1}^{0\prime}W_{1}^{0}}{T}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}w_{1,t}^{0}e_{1,it}\right]_{k}\right|\\
&=\sum_{k\in[r_{1}]}\max_{i\in I_{3}}\left\|\left[\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}w_{1,t,k}^{0}e_{1,it}\right]\right\|_{F}O_{p}(\eta_{N})\\
&=O_{p}(\eta_{N}^{2}),
\end{aligned}$$

where $w_{1,t}$ is the fact of $X_{1,it}$, r_1 is the dimension of $w_{1,t}$, and $w_{1,t,k}$ is the kth element of $w_{1,it}$, the second equality holds by the results in (A.110) and (A.112), and the last equality is by the fact that

$$\mathbb{P}\left(\max_{i\in I_3}\left\|\left[\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}w_{1,t,k}^{0}e_{1,it}\right]\right\|_F > c_{23}\sqrt{\frac{\log(N\vee T)}{T}}\xi_N\middle|\mathscr{D}\right)=o(1)$$

following analogous analysis as (A.116) with a positive constant c_{23}.

For the second term on the RHS of equation (A.123), we notice that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left\{f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}\left[e_{1,it}l_{1,i}^{0\prime}\frac{1}{N}\sum_{i^*\in N}l_{1,i^*}^{0}e_{1,i^*t}\right]\right\}\\
&=\frac{1}{NT}\sum_{i^*\in[N]}\sum_{t\in[T]}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}l_{1,i}^{0\prime}l_{1,i^*}^{0}e_{1,it}e_{1,i^*t}\\
&=\frac{1}{NT}\sum_{t\in[T]}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}l_{1,i}^{0\prime}l_{1,i}^{0}e_{1,it}^{2}+\frac{1}{NT}\sum_{i^*\neq i}\sum_{t\in[T]}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}l_{1,i}^{0\prime}l_{1,i^*}^{0}e_{1,it}e_{1,i^*t}
\end{aligned}\tag{A.124}$$

with

$$\begin{aligned}
\max_{i\in I_3}\left\|\frac{1}{NT}\sum_{t\in[T]}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}l_{1,i}^{0\prime}l_{1,i}^{0}e_{1,it}^{2}\right\| &\leq \frac{1}{N}\max_{i\in I_3}\left\|l_{1,i}^{0}\right\|_2^2\max_{t\in[T]}\left(\left\|v_{t,1}^{0}\right\|_2\left\|v_{t,1}^{0}\right\|_2\right)\frac{1}{T}\sum_{t\in[T]}e_{1,it}^{2}\\
&=O_p\left(\frac{\xi_N^2}{N}\right),
\end{aligned}$$

where the equality holds by Assumption 2.1(iv), Lemma A.13(i) and Theorem 2.2(ii). For the second term in (A.124), by Assumption 2.1(iii), $e_{j,it}$ is strong mixing across t and independent given fixed effect. We define $E_i^{vec}=(\tilde{e}_{1,1}',\cdots,\tilde{e}_{1,i-1}',\tilde{e}_{1,i+1}',\cdots,\tilde{e}_{1,N}')'$ with $\tilde{e}_{1,i^*}=(e_{1,i^*1}e_{1,i1},\cdots,e_{1,i^*T}e_{1,iT})$ for $i^*\neq i$. We can see E_i^{vec} will also be a strong mixing sequence, conditional on $\mathscr{D}$, which implies

$$\begin{aligned}
&\mathbb{P}\left(\max_{i\in I_3}\left\|\frac{1}{NT}\sum_{i^*\neq i}\sum_{t\in[T]}f_{it}(0)\dot{v}_{t,1}^{(1)}v_{t,1}^{0\prime}O_1^{(1)\prime}l_{1,i}^{0\prime}l_{1,i^*}^{0}e_{1,it}e_{1,i^*t}\right\|_F>c_{23}\eta_N^2\right)\\
&\leq\sum_{i\in I_3}\mathbb{E}\mathbb{P}\left(\left\|\frac{1}{NT}\sum_{i^*\neq i}\sum_{t\in[T]}f_{it}(0)\dot{v}_{t,1}^{(1)}v_{t,1}^{0\prime}O_1^{(1)\prime}l_{1,i}^{0\prime}l_{1,i^*}^{0}e_{1,it}e_{1,i^*t}\right\|_F>c_{23}\eta_N^2\middle|\mathscr{D}\right)=o(1),
\end{aligned}$$

where the equality holds by Lemma A.12(i). This implies

$$\begin{aligned}
&\max_{i\in I_3}\left\|O_1^{(1)}\left[\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}e_{1,it}\left(\hat{e}_{1,it}-e_{1,it}\right)\right]O_1^{(1)\prime}\right\|_F\\
&=\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)v_{t,1}^{0}v_{t,1}^{0\prime}e_{1,it}\left(\hat{e}_{1,it}-e_{1,it}\right)\right\|_F\\
&=O_p(\eta_N^2).
\end{aligned}$$

Therefore, we conclude that $\max_{i\in I_3}\left\|J_{1,i}^4\right\|_F = O_p\left(\eta_N^2\right)$. Then

$$\max_{i\in I_3}\left\|J_{1,i}\right\|_F = \left\|\begin{bmatrix} 0 & O_p(\eta_N) \\ O_p\left(\eta_N^2\right) & O_p\left(\eta_N^2\right)\end{bmatrix}\right\|_F.$$

Next, for $J_{2,i}^2$ and $J_{2,i}^3$, conditioning on $\mathscr{D}^{I_1\cup I_2}$ and following Bernstein's inequality in Lemma A.12(i), we have

$$\max_{i\in I_3}\left\|J_{2,i}^2\right\|_F = \max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\dot{v}_{t,0}^{(1)}\left(O_1^{(1)}v_{t,1}^0 - \dot{v}_{t,1}^{(1)}\right)' e_{1,it}\right\|_F = O_p\left(\eta_N^2\right),$$

$$\max_{i\in I_3}\left\|J_{2,i}^3\right\|_F = \max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\dot{v}_{t,1}^{(1)}\left(O_0^{(1)}v_{t,0}^0 - \dot{v}_{t,0}^{(1)}\right)' e_{1,it}\right\|_F = O_p\left(\eta_N^2\right),$$

which can be obtained by the similar arguments as Lemma A.24(i). These results, in conjunction with the fact that

$$\begin{aligned}\max_{i\in I_3}\left\|J_{2,i}^4\right\|_F &= \max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)\left(\dot{v}_{t,1}^{(1)} - O_1^{(1)}v_{t,1}^0\right)\left(O_1^{(1)}v_{t,1}^0 - \dot{v}_{t,1}^{(1)}\right)' e_{1,it}^2\right\|_F \\ &\leq \max_{i\in I_3, t\in[T]}\left(\frac{1}{T}\sum_{t=1}^{T} f_{it}(0)e_{1,it}^2\right)\max_{t\in[T]}\left\|\dot{v}_{t,1}^{(1)} - O_1^{(1)}v_{t,1}^0\right\|_2^2 = O_p\left(\eta_N^2\right),\end{aligned}$$

imply that $\max_{i\in I_3}\left\|J_{2,i}\right\|_F = O_p(\eta_N^2)$.

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

$$\begin{aligned}\max_{i\in I_3}\left\|J_{3,i}^1\right\|_F &= \max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[f_{it}\left(\tilde{\iota}_{it}\right) - f_{it}(0)\right]\dot{v}_{t,0}^{(1)}\left(O_0'v_{t,0}^0 - \dot{v}_{t,0}^{(1)}\right)'\right\|_F \\ &\leq \max_{t\in[T]}\left\|\dot{v}_{t,0}^{(1)}\right\|_2\left\|\dot{v}_{t,0}^{(1)}\left(O_0'v_{t,0}^0 - \dot{v}_{t,0}^{(1)}\right)'\right\|_2\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left|f_{it}\left(\tilde{\iota}_{it}\right) - f_{it}(0)\right| \\ &\lesssim O_p(\eta_N)\max_{i\in I_3}\frac{1}{T}\sum_{t\in[T]}\left|\tilde{\iota}_{it}\right| \\ &= O_p\left(\eta_N^2\right), \qquad \text{(A.125)}\end{aligned}$$

where the third line is by Lipschitz continuity of the density function, Theorem 2.2(ii), Assumption 2.2, and the fact that $|\tilde{\iota}_{it}|$ lies between 0 and $|\iota_{it}|$, and the last line is by the fact that $\max_{i\in I_3}\frac{1}{T}\sum_{t\in[T]}|\tilde{\iota}_{it}| = O_p\left(\eta_N\right)$ by (A.118) and (A.119). The bounds for other three blocks in $J_{3,i}$ can be established in the same manner. Hence, we have $\max_{i\in I_3}||J_{3,i}||_F = O_p(\eta_N^2)$.

Combining all results above yields the desired result. ■

Lemma A.24. *Under Assumptions 2.1-2.9, we have* *(i)* $\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^0\right)(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right\|_F=O_p\left(\eta_N^2\right)$,

(ii) $\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{v}_{t,1}^{(1)}-O_0^{(1)}v_{t,1}^0\right)(\hat{e}_{1,it}-e_{1,it})(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right\|_F=O_p\left(\eta_N^2\right)$,

(iii) $\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}e_{1,it}\left(\dot{v}_{t,1}^{(1)}-O_0^{(1)}v_{t,1}^0\right)(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right\|_F=O_p\left(\eta_N^2\right)$,

(iv) $\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}O_1^{(1)}v_{t,1}^0(\hat{e}_{1,it}-e_{1,it})(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right\|_F=O_p\left(\eta_N^2\right)$.

Proof (i) We notice that $\mathbb{E}\left[\left(\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^0\right)(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\middle|\mathscr{D}^{I_1\cup I_2}\right]=0$ by Assumption 2.1(ii). Define event $\mathscr{A}_{10}(M)=\left\{\max_{t\in[T]}\left\|\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^0\right\|_2\leq M\xi_N\right\}$ with $\mathbb{P}\left(\mathscr{A}_{10}(M)^c\right)\leq e$ for any $e>0$ by Theorem 2.2(ii). With some positive constant c_{24}, it follows that

$$
\begin{aligned}
&\mathbb{P}\left(\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^0\right)(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right\|_F>c_{24}\eta_N^2\right)\\
&\leq\mathbb{P}\left(\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^0\right)(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right\|_F>c_{24}\eta_N^2,\mathscr{A}_{10}(M)\right)+e\\
&\leq\sum_{i\in I_3}\mathbb{P}\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^0\right)(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right\|_F>c_{24}\eta_N^2,\mathscr{A}_{10}(M)\right)+e\\
&\leq\sum_{i\in I_3}\mathbb{E}\mathbb{P}\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^0\right)(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right\|_F>c_{24}\eta_N^2\middle|\mathscr{D}^{I_1\cup I_2}\right)\mathbf{1}\{\mathscr{A}_{10}(M)\}+e\\
&\leq\sum_{i\in I_3}\exp\left\{-\frac{c_{12}c_{24}^2T^2\eta_N^4}{M^2T\eta_N^2+c_{24}MT\eta_N^2\log T\log\log T}\right\}+e\\
&=o(1)+e.
\end{aligned}
\tag{A.126}
$$

Since e can be made arbitrarily small, this completes the proof of statement (i).

(ii) It's clear that

$$
\begin{aligned}
&\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\dot{v}_{t,1}^{(1)}-O_0^{(1)}v_{t,1}^0\right)(\hat{e}_{1,it}-e_{1,it})(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\right\|_F\\
&\leq\max_{i\in I_3,t\in[T]}\left|\hat{e}_{1,it}-e_{1,it}\right|\max_{t\in[T]}\left\|\dot{v}_{t,1}^{(1)}-O_0^{(1)}v_{t,1}^0\right\|_2=O_p(\eta_N^2),
\end{aligned}
$$

where the equality holds by Lemma A.21(iii) and Theorem 2.2(ii).

(iii) Noting that

$$
\mathbb{E}\left[e_{1,it}\left(\dot{v}_{t,1}^{(1)}-O_0^{(1)}v_{t,1}^0\right)(\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\})\middle|\mathscr{D}^{I_1\cup I_2}\right]=0
$$

by law of iterated expectation and Assumption 2.1(ii), we can obtain the desired result as in (A.126).

(iv) Note that

$$
\begin{aligned}
&\frac{1}{T}\sum_{t\in[T]} v_{t,1}^{0}\left(\hat{e}_{1,it}-e_{1,it}\right)\left(\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right) \\
&=\frac{1}{T^{2}}\sum_{s\in[T]}\sum_{t\in[T]} v_{t,1}^{0}\left(\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right) w_{j,t}^{0\prime}\left(\frac{W_{j}^{0\prime}W_{j}^{0}}{T}\right)^{-1} w_{j,s}^{0}e_{j,is} \\
&+\frac{1}{NT}\sum_{m\in[N]}\sum_{t\in[T]} v_{t,1}^{0}\left(\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right) l_{j,i}^{0\prime}l_{j,m}^{0}e_{j,ms}+\mathscr{R}_{j,it}. \qquad \text{(A.127)}
\end{aligned}
$$

For the first term on the RHS of (A.127), we have

$$
\begin{aligned}
&\max_{i\in I_3}\left\|\frac{1}{T^{2}}\sum_{s\in[T]}\sum_{t\in[T]} v_{t,1}^{0}\left(\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right) w_{j,t}^{0\prime}\left(\frac{W_{j}^{0\prime}W_{j}^{0}}{T}\right)^{-1} w_{j,s}^{0}e_{j,is}\right\|_{2} \\
&\leq \max_{i\in I_3}\left\|\frac{1}{T}\sum_{t\in[T]} v_{t,1}^{0}\left(\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right) w_{j,t}^{0\prime}\right\|_{F}\left\|\frac{1}{T}\sum_{s\in[T]}\left(\frac{W_{j}^{0\prime}W_{j}^{0}}{T}\right)^{-1} w_{j,s}^{0}e_{j,is}\right\|_{2} \\
&=O_p\left(\eta_N^2\right).
\end{aligned}
$$

For the second term on RHS of (A.127), by Assumption 2.9(ii), Bernstein's inequality in Lemma A.12(i) conditional on factors, we have

$$
\max_{i\in I_3}\left\|\frac{1}{NT}\sum_{m\in[N]}\sum_{t\in[T]} v_{t,1}^{0}\left(\tau-\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}\right) l_{j,i}^{0\prime}l_{j,m}^{0}e_{j,ms}\right\|_{2}=O_p\left(\eta_N^2\right).
$$

Then statement (iv) follows. ■

Lemma A.25. *Under Assumptions 2.1-2.9, we have*

(i) $\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T} e_{1,it}O_1^{(1)}v_{t,1}^{0}f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\right\|_2=O_p\left(\eta_N^2\right)$,

(ii) $\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T} O_0^{(1)}v_{t,0}^{0}f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\right\|_2=O_p(\eta_N)$.

Proof (i) Recall from (A.118), we have

$$
\begin{aligned}
&\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}-\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0} \\
&=\left(\hat{\mu}_{1,it}-\mu_{1,it}\right)u_{i,j}^{0\prime}v_{t,j}^{0}+\mu_{1,it}\left(\dot{u}_{i,1}^{(1)}-O_1^{(1)}u_{i,1}^{0}\right)O_1^{(1)}v_{t,1}^{0} \\
&+\mu_{1,it}\left(O_1^{(1)}u_{i,1}^{0}\right)'\left(\dot{v}_{t,1}^{(1)}-O_1^{(1)}v_{t,1}^{0}\right)+O_p\left(\eta_N^2\right) \quad \text{uniformly over } i\in I_3, t\in[T].
\end{aligned}
$$

Then

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} e_{1,it}O_1^{(1)}v_{t,1}^{0}f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right) \\
&=\frac{1}{T}\sum_{t=1}^{T} e_{1,it}O_1^{(1)}v_{t,1}^{0}f_{it}(0)u_{i,j}^{0\prime}v_{t,j}^{0}\left(\hat{\mu}_{1,it}-\mu_{1,it}\right)
\end{aligned}
$$

$$+\frac{1}{T}\sum_{t=1}^{T}e_{1,it}O_1^{(1)}v_{t,1}^0 f_{it}(0)\mu_{1,it}\left(\dot{u}_{i,1}^{(1)}-O_1^{(1)}u_{i,1}^0\right)'O_1^{(1)}v_{t,1}^0$$
$$+\frac{1}{T}\sum_{t=1}^{T}e_{1,it}O_1^{(1)}v_{t,1}^0 f_{it}(0)\mu_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)}-O_1^{(1)}v_{t,1}^0\right)+O_p\left(\eta_N^2\right). \quad \text{(A.128)}$$

First, note that $\mathbb{E}\left(e_{1,it}v_{t,1}^0 f_{it}(0)\mu_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)}-O_1^{(1)}v_{t,1}^0\right)\Big|\mathscr{D}^{I_1\cup I_2}\right)=0$ by Assumption 2.9(i). Following similar arguments used in the proof of Lemma A.21(i), we can show that the third term on the RHS of (A.128) is $O_p\left(\eta_N^2\right)$. By analogous arguments in (A.123) with the fact that $\hat{\mu}_{1,it}-\mu_{1,it}=e_{1,it}-\hat{e}_{1,it}$, we obtain that the first term on the RHS of (A.128) is $O_p\left(\eta_N^2\right)$. In addition

$$\begin{aligned}&\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}e_{1,it}v_{t,1}^0 f_{it}(0)\mu_{1,it}\left(\dot{u}_{i,1}^{(1)}-O_1^{(1)}u_{i,1}^0\right)'O_1^{(1)}v_{t,1}^0\right\|_2\\&\leq\max_{i\in I_3}||\dot{u}_{i,1}^{(1)}-O_1^{(1)}u_{i,1}^0||_2\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)e_{1,it}\mu_{1,it}v_{t,1}^0v_{t,1}^{0\prime}\right\|_F=O_p\left(\eta_N^2\right),\end{aligned} \quad \text{(A.129)}$$

where the equality holds by the fact that

$$\mathbb{P}\left(\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}f_{it}(0)e_{1,it}\mu_{1,it}v_{t,1}^0v_{t,1}^{0\prime}\right\|_F>c_{25}\sqrt{\frac{\log(N\vee T)}{T}}\xi_N\Big|\mathscr{D}\right)=o(1)$$

as in (A.116). Noted that Assumption 2.1(iv) implies factor and factor loading of $X_{j,it}$ is uniformly bounded for $\forall j\in[p]$, which indicates that $\mu_{j,it}$ is also uniformly bounded a.s.. This completes the proof statement (i).

(ii). Note that uniformly over $i\in I_3$, we have

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}O_0^{(1)}v_{t,0}^0 f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^0-\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)}\right)\\&=\frac{1}{T}\sum_{t=1}^{T}O_0^{(1)}v_{t,0}^0 f_{it}(0)\left(\hat{\mu}_{1,it}-\mu_{1,it}\right)u_{i,j}^{0\prime}v_{t,j}^0+\frac{1}{T}\sum_{t=1}^{T}O_0^{(1)}v_{t,0}^0 f_{it}\mu_{1,it}\left(\dot{u}_{i,1}^{(1)}-O_1^{(1)}u_{i,1}^0\right)O_1^{(1)}v_{t,1}^0\\&+\frac{1}{T}\sum_{t=1}^{T}O_0^{(1)}v_{t,0}^0 f_{it}\mu_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)}-O_1^{(1)}v_{t,1}^0\right)+O_p\left(\eta_N^2\right)\\&=O_p(\eta_N).\end{aligned} \quad \text{(A.130)}$$

This term will remain as the bias term in the linear expansion of $\hat{u}_{i,0}^{(3,1)}$. ■

Lemma A.26. *Under Assumptions 2.1-2.9, we have*

(i) $\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}e_{1,it}O_1^{(1)}v_{t,1}^0\left\{\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right\}\right]\right.\right.$

$$-\left(F_{it}(0)-F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]\right)\Bigg\}\Bigg\|_2=o_p\left((N\vee T)^{-\frac{1}{2}}\right),$$

(ii) $\max_{i\in I_3}\Bigg\|\frac{1}{T}\sum_{t=1}^T O_0^{(1)}v_{t,0}^0\Bigg\{\left[\mathbf{1}\{\varepsilon_{it}\le 0\}-\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right\}\right]$

$$-\left(F_{it}(0)-F_{it}\left[\iota_{it}\left(O_0^{(1)}u_{i,0}^0,O_1^{(1)}u_{i,1}^0,\dot{u}_{i,1}^{(1)}\right)\right]\right)\Bigg\}\Bigg\|_2=o_p\left((N\vee T)^{-\frac{1}{2}}\right).$$

Proof (i) We still assume $K_1=1$ for notation simplicity. Let $O^{(1)}=diag\left(O_0^{(1)},O_1^{(1)}\right)$. Recall from Theorem 2.2(iii) that with

$$D_i^I=O^{(1)}\frac{1}{T}\sum_{t=1}^T f_{it}(0)\begin{bmatrix}v_{t,0}^0v_{t,0}' & v_{t,0}^0v_{t,1}^{0\prime}X_{1,it}\\ v_{t,1}^0v_{t,0}^{0\prime}X_{1,it} & v_{t,1}^0v_{t,1}^{0\prime}X_{1,it}^2\end{bmatrix}O^{(1)\prime},$$

$$D_i^{II}=O^{(1)}\frac{1}{T}\sum_{t=1}^T[\tau-\mathbf{1}\{\varepsilon_{it}\le 0\}]\begin{bmatrix}v_{t,0}^0\\ v_{t,1}^0X_{1,it}\end{bmatrix}$$

$$\mathbb{J}_i\left(\{\dot{\Delta}_{t,v}\}_{t\in[T]}\right):=O^{(1)}\frac{1}{T}\sum_{t=1}^T\left[\mathbf{1}\{\varepsilon_{it}\le 0\}-\mathbf{1}\{\varepsilon_{it}\le\dot{\Delta}_{t,v}'\Psi_{it}^0\}\right]\begin{bmatrix}v_{t,0}^0\\ v_{t,1}^0X_{1,it}\end{bmatrix},$$

uniformly over $i\in I_3$, we have

$$\begin{aligned}
&\dot{\Delta}_{i,u}\\
&=\left[D_i^I\right]^{-1}\left[D_i^{II}+\mathbb{J}_i\left(\{\dot{\Delta}_{t,v}\}_{t\in[T]}\right)\right]+o_p\left((N\vee T)^{-\frac{1}{2}}\right)\\
&=\left(O^{(1)\prime}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T f_{it}(0)\begin{bmatrix}v_{t,0}^0v_{t,0}' & v_{t,0}^0v_{t,1}^{0\prime}X_{1,it}\\ v_{t,1}^0v_{t,0}^{0\prime}X_{1,it} & v_{t,1}^0v_{t,1}^{0\prime}X_{1,it}^2\end{bmatrix}\right)^{-1}\left[\frac{1}{T}\sum_{t=1}^T[\tau-\mathbf{1}\{\varepsilon_{it}\le 0\}]\begin{bmatrix}v_{t,0}^0\\ v_{t,1}^0X_{1,it}\end{bmatrix}\right.\\
&\left.+\frac{1}{T}\sum_{t=1}^T\left[\mathbf{1}\{\varepsilon_{it}\le 0\}-\mathbf{1}\{\varepsilon_{it}\le\dot{\Delta}_{t,v}'\Psi_{it}^0\}\right]\begin{bmatrix}v_{t,0}^0\\ v_{t,1}^0X_{1,it}\end{bmatrix}\right]+o_p\left((N\vee T)^{-\frac{1}{2}}\right)\\
&:=h_i+o_p\left((N\vee T)^{-\frac{1}{2}}\right).
\end{aligned}$$

Let $\mathfrak{l}=\left(\mathbf{0}_{K_0}',\mathbf{1}_{K_1}'\right)'$ with $\mathbf{0}_{K_0}$ being a $K_0\times 1$ vector of zeros and $\mathbf{1}_{K_1}$ a $K_1\times 1$ vector of ones. Let

$$h_i^I=\mathfrak{l}'h_i. \tag{A.131}$$

Then we have

$$O_1^{(1)\prime}\dot{u}_{i,1}^{(1)}-u_{i,1}^0=h_i^I+o_p\left((N\vee T)^{-\frac{1}{2}}\right)\quad\text{uniformly over } i\in I_3,$$

and $\max_{i\in I_3}||h_i^I||_2=O_p(\eta_N)$.

Combining (A.117)-(A.119), uniformly in $i \in I_3$ and $t \in [T]$ we have

$$
\begin{aligned}
&\iota_{it}\left(O_0^{(1)}u_{i,0}^0, O_1^{(1)}u_{i,1}^0, \dot{u}_{i,1}^{(1)}\right) \\
&= \left[\left(O_0^{(1)}u_{i,0}^0\right)'\left(\dot{v}_{t,0}^{(1)} - O_0^{(1)}v_{t,0}^0\right)\right] + \left(\hat{\mu}_{1,it}\dot{u}_{i,1}^{(1)\prime}\dot{v}_{t,1}^{(1)} - \mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^0\right) \\
&+ \left(\hat{e}_{1,it}u_{i,1}^{0\prime}O_1^{(1)}\dot{v}_{t,1}^{(1)} - e_{1,it}u_{i,1}^{0\prime}v_{t,1}^0\right) \\
&= \left(O_0^{(1)}u_{i,0}^0\right)'\left(\dot{v}_{t,0}^{(1)} - O_0^{(1)}v_{t,0}^0\right) + \left(\hat{\mu}_{1,it}^{(1)} - \mu_{1,it}\right)\left(O_1^{(1)}u_{i,1}^0\right)'O_1^{(1)}v_{t,1}^0 \\
&+ \mu_{1,it}\left(\dot{u}_{i,1}^{(1)} - O_1^{(1)}u_{i,1}^0\right)'O_1^{(1)}v_{t,1}^0 + \mu_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)} - O_1^{(1)}v_{t,1}^0\right) \\
&+ \left(\hat{e}_{1,it} - e_{1,it}\right)\left(O_1^{(1)}u_{i,1}^0\right)'O_1^{(1)}v_{t,1}^0 \\
&+ e_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)} - O_1^{(1)}v_{t,1}^0\right) + o_p\left((N \vee T)^{-\frac{1}{2}}\right) \\
&= \mu_{1,it}\left(\dot{u}_{i,1}^{(1)} - O_1^{(1)}u_{i,1}^0\right)'O_1^{(1)}v_{t,1}^0 + \left(O_0^{(1)}u_{i,0}^0\right)'\left(\dot{v}_{t,0}^{(1)} - O_0^{(1)}v_{t,0}^0\right) \\
&+ \left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)} - O_1^{(1)}v_{t,1}^0\right)X_{1,it} + o_p\left((N \vee T)^{-\frac{1}{2}}\right) \\
&= \mu_{1,it}v_{t,1}^{0\prime}h_i^I + \left\{\left(O_0^{(1)}u_{i,0}^0\right)'\left(\dot{v}_{t,0}^{(1)} - O_0^{(1)}v_{t,0}^0\right) + \left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)} - O_1^{(1)}v_{t,1}^0\right)X_{1,it}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&+\mathscr{R}_{\iota,it} \\
&:= \mu_{1,it}v_{t,1}^{0\prime}h_i^I + h_{it}^{II}\left(\dot{\Delta}_{t,v}\right) + \mathscr{R}_{\iota,it} \qquad \text{(A.132)} \\
&:= R_{\iota,it}^1\left(|\mu_{1,it}| + |e_{1,it}|\right) + R_{\iota,it}^2
\end{aligned}
$$

such that $\max_{i\in I_3, t\in[T]}|\mathscr{R}_{\iota,it}| = o_p\left((N \vee T)^{-\frac{1}{2}}\right)$, $\max_{i\in I_3, t\in[T]}\left|R_{\iota,it}^1\right| = O_p(\eta_N)$ and $\max_{i\in I_3, t\in[T]}\left|R_{\iota,it}^2\right| = O_p(\eta_N)$, where we use the fact that $\hat{\mu}_{1,it} - \mu_{1,it} + \hat{e}_{1,it} - e_{1,it} = 0$, and $h_{it}^{II}\left(\dot{\Delta}_{t,v}\right) = \Psi_{it}^{0\prime}\dot{\Delta}_{t,v}$ with $\Psi_{it} = \left((O_0^{(1)}u_{i,0}^0)', (O_1^{(1)}u_{i,1}^0X_{1,it})'\right)'$. Let

$$
\begin{aligned}
\hat{\mathbb{I}}_{3,it}^I = e_{1,it}v_{t,1}^0\Bigg\{ &\left[\mathbf{1}\{\varepsilon_{it} \le 0\} - \mathbf{1}\left\{\varepsilon_{it} \le \mu_{1,it}v_{t,1}^{0\prime}h_i^I + h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right\}\right] \\
&- \left(F_{it}(0) - F_{it}\left[\mu_{1,it}v_{t,1}^{0\prime}h_i^I + h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right]\right)\Bigg\}, \\
\hat{\mathbb{I}}_{3,it}^{II} = e_{1,it}v_{t,1}^0\Bigg\{ &\left[\mathbf{1}\left\{\varepsilon_{it} \le \mu_{1,it}v_{t,1}^{0\prime}h_i^I + h_{it}^{II}\left(\dot{\Delta}_{t,v}\right) + \mathscr{R}_{\iota,it}\right\} - \mathbf{1}\left\{\varepsilon_{it} \le \mu_{1,it}v_{t,1}^{0\prime}h_i^I + h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right\}\right] \\
&- \left(F_{it}\left[\mu_{1,it}v_{t,1}^{0\prime}h_i^I + h_{it}^{II}\left(\dot{\Delta}_{t,v}\right) + \mathscr{R}_{\iota,it}\right] - F_{it}\left[\mu_{1,it}v_{t,1}^{0\prime}h_i^I + h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right]\right)\Bigg\}.
\end{aligned}
$$

We first show that $\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^T \hat{\mathbb{I}}_{3,it}^I\right\|_2 = O_p\left(\eta_N^2\right)$.

For some sufficiently large constant M, define event $\mathscr{A}_{11}(M) = \left\{\max_{i\in I_3}\left\|h_i^I\right\|_2 \le M\eta_N\right\}$. We have $\mathbb{P}\left(\mathscr{A}_{11}^c(M)\right) \le e$ for any $e > 0$. For some positive large enough constant

c_{26}, we have

$$\mathbb{P}\left(\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\mathbb{I}}_{3,it}^{I}\right\|_2 > c_{26}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right)$$
$$\leq \mathbb{P}\left(\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\mathbb{I}}_{3,it}^{I}\right\|_2 > c_{26}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}, \mathscr{A}_{11}(M)\right)+e$$
$$\leq \mathbb{P}\left(\max_{i\in I_3}\sup_{\xi\in\Xi}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\mathbb{I}}_{3,it}^{I}(\xi)\right\|_2 > c_{26}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right)+e. \tag{A.133}$$

with $\Xi^1 := \left\{\xi \in \mathbb{R}^{(K_0+K_1)\times 1} : \|\xi\|_2 \leq M\eta_N\right\}$ and

$$\hat{\mathbb{I}}_{3,it}^{I}(\xi) = e_{1,it}v_{t,1}^{0}\Big\{\left[\mathbf{1}\{\varepsilon_{it}\leq 0\} - \mathbf{1}\left\{\varepsilon_{it}\leq \mu_{1,it}v_{t,1}^{0\prime}\xi + h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right\}\right]$$
$$- \left[F_{it}(0) - F_{it}\left(\mu_{1,it}v_{t,1}^{0\prime}\xi + h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right)\right]\Big\}.$$

Divide Ξ^1 into sub classes Ξ_s^1 for $s = 1,\ldots,n_{\Xi^1}$ such that $\left\|\xi - \bar{\xi}\right\|_2 < \frac{\varepsilon}{T}$ for $\forall\xi,\ \bar{\xi} \in \Xi_s^1$ and $n_{\Xi^1} \asymp T^{K_0+K_1}$. With analogous arguments from (A.14)-(A.17), for $\forall\xi_s \in \Xi_s^1$, we have

$$\left\|\frac{1}{T}\sum_{t\in[T]}\hat{\mathbb{I}}_{3,it}^{I}(\xi)\right\|_2 \leq \left\|\frac{1}{T}\sum_{t\in[T]}\hat{\mathbb{I}}_{3,it}^{I}(\xi_s)\right\|_2 + \left\|\frac{1}{T}\sum_{t\in[T]}\left[\hat{\mathbb{I}}_{3,it}^{I}(\xi) - \hat{\mathbb{I}}_{3,it}^{I}(\xi_s)\right]\right\|_2$$

with

$$\max_{i\in I_3, s\in[n_{\Xi^1}]}\sup_{\xi\in\Xi_s}\left\|\frac{1}{T}\sum_{t\in[T]}\left[\hat{\mathbb{I}}_{3,it}^{I}(\xi) - \hat{\mathbb{I}}_{3,it}^{I}(\xi_s)\right]\right\|_2$$
$$\leq \max_{i\in I_3}\frac{1}{T}\sum_{t\in[T]}\mathbb{E}\left[\left\|e_{1,it}v_{t,1}^{0}\right\|_2\mathbf{1}\left\{\left|\varepsilon_{it} - h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right| \leq \left|\mu_{1,it}\right|\left\|v_{t,1}^{0\prime}\right\|_2\frac{\varepsilon}{T}\right\}\middle|\mathscr{D}_e^{I_1\cup I_2}\right]$$
$$+ \max_{i\in I_3}\left\|\hat{\mathbb{I}}_{3,i}^{III}\right\|_2 + \max_{i\in I_3}\frac{1}{T}\sum_{t\in[T]}\left\|e_{1,it}v_{t,1}^{0}\right\|_2\left|\mu_{1,it}\right|\left\|v_{t,1}^{0\prime}\right\|_2\frac{\varepsilon}{T}$$
$$\leq \max_{i\in I_3}\left\|\hat{\mathbb{I}}_{3,i}^{III}\right\|_2 + O\left(\frac{\varepsilon}{T}\right)\ a.s. \tag{A.134}$$

where $\hat{\mathbb{I}}_{3,i}^{III} = \frac{1}{T}\sum_{t\in[T]}\hat{\mathbb{I}}_{3,it}^{III}$ and

$$\hat{\mathbb{I}}_{3,it}^{III} = \left\|e_{1,it}v_{t,1}^{0}\right\|_2\Big\{\mathbf{1}\left\{\left|\varepsilon_{it} - h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right| \leq \left|\mu_{1,it}\right|\left\|v_{t,1}^{0\prime}\right\|_2\frac{\varepsilon}{T}\right\}$$
$$- \mathbb{E}\left[\mathbf{1}\left\{\left|\varepsilon_{it} - h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right| \leq \left|\mu_{1,it}\right|\left\|v_{t,1}^{0\prime}\right\|_2\frac{\varepsilon}{T}\right\}\middle|\mathscr{D}_e^{I_1\cup I_2}\right]\Big\}.$$

Similarly to (A.17), we can show that

$$\mathbb{P}\left(\max_{i\in I_3}\left\|\hat{\mathbb{I}}_{3,it}^{III}\right\|_2 > c_{26}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right)=o(1), \quad \text{and} \tag{A.135}$$

$$\mathbb{P}\left(\max_{i\in I_3}\max_{s\in[n_{\Xi 1}]}\left\|\hat{\mathbb{I}}_{3,it}^{I}\left(\xi_s\right)\right\|_2 > c_{26}\xi_N^{\frac{5+\vartheta}{2+\vartheta}}\left(\frac{\log(N\vee T)}{N\wedge T}\right)^{\frac{1}{4+2\vartheta}}\sqrt{\frac{\log(N\vee T)}{T}}\right)=o(1). \tag{A.136}$$

Combining (A.133)-(A.135), we have

$$\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\mathbb{I}}_{3,it}^{I}\right\|_2 = o_p\left((N\vee T)^{-\frac{1}{2}}\right).$$

Next, we notice that

$$\begin{aligned}
&\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\mathbb{I}}_{3,it}^{II}\right\|_2\\
&\leq \max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|e_{1,it}v_{t,1}^{0}\right\|_2\mathbf{1}\left\{\left|\mu_{1,it}v_{t,1}^{0\prime}h_i^I+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right|\leq|\varepsilon_{it}|\leq\left|\mu_{1,it}v_{t,1}^{0\prime}h_i^I+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right|+|\mathscr{R}_{\iota,it}|\right\}\\
&+\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|e_{1,it}v_{t,1}^{0}\right\|_2\left|F_{it}\left[\mu_{1,it}v_{t,1}^{0\prime}h_i^I+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)+\mathscr{R}_{\iota,it}\right]-F_{it}\left[\mu_{1,it}v_{t,1}^{0\prime}h_i^I+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right]\right|\\
&\leq \max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|e_{1,it}v_{t,1}^{0}\right\|_2\mathbf{1}\left\{\left|\mu_{1,it}v_{t,1}^{0\prime}h_i^I+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right|\leq|\varepsilon_{it}|\leq\left|\mu_{1,it}v_{t,1}^{0\prime}h_i^I+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right|+|\mathscr{R}_{\iota,it}|\right\}\\
&+\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|e_{1,it}v_{t,1}^{0}\right\|_2|\mathscr{R}_{\iota,it}|\\
&=\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\hat{\mathbb{I}}_{3,it}^{IV}+o_p\left((N\vee T)^{-\frac{1}{2}}\right),
\end{aligned} \tag{A.137}$$

where the first inequality is by triangle inequality, the second inequality is by mean value theorem and the last line is by the uniform convergence rate for $\mathscr{R}_{\iota,it}$ with

$$\hat{\mathbb{I}}_{3,it}^{IV}=\left\|e_{1,it}v_{t,1}^{0}\right\|_2\mathbf{1}\left\{\left|\mu_{1,it}v_{t,1}^{0\prime}h_i^I+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right|\leq|\varepsilon_{it}|\leq\left|\mu_{1,it}v_{t,1}^{0\prime}h_i^I+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right|+|\mathscr{R}_{\iota,it}|\right\}.$$

Define the event $\mathscr{A}_{12}(M):=\left\{\max_{i\in I_3,t\in[T]}|\mathscr{R}_{\iota,it}|\leq M\eta_N^2\right\}$ with $\mathbb{P}\left\{\mathscr{A}_{12}^c(M)\right\}\leq e$ for any $e>0$. Then for a large enough constant c_{26}, similarly as (A.37), we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\mathbb{I}}_{3,it}^{IV}\right\|_2>c_{26}\eta_N^2\right)\leq\mathbb{P}\left(\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\mathbb{I}}_{3,it}^{IV}\right\|_2>c_{26}\eta_N^2,\mathscr{A}_{12}(M)\right)+e\\
&\leq\mathbb{P}\left(\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{1}}_{it}(h_i^I)>c_{26}\eta_N^2\right)+e
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{P}\left(\sup_{i\in I_3,\xi\in\Xi^1}\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{l}}_{it}(\xi)>c_{26}\eta_N^2\right)+2e\\
&\leq \mathbb{P}\left(\max_{i\in I_3}\sup_{\xi\in\Xi^1}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{l}}_{it}(\xi)-\bar{\tilde{\mathbf{l}}}_{it}(\xi)\right|>\frac{c_{26}\eta_N^2}{2}\right)\\
&+\mathbb{P}\left(\max_{i\in I_3}\sup_{\xi\in\Xi^1}\left|\frac{1}{T}\sum_{t=1}^{T}\bar{\tilde{\mathbf{l}}}_{it}(\xi)\right|>\frac{c_{26}\eta_N^2}{2}\right)+2e\\
&=\mathbb{E}\left[\mathbb{P}\left(\max_{i\in I_3}\sup_{\xi\in\Xi^1}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{l}}_{it}(\xi)-\bar{\tilde{\mathbf{l}}}_{it}(\xi)\right|>\frac{c_{26}\eta_N^2}{2}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)\right]\\
&+\mathbb{P}\left(\max_{i\in I_3}\sup_{\xi\in\Xi^1}\left|\frac{1}{T}\sum_{t=1}^{T}\bar{\tilde{\mathbf{l}}}_{it}(\xi)\right|>\frac{c_{26}\eta_N^2}{2}\right)+2e \qquad \text{(A.138)}
\end{aligned}
$$

where $\tilde{\mathbf{l}}_{it}(h_i^I):=\left\|e_{1,it}v_{t,1}^0\right\|_2\mathbf{1}\left\{\left|\mu_{1,it}v_{t,1}^{0\prime}h_i^I+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right|\leq|\varepsilon_{it}|\leq\left|\mu_{1,it}v_{t,1}^{0\prime}h_i^I+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right|+M\eta_N^2\right\}$ and $\bar{\tilde{\mathbf{l}}}_{it}(h_i^I)=\mathbb{E}\left[\tilde{\mathbf{l}}_{it}(h_i^I)\middle|\mathscr{D}_e^{I_1\cup I_2}\right]$. By analogous arguments for the first term on the RHS of inequality (A.133), we can show that

$$
\mathbb{E}\left[\mathbb{P}\left(\max_{i\in I_3}\sup_{\xi\in\Xi^1}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{l}}_{it}(\xi)-\bar{\tilde{\mathbf{l}}}_{it}(\xi)\right|>\frac{c_{26}\eta_N^2}{2}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)\right]=o(1).
$$

Besides, we observe that

$$
\begin{aligned}
&\max_{i\in I_3}\sup_{\xi\in\Xi^1}\left|\frac{1}{T}\sum_{t=1}^{T}\bar{\tilde{\mathbf{l}}}_{it}(\xi)\right|\\
&\leq\max_{i\in I_3}\sup_{\xi\in\Xi^1}\frac{1}{T}\sum_{t=1}^{T}\left\|e_{1,it}v_{t,1}^0\right\|_2\left|F_{it}\left[\left|\mu_{1,it}v_{t,1}^{0\prime}\xi+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right|+M\eta_N^2\right]-F_{it}\left[\left|\mu_{1,it}v_{t,1}^{0\prime}\xi+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)\right|\right]\right|\\
&\leq\max_{i\in I_3}\frac{1}{T}\sum_{t=1}^{T}\left\|e_{1,it}v_{t,1}^0\right\|_2M\eta_N^2,
\end{aligned}
$$

which yields

$$
\mathbb{P}\left(\max_{i\in I_3}\sup_{\xi\in\Xi^1}\left|\frac{1}{T}\sum_{t=1}^{T}\bar{\tilde{\mathbf{l}}}_{it}(\xi)\right|>\frac{c_{26}\eta_N^2}{2}\right)=o(1). \qquad \text{(A.139)}
$$

Combining (A.137)-(A.139) yields $\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\mathbb{I}}_{3,it}^{II}\right\|_2=O_p\left(\eta_N^2\right)=o_p\left((N\vee T)^{-\frac{1}{2}}\right)$ by Assumption 2.1(ix). This concludes the proof of statement (i).

(ii) The proof is analogous to that of part (i) and thus omitted. ■

Lemma A.27. *Under Assumptions 2.1-2.9, uniformly in $i\in I_3$, we can show that*

(i) $\max_{i\in I_3}\left\|\frac{1}{T}\sum\limits_{t=1}^{T}e_{1,it}O_1^{(1)}v_{t,1}^0\left\{\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{\varepsilon_{it}\leq\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right\}\right.\right.$

$$-\left(F_{it}(0)-F_{it}\left[\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right]\right)\Bigg\}\Bigg\|_2=o_p\left((N\vee T)^{-\frac{1}{2}}\right),$$

$$\text{(ii)}\ \max_{i\in I_3}\Bigg\|\frac{1}{T}\sum_{t=1}^{T}v_{t,0}^0\Bigg\{\mathbf{1}\{\varepsilon_{it}\le 0\}-\mathbf{1}\left\{\varepsilon_{it}\le\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right\}-$$
$$\left(F_{it}(0)-F_{it}\left[\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\right]\right)\Bigg\}\Bigg\|_2=o_p\left((N\vee T)^{-\frac{1}{2}}\right).$$

Proof As in (A.132), we can show that

$$\begin{aligned}
&\iota_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\dot{u}_{i,1}^{(1)}\right)\\
&=\mu_{1,it}v_{t,1}^{0\prime}h_i^I+\left[\left(\hat{u}_{i,0}^{(3,1)}-O_0^{(1)}u_{i,0}^0\right)'O_0^{(1)}v_{t,0}^0+e_{1,it}\left(\hat{u}_{i,1}^{(3,1)}-O_1^{(1)}u_{i,1}^0\right)'O_1^{(1)}v_{t,1}^0\right]\\
&+\left[\left(O_0^{(0)}u_{i,0}^0\right)'\left(\dot{v}_{t,0}^{(1)}-O_0^{(1)}v_{t,0}^0\right)+X_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)}-O_1^{(1)}v_{t,1}^0\right)\right]+\mathscr{R}_{\hat{\iota},it}\\
&:=\mu_{1,it}v_{t,1}^{0\prime}h_i^I+\left(\hat{u}_{i,0}^{(3,1)}-O_0^{(1)}u_{i,0}^0\right)'O_0^{(1)}v_{t,0}^0+e_{1,it}\left(\hat{u}_{i,1}^{(3,1)}-O_1^{(1)}u_{i,1}^0\right)'O_1^{(1)}v_{t,1}^0\\
&+h_{it}^{II}\left(\dot{\Delta}_{t,v}\right)+\mathscr{R}_{\hat{\iota},it},
\end{aligned}\tag{A.140}$$

where $\max_{i\in I_3,t\in[T]}\left|\mathscr{R}_{\hat{\iota},it}\right|=o_p\left((N\vee T)^{-\frac{1}{2}}\right)$. As in the proof of Theorem 2.2, we can show that $\max_{i\in I_3}\left\|\hat{u}_{i,j}^{(3,1)}-O_j^{(1)}u_{i,j}^0\right\|_2=O_p(\eta_N)$ for $\forall j\in[p]\cup\{0\}$. Then by changing the event set $\mathscr{A}_{11}(M)$ to

$$\left\{\max_{i\in I_3}\|h_i^I\|_2\le M\eta_N,\max_{i\in I_3}\left\|\hat{u}_{i,j}-O_j^{(1)}u_{i,j}^0\right\|_2\le M\eta_N\right\},$$

we can repeat the analysis in Lemma A.26 and obtain the desired results for statement (i). With some obvious adjustment, statement (ii) can be proved. ■

Lemma A.28. *Under Assumptions 2.1-2.9, for block matrices $\hat{D}_t^F$, D_t^F, $\hat{D}_t^J$ and D_t^J defined in Appendix A, we have*

$$\max_{t\in[T]}||\hat{D}_t^F-D_t^F||_F=O_p(\eta_N),\ \textit{and}\ \max_{t\in[T]}||\hat{D}_t^J-D_t^J||_F=\left\|\begin{bmatrix}O_p\left(\eta_N^2\right) & O_p(\eta_N)\\ O_p\left(\eta_N^2\right) & O_p\left(\eta_N^2\right)\end{bmatrix}\right\|_F.$$

Proof By analogous arguments as used in the proofs of Lemmas A.22 and A.23, we can prove the lemma. ■

Lemma A.29. *Under Assumptions 2.1-2.9, we have*

$$\text{(i)}\ \max_{t\in[T]}\left\|\frac{1}{N_3}\sum_{i\in I_3}O_{u,0}^{(1)}u_{i,0}^0f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^0-\hat{\mu}_{1,it}\hat{u}_{i,1}^{(3,1)\prime}\dot{v}_{t,1}^{(1)}\right)\right\|_2=O_p(\eta_N),$$

$$\text{(ii)}\ \max_{t\in[T]}\left\|\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}O_{u,1}^{(1)}u_{i,1}^0f_{it}(0)\left(\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^0-\hat{\mu}_{1,it}\hat{u}_{i,1}^{(3,1)\prime}\dot{v}_{t,1}^{(1)}\right)\right\|_2=o_p\left((N\vee T)^{-\frac{1}{2}}\right).$$

Proof (i) Note that

$$
\begin{aligned}
&\hat{\mu}_{1,it}\hat{u}_{i,1}^{(3,1)\prime}\dot{v}_{t,1}^{(1)}-\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}\\
&=(\hat{\mu}_{1,it}-\mu_{1,it})\left(\hat{u}_{i,1}^{(3,1)}-O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}\left(\dot{v}_{t,1}^{(1)}-O_{1}^{(1)}v_{t,1}^{0}\right)\\
&+(\hat{\mu}_{1,it}-\mu_{1,it})\left(O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}\left(\dot{v}_{t,1}^{(1)}-O_{1}^{(1)}v_{t,1}^{0}\right)\\
&+(\hat{\mu}_{1,it}-\mu_{1,it})\left(\hat{u}_{i,1}^{(3,1)}-O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}O_{1}^{(1)}v_{t,1}^{0}+(\hat{\mu}_{1,it}-\mu_{1,it})\left(O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}O_{1}^{(1)}v_{t,1}^{0}\\
&+\mu_{1,it}\left(\hat{u}_{i,1}^{(3,1)}-O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}\left(\dot{v}_{t,1}^{(1)}-O_{1}^{(1)}v_{t,1}^{0}\right)+\mu_{1,it}\left(\hat{u}_{i,1}^{(3,1)}-O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}O_{1}^{(1)}v_{t,1}^{0}\\
&+\mu_{1,it}\left(O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}\left(\dot{v}_{t,1}^{(1)}-O_{1}^{(1)}v_{t,1}^{0}\right)+\mu_{1,it}\left(O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}O_{1}^{(1)}v_{t,1}^{0}-\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}\\
&=(\hat{\mu}_{1,it}-\mu_{1,it})\left(O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}O_{1}^{(1)}v_{t,1}^{0}+\mu_{1,it}\left(\hat{u}_{i,1}^{(3,1)}-O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}O_{1}^{(1)}v_{t,1}^{0}\\
&+\mu_{1,it}\left(O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}\left(\dot{v}_{t,1}^{(1)}-O_{1}^{(1)}v_{t,1}^{0}\right)+\mu_{1,it}v_{t,1}^{0\prime}O_{1}^{(1)\prime}\left(O_{u,1}^{(1)}-O_{1}^{(1)}\right)u_{i,1}^{0}+O_{p}\left(\eta_{N}^{2}\right)\\
&=(\hat{\mu}_{1,it}-\mu_{1,it})u_{i,1}^{0\prime}v_{t,1}^{0}+\mu_{1,it}\left(\hat{u}_{i,1}^{(3,1)}-O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}O_{1}^{(1)}v_{t,1}^{0}\\
&+\mu_{1,it}\left(O_{1}^{(1)}u_{i,1}^{0}\right)^{\prime}\left(\dot{v}_{t,1}^{(1)}-O_{1}^{(1)}v_{t,1}^{0}\right)+\mu_{1,it}v_{t,1}^{0\prime}O_{1}^{(1)\prime}\left(O_{u,1}^{(1)}-O_{1}^{(1)}\right)u_{i,1}^{0}+O_{p}\left(\eta_{N}^{2}\right)\\
&=O_{p}(\eta_{N}),
\end{aligned}
\tag{A.141}
$$

uniformly over $i\in I_3$ and $t\in[T]$, where the last equality holds by the fact that $\left\|O_{u,1}^{(1)}-O_{1}^{(1)}\right\|_F=O_p(\eta_N)$. It follows that

$$
\max_{t\in[T]}\left\|\frac{1}{N_3}\sum_{i\in I_3}O_{u,0}^{(1)}u_{i,0}^{0}f_{it}(0)\left(\hat{\mu}_{1,it}\hat{u}_{i,1}^{(3,1)\prime}\dot{v}_{t,1}^{(1)}-\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}\right)\right\|_2=O_p(\eta_N).
$$

(ii) Observe that

$$
\begin{aligned}
&\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}O_{u,1}^{(1)}u_{i,1}^{0}f_{it}(0)\left(\hat{\mu}_{1,it}^{(1)}\hat{u}_{i,1}^{(3,1)\prime}\dot{v}_{t,1}^{(1)}-\mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^{0}\right)\\
&=O_{u,1}^{(1)}\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u_{i,1}^{0}f_{it}(0)\left(\hat{\mu}_{1,it}^{(1)}-\mu_{1,it}\right)u_{i,1}^{0\prime}v_{t,1}^{0}\\
&+O_{u,1}^{(1)}\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u_{i,1}^{0}f_{it}(0)\mu_{1,it}\left(\hat{u}_{i,1}^{(3,1)}-O_{u,1}^{(1)}u_{i,1}^{0}\right)^{\prime}O_{1}^{(1)}v_{t,1}^{0}\\
&+O_{u,1}^{(1)}\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u_{i,1}^{0}f_{it}(0)\mu_{1,it}\left(O_{1}^{(1)}u_{i,1}^{0}\right)^{\prime}\left(\dot{v}_{t,1}^{(1)}-O_{1}^{(1)}v_{t,1}^{0}\right)\\
&+O_{u,1}^{(1)}\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u_{i,1}^{0}f_{it}(0)\mu_{1,it}v_{t,1}^{(1)\prime}O_{1}^{(1)\prime}\left(O_{u,1}^{(1)}-O_{1}^{(1)}\right)u_{i,1}^{0}+O_{p}\left(\eta_{N}^{2}\right).
\end{aligned}
\tag{A.142}
$$

By similar arguments as used in (A.123), we can show that the first term on the RHS of (A.142) is $o_p\left((N\vee T)^{-\frac{1}{2}}\right)$ uniformly over t. For the second term, by inserting the linear expansion for $\hat{u}_{i,1}^{(3,1)} - O_{u,1}^{(1)} u_{i,1}^0$ in (A.41), we notice that

$$\frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)\mu_{1,it}\left(\hat{u}_{i,1}^{(3,1)} - O_{u,1}^{(1)}u_{i,1}^0\right)' O_1^{(1)} v_{t,1}^0$$
$$= \frac{1}{N_3 T}\sum_{i\in I_3}\sum_{t\in[T]} e_{1,it} f_{it}(0)\,\mu_{1,it} u_{i,1}^0 v_{t^*,1}^{0\prime}\hat{V}_{u_1}^{-1} e_{1,it^*} v_{t^*,1}^0 \left(\tau - \mathbf{1}\{\varepsilon_{it^*}\le 0\}\right) + o_p\left((N\vee T)^{-\frac{1}{2}}\right).$$

By arguments as used in (A.124), we can show the leading term in the last equality is $O_p\left(\sqrt{\frac{\log(N\vee T)}{NT}}\xi_N^2\right)$. Then the second term on the RHS of (A.142) is $o_p\left((N\vee T)^{-\frac{1}{2}}\right)$. For the third and the fourth terms on the RHS of (A.142), by conditional on $\mathscr{D}^{I_1\cup I_2}$, we notice that $\frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)\mu_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)} - O_1^{(1)}v_{t,1}^0\right)$ and

$$\frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)\mu_{1,it} v_{t,1}^{0\prime} O_1^{(1)\prime}\left(O_{u,1}^{(1)} - O_1^{(1)}\right)u_{i,1}^0$$

are both mean zero and the randomness depends only on $\{e_{1,it}f_{it}(0)\}$. By conditional Hoeffding's inequality, we can show that both these two terms are $o_p\left((N\vee T)^{-\frac{1}{2}}\right)$.

■

Define

$$\mathbb{J}_{it}\left(\dot{\Delta}_{t,v}\right) = \left[\mathbf{1}\{\varepsilon_{it}\le 0\} - \mathbf{1}\{\varepsilon_{it}\le \dot{\Delta}_{t,v}'\Psi_{it}^0\}\right]\begin{bmatrix} v_{t,0}^0 \\ v_{t,1}^0 X_{1,it}\end{bmatrix},$$

$$h_i^{I,1} = \frac{1}{T}\sum_{t=1}^T \mathbb{E}\left[f_{it}(0)\begin{bmatrix} v_{t,0}^0 v_{t,0}' & v_{t,0}^0 v_{t,1}^{0\prime}X_{1,it} \\ v_{t,1}^0 v_{t,0}^{0\prime}X_{1,it} & v_{t,1}^0 v_{t,1}^{0\prime}X_{1,it}^2\end{bmatrix}\middle|\mathscr{D}\right],$$

$$h_i^{I,2} = \frac{1}{T}\sum_{t=1}^T \mathbb{E}\left(\mathbb{J}_i\left(\dot{\Delta}_{t,v}\right)\middle|\mathscr{D}^{I_1\cup I_2}\right),$$

$$h_{it}^{III} = v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0\left(\tau - 1\{\varepsilon_{it^*}\le 0\}\right) + X_{1,it}v_{t,1}^{0\prime}\hat{V}_{u_1}^{-1}\frac{1}{T}\sum_{t^*=1}^T e_{1,it^*}v_{t^*,1}^0\left(\tau - 1\{\varepsilon_{it^*}\le 0\}\right)$$
$$-\left(v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0 f_{it^*}(0)\mu_{1,it^*}v_{t^*,1}^{0\prime}\right)\mathfrak{l}'\left(O_1^{(1)}\right)^{-1}\left(h_i^{I,1}\right)^{-1}\frac{1}{T}\sum_{t^*=1}^T\left[\tau - \mathbf{1}\{\varepsilon_{it^*}\le 0\}\right]\begin{bmatrix} v_{t^*,0}^0 \\ v_{t^*,1}^0 X_{1,it^*}\end{bmatrix}$$
$$-\left(v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0\mathbb{E}\left[f_{it^*}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\mu_{1,it^*}v_{t^*,1}^{0\prime}\right)\mathfrak{l}'\left(O_1^{(1)}\right)^{-1}\left(h_i^{I,1}\right)^{-1}h_i^{I,2}$$
$$= \mathscr{R}_{h,it}^1 + X_{1,it}\mathscr{R}_{h,it}^2, \tag{A.143}$$

with $\max_{i\in I_3, t\in[T]}\left|\mathscr{R}_{h,it}^a\right| = O_p(\eta_N)$ for $a\in\{1,2\}$. Note that $\mathbb{E}\left[f_{it}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right] = \mathbb{E}\left[f_{it}(0)\middle|\mathscr{D}\right]$.

Assumption A.1. *Let $\mathscr{F}_{it}(\cdot)$ and $f_{it|h_{it}^{III}}(\cdot)$ be the conditional CDF and PDF of*

ε_{it} *given* $\mathscr{D}_e$ *and* h_{it}^{III}.

(i) *The derivative of the density* $f_{it|h_{it}^{III}}$ *is uniformly bounded in absolute value.*

(ii) $\max_{i\in[N],t\in[T]}\left|f_{it|h_{it}^{III}}(0)-f_{it|h_{it}^{III}=0}(0)\right|\leq C\left|h_{it}^{III}\right|$ *for some Lipschitz constant* $C>0$.

Lemma A.30. *Under Assumptions 2.1-2.9 and Assumption A.1, we have*

(i) $\max_{t\in[T]}\Bigg\|\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}O_{u,1}^{(1)}u_{i,1}^0\Bigg\{\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{\varepsilon_{it}\leq\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime -1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime -1}v_{t,1}^0\right)\right\}\right]$
$-\left(F_{it}(0)-F_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime -1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime -1}v_{t,1}^0\right)\right]\right)\Bigg\}\Bigg\|_2=o_p\left((N\vee T)^{-\frac{1}{2}}\right)$

(ii) $\max_{t\in[T]}\Bigg\|\frac{1}{N_3}\sum_{i\in I_3}O_{u,0}^{(1)}u_{i,0}^0\Bigg\{\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{\varepsilon_{it}\leq\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime -1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime -1}v_{t,1}^0\right)\right\}\right]$
$-\left(F_{it}(0)-F_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime -1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime -1}v_{t,1}^0\right)\right]\right)\Bigg\}\Bigg\|_2=O_p(\eta_N).$

Proof (i) Recall from (A.131) that

$$h_i^I=\mathfrak{l}'\left(O_1^{(1)}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T f_{it}(0)\begin{bmatrix}v_{t,0}^0v_{t,0}' & v_{t,0}^0v_{t,1}^{0\prime}X_{1,it}\\ v_{t,1}^0v_{t,0}^{0\prime}X_{1,it} & v_{t,1}^0v_{t,1}^{0\prime}X_{1,it}^2\end{bmatrix}\right)^{-1}\cdot$$
$$\left[\frac{1}{T}\sum_{t=1}^T[\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\}]\begin{bmatrix}v_{t,0}^0\\ v_{t,1}^0X_{1,it}\end{bmatrix}+\frac{1}{T}\sum_{t=1}^T\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\{\varepsilon_{it}\leq\dot{\Delta}_{t,v}'\Psi_{it}^0\}\right]\begin{bmatrix}v_{t,0}^0\\ v_{t,1}^0X_{1,it}\end{bmatrix}\right].$$

By Bernstein's inequality in Lemma A.12, we can show that

$$\max_{i\in I_3}\Bigg\|\frac{1}{T}\sum_{t=1}^T\Bigg\{f_{it}(0)\begin{bmatrix}v_{t,0}^0v_{t,0}' & v_{t,0}^0v_{t,1}^{0\prime}X_{1,it}\\ v_{t,1}^0v_{t,0}^{0\prime}X_{1,it} & v_{t,1}^0v_{t,1}^{0\prime}X_{1,it}^2\end{bmatrix}$$
$$-\mathbb{E}\left(f_{it}(0)\begin{bmatrix}v_{t,0}^0v_{t,0}' & v_{t,0}^0v_{t,1}^{0\prime}X_{1,it}\\ v_{t,1}^0v_{t,0}^{0\prime}X_{1,it} & v_{t,1}^0v_{t,1}^{0\prime}X_{1,it}^2\end{bmatrix}\middle|\mathscr{D}\right)\Bigg\}\Bigg\|_F=O_p\left(\sqrt{\frac{\log(N\vee T)}{T}}\xi_N^2\right),\text{ and}$$
$$\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t=1}^T[\tau-\mathbf{1}\{\varepsilon_{it}\leq 0\}]\begin{bmatrix}v_{t,0}^0\\ v_{t,1}^0X_{1,it}\end{bmatrix}\right\|_2=O_p\left(\sqrt{\frac{\log(N\vee T)}{T}}\xi_N\right).$$

Besides, we observe that

$$\max_{i\in I_3,t\in[T]}\left\|Var\left(\mathbb{J}_{it}\left(\dot{\Delta}_{t,v}\right)\middle|\mathscr{D}^{I_1\cup I_2}\right)\right\|_F$$
$$\leq\max_{i\in I_3,t\in[T]}\left\|\mathbb{E}\left(\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\{\varepsilon_{it}\leq\dot{\Delta}_{t,v}'\Psi_{it}^0\}\right]\begin{bmatrix}v_{t,0}^0v_{t,0}' & v_{t,0}^0v_{t,1}^{0\prime}X_{1,it}\\ v_{t,1}^0v_{t,0}^{0\prime}X_{1,it} & v_{t,1}^0v_{t,1}^{0\prime}X_{1,it}^2\end{bmatrix}\middle|\mathscr{D}^{I_1\cup I_2}\right)\right\|_F$$

$$= \max_{i\in I_3, t\in[T]} \left\| \mathbb{E}\left(\left[F_{it}(0) - F_{it}\left(\dot{\Delta}_{t,v}'\Psi_{it}^0\right)\right] \begin{bmatrix} v_{t,0}^0 v_{t,0}' & v_{t,0}^0 v_{t,1}^{0\prime} X_{1,it} \\ v_{t,1}^0 v_{t,0}^{0\prime} X_{1,it} & v_{t,1}^0 v_{t,1}^{0\prime} X_{1,it}^2 \end{bmatrix} \middle| \mathscr{D}^{I_1\cup I_2} \right) \right\|_F$$
$$= \max_{t\in[T]} \left\|\dot{\Delta}_{t,v}\right\|_2 \max_{i\in I_3, t\in[T]} \left\| \mathbb{E}\left(f_{it}(\tilde{s}_{it})\Psi_{it}^0 \begin{bmatrix} v_{t,0}^0 v_{t,0}' & v_{t,0}^0 v_{t,1}^{0\prime} X_{1,it} \\ v_{t,1}^0 v_{t,0}^{0\prime} X_{1,it} & v_{t,1}^0 v_{t,1}^{0\prime} X_{1,it}^2 \end{bmatrix} \middle| \mathscr{D}^{I_1\cup I_2} \right) \right\|_F$$
$$= O_p(\eta_N), \tag{A.144}$$

where the first equality holds by law of iterated expectations, the second equality is by mean-value theorem for $|\tilde{s}_{it}|$ lies between 0 and $\left|\dot{\Delta}_{t,v}'\Psi_{it}^0\right|$ and the last line is by Theorem 2.2(ii) and Assumption 2.1(iv). Similarly, for any $\vartheta > 0$, we also have

$$\max_{i\in I_3, t\in[T]} \sum_{s=t+1}^{T} \left\| Cov\left(\mathbb{J}_{it}\left(\dot{\Delta}_{t,v}\right), \mathbb{J}_{is}\left(\dot{\Delta}_{t,v}\right)' \middle| \mathscr{D}^{I_1\cup I_2} \right) \right\|_F = O_p\left(\eta_N^{\frac{2}{2+\vartheta}}\right).$$

By similar arguments as used in (A.12) and (A.13), we have

$$\max_{i\in I_3} \left\| \frac{1}{T}\sum_{t=1}^{T} \left\{ \mathbb{J}_{it}\left(\dot{\Delta}_{t,v}\right) - \mathbb{E}\left(\mathbb{J}_{it}\left(\dot{\Delta}_{t,v}\right) \middle| \mathscr{D}^{I_1\cup I_2} \right) \right\} \right\|_2 = o_p\left((N\vee T)^{-1/2}\right). \tag{A.145}$$

Together with the fact that $\max_{i\in I_3}\left\|\mathbb{J}_i\left(\dot{\Delta}_{t,v}\right)\right\|_2 = O_p(\eta_N)$, uniformly over $t\in[T]$, we have

$$h_i^I = \mathfrak{l}'\left(O_1^{(1)}\right)^{-1}\left(h_i^{I,1}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T}\left[\tau - \mathbf{1}\{\varepsilon_{it}\le 0\}\right] \begin{bmatrix} v_{t,0}^0 \\ v_{t,1}^0 X_{1,it} \end{bmatrix}$$
$$+ \mathfrak{l}'\left(O_1^{(1)}\right)^{-1}\left(h_i^{I,1}\right)^{-1} h_i^{I,2} + o_p\left((N\vee T)^{-1/2}\right).$$

By Assumption 2.1(iv) and Lemma A.13, we have $\max_{i\in I_3}\left\|h_i^{I,1}\right\|_F = O(1)$ *a.s.*. Like (A.144), we can show that $\max_{i\in I_3}\left\|h_i^{I,2}\right\|_2 = O_p(\eta_N)$.

Similarly as (A.141), uniformly in $i\in I_3$ and $t\in[T]$ we have

$$\rho_{it}\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \left(O_{u,0}^{(1)}\right)^{\prime -1} v_{t,0}^0, \left(O_{u,1}^{(1)}\right)^{\prime -1} v_{t,1}^0\right)$$
$$= \hat{u}_{i,0}^{(3,1)\prime}\left(O_{u,0}^{(1)}\right)^{\prime -1} v_{t,0}^0 - u_{i,0}^{0\prime} v_{t,0}^0 + \hat{\mu}_{1,it}\hat{u}_{i,1}^{(3,1)\prime}\dot{v}_{t,1}^{(1)} - \mu_{1,it}u_{i,1}^{0\prime}v_{t,1}^0$$
$$+ \hat{e}_{1,it}\hat{u}_{i,1}^{(3,1)\prime}\left(O_{u,1}^{(1)}\right)^{\prime -1} v_{t,1}^0 - e_{1,it}u_{i,1}^{0\prime}v_{t,1}^0$$
$$= v_{t,0}^{0\prime} O_{u,0}^{-1}\left(\hat{u}_{i,0}^{(3,1)} - O_{u,0}^{(1)}u_{i,0}^0\right) + X_{1,it}v_{t,1}^{0\prime}O_{u,1}^{(1)-1}\left(\hat{u}_{i,1}^{(3,1)} - O_{u,1}^{(1)}u_{i,1}^0\right)$$
$$+ \mu_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)} - O_1^{(1)}v_{t,1}^0\right)$$
$$+ \mu_{1,it}u_{i,1}^{0\prime}O_{u,1}^{(1)\prime}\left(O_{u,1}^{(1)} - O_1^{(1)}\right)v_{t,1}^0 + o_p\left((N\vee T)^{-1/2}\right)$$
$$= v_{t,0}^{0\prime}\left(O_0^{(1)}\right)^{-1}\left(\hat{u}_{i,0}^{(3,1)} - O_{u,0}^{(1)}u_{i,0}^0\right) + X_{1,it}v_{t,1}^{0\prime}\left(O_1^{(1)}\right)^{-1}\left(\hat{u}_{i,1}^{(3,1)} - O_{u,1}^{(1)}u_{i,1}^0\right)$$

$$+\mu_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot v_{t,1}^{(1)}-O_1^{(1)}v_{t,1}^0\right)$$
$$+\mu_{1,it}u_{i,1}^{0\prime}O_1^{(1)\prime}\left(O_{u,1}^{(1)}-O_1^{(1)}\right)v_{t,1}^0+o_p\left((N\vee T)^{-1/2}\right), \tag{A.146}$$

where the last line is by the fact that $\left\|O_{u,0}^{(1)}-O_0^{(1)}\right\|_F=O_p(\eta_N)$ and $\left\|O_{u,1}^{(1)}-O_1^{(1)}\right\|_F=O_p(\eta_N)$. Combining (A.41), (A.42) and Theorem 2.2(iii), we have

$$\rho_{it}\left(\hat u_{i,0}^{(3,1)},\hat u_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)^{\prime-1}v_{t,0}^0,\left(O_{u,1}^{(1)}\right)^{\prime-1}v_{t,1}^0\right)$$
$$=v_{t,0}^{0\prime}\hat V_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0\left(\tau-1\left\{\varepsilon_{it^*}\le 0\right\}\right)$$
$$+v_{t,0}^{0\prime}\hat V_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0 f_{it^*}(0)\left(\mu_{1,it^*}u_{i,1}^{0\prime}v_{t^*,1}^0-\hat\mu_{1,it^*}^{(1)}\dot u_{i,1}^{(1)\prime}\dot v_{t^*,1}^{(1)}\right)$$
$$+v_{t,0}^{0\prime}\hat V_{u_0,i}^{-1}\frac{1}{T}\sum_{t=1}^T f_{it^*}(0)v_{t^*,0}^0 v_{t^*,1}^{0\prime}u_{i,1}^0\left(e_{1,it^*}-\hat e_{1,it^*}\right)$$
$$+X_{1,it}v_{t,1}^{0\prime}\hat V_{u_1}^{-1}\frac{1}{T}\sum_{t^*=1}^T e_{1,it^*}v_{t^*,1}^0\left(\tau-1\left\{\varepsilon_{it^*}\le 0\right\}\right)$$
$$+\mu_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot v_{t,1}^{(1)}-O_1^{(1)}v_{t,1}^0\right)+\mu_{1,it}u_{i,1}^{0\prime}O_1^{(1)\prime}\left(O_{u,1}^{(1)}-O_1^{(1)}\right)v_{t,1}^0$$
$$+o_p\left((N\vee T)^{-1/2}\right)$$
$$=v_{t,0}^{0\prime}\hat V_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0\left(\tau-1\left\{\varepsilon_{it^*}\le 0\right\}\right)$$
$$+X_{1,it}v_{t,1}^{0\prime}\hat V_{u_1}^{-1}\frac{1}{T}\sum_{t^*=1}^T e_{1,it^*}v_{t^*,1}^0\left(\tau-1\left\{\varepsilon_{it^*}\le 0\right\}\right)$$
$$-v_{t,0}^{0\prime}\hat V_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0 f_{it^*}(0)\mu_{1,it^*}\left(\dot u_{i,1}^{(1)}-O_1^{(1)}u_{i,1}^0\right)'O_1^{(1)}v_{t^*,1}^0$$
$$-v_{t,0}^{0\prime}\hat V_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0 f_{it^*}(0)\left(\hat\mu_{1,it^*}-\mu_{1,it^*}\right)u_{i,1}^{0\prime}v_{t^*,1}^0$$
$$-v_{t,0}^{0\prime}\hat V_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0 f_{it^*}(0)\mu_{1,it^*}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot v_{t^*,1}^{(1)}-O_1^{(1)}v_{t^*,1}^0\right)$$
$$+v_{t,0}^{0\prime}\hat V_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T f_{it^*}(0)v_{t^*,0}^0 v_{t^*,1}^{0\prime}u_{i,1}^0\left(e_{1,it^*}-\hat e_{1,it^*}\right)$$
$$+\mu_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot v_{t,1}^{(1)}-O_1^{(1)}v_{t,1}^0\right)$$
$$+\mu_{1,it}u_{i,1}^{0\prime}O_1^{(1)\prime}\left(O_{u,1}^{(1)}-O_1^{(1)}\right)v_{t,1}^0+o_p\left((N\vee T)^{-1/2}\right)$$
$$=v_{t,0}^{0\prime}\hat V_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0\left(\tau-1\left\{\varepsilon_{it^*}\le 0\right\}\right)+X_{1,it}v_{t,1}^{0\prime}\hat V_{u_1}^{-1}\frac{1}{T}\sum_{t^*=1}^T e_{1,it^*}v_{t^*,1}^0\left(\tau-1\left\{\varepsilon_{it^*}\le 0\right\}\right)$$
$$-v_{t,0}^{0\prime}\hat V_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^T v_{t^*,0}^0\mathbb{E}\left[f_{it^*}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\mu_{1,it^*}v_{t^*,1}^{0\prime}h_i^I+h_{it}^{IV}+o_p\left((N\vee T)^{-1/2}\right)$$

$$
\begin{aligned}
&- v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^{T} v_{t^*,0}^{0}\left(f_{it^*}(0)-\mathbb{E}\left[f_{it^*}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\right)\mu_{1,it}\left(O_1^{(1)}u_{i,1}^{0}\right)'\left(\dot{v}_{t^*,1}^{(1)}-O_1^{(1)}v_{t^*,1}^{0}\right)\\
&- v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^{T} v_{t^*,0}^{0}\left(f_{it^*}(0)-\mathbb{E}\left[f_{it^*}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\right)\mu_{1,it^*}v_{t^*,1}^{0\prime}h_i^I\\
&:= h_{it}^{III}+h_{it}^{IV}+\mathscr{R}_{\rho,it} \qquad\qquad (A.147)
\end{aligned}
$$

such that $\max_{i\in I_3,t\in[T]}\left|\mathscr{R}_{\rho,it}\right| = O_p(\eta_N^2)$, where the first equality is by inserting the linear expansion for $\hat{u}_{i,0}^{(3,1)} - O_{u,0}^{(1)}u_{i,0}^{0}$ in (A.41) and $\hat{u}_{i,1}^{(3,1)} - O_{u,1}^{(1)}u_{i,1}^{0}$ in (A.42), the second equality is by inserting the linear expansion for $\dot{u}_{i,0}^{(1)} - O_{u,0}^{(1)}u_{i,0}^{0}$, the third equality is by the fact that $\hat{\mu}_{1,it}-\mu_{1,it} = e_{1,it}-\hat{e}_{1,it}$ which leads to the cancelling of the fourth and sixth terms in the second equality, the last equality is by the definition in (A.143) and

$$
\begin{aligned}
h_{it}^{IV} &= -v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^{T} v_{t^*,0}^{0}\mathbb{E}\left[f_{it^*}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\mu_{1,it^*}\left(O_1^{(1)}u_{i,1}^{0}\right)'\left(\dot{v}_{t^*,1}^{(1)}-O_1^{(1)}v_{t^*,1}^{0}\right)\\
&\quad+\mu_{1,it}\left(O_1^{(1)}u_{i,1}^{0}\right)'\left(\dot{v}_{t,1}^{(1)}-O_1^{(1)}v_{t,1}^{0}\right)+\mu_{1,it}u_{i,1}^{0\prime}O_1^{(1)\prime}\left(O_{u,1}^{(1)}-O_1^{(1)}\right)v_{t,1}^{0}\\
&= \mathscr{R}_{h,it}^{3}+\mu_{1,it}\mathscr{R}_{h,it}^{4}
\end{aligned}
$$

with $\max_{i\in I_3,t\in[T]}\left|\mathscr{R}_{h,it}^{a}\right| = O_p(\eta_N)$ for $a\in\{3,4\}$, and the last equality holds by the fact that

$$
\begin{aligned}
&\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t^*=1}^{T} v_{t^*,0}^{0}\left(f_{it^*}(0)-\mathbb{E}\left[f_{it^*}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\right)\mu_{1,it^*}\left(O_1^{(1)}u_{i,1}^{0}\right)'\left(\dot{v}_{t^*,1}^{(1)}-O_1^{(1)}v_{tv,1}^{0}\right)\right\|_2\\
&= o_p\left((N\vee T)^{-1/2}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\max_{i\in I_3}\left\|\frac{1}{T}\sum_{t^*=1}^{T} v_{t,0}^{0}\left(f_{it^*}(0)-\mathbb{E}\left[f_{it^*}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\right)\mu_{1,it^*}v_{t^*,1}^{0\prime}h_i^I\right\|_2\\
&\leq \max_{i\in I_3}\left|\frac{1}{T}\sum_{t^*=1}^{T} v_{t^*,0}^{0}\left(f_{it^*}(0)-\mathbb{E}\left[f_{it^*}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\right)\mu_{1,it^*}v_{t^*,1}^{0\prime}\right|\max_{i\in I_3}||h_i^I||_2\\
&= o_p\left((N\vee T)^{-1/2}\right)
\end{aligned}
$$

by conditional Bernstein's inequality given $\mathscr{D}^{I_1\cup I_2}$ and similar arguments as used in (A.145).

We notice that

$$
\frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0}\left\{\left[\mathbf{1}\{\varepsilon_{it}\leq 0\}-\mathbf{1}\left\{\varepsilon_{it}\leq \rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\left(O_{u,0}^{(1)}\right)'^{-1}v_{t,0}^{0},\left(O_{u,1}^{(1)}\right)'^{-1}v_{t,1}^{0}\right)\right\}\right]\right.
$$

$$
\begin{aligned}
&-\left(F_{it}(0)-F_{it}\left[\rho_{it}\left(\hat{u}^{(3,1)}_{i,0},\hat{u}^{(3,1)}_{i,1},\left(O^{(1)}_{u,0}\right)^{\prime-1}v^0_{t,0},\left(O^{(1)}_{u,1}\right)^{\prime-1}v^0_{t,1}\right)\right]\right)\Bigg\}\\
&=\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u^0_{i,1}\left\{\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq h^{III}_{it}+h^{IV}_{it}\right\}\right]-\left(F_{it}(0)-F_{it}\left[h^{III}_{it}+h^{IV}_{it}\right]\right)\right\}\\
&+\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u^0_{i,1}\Bigg\{\left[\mathbf{1}\left\{\varepsilon_{it}\leq h^{III}_{it}+h^{IV}_{it}\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq h^{III}_{it}+h^{IV}_{it}+\mathscr{R}_{\rho,it}\right\}\right]\\
&-\left[F_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)-F_{it}\left(h^{III}_{it}+h^{IV}_{it}+\mathscr{R}_{\rho,it}\right)\right]\Bigg\}\\
&=\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u^0_{i,1}\left\{\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq h^{III}_{it}+h^{IV}_{it}\right\}\right]-\left[\mathscr{F}_{it}(0)-\mathscr{F}_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)\right]\right\}\\
&+\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u^0_{i,1}\Bigg\{\left[\mathbf{1}\left\{\varepsilon_{it}\leq h^{III}_{it}+h^{IV}_{it}\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq h^{III}_{it}+h^{IV}_{it}+\mathscr{R}_{\rho,it}\right\}\right]\\
&-\left[\mathscr{F}_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)-\mathscr{F}_{it}\left(h^{III}_{it}+h^{IV}_{it}+\mathscr{R}_{\rho,it}\right)\right]\Bigg\}\\
&+\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u^0_{i,1}\left\{\left[\mathscr{F}_{it}(0)-\mathscr{F}_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)\right]-\left[F_{it}(0)-F_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)\right]\right\}\\
&+\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u^0_{i,1}\left[\mathscr{F}_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)-\mathscr{F}_{it}\left(h^{III}_{it}+h^{IV}_{it}+\mathscr{R}_{\rho,it}\right)\right]\\
&-\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u^0_{i,1}\left[F_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)-F_{it}\left(h^{III}_{it}+h^{IV}_{it}+\mathscr{R}_{\rho,it}\right)\right]\\
&:=\mathbb{I}^{I}_{6,t}+\mathbb{I}^{II}_{6,t}-\mathbb{I}^{III}_{6,t}+\mathbb{I}^{IV}_{6,t}+o_p\left((N\vee T)^{-\frac{1}{2}}\right)
\end{aligned}
\tag{A.148}
$$

such that $\mathbb{I}^{I}_{6,t}:=\frac{1}{N_3}\sum_{i\in I_3}\mathbb{I}^{I}_{6,it}$, $\mathbb{I}^{II}_{6,t}:=\frac{1}{N_3}\sum_{i\in I_3}\mathbb{I}^{II}_{6,it}$, $\mathbb{I}^{III}_{6,t}:=\frac{1}{N_3}\sum_{i\in I_3}\mathbb{I}^{III}_{6,it}$, $\mathbb{I}^{IV}_{6,t}:=\frac{1}{N_3}\sum_{i\in I_3}\mathbb{I}^{IV}_{6,it}$, with

$$
\begin{aligned}
\mathbb{I}^{I}_{6,it}&=e_{1,it}u^0_{i,1}\left\{\left[\mathbf{1}\left\{\varepsilon_{it}\leq 0\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq h^{III}_{it}+h^{IV}_{it}\right\}\right]-\left[\mathscr{F}_{it}(0)-\mathscr{F}_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)\right]\right\},\\
\mathbb{I}^{II}_{6,it}&=e_{1,it}u^0_{i,1}\Bigg\{\left[\mathbf{1}\left\{\varepsilon_{it}\leq h^{III}_{it}+h^{IV}_{it}\right\}-\mathbf{1}\left\{\varepsilon_{it}\leq h^{III}_{it}+h^{IV}_{it}+\mathscr{R}_{\rho,it}\right\}\right]\\
&\quad-\left[\mathscr{F}_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)-\mathscr{F}_{it}\left(h^{III}_{it}+h^{IV}_{it}+\mathscr{R}_{\rho,it}\right)\right]\Bigg\},
\end{aligned}
\tag{A.149}
$$

$$
\begin{aligned}
\mathbb{I}^{III}_{6,it}&=e_{1,it}u^0_{i,1}\left[F_{it}(0)-F_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)\right],\\
\mathbb{I}^{IV}_{6,it}&=e_{1,it}u^0_{i,1}\left[\mathscr{F}_{it}(0)-\mathscr{F}_{it}\left(h^{III}_{it}+h^{IV}_{it}\right)\right].
\end{aligned}
$$

The last line in (A.148) is due to the fact that the last two terms in the second equality is $o_p\left((N\vee T)^{-\frac{1}{2}}\right)$ by mean-value theorem, Assumption (2.1)(iv), and Assumption (2.1)(viii) and the union bound for $\mathscr{R}_{\rho,it}$.

For $\mathbb{I}^{I}_{6,t}$ and $\mathbb{I}^{II}_{6,t}$, conditional on $\mathscr{D}^{I_1\cup I_2}_{e}$ and h^{III}_{it}, the randomness is from ε_{it}, and we observe that $\mathbb{I}_{6,it}^{I}$ and $\mathbb{I}_{6,it}^{II}$ are independent over i by conditioning on $\mathscr{D}_e^{I_1\cup I_2}$ and h_{it}^{III}. Therefore, we obtain that $\max_{t\in[T]}\left\|\mathbb{I}_{6,t}^{I}\right\|_2 = o_p\left((N\vee T)^{-\frac{1}{2}}\right)$ and $\max_{t\in[T]}\left\|\mathbb{I}_{6,t}^{II}\right\|_2 = o_p\left((N\vee T)^{-\frac{1}{2}}\right)$ by conditional Bernstein's inequality for independent sequence given $\mathscr{D}_e^{I_1\cup I_2}$ and h_{it}^{III}.

For $\mathbb{I}_{6,t}^{III}$, we notice that

$$\begin{aligned}
\mathbb{I}_{6,t}^{III} &= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0\left[F_{it}(0)-F_{it}\left(h_{it}^{III}+h_{it}^{IV}\right)\right] \\
&= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(\dot{s}_{it})\left(h_{it}^{III}+h_{it}^{IV}\right) \\
&= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)\left(h_{it}^{III}+h_{it}^{IV}\right) + \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0\left[f_{it}(\dot{s}_{it})-f_{it}(0)\right]\left(h_{it}^{III}+h_{it}^{IV}\right) \\
&= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)\left(h_{it}^{III}+h_{it}^{IV}\right) + o_p\left((N\vee T)^{-\frac{1}{2}}\right) \quad \text{uniformly}, \qquad \text{(A.150)}
\end{aligned}$$

where the second line is by mean-value theorem with $|\dot{s}_{it}|$ lies between 0 and $\left|h_{it}^{III}+h_{it}^{IV}\right|$ and the last line is by Assumption 2.1(viii). By inserting h_{it}^{III} and h_{it}^{IV}, we have

$$\begin{aligned}
&\frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)\left(h_{it}^{III}+h_{it}^{IV}\right) \\
&= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^{T} v_{t^*,0}^0\left(\tau-1\left\{\varepsilon_{it^*}\le 0\right\}\right) \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)X_{1,it}v_{t,1}^{0\prime}\hat{V}_{u_1}^{-1}\frac{1}{T}\sum_{t^*=1}^{T} e_{1,it^*}v_{t^*,1}^0\left(\tau-1\left\{\varepsilon_{it^*}\le 0\right\}\right) \\
&- \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)\left(v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^{T} v_{t^*,0}^0 f_{it^*}(0)\mu_{1,it^*}v_{t^*,1}^{0\prime}\right)\mathfrak{l}'\left(O_1^{(1)}\right)^{-1}\left(h_i^{I,1}\right)^{-1}\cdot \\
&\frac{1}{T}\sum_{t^*=1}^{T}\left[\tau-\mathbf{1}\left\{\varepsilon_{it^*}\le 0\right\}\right]\begin{bmatrix} v_{t^*,0}^0 \\ v_{t^*,1}^0 X_{1,it^*}\end{bmatrix} \\
&- \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)\left(v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^{T} v_{t^*,0}^0\mathbb{E}\left[f_{it^*}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\mu_{1,it^*}v_{t^*,1}^{0\prime}\right)\cdot \\
&\mathfrak{l}'\left(O_1^{(1)}\right)^{-1}\left(h_i^{I,1}\right)^{-1}h_i^{I,2} \\
&- \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t^*=1}^{T} v_{t^*,0}^0\mathbb{E}\left[f_{it^*}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\mu_{1,it^*}\cdot \\
&\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t^*,1}^{(1)}-O_1^{(1)}v_{t^*,1}^0\right) \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)\mu_{1,it}\left(O_1^{(1)}u_{i,1}^0\right)'\left(\dot{v}_{t,1}^{(1)}-O_1^{(1)}v_{t,1}^0\right) \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^0 f_{it}(0)\mu_{1,it}u_{i,1}^{0\prime}O_1^{(1)\prime}\left(O_{u,1}^{(1)}-O_1^{(1)}\right)v_{t,1}^0
\end{aligned}$$

$$:= \sum_{m\in[7]} \mathbb{I}_{6,t}^{III,m}. \tag{A.151}$$

$\mathbb{I}_{6,t}^{III,1}$, $\mathbb{I}_{6,t}^{III,2}$, $\mathbb{I}_{6,t}^{III,3}$ can be analyzed in the same manner, and we take $\mathbb{I}_{6,t}^{III,1}$ for instance. Noticed that

$$\mathbb{I}_{6,t}^{III,1} = \frac{1}{N_3 T} \sum_{i\in I_3} \sum_{t^*=1}^{T} e_{1,it} u_{i,1}^{0} f_{it}(0) v_{t,0}^{0\prime} \hat{V}_{u_0,i}^{-1} v_{t^*,0}^{0} \left(\tau - 1\left\{\varepsilon_{it^*} \leq 0\right\}\right),$$

it is clear that conditioning on $\mathscr{D}_e$, $\mathbb{I}_{6,t}^{III,1}$ is mean zero by Assumption 2.1(ii) and the randomness is from ε_{it} which is strong mixing. With the similar arguments as the second term in (A.124), we obtain that $\max_{t\in[T]} \left\|\mathbb{I}_{6,t}^{III,1}\right\|_2 = O_p\left(\sqrt{\frac{\log(N\vee T)}{NT}}\xi_N\right)$, and by Assumption 2.1(ix), it follows that

$$\max_{t\in[T]} \left\|\mathbb{I}_{6,t}^{III,1}\right\|_2 = o_p\left((N\vee T)^{-1/2}\right).$$

We can also analyze $\mathbb{I}_{6,t}^{III,4}$ and $\mathbb{I}_{6,t}^{III,5}$ in the same manner. Take $\mathbb{I}_{6,t}^{III,4}$ for instance. Note that

$$\begin{aligned}
-\mathbb{I}_{6,t}^{III,4} =& \frac{1}{N_3 T} \sum_{i\in I_3} \sum_{t^*=1}^{T} e_{1,it} u_{i,1}^{0} f_{it}(0) \left(v_{t,0}^{0\prime} \hat{V}_{u_0,i}^{-1} v_{t,0}^{0} \mathbb{E}\left[f_{it^*}(0) \middle| \mathscr{D}^{I_1\cup I_2}\right] \mu_{1,it^*} v_{t^*,1}^{0\prime} \right) \cdot \\
& \mathfrak{l}' \left(O_1^{(1)}\right)^{-1} \left(h_i^{I,1}\right)^{-1} h_i^{I,2}
\end{aligned}$$

which is mean zero by conditioning on $\mathscr{D}^{I_1\cup I_2}$ owing to Assumption 2.9(i) and the fact that $h_i^{I,1}$ and $h_i^{I,2}$ are fixed given $\mathscr{D}^{I_1\cup I_2}$. Then by similar arguments for $\mathbb{I}_{6,t}^{III,1}$ above, we have

$$\max_{t\in[T]} \left\|\mathbb{I}_{6,t}^{III,4}\right\| = o_p\left((N\vee T)^{-1/2}\right). \tag{A.152}$$

For $\mathbb{I}_{6,t}^{III,6}$ and $\mathbb{I}_{6,t}^{III,7}$, we have

$$\begin{aligned}
\max_{t\in[T]} \left\|\mathbb{I}_{6,t}^{III,6}\right\|_2 &\leq \max_{t\in[T]} \left\| \frac{1}{N_3} \sum_{i\in I_3} e_{1,it} u_{i,1}^{0} f_{it}(0) \mu_{1,it} u_{i,1}^{0\prime} \right\|_2 \max_{t\in[T]} \left\| \dot{v}_{t,1}^{(1)} - O_1^{(1)} v_{t,1}^{0} \right\|_2 \\
&\leq \max_{t\in[T]} \left\| \frac{1}{N_3} \sum_{i\in I_3} e_{1,it} u_{i,1}^{0} f_{it}(0) \mu_{1,it} u_{i,1}^{0\prime} \right\|_2 O_p(\eta_N) \\
&\leq O_p\left(\sqrt{\frac{\log(N\vee T)}{N}}\xi_N\right) O_p(\eta_N) = o_p\left((N\vee T)^{-1/2}\right), \quad \text{(A.153)} \\
\max_{t\in[T]} \left\|\mathbb{I}_{6,t}^{III,7}\right\|_2 &\leq \max_{t\in[T]} \left\| \frac{1}{N_3} \sum_{i\in I_3} e_{1,it} u_{i,1}^{0} f_{it}(0) \mu_{1,it} u_{i,1}^{0\prime} \right\|_2 \max_{t\in[T]} \left\| v_{t,1}^{0} \right\|_2 \left\| O_{u,1}^{(1)} - O_1^{(1)} \right\|_F \\
&\leq O_p\left(\sqrt{\frac{\log(N\vee T)}{N}}\xi_N\right) O_p(\eta_N) = o_p\left((N\vee T)^{-1/2}\right), \quad \text{(A.154)}
\end{aligned}$$

where the second inequality is by Theorem 2.2(ii) and the third inequality is by Hoeffding's inequality conditional on $\mathscr{D}$ and the last line combines Lemma A.13(i) and the fact that $\left\|O_{u,1}^{(1)}-O_{1}^{(1)}\right\|_F = O_p(\eta_N)$. Combining (A.150)-(A.153), we have $\max_{t\in[T]}\left\|\mathbb{I}_{6,it}^{III}\right\|_2 = o_p\left((N\vee T)^{-1/2}\right)$.

Last, we analyze $\mathbb{I}_{6,t}^{IV}$. Like (A.150), we have

$$
\begin{aligned}
\mathbb{I}_{6,t}^{IV} &= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0}\left[\mathscr{F}_{it}(0)-\mathscr{F}_{it}\left(h_{it}^{III}+h_{it}^{IV}\right)\right] \\
&= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} f_{it|h_{it}^{III}}(\dot{s}_{it})\left(h_{it}^{III}+h_{it}^{IV}\right) \\
&= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} f_{it|h_{it}^{III}}(0)\left(h_{it}^{III}+h_{it}^{IV}\right) \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0}\left[f_{it|h_{it}^{III}}(\dot{s}_{it})-f_{it|h_{it}^{III}}(0)\right]\left(h_{it}^{III}+h_{it}^{IV}\right) \\
&= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} f_{it|h_{it}^{III}}(0)h_{it}^{III} + \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} f_{it|h_{it}^{III}}(0)h_{it}^{IV} + o_p\left((N\vee T)^{-\frac{1}{2}}\right) \\
&= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} f_{it|h_{it}^{III}}(0)\left[h_{it}^{III}-\mathbb{E}\left(h_{it}^{III}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)\right] \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} f_{it|h_{it}^{III}}(0)\mathbb{E}\left(h_{it}^{III}\middle|\mathscr{D}_e^{I_1\cup I_2}\right) \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} h_{it}^{IV}\left[f_{it|h_{it}^{III}}(0)-\mathbb{E}\left(f_{it|h_{it}^{III}}(0)|\mathscr{D}_e^{I_1\cup I_2}\right)\right] \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} h_{it}^{IV}\mathbb{E}\left(f_{it|h_{it}^{III}}(0)|\mathscr{D}_e^{I_1\cup I_2}\right) + o_p\left((N\vee T)^{-\frac{1}{2}}\right) \\
&= \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0}\left[f_{it|h_{it}^{III}}(0)-f_{it|h_{it}^{III}=0}(0)\right]\left[h_{it}^{III}-\mathbb{E}\left(h_{it}^{III}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)\right] \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} f_{it|h_{it}^{III}=0}(0)\left[h_{it}^{III}-\mathbb{E}\left(h_{it}^{III}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)\right] \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0}\mathbb{E}\left(h_{it}^{III}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)\left[f_{it|h_{it}^{III}}(0)-\mathbb{E}\left(f_{it|h_{it}^{III}}(0)|\mathscr{D}_e^{I_1\cup I_2}\right)\right] \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0}\mathbb{E}\left(h_{it}^{III}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)\mathbb{E}\left(f_{it|h_{it}^{III}}(0)|\mathscr{D}_e^{I_1\cup I_2}\right) \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} h_{it}^{IV}\left[f_{it|h_{it}^{III}}(0)-\mathbb{E}\left(f_{it|h_{it}^{III}}(0)|\mathscr{D}_e^{I_1\cup I_2}\right)\right] \\
&+ \frac{1}{N_3}\sum_{i\in I_3} e_{1,it}u_{i,1}^{0} h_{it}^{IV}\mathbb{E}\left(f_{it|h_{it}^{III}}(0)|\mathscr{D}_e^{I_1\cup I_2}\right) + o_p\left((N\vee T)^{-\frac{1}{2}}\right) \\
&:= \sum_{m\in[6]}\mathbb{I}_{6,t}^{IV,m} + o_p\left((N\vee T)^{-\frac{1}{2}}\right),
\end{aligned}
\tag{A.155}
$$

where all the $o_p\left((N\vee T)^{-\frac{1}{2}}\right)$ terms hold uniformly over $t\in[T]$ and $o_p\left((N\vee T)^{-\frac{1}{2}}\right)$ term in the fourth equality is by mean-value theorem, Assumption A.1(i) and the fact that $|\dot{s}_{it}|$ lies between 0 and $\left|h_{it}^{III}+h_{it}^{IV}\right|$.

For $\mathbb{I}_{6,t}^{IV,1}$, with Assumption A.1(ii), (A.143) and the fact that

$$
\begin{aligned}
&\max_{i\in I_3,t\in[T]}\left|\mathbb{E}\left(h_{it}^{III}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)\right| \\
&=\max_{i\in I_3,t\in[T]}\left|\mathbb{E}\left[\left(v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t=1}^{T}v_{t,0}^{0}\mathbb{E}\left[f_{it}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\mu_{1,it}v_{t,1}^{0\prime}\right)\mathfrak{l}'\left(O_1^{(1)}\right)^{-1}\left(h_i^{I,1}\right)^{-1}h_i^{I,2}\middle|\mathscr{D}_e^{I_1\cup I_2}\right]\right| \\
&=\max_{i\in I_3,t\in[T]}\left|\left(v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t=1}^{T}v_{t,0}^{0}\mathbb{E}\left[f_{it}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\mu_{1,it}v_{t,1}^{0\prime}\right)\mathfrak{l}'\left(O_1^{(1)}\right)^{-1}\left(h_i^{I,1}\right)^{-1}h_i^{I,2}\right| \\
&=O_p(\eta_N),
\end{aligned}
$$

we have

$$
\begin{aligned}
\max_{t\in[T]}\left\|\mathbb{I}_{6,t}^{IV,1}\right\|_2 &\lesssim \max_{t\in[T]}\frac{1}{N_3}\sum_{i\in I_3}\left\|e_{1,it}u_{i,1}^0\right\|_2\left|h_{it}^{III}\right|\left(\left|h_{it}^{III}\right|+\left|\mathbb{E}\left(h_{it}^{III}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)\right|\right) \\
&=O_p(\eta_N^2). \qquad\qquad (A.156)
\end{aligned}
$$

For $\mathbb{I}_{6,t}^{IV,2}$, $\mathbb{I}_{6,t}^{IV,3}$ and $\mathbb{I}_{6,t}^{IV,5}$, conditioning on $\mathscr{D}_e^{I_1\cup I_2}$, the randomness is only from h_{it}^{III}, which is independent across i, and $\mathbb{I}_{6,t}^{IV,2}$, $\mathbb{I}_{6,t}^{IV,3}$ and $\mathbb{I}_{6,t}^{IV,5}$ are zero mean by conditioning on $\mathscr{D}_e^{I_1\cup I_2}$. Similar to the arguments for $\mathbb{I}_{6,t}^{I}$ and $\mathbb{I}_{6,t}^{II}$ in (A.149), we have

$$
\max_{t\in[T]}\left\|\mathbb{I}_{6,t}^{IV,m}\right\|_2=o_p\left((N\vee T)^{-\frac{1}{2}}\right), m\in\{2,3,5\}. \qquad (A.157)
$$

For $\mathbb{I}_{6,t}^{IV,4}$, by inserting $\mathbb{E}\left(h_{it}^{III}\middle|\mathscr{D}_e^{I_1\cup I_2}\right)$ and the fact that $\mathbb{E}\left(f_{it|h_{it}^{III}}(0)|\mathscr{D}_e^{I_1\cup I_2}\right)=f_{it}(0)$, it yields

$$
\begin{aligned}
\mathbb{I}_{6,t}^{IV,4}&=-\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}f_{it}(0)u_{i,1}^0\left(v_{t,0}^{0\prime}\hat{V}_{u_0,i}^{-1}\frac{1}{T}\sum_{t=1}^{T}v_{t,0}^{0}\mathbb{E}\left[f_{it}(0)\middle|\mathscr{D}^{I_1\cup I_2}\right]\mu_{1,it}v_{t,1}^{0\prime}\right)\cdot \\
&\quad \mathfrak{l}'\left(O_1^{(1)}\right)^{-1}\left(h_i^{I,1}\right)^{-1}h_i^{I,2} \\
&=\mathbb{I}_{6,t}^{III,4}=o_p\left((N\vee T)^{-1/2}\right)\quad\text{uniformly}, \qquad (A.158)
\end{aligned}
$$

where the last equality is by (A.152).

For $\mathbb{I}_{6,t}^{IV,6}$, we notice that

$$
\mathbb{I}_{6,t}^{IV,6}=\frac{1}{N_3}\sum_{i\in I_3}e_{1,it}u_{i,1}^0h_{it}^{IV}\mathbb{E}\left(f_{it|h_{it}^{III}}(0)|\mathscr{D}_e^{I_1\cup I_2}\right)
$$

$$= \frac{1}{N_3} \sum_{i \in I_3} e_{1,it} f_{it}(0) u_{i,1}^0 h_{it}^{IV} = \sum_{m \in \{5,6,7\}} \mathbb{I}_{6,t}^{III,m} = o_p\left((N \vee T)^{-1/2}\right) \quad \text{uniformly.} \tag{A.159}$$

Combining (A.155)-(A.159) yields $\max_{t \in [T]} \left\| \mathbb{I}_{6,it}^{III} \right\|_2 = o_p\left((N \vee T)^{-1/2}\right)$, which leads to the desired result in statement (i).

(ii) As in (A.148), we have

$$\frac{1}{N_3} \sum_{i \in I_3} O_{u,0}^{(1)} u_{i,0}^0 \Bigg\{ \left[\mathbf{1}\{\varepsilon_{it} \le 0\} - \mathbf{1}\left\{ \varepsilon_{it} \le \rho_{it}\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \left(O_{u,0}^{(1)}\right)^{\prime -1} v_{t,0}^0, \left(O_{u,1}^{(1)}\right)^{\prime -1} v_{t,1}^0 \right) \right\} \right]$$
$$- \left(F_{it}(0) - F_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)}, \hat{u}_{i,1}^{(3,1)}, \left(O_{u,0}^{(1)}\right)^{\prime -1} v_{t,0}^0, \left(O_{u,1}^{(1)}\right)^{\prime -1} v_{t,1}^0 \right) \right] \right) \Bigg\}$$
$$= \frac{1}{N_3} \sum_{i \in I_3} u_{i,1}^0 \left\{ \left[\mathbf{1}\{\varepsilon_{it} \le 0\} - \mathbf{1}\left\{ \varepsilon_{it} \le h_{it}^{III} + h_{it}^{IV} \right\} \right] - \left[\mathscr{F}_{it}(0) - \mathscr{F}_{it}\left(h_{it}^{III} + h_{it}^{IV} \right) \right] \right\}$$
$$+ \frac{1}{N_3} \sum_{i \in I_3} u_{i,1}^0 \Bigg\{ \left[\mathbf{1}\left\{ \varepsilon_{it} \le h_{it}^{III} + h_{it}^{IV} \right\} - \mathbf{1}\left\{ \varepsilon_{it} \le h_{it}^{III} + h_{it}^{IV} + \mathscr{R}_{\rho,it} \right\} \right]$$
$$- \left[\mathscr{F}_{it}\left(h_{it}^{III} + h_{it}^{IV} \right) - \mathscr{F}_{it}\left(h_{it}^{III} + h_{it}^{IV} + \mathscr{R}_{\rho,it} \right) \right] \Bigg\}$$
$$+ \frac{1}{N_3} \sum_{i \in I_3} u_{i,1}^0 \left\{ \left[\mathscr{F}_{it}(0) - \mathscr{F}_{it}\left(h_{it}^{III} + h_{it}^{IV} \right) \right] - \left[F_{it}(0) - F_{it}\left(h_{it}^{III} + h_{it}^{IV} \right) \right] \right\}$$
$$+ \frac{1}{N_3} \sum_{i \in I_3} u_{i,1}^0 \left[\mathscr{F}_{it}\left(h_{it}^{III} + h_{it}^{IV} \right) - \mathscr{F}_{it}\left(h_{it}^{III} + h_{it}^{IV} + \mathscr{R}_{\rho,it} \right) \right]$$
$$- \frac{1}{N_3} \sum_{i \in I_3} u_{i,1}^0 \left[F_{it}\left(h_{it}^{III} + h_{it}^{IV} \right) - F_{it}\left(h_{it}^{III} + h_{it}^{IV} + \mathscr{R}_{\rho,it} \right) \right]$$
$$= \frac{1}{N_3} \sum_{i \in I_3} u_{i,1}^0 \left\{ \left[\mathbf{1}\{\varepsilon_{it} \le 0\} - \mathbf{1}\left\{ \varepsilon_{it} \le h_{it}^{III} + h_{it}^{IV} \right\} \right] - \left[\mathscr{F}_{it}(0) - \mathscr{F}_{it}\left(h_{it}^{III} + h_{it}^{IV} \right) \right] \right\}$$
$$+ \frac{1}{N_3} \sum_{i \in I_3} u_{i,1}^0 \Bigg\{ \left[\mathbf{1}\left\{ \varepsilon_{it} \le h_{it}^{III} + h_{it}^{IV} \right\} - \mathbf{1}\left\{ \varepsilon_{it} \le h_{it}^{III} + h_{it}^{IV} + \mathscr{R}_{\rho,it} \right\} \right]$$
$$- \left[\mathscr{F}_{it}\left(h_{it}^{III} + h_{it}^{IV} \right) - \mathscr{F}_{it}\left(h_{it}^{III} + h_{it}^{IV} + \mathscr{R}_{\rho,it} \right) \right] \Bigg\}$$
$$+ \frac{1}{N_3} \sum_{i \in I_3} u_{i,1}^0 \left\{ \left[\mathscr{F}_{it}(0) - \mathscr{F}_{it}\left(h_{it}^{III} + h_{it}^{IV} \right) \right] - \left[F_{it}(0) - F_{it}\left(h_{it}^{III} + h_{it}^{IV} \right) \right] \right\}$$
$$+ o_p\left((N \vee T)^{-1/2}\right),$$

where the last equality holds by similar arguments as used in the last line in (A.148). We can show the first and second terms are $o_p\left((N \vee T)^{-1/2}\right)$ by similar arguments for $\mathbb{I}_{6,t}^{I}$ and $\mathbb{I}_{6,t}^{II}$. The third term is $O_p(\eta_N)$ by mean-value theorem and Assumption 2.1(viii). Compared with $\mathbb{I}_{6,t}^{III}$ and $\mathbb{I}_{6,t}^{IV}$ in the proof of statement (i), the third term here is not mean zero, and converges to zero at the rate η_N. ■

Lemma A.31. *Under Assumptions 2.1-2.9 and Assumption A.1, we have*

$$\text{(i)}\ \max_{t\in[T]}\left\|\frac{1}{N_3}\sum_{i\in I_3} e_{1,it}O_{u,1}^{(1)}u_{i,1}^{0}\left\{\mathbf{1}\{\varepsilon_{it}\le 0\}-\mathbf{1}\left\{\varepsilon_{it}\le\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\hat{v}_{t,0}^{(3,1)},\hat{v}_{t,1}^{(3,1)}\right)\right\}\right.\right.$$
$$\left.\left.-\left(F_{it}(0)-F_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\hat{v}_{t,0}^{(3,1)},\hat{v}_{t,1}^{(3,1)}\right)\right]\right)\right\}\right\|_2=o_p\left((N\vee T)^{-\frac{1}{2}}\right),$$

$$\text{(ii)}\ \max_{t\in[T]}\left\|\frac{1}{N_3}\sum_{i\in I_3} O_{u,0}^{(1)}u_{i,0}^{0}\left\{\mathbf{1}\{\varepsilon_{it}\le 0\}-\mathbf{1}\left\{\varepsilon_{it}\le\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\hat{v}_{t,0}^{(3,1)},\hat{v}_{t,1}^{(3,1)}\right)\right\}\right.\right.$$
$$\left.\left.-\left(F_{it}(0)-F_{it}\left[\rho_{it}\left(\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)},\hat{v}_{t,0}^{(3,1)},\hat{v}_{t,1}^{(3,1)}\right)\right]\right)\right\}\right\|_2=O_p(\eta_N).$$

Proof To handle the correlation between $\{\varepsilon_{it},e_{it}\}$ and $\left\{\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)}\right\}$, we follow similar arguments as used in the proof of Lemma A.27 by putting $\left\{\hat{u}_{i,0}^{(3,1)},\hat{u}_{i,1}^{(3,1)}\right\}$ in a parameter set. Then by similar arguments as used in the proof of Lemma A.30, we can obtain the desired results. ■

Lemma A.32. *Under Assumptions 2.1-2.9 and Assumption A.1, we have*

$$\max_{t\in[T]}\left\|O_{v_1,t}^{(1)}-\left(O_{u,1}^{(1)\prime}\right)^{-1}\right\|_F=o_p\left((N\vee T)^{-\frac{1}{2}}\right).$$

Proof Recall that

$$O_{v_1,t}^{(1)}=\left\{I_{K_1}+\left(O_{u,1}^{(1)\prime}\right)^{-1}\left[\hat{V}_{v_1,t}^{I}\right]^{-1}\left[\frac{1}{N_3}\sum_{i\in I_3}f_{it}(0)e_{1,it}^2u_{i,1}^0\left(O_{u,1}^{(1)}u_{i,1}^0-\hat{u}_{i,1}^{(3,1)}\right)'\right]\right\}\left(O_{u,1}^{(1)\prime}\right)^{-1}.$$

Then

$$O_{v_1,t}^{(1)}-\left(O_{u,1}^{(1)\prime}\right)^{-1}=\left(O_{u,1}^{(1)\prime}\right)^{-1}\left[\hat{V}_{v_1,t}^{3}\right]^{-1}\left[\frac{1}{N_3}\sum_{i\in I_3}f_{it}(0)e_{1,it}^2u_{i,1}^0\left(O_{u,1}^{(1)}u_{i,1}^0-\hat{u}_{i,1}^{(3,1)}\right)'\right].$$

Note that

$$\begin{aligned}
&\frac{1}{N_3}\sum_{i\in I_3}f_{it}(0)e_{1,it}^2u_{i,1}^0\left(O_{u,1}^{(1)}u_{i,1}^0-\hat{u}_{i,1}^{(3,1)}\right)'\\
&=\frac{1}{N_3}\sum_{i\in I_3}f_{it}(0)e_{1,it}^2\left\{O_1^{(1)}\hat{V}_{u_1}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_{1,it}v_{t,1}^0\left(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)+\mathscr{R}_{i,u}^1\right\}u_{i,1}^{0\prime}\\
&=\frac{1}{N_3}\sum_{i\in I_3}f_{it}(0)e_{1,it}^2O_1^{(1)}\hat{V}_{u_1}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_{1,it}v_{t,1}^0\left(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\}\right)u_{i,1}^{0\prime}+o_p\left((N\vee T)^{-\frac{1}{2}}\right)\\
&=\frac{1}{N_3T}\sum_{i\in I_3}\sum_{t=1'}^{T}f_{it}(0)e_{1,it}^2O_1^{(1)}\hat{V}_{u_1}^{-1}e_{1,it^*}v_{t^*,1}^0\left(\tau-1\{\varepsilon_{it^*}\le 0\}\right)u_{i,1}^{0\prime}+o_p\left((N\vee T)^{-\frac{1}{2}}\right)
\end{aligned}$$

$$= o_p\left((N \vee T)^{-\frac{1}{2}}\right) \quad \text{uniformly over } t \in [T],$$

where the second equality is by uniform convergence rate of $\mathscr{R}^1_{i,u}$ and the last line follows by similar arguments as in (A.124) by Bernstein's inequality conditional on $\mathscr{D}_e$. Then the result follows by noting that $O_{u,1}$ is bounded and $\hat{V}^I_{v_1,t}$ is bounded uniformly over $t \in [T]$. ■

A.2.4 Lemmas for the Consistent Estimation of the Asymptotic Variances

Recall that

$$\hat{\mathbb{V}}_{u_j} = \frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]} k_{h_N}(\hat{\varepsilon}_{it})\hat{e}^2_{j,it}\hat{\mathrm{v}}_{t,t,j}, \quad \hat{\mathbb{V}}_{v_j} = \frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]} k_{h_N}(\hat{\varepsilon}_{it})\hat{e}^2_{j,it}\hat{\mathrm{u}}_{i,i,j},$$

$$\hat{\Omega}_{u_j} = \frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]} \tau(1-\tau)\hat{e}^2_{j,it}\hat{\mathrm{v}}_{t,t,j}$$

$$+\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\mathrm{v}}_{t,s,j}\left[\tau-K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right)\right]\left[\tau-K\left(\frac{\hat{\varepsilon}_{is}}{h_N}\right)\right]$$

$$+\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1+T_1}^{T}\sum_{s=t-T_1}^{t-1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\mathrm{v}}_{t,s,j}\left[\tau-K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right)\right]\left[\tau-K\left(\frac{\hat{\varepsilon}_{is}}{h_N}\right)\right],$$

$$\hat{\Omega}_{v_j} = \tau(1-\tau)\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\hat{e}^2_{j,it}\hat{\mathrm{u}}_{i,i,j},$$

$$\hat{\Sigma}_{u_j} = \left(\hat{\mathbb{V}}_{u_j}\right)^{-1}\hat{\Omega}_{u_j}\left(\hat{\mathbb{V}}_{u_j}\right)^{-1}, \quad \hat{\Sigma}_{v_j} = \left(\hat{\mathbb{V}}_{v_j}\right)^{-1}\hat{\Omega}_{v_j}\left(\hat{\mathbb{V}}_{v_j}\right)^{-1},$$

$$V_{u_j} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[f_{it}(0)e^2_{j,it}v^0_{t,j}v^{0\prime}_{t,j}\right], \quad V^{(a)}_{v_j} = \frac{1}{N_a}\sum_{i\in I_a}\mathbb{E}\left[f_{it}(0)e^2_{j,it}\right]u^0_{i,j}u^{0\prime}_{i,j},$$

$$\Omega_{u_j} = Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_{j,it}v^0_{t,j}(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\})\right],$$

$$\Omega_{v_j} = \tau(1-\tau)\frac{1}{N}\sum_{i\in I_3}\mathbb{E}\left(e^2_{j,it}u^0_{i,j}u^{0\prime}_{i,j}\right),$$

$$\Sigma_{u_j} = O^{(1)}_j V^{-1}_{u_j}\Omega_{u_j}V^{-1}_{u_j}O^{(1)\prime}_j, \quad \Sigma_{v_j} = O^{(1)}_j V^{-1}_{v_j}\Omega_{v_j}V^{-1}_{v_j}O^{(1)\prime}_j.$$

Lemma A.33. *Under Assumptions 2.1-2.10 and Assumption A.1,* $\hat{\Sigma}_{u_j} = \Sigma_{u_j} + o_p(1)$ *and* $\hat{\Sigma}_{v_j} = \Sigma_{v_j} + o_p(1)$.

Proof First, we show that $\hat{\mathbb{V}}_{u_j} = O^{(1)}_j V_{u_j} O^{(1)\prime}_j + o_p(1)$. Note that

$$\max_{i\in I_3,t\in[T]}\left|\hat{\varepsilon}_{it}-\varepsilon_{it}\right| \le \max_{i\in I_3,t\in[T]}\left|\hat{\Theta}_{0,it}-\Theta^0_{0,it}\right| + \max_{i\in I_3,t\in[T]}\sum_{j\in[p]}\left|X_{j,it}\right|\left|\hat{\Theta}_{j,it}-\Theta^0_{j,it}\right|$$

$$= R^1_{\varepsilon,it} + \max_{i\in I_3,t\in[T],j\in[p]}\left|X_{j,it}\right|R^2_{\varepsilon,it}, \text{ and}$$

$$\max_{i\in I_3,t\in[T]}\left|k_{h_N}(\hat{\varepsilon}_{it})-k_{h_N}(\varepsilon_{it})\right|=\frac{1}{h_N}\max_{i\in I_3,t\in[T]}\left|k(\frac{\hat{\varepsilon}_{it}}{h_N})-k(\frac{\varepsilon_{it}}{h_N})\right|\lesssim\frac{1}{h_N^2}\max_{i\in I_3,t\in[T]}\left|\hat{\varepsilon}_{it}-\varepsilon_{it}\right|$$
$$=R^1_{k,it}+\max_{i\in I_3,t\in[T],j\in[p]}\left|X_{j,it}\right|R^2_{k,it}, \tag{A.160}$$

where $\max_{i\in I_3,t\in[T]}\left|R^1_{\varepsilon,it}\right|=O_p(\eta_N)$, $\max_{i\in I_3,t\in[T]}\left|R^2_{\varepsilon,it}\right|=O_p\left(\frac{\log N\vee T}{N\wedge T}\right)$, $\max_{i\in I_3,t\in[T]}\left|R^1_{k,it}\right|=O_p(\eta_N h_N^{-2})$ and $\max_{i\in I_3,t\in[T]}\left|R^2_{k,it}\right|=O_p(\frac{\log N\vee T}{N\wedge T}h_N^{-2})$ by (A.48) and Assumption 2.1(iv). Let

$$\mathrm{v}^0_{t,t,j}=\frac{1}{6}\sum_{a\in[3]}\sum_{b\in[3]\setminus\{a\}}O_j^{(b)}v^0_{t,j}v^{0\prime}_{t,j}O_j^{(b)\prime}$$

and recall that $\hat{\mathrm{v}}_{t,t,j}=\frac{1}{6}\sum_{a\in[3]}\sum_{b\in[3]\setminus\{a\}}\hat{v}^{(a,b)}_{t,j}\hat{v}^{(a,b)\prime}_{t,j}$. With Theorem 2.3, it is clear that

$$\begin{aligned}
&\max_{t\in[T]}\left\|\hat{\mathrm{v}}_{t,t,j}-\mathrm{v}^0_{t,t,j}\right\|_F\\
&=\frac{1}{6}\sum_{a\in[3]}\sum_{b\in[3]\setminus\{a\}}\left(\hat{v}^{(a,b)}_{t,j}\hat{v}^{(a,b)\prime}_{t,j}-O_j^{(b)}v^0_{t,j}v^{0\prime}_{t,j}O_j^{(b)\prime}\right)\\
&=\frac{1}{6}\sum_{a\in[3]}\sum_{b\in[3]\setminus\{a\}}\left[\left(\hat{v}^{(a,b)}_{t,j}-O_j^{(b)}v^0_{t,j}\right)\left(\hat{v}^{(a,b)}_{t,j}-O_j^{(b)}v^0_{t,j}\right)'+O_j^{(b)}v^0_{t,j}\left(\hat{v}^{(a,b)}_{t,j}-O_j^{(b)}v^0_{t,j}\right)'\right.\\
&\left.+\left(\hat{v}^{(a,b)}_{t,j}-O_j^{(b)}v^0_{t,j}\right)\left(O_j^{(b)}v^0_{t,j}\right)'\right]\\
&=O_p\left(\sqrt{\frac{\log N\vee T}{N}}\right).
\end{aligned}$$

Let $\mathrm{v}^0_{t,s,j}=\frac{1}{6}\sum_{a\in[3]}\sum_{b\in[3]\setminus\{a\}}O_j^{(b)}v^0_{t,j}v^{0\prime}_{s,j}O_j^{(b)\prime}$ and recall that

$$\hat{\mathrm{v}}_{t,s,j}=\frac{1}{6}\sum_{a\in[3]}\sum_{b\in[3]\setminus\{a\}}\hat{v}^{(a,b)}_{t,j}\hat{v}^{(a,b)\prime}_{s,j},$$

similarly as above, we have $\max_{t\in[T],s\in[T]}\left\|\hat{\mathrm{v}}_{t,s,j}-\mathrm{v}^0_{t,s,j}\right\|_F=O_p\left(\sqrt{\frac{\log N\vee T}{N}}\right)$. It follows that

$$\begin{aligned}
\hat{\mathbb{V}}_{u_j}&=\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}k_{h_N}(\hat{\varepsilon}_{it})\hat{e}^2_{j,it}\hat{\mathrm{v}}_{t,t,j}\\
&=\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}k_{h_N}(\varepsilon_{it})e^2_{j,it}\mathrm{v}^0_{t,t,j}+\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}k_{h_N}(\varepsilon_{it})e^2_{j,it}\left(\hat{\mathrm{v}}_{t,t,j}-\mathrm{v}^0_{t,t,j}\right)\\
&+\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}k_{h_N}(\varepsilon_{it})\left(\hat{e}^2_{j,it}-e^2_{j,it}\right)\mathrm{v}^0_{t,t,j}\\
&+\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}k_{h_N}(\varepsilon_{it})\left(\hat{e}^2_{j,it}-e^2_{j,it}\right)\left(\hat{\mathrm{v}}_{t,t,j}-\mathrm{v}^0_{t,t,j}\right)
\end{aligned}$$

$$+\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\left[k_{h_N}(\hat{\varepsilon}_{it})-k_{h_N}(\varepsilon_{it})\right]e_{j,it}^2\mathrm{v}_{t,t,j}^0$$
$$+\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\left[k_{h_N}(\hat{\varepsilon}_{it})-k_{h_N}(\varepsilon_{it})\right]e_{j,it}^2\left(\hat{\mathrm{v}}_{t,t,j}-\mathrm{v}_{t,t,j}^0\right)$$
$$+\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\left[k_{h_N}(\hat{\varepsilon}_{it})-k_{h_N}(\varepsilon_{it})\right]\left(\hat{e}_{j,it}^2-e_{j,it}^2\right)\mathrm{v}_{t,t,j}^0$$
$$+\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\left[k_{h_N}(\hat{\varepsilon}_{it})-k_{h_N}(\varepsilon_{it})\right]\left(\hat{e}_{j,it}^2-e_{j,it}^2\right)\left(\hat{\mathrm{v}}_{t,t,j}-\mathrm{v}_{t,t,j}^0\right)$$
$$=\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}k_{h_N}(\varepsilon_{it})e_{j,it}^2\mathrm{v}_{t,t,j}^0+O_p\left(\eta_N h_N^{-2}\right)$$
$$=\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}k_{h_N}(\varepsilon_{it})e_{j,it}^2\mathrm{v}_{t,t,j}^0+o_p(1), \tag{A.161}$$

where the last two lines combines (A.47), (A.160), Assumption 2.10(ii) and facts that $\hat{e}_{j,it}^2-e_{j,it}^2=\left(\hat{e}_{j,it}-e_{j,it}\right)^2+e_{j,it}\left(\hat{e}_{j,it}-e_{j,it}\right)=O_p(\eta_N^2)+e_{j,it}O_p(\eta_N)$ uniformly by Lemma A.21 and $\max_{i\in I_3,t\in[T]}\left|k_{h_N}(\varepsilon_{it})\right|=O(h_N^{-1})$. By Bernstein's inequality, we obtain that

$$\left\|\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\left[k_{h_N}(\varepsilon_{it})e_{j,it}^2-\mathbb{E}\left(k_{h_N}(\varepsilon_{it})e_{j,it}^2\middle|\mathscr{D}\right)\right]v_{t,j}^0v_{t,j}^{0\prime}\right\|_F=O_p\left(\sqrt{\frac{\log N}{NT}\frac{\xi_N^2}{h_N}}\right),$$
$$\left\|\frac{1}{T}\sum_{t\in[T]}\mathbb{E}\left[f_{it}(0)e_{j,it}^2\middle|\mathscr{D}\right]v_{t,j}^0v_{t,j}^{0\prime}-\mathbb{E}\left[f_{it}(0)e_{j,it}^2v_{t,j}^0v_{t,j}^{0\prime}\right]\right\|_F=O_p\left(\sqrt{\frac{\log T}{T}}\xi_N^2\right). \tag{A.162}$$

Besides, by Assumption 2.10(i), we observe that

$$\mathbb{E}\left[k_{h_N}(\varepsilon_{it})\middle|\mathscr{D}_e\right]=f_{it}(0)+O(h_N^m), \tag{A.163}$$

together with Assumption 2.10(v), and it gives

$$\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\mathbb{E}\left(k_{h_N}(\varepsilon_{it})e_{j,it}^2\middle|\mathscr{D}\right)v_{t,j}^0v_{t,j}^{0\prime}=\frac{1}{T}\sum_{t\in[T]}\mathbb{E}\left[f_{it}(0)e_{j,it}^2\middle|\mathscr{D}\right]v_{t,j}^0v_{t,j}^{0\prime}+O(h_N^m). \tag{A.164}$$

Combining (A.161)-(A.164) and Assumption 2.10(ii), we obtain that

$$\hat{\mathbb{V}}_{u_j}=O_j^{(1)}V_{u_j}O_j^{(1)\prime}+o_p(1).$$

By analogous analysis and Assumption 2.10(v), we can also show that $\hat{\mathbb{V}}_{v_j}^{(1)}=O_j^{(1)}V_{v_j}O_j^{(1)\prime}+o_p(1)$.

Next, we show the consistency of $\hat{\Omega}_{u_j}$. With the restriction for T_1 in Assumption

2.10(iii), we first note that

$$\Omega_{u_j} = Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_{j,it}v_{t,j}^{0}(\tau - \mathbf{1}\{\varepsilon_{it}\le 0\})\right]$$

$$= \tau(1-\tau)\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(e_{j,it}^2 v_{t,j}^0 v_{t,j}^{0\prime}\right)$$

$$+\frac{1}{T}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\mathbb{E}\left[e_{j,it}e_{j,is}v_{t,j}^0 v_{s,j}^{0\prime}(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\})(\tau-\mathbf{1}\{\varepsilon_{is}\le 0\})\right]$$

$$+\frac{1}{T}\sum_{t=1+T_1}^{T}\sum_{s=t-T_1}^{t-1}\mathbb{E}\left[e_{j,it}e_{j,is}v_{t,j}^0 v_{s,j}^{0\prime}(\tau-\mathbf{1}\{\varepsilon_{it}\le 0\})(\tau-\mathbf{1}\{\varepsilon_{is}\le 0\})\right]+O(\alpha^{T_1})$$

$$= \tau(1-\tau)\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(e_{j,it}^2 v_{t,j}^0 v_{t,j}^{0\prime}\right)+\frac{1}{T}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\mathbb{E}\left[e_{j,it}e_{j,is}v_{t,j}^0 v_{s,j}^{0\prime}\left(F_{i,ts}(0,0)-\tau^2\right)\right]$$

$$+\frac{1}{T}\sum_{t=1+T_1}^{T}\sum_{s=t-T_1}^{t-1}\mathbb{E}\left[e_{j,it}e_{j,is}v_{t,j}^0 v_{s,j}^{0\prime}\left(F_{i,ts}(0,0)-\tau^2\right)\right]+o(1),$$

where the second equality is by Assumption 2.1(iii), the third equality is by Assumption 2.1(vii) and Assumption 2.10(iii). Compared to $\hat{\Omega}_{u_j}$, what remains to show are

$$\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\hat{e}_{j,it}^2\hat{\text{v}}_{t,t,j} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(e_{j,it}^2 \text{v}_{t,t,j}^0\right)+o_p(1), \tag{A.165}$$

$$\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\text{v}}_{t,s,j}\left[\tau-K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right)\right]\left[\tau-K\left(\frac{\hat{\varepsilon}_{is}}{h_N}\right)\right]$$
$$=\frac{1}{T}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\mathbb{E}\left[e_{j,it}e_{j,is}\text{v}_{t,s,j}^0\left(F_{i,ts}(0,0)-\tau^2\right)\right]+o_p(1), \tag{A.166}$$

$$\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1+T_1}^{T}\sum_{s=t-T_1}^{t-1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\text{v}}_{t,s,j}\left[\tau-K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right)\right]\left[\tau-K\left(\frac{\hat{\varepsilon}_{is}}{h_N}\right)\right]$$
$$=\frac{1}{T}\sum_{t=1+T_1}^{T}\sum_{s=t-T_1}^{t-1}\mathbb{E}\left[e_{j,it}e_{j,is}\text{v}_{t,s,j}^0\left(F_{i,ts}(0,0)-\tau^2\right)\right]+o_p(1). \tag{A.167}$$

For (A.165), like (A.161), we notice that

$$\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\hat{e}_{j,it}^2\hat{\text{v}}_{t,t,j} = \frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}e_{j,it}^2\text{v}_{t,t,j}^0+o_p(1)$$
$$=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(e_{j,it}^2\text{v}_{t,t,j}^0\right)+\frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\left[e_{j,it}^2\text{v}_{t,t,j}^0-\mathbb{E}\left(e_{j,it}^2\text{v}_{t,t,j}^0\right)\right]+o_p(1)$$
$$=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(e_{j,it}^2\text{v}_{t,t,j}^0\right)+o_p(1), \tag{A.168}$$

where the first equality is by Lemma A.21 and (A.47), and the second equality is by Bernstein's inequality and Assumption 2.10(v).

For (A.166), we observe that

$$\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\mathrm{v}}_{t,s,j}\left[\tau-K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right)\right]\left[\tau-K\left(\frac{\hat{\varepsilon}_{is}}{h_N}\right)\right]$$
$$=\tau^2\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\mathrm{v}}_{t,s,j}-\tau\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\mathrm{v}}_{t,s,j}K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right)$$
$$-\tau\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\mathrm{v}}_{t,s,j}K\left(\frac{\hat{\varepsilon}_{is}}{h_N}\right)$$
$$+\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\mathrm{v}}_{t,s,j}K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right)K\left(\frac{\hat{\varepsilon}_{is}}{h_N}\right) \tag{A.169}$$

such that

$$\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\mathrm{v}}_{t,s,j}=\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}+O_p\left(T_1\eta_N\right)$$
$$=\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\mathbb{E}\left(e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}\right)$$
$$+\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\left[e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}-\mathbb{E}\left(e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}\right)\right]+O_p\left(T_1\eta_N\right)$$
$$=\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\mathbb{E}\left(e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}\right)+o_p(1), \tag{A.170}$$

where the first equality is by the similar arguments as (A.161) and the last line combines Assumption 2.10(iii) and the fact that the second term in the second equality can be shown to be $o_p(1)$ by Bernstein's inequality. Furthermore, with the fact that $\max_{i\in I_3,t\in[T]}\left|K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right)-K\left(\frac{\varepsilon_{it}}{h_N}\right)\right|\lesssim\frac{1}{h_N}\max_{i\in I_3,t\in[T]}\left|\hat{\varepsilon}_{it}-\varepsilon_{it}\right|=O_p\left(\eta_N h_N^{-1}\right)$ and by the analogous arguments as above, we can show that

$$\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\hat{e}_{j,it}\hat{e}_{j,is}\hat{\mathrm{v}}_{t,s,j}K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right)$$
$$=\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}K\left(\frac{\varepsilon_{it}}{h_N}\right)+O_p\left(T_1\eta_N h_N^{-1}\right)$$
$$=\frac{1}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}\mathbb{E}\left[K\left(\frac{\varepsilon_{it}}{h_N}\right)\middle|\mathscr{D}_e\right]+o_p(1)$$
$$=\frac{\tau}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}+O(h_N^m)+o_p(1)$$
$$=\frac{\tau}{NT}\sum_{i\in[N]}\sum_{t=1}^{T-T_1}\sum_{s=t+1}^{t+T_1}\mathbb{E}\left[e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}\right]+O(h_N^m)+o_p(1)$$

$$= \frac{\tau}{NT} \sum_{i\in[N]} \sum_{t=1}^{T-T_1} \sum_{s=t+1}^{t+T_1} \mathbb{E}\left[e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}\right] + o_p(1), \quad (A.171)$$

where the first equality is similar as (A.161), the second equality is by addition and subtracting and Assumption 2.10(iii), the third equality is by the fact that $\mathbb{E}\left[K\left(\frac{\varepsilon_{it}}{h_N}\right)|\mathscr{D}_e\right] = \tau + O(h_N^m)$ by the calculation of nonparametric kernel estimator which can be found in Galvao and Kato (2016). The last equality is similar as the second equality and combines Assumption 2.10(i) and Assumption 2.10(ii).

Moreover, similarly as (A.170), with the fact that $\mathbb{E}\left[K\left(\frac{\varepsilon_{it}}{h_N}\right)K\left(\frac{\varepsilon_{is}}{h_N}\right)|\mathscr{D}_e\right] = F_{i,ts}(0,0) + O(h_N^m)$, we can show that

$$\begin{aligned}
&\frac{1}{NT} \sum_{i\in[N]} \sum_{t=1}^{T-T_1} \sum_{s=t+1}^{t+T_1} \hat{e}_{j,it}\hat{e}_{j,is}\hat{\mathrm{v}}_{t,s,j} K\left(\frac{\hat{\varepsilon}_{it}}{h_N}\right) K\left(\frac{\hat{\varepsilon}_{is}}{h_N}\right) \\
&= \frac{1}{NT} \sum_{i\in[N]} \sum_{t=1}^{T-T_1} \sum_{s=t+1}^{t+T_1} e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j} K\left(\frac{\varepsilon_{it}}{h_N}\right) K\left(\frac{\varepsilon_{is}}{h_N}\right) + O_p\left(\sqrt{\frac{\log(N\vee T)}{N\wedge T}}\frac{T_1\xi_N^2}{h_N^2}\right) \\
&= \frac{1}{NT} \sum_{i\in[N]} \sum_{t=1}^{T-T_1} \sum_{s=t+1}^{t+T_1} \mathbb{E}\left[e_{j,it}e_{j,is}\mathrm{v}^0_{t,s,j}F_{i,ts}(0,0)\right] + o_p(1). \quad (A.172)
\end{aligned}$$

Combining (A.169)-(A.172), we complete the proof for (A.166). By the analogous arguments, we can show the proof for (A.167), which yields $\hat{\Omega}_{u_j} = \Omega_{u_j} + o_p(1)$. ■

A.3 Algorithm for Low-rank Estimation

In this section, we provide the algorithm for the case of low-rank estimation with two regressors, the case of more than two regressors is self-evident. To solve the regularized quantile regression, let the optimization problem with two regressors be as

$$\min_{\Theta_0,\Theta_1,\Theta_2} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau\left(y_{it} - \Theta_{0,it} - x_{1,it}\Theta_{1,it} - x_{2,it}\Theta_{2,it}\right) + \nu_0 \|\Theta_0\|_* + \nu_1 \|\Theta_1\|_* + \nu_2 \|\Theta_2\|_*.$$

As in Belloni et al. (2023), the above minimization problem is equivalent to the following one:

$$\begin{aligned}
&\min_{\Theta_0,\Theta_1,\Theta_2,V,W,Z_{\Theta_0},Z_{\Theta_1},Z_{\Theta_2}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(V_{it}) + \nu_0 \|\Theta_0\|_* + \nu_1 \|Z_{\Theta_1}\|_* + \nu_2 \|Z_{\Theta_2}\|_* \\
&s.t.\ V = W,\ W = Y - X_1 \odot \Theta_1 - X_2 \odot \Theta_2 - Z_{\Theta_0}, \\
&Z_{\Theta_0} - \Theta_0 = 0,\ Z_{\Theta_1} - \Theta_1 = 0,\ Z_{\Theta_2} - \Theta_2 = 0.
\end{aligned}$$

As our theoretical results show, ν_0, ν_1 and ν_2 converge to zero at rate $\frac{\sqrt{N}\vee\sqrt{T}}{NT}$. The augmented Lagrangian is

$$\begin{aligned}
&\mathscr{L}\left(V,W,\Theta_0,Z_{\Theta_0},\Theta_1,Z_{\Theta_1},\Theta_2,Z_{\Theta_2},U_v,U_w,U_{\Theta_0},U_{\Theta_1},U_{\Theta_2}\right)\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau(V_{it})+\nu_0\|\Theta_0\|_*+\nu_1\|Z_{\Theta_1}\|_*+\nu_2\|Z_{\Theta_2}\|_*+\frac{\rho}{2NT}\|V-W+U_v\|_F^2\\
&+\frac{\rho}{2NT}\left\|W-Y+X_1\odot\Theta_1+X_2\odot\Theta_2+Z_{\Theta_0}+U_W\right\|_F^2+\frac{\rho}{2NT}\left\|Z_{\Theta_0}-\Theta_0+U_{\Theta_0}\right\|_F^2\\
&+\frac{\rho}{2NT}\|Z_{\Theta_1}-\Theta_1+U_{\Theta_1}\|_F^2+\frac{\rho}{2NT}\|Z_{\Theta_2}-\Theta_2+U_{\Theta_2}\|_F^2,
\end{aligned}$$

where $\rho>0$ is the penalty parameter.

By ADMM algorithm, similarly as Belloni et al. (2023), updates are as follows:

$$V^{k+1}\leftarrow\arg\min_V\left\{\frac{1}{NT}\sum_i\sum_t\rho_\tau(V_{it})+\frac{\rho}{2NT}\left\|V-W^k+U_v^k\right\|_F^2\right\}\tag{A.173}$$

$$\begin{aligned}
(\Theta_1^{k+1},\Theta_2^{k+1})\leftarrow\arg\min_{\Theta_1,\Theta_2}\Big\{&\left\|W^k-Y+X_1\odot\Theta_1+X_2\odot\Theta_2+Z_{\Theta_0}^k+U_W^k\right\|_F^2\\
&+\left\|Z_{\Theta_1}^k-\Theta_1+U_{\Theta_1}^k\right\|_F^2+\left\|Z_{\Theta_2}^k-\Theta_2+U_{\Theta_2}^k\right\|_F^2\Big\}
\end{aligned}\tag{A.174}$$

$$\Theta_0^{k+1}\leftarrow\arg\min_{\Theta_0}\left\{\frac{1}{2}\left\|Z_{\Theta_0}^k-\Theta_0+U_{\Theta_0}^k\right\|_F^2+\frac{\nu_0 NT}{\rho}\|\Theta_0\|_*\right\}\tag{A.175}$$

$$Z_{\Theta_1}^{k+1}\leftarrow\arg\min_{Z_{\Theta_1}}\left\{\frac{1}{2}\left\|\Theta_1^{k+1}-U_{\Theta_1}^k-Z_{\Theta_1}\right\|_F^2+\frac{\nu_1 NT}{\rho}\|Z_{\Theta_1}\|_*\right\}$$

$$Z_{\Theta_2}^{k+1}\leftarrow\arg\min_{Z_{\Theta_2}}\left\{\frac{1}{2}\left\|\Theta_2^{k+1}-U_{\Theta_2}^k-Z_{\Theta_2}\right\|_F^2+\frac{\nu_2 NT}{\rho}\|Z_{\Theta_2}\|_*\right\}$$

$$\begin{aligned}
(Z_{\Theta_0}^{k+1},W^{k+1})\leftarrow\arg\min_{Z_{\Theta_0},W}\Big\{&\left\|V^{k+1}-W+U_v^k\right\|_F^2\\
&+\left\|W-Y+X_1\odot\Theta_1^{k+1}+X_2\odot\Theta_2^{k+1}+Z_{\Theta_0}+U_W^k\right\|_F^2+\left\|Z_{\Theta_0}-\Theta_0^{k+1}+U_{\Theta_0}^k\right\|_F^2\Big\}
\end{aligned}$$

$$U_v^{k+1}\leftarrow V^{k+1}-W^{k+1}+U_v^k$$

$$U_W^{k+1}\leftarrow W^{k+1}-Y+X_1\odot\Theta_1^{k+1}+X_2\odot\Theta_2^{k+1}+Z_{\Theta_0}^{k+1}+U_W^k$$

$$U_{\Theta_0}^{k+1}\leftarrow Z_{\Theta_0}^{k+1}-\tilde{\Theta_0}^{k+1}+U_{\Theta_0}^k$$

$$U_{\Theta_1}^{k+1}\leftarrow Z_{\Theta_1}^{k+1}-\Theta_1^{k+1}+U_{\Theta_1}^k$$

$$U_{\Theta_2}^{k+1}\leftarrow Z_{\Theta_2}^{k+1}-\Theta_2^{k+1}+U_{\Theta_2}^k$$

For (A.173), by Ali et al. (2016),

$$V^{k+1}\leftarrow P_+\left(W^k-U_v^k-\frac{\tau}{\rho}\iota_N\iota_T'\right)+P_-\left(W^k-U_v^k-\frac{(1-\tau)}{\rho}\iota_N\iota_T'\right).$$

where ι_N is the $N \times 1$ all-ones vector, and same for ι_T. For (A.174), first order condition gives

$$\Theta_{1,it}^{k+1} = \frac{\left(1+x_{2,it}^2\right)\left(Z_{\Theta_1,it}^k + U_{\Theta_1,it}^k - A_{it}x_{1,it}\right) - x_{1,it}x_{2,it}\left(Z_{\Theta_2,it}^k + U_{\Theta_2,it}^k - A_{it}x_{2,it}\right)}{1+x_{1,it}^2+x_{2,it}^2},$$

$$\Theta_{2,it}^{k+1} = \frac{\left(1+x_{1,it}^2\right)\left(Z_{\Theta_2,it}^k + U_{\Theta_2,it}^k - A_{it}x_{2,it}\right) - x_{1,it}x_{2,it}\left(Z_{\Theta_1,it}^k + U_{\Theta_1,it}^k - A_{it}x_{1,it}\right)}{1+x_{1,it}^2+x_{2,it}^2},$$

where

$$A := W^k + Z_{\Theta_0}^k + U_W^k - Y.$$

To solve (A.175), by singular value thresholding estimations, the update for Θ_0^{k+1} is

$$\Theta_0^{k+1} \leftarrow P_0 D_{0,\frac{\nu_1 NT}{\rho}} Q_0',$$

where $Z_{\Theta_0}^k + U_{\Theta_0}^k = P_0 D_0 Q_0'$, and $D_{0,\frac{\nu_0}{\rho},ii} = \max(D_{0,ii} - \frac{\nu_0}{\rho}, 0)$. Similarly for $Z_{\Theta_1}^{k+1}$ and $Z_{\Theta_2}^{k+1}$,

$$Z_{\Theta_1}^{k+1} \leftarrow P_1 D_{1,\frac{\nu_1 NT}{\rho}} Q_1',$$

$$Z_{\Theta_2}^{k+1} \leftarrow P_2 D_{2,\frac{\nu_2 NT}{\rho}} Q_2',$$

where $\Theta_1^{k+1} - U_{\Theta_1}^k = P_1 D_1 Q_1'$, $\Theta_2^{k+1} - U_{\Theta_2}^k = P_2 D_2 Q_2'$, $D_{1,\frac{\nu_1}{\rho},ii} = \max\left(D_{1,ii} - \frac{\nu_1}{\rho}, 0\right)$, and $D_{2,\frac{\nu_2}{\rho},ii} = \max\left(D_{2,ii} - \frac{\nu_2}{\rho}, 0\right)$.

Finally, let $\tilde{A} := -Y + X_1 \odot \Theta_1^{k+1} + X_2 \odot \Theta_2^{k+1} + U_W^{k+1}$, $\tilde{B} := -V^{k+1} - U_v^k$, $\tilde{C} := -\tilde{\Theta}_0^{k+1} + U_{\Theta_0}^k$, then

$$Z_{\Theta_0}^{k+1} \leftarrow \frac{-\tilde{A} - 2\tilde{C} + \tilde{B}}{3},$$

$$W^{k+1} \leftarrow -\tilde{A} - \tilde{C} - 2Z_{\Theta_0}^{k+1}.$$

Appendix B

Technical Results for Chapter 3

B.1 Proofs of the Main Results

B.1.1 Proof of Lemma 3.1

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

Let $A_{j,i}^{(\ell)} = \sum_{k=1}^{K_\ell} \alpha_{k,j}^{(\ell)} \mathbf{1}\{i \in G_{k,j}^{(\ell)}\}$, $A_{j,i} = (A_{j,i}^{(1)}, A_{j,i}^{(2)})'$ and $A_j = \left(A_{j,1}, \cdots, A_{j,N}\right)' \in \mathbb{R}^{N\times 2}$. 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

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

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

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

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

Now, let $U_j = \mathscr{U}_j \Sigma_j / \sqrt{T}$ and $V_j = \sqrt{T} \mathscr{V}_j$. 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. ■

B.1.2 Proof of Theorem 3.1

Proof of Statement (i).

Let

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

By Lemma B.4, we notice that

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

Recall from (3.11) that

$$\begin{aligned} \mathscr{R}(C_1, C_2) := \Bigg\{ (\{\Delta_{\Theta_j}\}_{j\in[p]\cup\{0\}}) : \sum_{j\in[p]\cup\{0\}} \left\| \mathscr{P}_j^\perp(\Delta_{\Theta_j}) \right\|_* \le C_1 \sum_{j\in[p]\cup\{0\}} \left\| \mathscr{P}_j(\Delta_{\Theta_j}) \right\|_*, \\ \sum_{j\in[p]\cup\{0\}} \left\| \Theta_j \right\|^2 \ge C_2 \sqrt{NT} \Bigg\}. \end{aligned}$$

When $\{\tilde{\Delta}_{\Theta_j}\}_{j\in[p]\cup\{0\}} \in \mathscr{R}(3)$ and $\{\tilde{\Delta}_{\Theta_j}\}_{j\in[p]\cup\{0\}} \notin \mathscr{R}(3, C_2)$, we have $\sum_{j\in[p]\cup\{0\}} \left\| \tilde{\Delta}_{\Theta_j} \right\|^2 < C_2 \sqrt{NT}$, which gives

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

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

Define the event

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

for some positive constant c_3. By Lemma B.3, we have $\mathbb{P}(\mathscr{A}_{1,N}^c(c_3)) \leq \varepsilon$ for any $\varepsilon > 0$. By the definition of $\{\tilde{\Theta}_j\}_{j\in[p]\cup\{0\}}$ in (3.3), we have

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

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

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

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.31), the first equality holds by the fact that $\left\|\tilde{\Delta}_{\Theta_j}\right\|_* = \|\mathscr{P}_j(\tilde{\Delta}_{\Theta_j})\|_* + \left\|\mathscr{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_*$ (see, e.g., Lemma D.2(i) in Chernozhukov et al. (2019)) and that $\nu_j = \frac{4c_3\left(\sqrt{N} \vee \sqrt{T \log T}\right)}{NT}$. It follows that

$$C_3 \sum_{j\in[p]\cup\{0\}} \|\tilde{\Delta}_{\Theta_j}\|^2$$

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

where the first inequality holds by Assumption 3.4, the second inequality is by (B.1), the third inequality is by the fact that $\left\| \mathscr{P}_j(\tilde{\Delta}_{\Theta_j}) \right\|_* \leq rank(\mathscr{P}_j(\tilde{\Delta}_{\Theta_j})) \left\| \mathscr{P}_j(\tilde{\Delta}_{\Theta_j}) \right\|$ with $rank(\mathscr{P}_j(\tilde{\Delta}_{\Theta_j})) \leq 2\bar{r}$ by Lemma D.2(ii) in Chernozhukov et al. (2019), the fourth inequality is by the fact that $\left\| \tilde{\Delta}_{\Theta_j} \right\| = \left\| \mathscr{P}_j(\tilde{\Delta}_{\Theta_j}) \right\| + \left\| \mathscr{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j}) \right\|$ (see, e.g., Lemma D.2(ii) in Chernozhukov et al. (2019)), and the last inequality holds by Jensen inequality. Consequently, we can conclude that

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

In addition, $\max_{k\in[r_j]} \left| \tilde{\sigma}_{k,j} - \sigma_{k,j} \right| = O_p(\eta_{N,1})$ 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 $\tilde{D}_j = \frac{1}{NT}\tilde{\Theta}_j'\tilde{\Theta}_j = \hat{\tilde{\mathscr{V}}}_j\hat{\tilde{\Sigma}}_j\hat{\tilde{\mathscr{V}}}_j'$, and $D_j^0 = \frac{1}{NT}\Theta_j^{0\prime}\Theta_j^0 = \mathscr{V}_j^0\Sigma_j^0\mathscr{V}_j^{0\prime}$. Define the event

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

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

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

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,

$$
\left\| \mathscr{V}_j^0 - \hat{\tilde{\mathscr{V}}}_j O_j \right\| \leq \sqrt{r_j} \left\| \mathscr{V}_j^0 - \hat{\tilde{\mathscr{V}}}_j O_j \right\|_{op} \leq \sqrt{r_j} \frac{2\sqrt{2}M^2\eta_{N,1}}{\sigma_{K_j,1}^2 - 2M^2\eta_{N,1}}
$$

$$\leq \sqrt{r_j}\frac{2\sqrt{2}M^2\eta_{N,1}}{c_\sigma^2 - 2M^2\eta_{N,1}} \leq \sqrt{r_j}\frac{2\sqrt{2}M^2\eta_{N,1}}{C_6 c_\sigma^2} \leq C_7\eta_{N,1} \tag{B.2}$$

for $C_7 = \frac{2\sqrt{2}M^2\sqrt{\bar{r}}}{C_6 c_\sigma^2}$, 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 $\left\|V_j^0 - \tilde{V}_j O_j\right\| \leq C_7\sqrt{T}\eta_{N,1}$ 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 3.1(i), it implies $\frac{1}{\sqrt{T}}\left\|V_j^0 - \tilde{V}_j O_j\right\| = O_p\left(T\eta_{N,1}\right)$.

Proof of Statement (ii).

Define

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

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

Let $\tilde{Y}_{it} := Y_{it} - \left(O_0 u_{i,0}^0\right)'\tilde{v}_{t,0} - \sum_{j=1}^{p}\left(O_j u_{i,j}^0\right)'\tilde{v}_{t,j}X_{j,it}$. By the definition of $\left\{\dot{u}_{i,j}\right\}$ in (3.4), we have

$$\begin{aligned}
0 &\geq \frac{1}{T}\sum_{t\in[T]}\left(Y_{it} - \dot{u}_{i,0}'\tilde{v}_{t,0} - \sum_{j=1}^{p}\dot{u}_{i,j}'\tilde{v}_{t,j}X_{j,it}\right)^2 - \frac{1}{T}\sum_{t\in[T]}\tilde{Y}_{it}^2 \\
&= \frac{1}{T}\sum_{t\in[T]}\left(\tilde{Y}_{it} - \left(\dot{u}_{i,0} - O_0 u_{i,0}^0\right)'\tilde{v}_{t,0} - \sum_{j\in[p]}\left(\dot{u}_{i,j} - O_j u_{i,j}^0\right)'\tilde{v}_{t,j}X_{j,it}\right)^2 - \frac{1}{T}\sum_{t\in[T]}\tilde{Y}_{it}^2 \\
&= \frac{1}{T}\sum_{t\in[T]}\left[\left(\dot{\Delta}_{i,u}'\tilde{\phi}_{it}\right)^2 - 2\left(\dot{\Delta}_{i,u}'\tilde{\phi}_{it}\right)\left(Y_{it} - u_i^{0\prime}\tilde{\phi}_{it}\right)\right],
\end{aligned}$$

which implies

$$\begin{aligned}
&\left\|\dot{\Delta}_{i,u}\right\|_2^2\lambda_{\min}\left(\frac{1}{T}\sum_{t\in[T]}\tilde{\phi}_{it}\tilde{\phi}_{it}'\right) \leq \frac{1}{T}\sum_{t\in[T]}\left(\dot{\Delta}_{i,u}'\tilde{\phi}_{it}\right)^2 \leq \frac{2}{T}\sum_{t\in[T]}\dot{\Delta}_{i,u}'\tilde{\phi}_{it}\left(Y_{it} - u_i^{0\prime}\tilde{\phi}_{it}\right) \\
&= \frac{2}{T}\sum_{t\in[T]}\dot{\Delta}_{i,u}'\tilde{\phi}_{it}\left[e_{it} - u_i^{0\prime}\left(\tilde{\phi}_{it} - \phi_{it}^0\right)\right] \\
&= 2\left\{\frac{1}{T}\sum_{t\in[T]}\phi_{it}^0 e_{it}\right\}'\dot{\Delta}_{i,u} + 2\left\{\frac{1}{T}\sum_{t\in[T]}\left(\tilde{\phi}_{it} - \phi_{it}^0\right)e_{it}\right\}'\dot{\Delta}_{i,u} \\
&- \frac{2}{T}\sum_{t\in[T]}\tilde{\phi}_{it}'\dot{\Delta}_{i,u}\left[u_i^{0\prime}\left(\tilde{\phi}_{it} - \phi_{it}^0\right)\right] \\
&:= 2G_{i,1}'\dot{\Delta}_{i,u} + 2G_{i,2} - 2G_{i,3}.
\end{aligned} \tag{B.3}$$

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 (conditional) strong mixing sequence by Assumption 3.1(iii). Besides, we observe that

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

by Lemma B.7(ii) and Assumption 3.1(v). Following similar arguments, we have

$$\begin{aligned}
&\max_{i,t} \sum_{s=t+1}^{T} \left\|Cov\left(\phi_{it}^0 e_{it}, \phi_{is}^0 e_{is} | \mathscr{D}\right)\right\| \\
&\leq 8 \max_{t} \sum_{s=t+1}^{T} \left[\mathbb{E}\left(\left\|\phi_{it}^0 e_{it}\right\|_2^q | \mathscr{D}\right)\right]^{1/q} \left[\left\|\mathbb{E}\left(\phi_{is}^0 e_{is}\right\|_2^q | \mathscr{D}\right)\right]^{1/q} (\alpha(t-s))^{1-2/q} \\
&= O(1)\ a.s,
\end{aligned}$$

where the first inequality is by the conditional Davydov's inequality that says

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

for any conditional 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 Lemma A.4 in Su and Chen (2013b).

Following this, for some constant C_8, we have

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

and $\max_{i,t} \left\|\phi_{it}^0 e_{it}\right\|_{\max} \leq C_8 (NT)^{1/q}$ by Lemma B.7(i) and Assumption 3.1(iv). Define $\mathscr{A}_{3,N}(M) = \{\max_{i,t} \left\|\phi_{it}^0 e_{it}\right\| \leq M(NT)^{1/q}\}$ and $\mathscr{A}_{3,N,i}(M) = \{\max_t \left\|\phi_{it}^0 e_{it}\right\| \leq M(NT)^{1/q}\}$ for a large enough constant M. For a positive constant C_9, it yields that

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

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

$$\max_i \frac{\left|G_{i,1}'\dot{\Delta}_{i,u}\right|}{\left\|\dot{\Delta}_{i,u}\right\|} \le \max_i \left\|G_{i,1}\right\| = O_p(\sqrt{(\log N)/T}(NT)^{\frac{1}{q}}). \tag{B.4}$$

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

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

where the second inequality holds by Cauchy's inequality and the last equality is by Lemma B.7(iv) and Assumption 3.1(iv)

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

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

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

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

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

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

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

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

$$\left\|\dot{\Delta}_{t,v}\right\|^2 \lambda_{\min}\left(\frac{1}{N}\sum_{i\in[N]}\dot{\psi}_{it}\dot{\psi}_{it}'\right)$$

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

By the fact that

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

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

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

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

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 (B.5) and (B.6). It follows that

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

Proof of Statement (iii).

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

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

which implies

$$\max_{i,t}\left|\dot{\Theta}_{j,it}-\Theta_{j,it}^{0}\right|$$
$$\leq \max_{i}\left\|\dot{u}_{i,j}-O_{j}u_{i,j}^{0}\right\|\max_{t}\left\|\dot{v}_{t,j}-O_{j}v_{t,j}^{0}\right\|$$

$$+\max_i \left\| O_j u^0_{i,j} \right\| \max_t \left\| \dot{v}_{t,j} - O_j v^0_{t,j} \right\| + \max_i \left\| \dot{u}_{i,j} - O_j u^0_{i,j} \right\| \max_t \left\| O_j v^0_{t,j} \right\|$$

$$= O_p(\eta_{N,2}),$$

where the last equality combines results from Theorem 3.1(ii) and Lemma B.7(i). ■

B.1.3 Proof of Theorem 3.2

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

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

$$\bar{\dot{\Theta}}^{(1s)}_{j,i} = \frac{1}{s}\sum_{t=1}^{s}\dot{\Theta}_{j,it} = \frac{1}{s}\sum_{t=1}^{s}\left[\Theta^0_{j,it} + (\dot{\Theta}_{j,it} - \Theta^0_{j,it})\right] = \alpha^{(1)}_{g_i^{(1)},j} + \bar{\Delta}_{s,i}(j),$$

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

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

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

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

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

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

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

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

Obviously,

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

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 $\mathbb{P}(\exists s<T_1, L(s)-L(T_1)<0)\to 0$. By (B.9) and (B.10), we observe that

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

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 $0<\frac{1}{T}\leq\kappa_s\leq\frac{T_1-2}{T}\asymp 1$. We analyze the five terms in (B.11) in turn.

For $A_1(s)$, we have

$$\begin{aligned}
A_1(s) &= \frac{1}{pNT}\sum_{j\in[p]}\sum_{i\in[N]}\frac{(T_1-s)(T-T_1)}{T-s}(\alpha^{(1)}_{g_i^{(1)},j}-\alpha^{(2)}_{g_i^{(2)},j})^2 \\
&= \frac{T_1-s}{T-s}(1-\tau_T)\frac{1}{pN}\sum_{j\in[p]}\sum_{i\in[N]}(\alpha^{(1)}_{g_i^{(1)},j}-\alpha^{(2)}_{g_i^{(2)},j})^2 \\
&= \frac{T_1-s}{T-s}(1-\tau_T)\frac{1}{pN}\sum_{i\in[N]}\left\|\alpha^{(1)}_{g_i^{(1)}}-\alpha^{(2)}_{g_i^{(2)}}\right\|^2 \\
&= \kappa_s\frac{(1-\tau_T)}{1-\frac{s}{T}}D_{N\alpha}=\kappa_s O(\zeta_{NT}^2),
\end{aligned} \tag{B.12}$$

where $D_{N\alpha}:=\frac{1}{pN}\sum_{i\in[N]}\left\|\alpha^{(1)}_{g_i^{(1)}}-\alpha^{(2)}_{g_i^{(2)}}\right\|^2$ and the last equality holds by Assumption 3.6.

For $A_2(s)$, noting that

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

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

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

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

where the second last equality holds by the fact that

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

from Theorem 3.1(iii) and

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

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

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

we have

$$
\begin{aligned}
A_3(s) &= \frac{1}{pNT} \sum_{j\in[p]} \sum_{i\in[N]} \left\{ \sum_{t=s+1}^{T} \left[\Delta_{it}(j) - \bar{\Delta}_{s_+,i}(j)\right]^2 - \sum_{t=T_1+1}^{T} \left[\Delta_{it}(j) - \bar{\Delta}_{T_{1+},i}(j)\right]^2 \right\} \\
&= \frac{1}{pNT} \sum_{j\in[p]} \sum_{i\in[N]} \left\{ \sum_{t=s+1}^{T} \Delta_{it}^2(j) - (T-s)\bar{\Delta}_{s_+,i}^2(j) - \sum_{t=T_1+1}^{T} \Delta_{it}^2(j) + (T-T_1)\bar{\Delta}_{T_{1+},i}(j) \right\}
\end{aligned}
$$

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

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

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

Last, we notice that

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

where the first inequality holds by Cauchy-Schwarz inequality.

Combining (B.11), (B.12), (B.13), (B.16), (B.18) and Assumption 3.6(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 3.6(i). This implies that

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

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

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

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

It follows that

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

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

As in (B.8), we obtain that

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

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

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

$$+2\frac{T_1\,(s-T_1)}{s}(\alpha^{(1)}_{g_i^{(1)},j}-\alpha^{(2)}_{g_i^{(2)},j})\left[\frac{1}{T_1}\sum_{t=s+1}^{T_1}\Delta_{it}(j)-\frac{1}{s-T_1}\sum_{t=T_1+1}^{T}\Delta_{it}(j)\right]$$

and

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

It follows that

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

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

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

and $B_4(s)=\bar{\kappa}_sO_p(\eta_{N,2}\zeta_{NT})=\bar{\kappa}_so_p(\zeta^2_{NT})$. It follows that for any $s>T_1$,

$$\text{plim}_{(N,T)\to\infty}\frac{1}{\bar{\kappa}_s\zeta^2_{NT}}[L(s)-L(T_1)]=\text{plim}_{(N,T)\to\infty}\frac{T_1T}{Ts}\frac{1}{\zeta^2_{NT}}D_{N\alpha}\geq\tau D_\alpha>0.$$

This implies that

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

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

B.1.4 Proof of Theorem 3.3

By Theorem 3.2, $\mathbb{P}(\hat{T}_1 = T_1) \to 1$. It follows that we can prove Theorem 3.3 by conditioning on the event that $\{\hat{T}_1 = T_1\}$. Below we prove the theorem under the event that $\{\hat{T}_1 = T_1\}$.

First, we define $\dot{\Theta}_{j,i}^{0,(1)} = (\dot{\Theta}_{j,i1}, \cdots, \dot{\Theta}_{j,iT_1})'$, $\dot{\Theta}_{j,i}^{0,(2)} = (\dot{\Theta}_{j,i,T_1+1}, \cdots, \dot{\Theta}_{j,iT})'$, $\dot{\beta}_i^{0,(1)} = \frac{1}{\sqrt{T_1}}(\dot{\Theta}_{1,i}^{0,(1)\prime}, \cdots, \dot{\Theta}_{p,i}^{0,(1)\prime})'$, 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 (3.7) and (3.8), we further define

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

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

(i) Under the case with $m = K^{(\ell)}$, Theorem 3.3(i.a) is from the combination of Lemma B.9 for the consistency of the membership estimates via K-means algorithm and Theorem 3.2 for the consistency of the break point estimator.

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

$$\mathbb{P}\left(\hat{K}^{(\ell)} \le K^{(\ell)}\right) \ge \mathbb{P}\left(\hat{\Gamma}_{K^{(\ell)}}^{(\ell)} \le z_\varsigma\right),$$

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

$$\mathbb{P}\left(\hat{\Gamma}_{K^{(\ell)}}^{(\ell)} \le z_\varsigma\right) \ge 1 - \varsigma + o(1).$$

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 3.3(i.1), we notice that 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 $\mathscr{T}_1 = [T_1]$ and $\mathscr{T}_2 = [T] \backslash [T_1]$. Define $\mathscr{T}_{1,-1} = \mathscr{T}_1 \backslash \{T_1\}$, $\mathscr{T}_{2,-1} = \mathscr{T}_2 \backslash \{T\}$, $\mathscr{T}_{1,j} = \{1+j, \cdots, T_1\}$ and $\mathscr{T}_{2,j} = \{T_1+1+j, \cdots, T\}$ for some specific $j \in \mathscr{T}_{\ell,-1}$. Let

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

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

$$\hat{\bar{\theta}}^{0,(\ell)}_{k,K^{(\ell)}} = \frac{1}{|G_k^{(\ell)}|} \sum_{i\in G_k^{(\ell)}} \hat{\theta}^{0,(\ell)}_{i,k,K^{(\ell)}}, \qquad \hat{S}^{0,(\ell)}_{ii,k,K^{(\ell)}} = \frac{1}{T_\ell} (X_i^{(\ell)})' M_{\hat{F}^{0,(\ell)}_{k,K^{(\ell)}}} X_i^{(\ell)},$$

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

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

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

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

where

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

By Lemma B.20, we notice that $\hat{\Gamma}^{0,(\ell)}_{k,K^{(\ell)}} \rightsquigarrow \mathbb{N}(0,1)$ owing to the fact that the slope coefficient $\alpha_k^{(\ell)}$ is homogeneous across $i \in G_k^{(\ell)}$ $\forall k \in [K^{(\ell)}]$. Furthermore, $\left\{\hat{\Gamma}^{0,(\ell)}_{k,K^{(\ell)}},\ k \in [K^{(\ell)}]\right\}$ are asymptotically independent under Assumption 3.1(i). It follows that

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

where $\mathscr{Z}$ is the maximum of m independent $\chi^2(1)$ random variables. Then Theorem 3.3(i.b) follows.

(ii) When $m < K^{(\ell)}$, Theorem 3.3(i.1) does not hold and we can not change the estimated group membership $\hat{\mathscr{G}}^{(\ell)}_{K^{(\ell)}}$ to the true group membership $\mathscr{G}^{(\ell)}$. To get around of this issue, we define the "pseudo groups". For $m < K^{(\ell)}$, let $\mathbb{G}_m^{(\ell)} := \{G^{(\ell)}_{1,m}, \cdots, G^{(\ell)}_{m,m}\}$ such that $[N] = G^{(\ell)}_{1,m} \cup \cdots \cup G^{(\ell)}_{m,m}$, which indicates one possible partition of the set $[N]$. We further define $\mathscr{G}_m^{(\ell)}$ to be the collection of all possible $\mathbb{G}_m^{(\ell)}$.

By Theorem 3.3(i.c), we can conclude that $\mathbb{P}\left(\hat{K}^{(\ell)} \neq K^{(\ell)}\right) \leq \varsigma + o(1)$ provided we can show that $\hat{\Gamma}_m^{(\ell)} \to \infty$ when $m < K^{(\ell)}$. By Lemma B.10, we notice that $\hat{\mathscr{G}}_m^{(\ell)} \in$

$\mathbb{G}_m^{(\ell)}$ w.p.a.1. Conditioning on the event $\{\hat{\mathscr{G}}_m^{(\ell)} \in \mathbb{G}_m^{(\ell)}\} \cap \{\hat{T}_1 = T_1\}$, we have

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

where

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

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 $|\mathbb{G}_m^{(\ell)}| = m^{K^{(\ell)}}$ which is a constant since $K^{(\ell)}$ is a constant, we can show that $\hat{\Gamma}_m^{(\ell)} \to \infty$ by showing that $\hat{\Gamma}_m^{0,(\ell)}(\mathscr{G}_m^{(\ell)}) \to \infty$ for any possible realization $\mathscr{G}_m^{(\ell)}$. 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

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

such that

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

by Assumption 3.7(iii). Then $|\hat{\Gamma}_{k,m}^{0,(\ell)}(G_{k,m}^{(\ell)})| / (\log N)^{1/2} \to \infty$ for some $k \in [m]$ by Lemma B.21. By the definition of $\hat{\Gamma}_m^{0,(\ell)}(\mathscr{G}_m^{(\ell)})$, we have $\hat{\Gamma}_m^{0,(\ell)}(\mathscr{G}_m^{(\ell)}) / (\log N)^{1/2} \to \infty$, which yields $\hat{\Gamma}_m^{(\ell)} / \log N \to \infty$ w.p.a.1 for $m < K^{(\ell)}$ and $\mathbb{P}(\hat{K}^{(\ell)} \neq K^{(\ell)}) \leq \varsigma + o(1)$ as z_ς diverges to infinity at rate $\log N$ as $\varsigma = \varsigma_N \to 0$ at some rate N^{-c} for some $c > 0$. ■

B.1.5 Proof of Theorem 3.4

To show Theorem 3.4, we can directly derive the asymptotic distribution for the oracle estimator $\hat{\alpha}_k^{*(\ell)}$ by combining Theorems 3.2 and 3.3.

The asymptotic distribution theory for 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 3.1* and 3.2–3.9, which is for the dynamic linear panel model, Theorem 3.4 extends Theorem 4.3 in Moon and Weidner (2017) to allow for multiple groups.

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 $\mathbb{C}_{NT,k,j}^{(\ell)} = \mathbb{C}_1(\Lambda_k^{0,(\ell)}, F^{0,(\ell)}, \mathbb{X}_{j,k}^{(\ell)}, E_k^{(\ell)}) + \mathbb{C}_2(\Lambda_k^{0,(\ell)}, F^{0,(\ell)}, \mathbb{X}_{j,k}^{(\ell)}, E_k^{(\ell)})$, where

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

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

By Lemma B.11, we have $\sqrt{N_k^{(\ell)} T_\ell}(\hat{\alpha}_k^{(\ell)} - \alpha_k^{(\ell)}) = \mathbb{W}_{NT,k}^{(\ell)-1} \mathbb{C}_{NT,k}^{(\ell)} + o_p(1)$.

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 3.1. For Lemma B.1, we need to extend it to allow for serially correlated errors in non-dynamic panels in Lemma B.12. Hence, by Lemma B.12 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

[1] It is well known that one cannot allow for serially correlated errors in dynamic panels in general to avoid the endogeneity issue.

readily show that

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

which yields the final distributional results in Theorem 3.4 by stacking all subgroups of parameter estimators into a large vector and resorting to the Cramer-Wold device.

B.1.6 Proof of Theorem 3.5

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

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

It follows that

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

where the last line is by Lemma B.7(i) and Theorem 3.1(ii).

B.2 Technical Lemmas

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

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

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 $\max_i\|A_i\|\leq\sqrt{m}$ almost surely and $\max_i\|\Sigma_i\|_{op}\leq M$ for some positive constant M.*

Then for every $t > 0$, with probability $1 - 2T\exp\left(-c_1 t^2\right)$, we have

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

where c_1 is an absolute constant.

Proof The proof follows the arguments as used in the proof of Theorem 5.41 in Vershynin (2010). Define $Z_i := \frac{1}{N}\left(A_i A_i' - \Sigma_i\right) \in \mathbb{R}^{T\times T}$, 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 follows:

$$\|Z_i\|_{op} \le \frac{1}{N}\left(\left\|A_iA_i'\right\|_{op} + \|\Sigma_i\|_{op}\right) \le \frac{1}{N}\left(\|A_i\|_2^2 + \|\Sigma_i\|_{op}\right) \le \frac{m+M}{N} \quad a.s. \tag{B.23}$$

uniformly over i. Moreover, note that

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

and

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

We then obtain that

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

uniformly over i, and

$$\left\|\sum_{i\in[N]}\mathbb{E}\left(Z_i^2\right)\right\|_{op} \le N\max_i \left\|\mathbb{E}\left(Z_i^2\right)\right\|_{op} \le \frac{mM+M^2}{N} \quad a.s. \tag{B.24}$$

Define $\varepsilon = \max(\sqrt{M}\delta, \delta^2)$ with $\delta = t\sqrt{\frac{m+M}{N}}$. Combining (B.23) and (B.24), by matrix Bernstein's inequality, we have

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

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

$$\left\|\frac{1}{N}A'A-\frac{1}{N}\sum_{i\in[N]}\Sigma_i\right\|_{op}\leq\max\left(\sqrt{M}\delta,\delta^2\right) \tag{B.25}$$

with probability $1-\exp\left(-ct^2\right)$. Combining the fact that $\|\Sigma_i\|_{op}\leq M$ uniformly over i and (B.25), we show that

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

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

Lemma B.3. *Recall that* $X_j=\left\{X_{j,it}\right\}$ *and* $E=\{e_{it}\}$. *Under Assumption 3.1,* $\forall j\in[p]$, *we have* $\left\|X_j\odot E\right\|_{op}=O_p(\sqrt{N}+\sqrt{T\log T})$ *and* $\|E\|_{op}=O_p(\sqrt{N}+\sqrt{T\log T})$.

Proof We focus on $\left\|X_j\odot E\right\|_{op}$ as the result for $\|E\|_{op}$ can be derived in the same manner. We first note that, conditional on $\left\{V_j^0\right\}_{j\in[p]\cup\{0\}}$, the rows of $X_j\odot E$ 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

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

where the second equality holds by Assumption 3.1(ii). Therefore, to apply Lemma B.2 conditionally on $\mathscr{D}$, we only need to upper bound $\|A_i\|$ and $||\mathbb{E}[A_iA_i'\big|\mathscr{D}]||_{op}$.

First, under Assumption 3.1, we have $\frac{1}{T}\sum_{t\in[T]}\left(X_{j,it}e_{it}\right)^2\leq C\,a.s.$ by Assumption

3.1(iv), which implies

$$||A_i|| = \left\| X_{j,i} \odot E_i \right\| \leq C\sqrt{T} \; a.s. \tag{B.26}$$

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

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

By Davydov's inequality for conditional strong mixing sequence (e.g., Lemma 4.3 in Su and Chen (2013b)), we can show that

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

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

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

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

$$\max_i \|\Sigma_i\|_{op} \leq \sqrt{\|\Sigma_i\|_1 \|\Sigma_i\|_\infty} \leq c_2 \; a.s. \tag{B.27}$$

Combining (B.26)-(B.27) and using Lemma B.2 with $t = \sqrt{\log T}$, we obtain the desired result. ■

Recall that

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

Lemma B.4. *Suppose Assumptions 3.1–3.3 hold. Then* $\{\tilde{\Delta}_{\Theta_j}\}_{j\in[p]\cup\{0\}} \in \mathscr{R}(3)$ *w.p.a.1.*

Proof Let A^c denote the complement of A. Define event

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

Then exists a positive constant c_3 such that $\mathbb{P}(\mathscr{A}_{1,N}^c(c_3)) \leq \varepsilon$ for any $\varepsilon > 0$ by Lemma B.3. Under event $\mathscr{A}_{1,N}(c_3)$, by the definition of $\tilde{\Theta}_j$ in (3.3), we notice that

$$0 \leq \frac{1}{NT}\left\|Y - \Theta_0^0 - \sum_{j\in[p]} X_j \odot \Theta_j^0\right\|^2 - \frac{1}{NT}\left\|Y - \tilde{\Theta}_0 - \sum_{j\in[p]} X_j \odot \tilde{\Theta}_j\right\|^2$$
$$+ \sum_{j\in[p]\cup\{0\}} \nu_j \left(\|\Theta_j^0\|_* - \|\tilde{\Theta}_j\|_*\right) \tag{B.28}$$

and

$$\frac{1}{NT}\left\|Y - \Theta_0^0 - \sum_{j\in[p]} X_j \odot \Theta_j^0\right\|^2 - \frac{1}{NT}\left\|Y - \tilde{\Theta}_0 - \sum_{j\in[p]} X_j \odot \tilde{\Theta}_j\right\|^2$$
$$= \frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\left\{e_{it}^2 - \left[e_{it} - \left(\tilde{\Delta}_{\Theta_0,it} + \sum_{j\in[p]} X_{j,it}\tilde{\Delta}_{\Theta_j,it}\right)\right]^2\right\}$$
$$= \frac{2}{NT}tr\left(E'\tilde{\Delta}_{\Theta_0}\right) + \sum_{j\in[p]}\frac{2}{NT}tr\left(\left(E\odot X_j\right)'\tilde{\Delta}_{\Theta_j}\right) - \frac{1}{NT}\sum_{i\in[N]}\sum_{t\in[T]}\left(\tilde{\Delta}_{\Theta_0,it} + \sum_{j\in[p]} X_{j,it}\tilde{\Delta}_{\Theta_j,it}\right)^2$$
$$\leq \frac{2}{NT}\left|tr\left(E'\tilde{\Delta}_{\Theta_0}\right)\right| + \sum_{j\in[p]}\frac{2}{NT}\left|tr\left(\left(E\odot X_j\right)'\tilde{\Delta}_{\Theta_j}\right)\right|$$
$$\leq \frac{2}{NT}\|E\|_{op}\|\tilde{\Delta}_{\Theta_0}\|_* + \sum_{j\in[p]}\frac{2}{NT}\|E\odot X_j\|_{op}\|\tilde{\Delta}_{\Theta_j}\|_*$$
$$\leq 2c_3\sum_{j\in[p]\cup\{0\}}\frac{\sqrt{N}\vee\sqrt{T\log T}}{NT}\|\tilde{\Delta}_{\Theta_j}\|_*, \tag{B.29}$$

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

Combining (B.28) and (B.29), we have

$$0 \leq \sum_{j\in[p]\cup\{0\}}\left\{\frac{2c_3(\sqrt{N}\vee\sqrt{T\log T})}{NT}\|\tilde{\Delta}_{\Theta_j}\|_* + \nu_j\left(\|\Theta_j^0\|_* - \|\tilde{\Theta}_j\|_*\right)\right\} \; w.p.a.1. \tag{B.30}$$

Besides, we can show that

$$\|\tilde{\Theta}_j\|_* = \|\tilde{\Delta}_{\Theta_j} + \Theta_j^0\|_* = \left\|\Theta_j^0 + \mathscr{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j}) + \mathscr{P}_j(\tilde{\Delta}_{\Theta_j})\right\|_*$$
$$\geq \left\|\Theta_j^0 + \mathscr{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_* - \|\mathscr{P}_j(\tilde{\Delta}_{\Theta_j})\|_*$$
$$= \|\Theta_j^0\|_* + \left\|\mathscr{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_* - \|\mathscr{P}_j(\tilde{\Delta}_{\Theta_j})\|_*, \tag{B.31}$$

where the second equality holds by Lemma D.2(i) in Chernozhukov et al. (2019), the first inequality is by triangle inequality and the last equality is by the construction of the spaces $\mathscr{P}_j^{\perp}$ and $\mathscr{P}_j$. Then combining (B.30) and (B.31), w.p.a.1, we have

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

and

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

If we set $\nu_j = \frac{4c_3(\sqrt{N}\vee\sqrt{T\log T})}{NT}$, we obtain the final result $\sum_{j\in[p]\cup\{0\}} \left\|\mathscr{P}_j^{\perp}(\tilde{\Delta}_{\Theta_j})\right\|_* \leq 3\sum_{j\in[p]\cup\{0\}} \|\mathscr{P}_j(\tilde{\Delta}_{\Theta_j})\|_*$. ■

Lemma B.5. *Consider a sequence of random variables $\{B_i, i\in[n]\}$.*

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

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

(ii) Suppose $\{B_i, i\in[n]\}$ is an m.d.s. 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 $\left|\sum_{i\in[n]}\mathbb{E}_{i-1}(B_i^2)\right|\leq\sigma^2$. 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 the partial sum of an independent sequence and the Freedman inequality for the partial sum of an 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) $\mathbb{P}\left\{\left|\sum_{t=1}^{T}\Upsilon_t\right| > \varepsilon\right\} \leq \exp\left\{-\frac{c_4\varepsilon^2}{\upsilon_0^2T+M_T^2+\varepsilon M_T(\log T)^2}\right\}$,

where $\upsilon_0^2 = \sup_{t\in[T]}[Var(\Upsilon_t)+2\sum_{s>t}|Cov(\Upsilon_t,\Upsilon_s)|]$.

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 3.1–3.4 hold, for* $j \in \{0,\cdots,p\}$*, we have*

(i) $\max_i \left\|u_{i,j}^0\right\| \leq M$ *and* $\max_t \left\|v_{t,j}^0\right\| \leq \frac{M}{\sigma_{K_j,j}} \leq \frac{M}{c_\sigma}$,

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

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

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

Proof (i) Recall that $\frac{1}{\sqrt{NT}}\Theta_j^0 = \mathscr{U}_j^0\Sigma_j^0\mathscr{V}_j^{0\prime}$, $U_j^0 = \sqrt{N}\mathscr{U}_j^0\Sigma_j^0$ and $V_j = \sqrt{T}\mathscr{V}_j^0$. Let $[A]_{i\cdot}$ and $[A]_{\cdot t}$ denote the i-th row and t-th column of A, respectively. Note that

$$\frac{1}{\sqrt{T}}\Theta_j^0\mathscr{V}_j^0 = \sqrt{N}\mathscr{U}_j^0\Sigma_j^0 = U_j^0, \text{ and } \frac{1}{\sqrt{N}}\mathscr{U}_j^{0\prime}\Theta_j^0 = \sqrt{T}\Sigma_j^0\mathscr{V}_j^{0\prime} = \Sigma_j^0V_j^{0\prime}. \quad \text{(B.32)}$$

Hence, it's easy to see that

$$\|u_{i,j}^0\| = \frac{1}{\sqrt{T}}\left\|\left[\Theta_j^0\mathscr{V}_j^0\right]_{i\cdot}\right\| \leq \frac{1}{\sqrt{T}}\left\|\left[\Theta_j\right]_{i\cdot}\right\| \leq M,$$

where the first inequality is due to the fact that $\mathscr{V}_j$ is the unitary matrix and the last inequality holds by Assumption 3.2. Since the upper bound M is not dependent on i, this result holds uniformly. Analogously, we see that

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

(ii) As in (B.32), we have

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

and

$$\|O_j'\tilde{v}_{t,j}\| \leq \frac{1}{\sqrt{N}}\frac{1}{\tilde{\sigma}_{K_j,j}^{(1)}}\left\|\left[\tilde{\mathscr{U}}_j^{(1)\prime}\tilde{\Theta}_j\right]_{\cdot t}\right\| \leq \frac{1}{\sqrt{N}}\frac{1}{\tilde{\sigma}_{K_j,j}^{(1)}}\left\|\left[\tilde{\Theta}_j\right]_{\cdot t}\right\| \leq \frac{2M}{c_\sigma},$$

where the last inequality holds due to the constrained optimization in (3.3) and the fact that $\max_{k\in[K_j]}|\tilde{\sigma}_{k,j}^{-1}-\sigma_{k,j}^{-1}|\leq\sigma_{K_j,j}^{-1}$ w.p.a.1.

(iii) Note that

$$\begin{aligned}
\max_i\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\phi}_{it}\right\|^2 &\leq \max_i\left\{\frac{1}{T}\sum_{t\in[T]}\left\|O_0'\tilde{v}_{t,0}\right\|^2+\sum_{j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left\|O_j'\tilde{v}_{t,j}\right\|^2\left|X_{j,it}\right|^2\right\}\\
&\leq \max_{t\in[T],j\in[p]\cup\{0\}}\left\|O_j'\tilde{v}_{t,j}\right\|^2\left\{1+\max_i\sum_{j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left|X_{j,it}\right|^2\right\}\\
&\leq \tfrac{4M^2}{c_\sigma^2}(1+pC)
\end{aligned}$$

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

(iv) Note that

$$\begin{aligned}
&\max_i\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{\phi}_{it}-\phi_{it}^0\right\|^2\\
&\leq\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{O}_0^{(1)\prime}\tilde{v}_{t,0}-v_{t,0}^0\right\|^2+p\max_{t\in[T],j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left|X_{j,it}\right|^2\left\|\tilde{O}_j^{(1)\prime}\tilde{v}_{t,j}-v_{t,j}^0\right\|^2\\
&\lesssim\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{O}_0^{(1)\prime}\tilde{v}_{t,0}-v_{t,0}^0\right\|^2+p(NT)^{2/q}\max_{j\in[p]}\frac{1}{T}\sum_{t\in[T]}\left\|\tilde{O}_j^{(1)\prime}\tilde{v}_{t,j}-v_{t,j}^0\right\|^2\\
&=\frac{1}{T}\left\|O_0\tilde{V}_0-V_0^0\right\|^2+p(NT)^{2/q}\max_{j\in[p]}\frac{1}{T}\left\|O_j\tilde{V}_j-V_j^0\right\|^2=O_p(\eta_{N,1}^2(NT)^{2/q}),
\end{aligned}$$

where the second inequality is by Assumption 3.1(v) and the last equality holds by Theorem 3.1(ii). ■

Lemma B.8. *Under Assumptions 3.1–3.5, we have* $\min_{i\in[N]}\lambda_{\min}(\tilde{\Phi}_i)\geq\frac{c_\phi}{2}$ *w.p.a.*1*, and* $\min_{t\in[T]}\lambda_{\min}(\tilde{\Psi}_t)\geq\frac{c_\phi}{2}$ *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 $\phi_{it}^0=[v_{t,0}^{0\prime},v_{t,1}^{0\prime}X_{1,it},\cdots,v_{t,p}^{0\prime}X_{p,it}]'$ 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

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

where the last equality holds by Lemma B.7(i) and Assumption 3.1(iv). It follows that

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

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

Lemma B.9. *Under Assumptions 3.1–3.7, we have* $\max_i \mathbf{1}\{\hat{g}_{i,K^{(\ell)}}^{0,(\ell)} \neq g_i^{(\ell)}\} = 0$ *w.p.a.1, where* $\hat{g}_{i,K^{(\ell)}}^{0,(\ell)}$ *is defined in* (B.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 $\alpha_k^{(\ell)} = (\alpha_{k,1}^{(\ell)}, \cdots, \alpha_{k,p}^{(\ell)})'$. Then we have

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

and

$$\max_{k\in[K^{(\ell)}]} \left\| \frac{1}{\sqrt{T_\ell}} \alpha_k^{(\ell)} \otimes \iota_{T_\ell} \right\| = \max_{k\in[K^{(\ell)}]} \frac{1}{\sqrt{T_\ell}} \sqrt{T_\ell \sum_{j=1}^{p} (\alpha_{k,j}^{(\ell)})^2} \leq \sqrt{p} \max_{k\in[K^{(\ell)}], j\in[p]} |\alpha_{k,j}^{(\ell)}| \leq \sqrt{p} M, \tag{B.33}$$

where the last inequality is due to Assumption 3.2.

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

$$\max_i \left\| \dot{\beta}_i^{0,(\ell)} - \beta_i^{0,(\ell)} \right\| = \frac{1}{\sqrt{T_\ell}} \max_i \left\| \dot{\Theta}_i^{(\ell)} - \Theta_i^{0,(\ell)} \right\| = \frac{1}{\sqrt{T_\ell}} \max_i \sqrt{\sum_{j=1}^{p} \sum_{t\in\mathscr{T}_\ell} (\dot{\Theta}_{j,it}^{(\ell)} - \Theta_{j,it}^{0,(\ell)})^2}$$

$$\leq \sqrt{p} \max_{j\in[p], i\in[N], t\in[T]} \left| \dot{\Theta}_{j,it} - \Theta_{j,it}^0 \right| \leq c_5 \eta_{N,2} \quad w.p.a.1, \tag{B.34}$$

with c_5 being some positive large enough constant, and the last inequality holds by Theorem 3.1(iii).

Third, we also observe that

$$\min_{1\leq k_s < k_{s^*} \leq K^{(\ell)}} \frac{1}{\sqrt{T_\ell}} \left\| \alpha_{k_s}^{(\ell)} \otimes \iota_{T_\ell} - \alpha_{k_{s^*}}^{(\ell)} \otimes \iota_{T_\ell} \right\| = \min_{1\leq k_s < k_{s^*} \leq K^{(\ell)}} \sqrt{\sum_{j=1}^{p} (\alpha_{k_s,j}^{(\ell)} - \alpha_{k_{s^*},j}^{(\ell)})^2} \geq C_5, \tag{B.35}$$

where the last inequality holds by Assumption 3.7(i).

Combining (B.33), (B.34) and (B.35), we obtain that $\mathbb{P}\left(\max_i \mathbf{1}\{\hat{g}_{i,K^{(\ell)}}^{0,(\ell)} \neq g_i^{(\ell)}\} = 0\right) \to 1$ 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 3.7, 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.1. *Fix $K^{(\ell)} > 1$ and $1 < m \le K^{(\ell)}$. Define a $K^{(\ell)} \times p$ matrix $\alpha^{(\ell)} = (\alpha_1^{(\ell)}, \cdots, \alpha_K^{(\ell)})'$. Let $d_{K^{(\ell)}}(\alpha^{(\ell)})$ be the minimum pairwise distance of all $K^{(\ell)}$ rows and $\alpha_k^{(\ell)}$ and $\alpha_l^{(\ell)}$ be the pair that satisfies $||\alpha_k^{(\ell)} - \alpha_l^{(\ell)}|| = d_{K^{(\ell)}}(\alpha^{(\ell)})$ (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.10. *Recall that $\hat{\mathscr{G}}_m^{(\ell)}$ is the estimated group structure from K-means algorithm with m groups. Under Assumptions 3.1-3.7 and the event $\{\hat{T}_1 = T_1\}$, w.p.a.1, for each $1 < m < K^{(\ell)}$, $\hat{\mathscr{G}}_m^{(\ell)}$ enjoys the NSP defined in Definition 3.1.*

Proof By Theorem 4.1 in Jin et al. (2022), Lemma B.10 is proved if we ensure all conditions in their Theorem 4.1 hold. We now apply their Theorem 4.1 with $\hat{x}_i = \dot{\beta}_i^{0,(\ell)}$, $x_i = \beta_i^{0,(\ell)}$ and $u_k = \frac{1}{\sqrt{T_\ell}}\alpha_k^{(\ell)} \otimes \iota_{T_\ell}$ for $k \in [K^{(\ell)}]$. By the definition of $d_m\left(\alpha^{(\ell)}\right)$ in Definition B.1, we notice that $d_m\left(\alpha^{(\ell)}\right) \ge d_{K^{(\ell)}}\left(\alpha^{(\ell)}\right)$ such that $d_{K^{(\ell)}}\left(\alpha^{(\ell)}\right) \ge C_5$ by Assumption 3.7(i). With (B.34) shown above and Assumption 3.2, we have

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

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

Lemma B.11. *Under Assumptions 3.1, 3.6(ii), 3.7(ii), 3.8 and 3.9(i)-(iii), for $\ell = \{1,2\}$ and $k \in [K^{(\ell)}]$, we have $\hat{\alpha}_k^{(\ell)} \xrightarrow{p} \alpha_k^{(\ell)}$ and*

$$\sqrt{N_k^{(\ell)} T_\ell}(\hat{\alpha}_k^{(\ell)} - \alpha_k^{(\ell)}) = \mathbb{W}_{NT,k}^{(\ell)-1}\mathbb{C}_{NT,k}^{(\ell)} + o_p(1).$$

Proof The result in the lemma combines those in Theorem 4.1 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 our Assumptions 3.8 and

3.9(ii)-(iii) are the same as their Assumptions 1 and 4.

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

$$\frac{1}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathscr{T}_\ell}X_{j,it}e_{it}\xrightarrow{p}0,\ \forall k\in[K^{(\ell)}],\ \ell\in\{1,2\}.$$

Fix a specific k and ℓ. We can show that

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathscr{T}_\ell}X_{j,it}e_{it}\middle|\mathscr{D}\right)^2\\
&=\frac{1}{\left(N_k^{(\ell)}T_\ell\right)^2}\sum_{i_1\in G_k^{(\ell)}}\sum_{i_2\in G_k^{(\ell)}}\sum_{t_1\in\mathscr{T}_\ell}\sum_{t_2\in\mathscr{T}_\ell}\mathbb{E}\left(X_{j,i_1t_1}X_{j,i_2t_2}e_{i_1t_1}e_{i_2t_2}|\mathscr{D}\right)\\
&=\frac{1}{\left(N_k^{(\ell)}T_\ell\right)^2}\sum_{i\in G_k^{(\ell)}}\sum_{t_1\in\mathscr{T}_\ell}\sum_{t_2\in\mathscr{T}_\ell}\mathbb{E}\left(X_{j,it_1}X_{j,it_2}e_{it_1}e_{it_2}|\mathscr{D}\right)\\
&=\frac{1}{\left(N_k^{(\ell)}T_\ell\right)^2}\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathscr{T}_\ell}\mathbb{E}\left(X_{j,it}^2e_{it}^2|\mathscr{D}\right)\\
&+\frac{2}{\left(N_k^{(\ell)}T_\ell\right)^2}\sum_{i\in G_k^{(\ell)}}\sum_{t_1\in\mathscr{T}_\ell}\sum_{t_2\in\mathscr{T}_\ell,t_2>t_1}\mathbb{E}\left(X_{j,it_1}X_{j,it_2}e_{it_1}e_{it_2}|\mathscr{D}\right)\\
&\leq\frac{M}{N_k^{(\ell)}T_\ell}+\frac{16}{\left(N_k^{(\ell)}T_\ell\right)^2}\max_{i\in G_k^{(\ell)}}\max_{t\in\mathscr{T}_L}\left(\mathbb{E}\left|X_{j,it}e_{it}\right|^q\right)^{2/q}\sum_{i\in G_k^{(\ell)}}\sum_{t_1\in\mathscr{T}_\ell}\sum_{t_2\in\mathscr{T}_\ell,t_2>t_1}[\alpha(t_2-t_1)]^{1-2/q}\\
&=O\left(\frac{1}{NT}\right),
\end{aligned}\tag{B.36}$$

where the second equality holds by Assumption 3.1(i) with the conditional independence sequence for $i_1\neq i_2$, the first inequality combines Assumption 3.1(ii), (iii), (v), and the Davydov's inequality for strong mixing sequence in Lemma 4.3, Su and Chen (2013b), and the last equality is by Assumption 3.1(iii), (v), Assumption 3.6(ii) and Assumption 3.7(ii). Following this, it yields that

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

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

$$\left\|E_k^{(\ell)}\right\|_{op}=O_p(\sqrt{N}+\sqrt{T\log T}),\tag{B.37}$$

which, in conjunction with Assumption 3.9(i), implies that Assumption 3* in Moon and Weidner (2017) is satisfied. ∎

For $j\in[p]$, recall that $X_{j,i}^{(1)}=\left(X_{j,i1},\cdots,X_{j,iT_1}\right)'$, $X_{j,i}^{(2)}=\left(X_{j,i(T_1+1)},\cdots,X_{j,iT}\right)'$,

$e_i^{(1)} = (e_{i1}, \cdots, e_{iT_1})'$, $e_i^{(2)} = \left(e_{i(T_1+1)}, \cdots, e_{iT}\right)'$, $\tilde{X}_{j,it} = X_{j,it} - \mathbb{E}\left(X_{j,it}|\mathscr{D}\right)$. 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

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

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

Lemma B.12. *Under Assumptions 3.1, 3.2, 3.6(ii), 3.7(ii), 3.8 and 3.9, 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) $\frac{1}{N_k^{(\ell)} T_\ell} \sum_{i \in G_k^{(\ell)}} \sum_{t \in \mathscr{T}_\ell} \left[e_{it}^2 \mathfrak{X}_{k,it}^{(\ell)} \mathfrak{X}_{k,it}^{(\ell)\prime} - \mathbb{E}\left(e_{it}^2 \mathfrak{X}_{k,it}^{(\ell)} \mathfrak{X}_{k,it}^{(\ell)\prime}|\mathscr{D}\right)\right] = o_p(1)$,

(x) $\frac{1}{N_k^{(\ell)} T_\ell} \sum_{i \in G_k^{(\ell)}} \sum_{t \in \mathscr{T}_\ell} e_{it}^2 \left(\mathfrak{X}_{k,it}^{(\ell)} \mathfrak{X}_{k,it}^{(\ell)\prime} - \mathscr{X}_{it} \mathscr{X}_{it}'\right) = o_p(1)$.

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

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

where the fourth line is by Lemma B.7(i) and the last line combines Assumption 3.1(v) and Davydov's inequality for conditional strong mixing sequences, similarly as (B.36). It follows that

$$\left\|P_{F^{0,(\ell)}}E_k^{(\ell)\prime}P_{\Lambda_k^{0,(\ell)}}\right\|$$
$$\leq \left\|F^{0,(\ell)}\right\|\left\|F^{0,(\ell)\prime}F^{0,(\ell)}\right\|\left\|F^{0,(\ell)\prime}E_k^{(\ell)\prime}\Lambda_k^{0,(\ell)}\right\|\left\|(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)})^{-1}\right\|\left\|\Lambda_k^{0,(\ell)\prime}\right\|$$
$$= O(T^{1/2})O_p(T^{-1})O_p(\sqrt{NT})O_p(N^{-1})O(N^{1/2}) = O_p(1), \tag{B.38}$$

where the first equality holds by Assumptions 3.2 and 3.8.

Moreover, we have

$$\left\|P_{\Lambda_k^{0,(\ell)}}\tilde{\mathbb{X}}_{j,k}^{(\ell)}\right\| \leq \left\|\Lambda_k^{0,(\ell)}\right\|\left\|(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)})^{-1}\right\|\left\|\Lambda_k^{0,(\ell)\prime}\tilde{\mathbb{X}}_{j,k}^{(\ell)}\right\| = O(N^{1/2})O_p(N^{-1})O_p(\sqrt{NT}) = O_p(T^{1/2}), \tag{B.39}$$

where the first equality holds by Assumptions 3.2 and 3.8(i) and the fact that

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

Then we are ready to show that

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

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

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

where the last line holds by the fact that

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

where the second equality is by Assumption 3.1(i) and the last line holds by Assumption 3.1(iii) and (v), and Davydov's inequality.

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

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

where the last equality holds by Assumption 3.9(iv). It follows that

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

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

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

where the last line combines (B.37),(B.38), the fact that $||\mathbb{X}^{(\ell)}_{j,k}|| = O_p((NT)^{1/2})$ by Assumption 3.8(ii), and $||F^{0,(\ell)}(F^{0,(\ell)\prime}F^{0,(\ell)})^{-1}(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)})^{-1}\Lambda_k^{0,(\ell)\prime}|| = O_p((NT)^{-1/2})$ by Assumptions 3.2 and 3.8(i).

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

(vi) First, we note that

$$\mathbb{E}\left(\left\|\Lambda_k^{0,(\ell)\prime}E_k^{(\ell)}\mathbb{X}^{(\ell)\prime}_{j,k}\right\|\middle|\mathscr{D}\right)$$

$$
\begin{aligned}
&= \mathbb{E}\left[\sum_{j_1=1}^{r_0}\sum_{m\in G_k^{(\ell)}}\left(\sum_{i_1\in G_k^{(\ell)}}\sum_{t_1\in\mathscr{T}_\ell}\lambda^0_{i_1,j_1}e_{i_1t_1}X_{j,mt_1}\right)^2\middle|\mathscr{D}\right] \\
&= \sum_{j_1=1}^{r_0}\sum_{m\in G_k^{(\ell)}}\sum_{i_1\in G_k^{(\ell)}}\sum_{i_2\in G_k^{(\ell)}}\sum_{t_1\in\mathscr{T}_\ell}\sum_{t_2\in\mathscr{T}_\ell}\lambda^0_{i_1,j_1}\lambda^0_{i_2,j_1}\mathbb{E}\left(e_{i_1t_1}X_{j,mt_1}e_{i_2t_2}X_{j,mt_2}|\mathscr{D}\right) \\
&= \sum_{j_1=1}^{r_0}\sum_{m\in G_k^{(\ell)}}\sum_{i_1\in G_k^{(\ell)}}\sum_{t_1\in\mathscr{T}_\ell}\sum_{t_2\in\mathscr{T}_\ell}\left(\lambda^0_{i_1,j_1}\right)^2\mathbb{E}\left(e_{i_1t_1}X_{j,mt_1}e_{i_1t_2}X_{j,mt_2}|\mathscr{D}\right) \\
&\lesssim \sum_{m\in G_k^{(\ell)}}\sum_{i_1\in G_k^{(\ell)}}\sum_{t\in\mathscr{T}_\ell}\mathbb{E}\left(e^2_{i_1t}X^2_{j,mt}|\mathscr{D}\right) \\
&+2\sum_{m\in G_k^{(\ell)}}\sum_{i_1\in G_k^{(\ell)}}\sum_{t_1\in\mathscr{T}_\ell}\sum_{t_2\in\mathscr{T}_\ell,t_2>t_1}\left|Cov\left(e_{i_1t_1}X_{j,mt_1},e_{i_1t_2}X_{j,mt_2}|\mathscr{D}\right)\right| \\
&= O_p(N^2T),
\end{aligned}
$$

which leads to the result that

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

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

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

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

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

where the last equality holds by the fact that

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

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

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

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

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

(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.39), we can also show that $||\tilde{\mathbb{X}}_{j,k}^{(\ell)}P_{F^{0,(\ell)}}||=O_p(N^{1/2})$. 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. ■

B.3 Estimation of Panels with IFEs and Heterogeneous Slopes

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

$$Y_{it}=\begin{cases}\lambda_i^{0\prime}f_t^0+X_{it}'\theta_i^{0,(1)}+e_{it},\ t\in\{1,\cdots,T_1\},\\ \lambda_i^{0\prime}f_t^0+X_{it}'\theta_i^{0,(2)}+e_{it},\ t\in\{T_1+1,\cdots,T\}.\end{cases}\tag{B.40}$$

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

Let $X_i^{(1)} = (X_{i1}, \cdots, X_{iT_1})'$, $X_i^{(2)} = \left(X_{i(T_1+1)}, \cdots, X_{iT}\right)'$, $e_i^{(1)} = (e_{i1}, \cdots, e_{iT_1})'$, $e_i^{(2)} = \left(e_{i(T_1+1)}, \cdots, e_{iT}\right)'$, $F^{0,(1)} = \left(f_1^0, \cdots, f_{T_1}^0\right)'$, and $F^{0,(2)} = \left(f_{T_1+1}^0, \cdots, f_T^0\right)'$. 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

$$\left(\left\{\hat{\theta}_i^{(\ell)}\right\}_{i\in\mathscr{N}}, \hat{F}^{(\ell)}\right) = \underset{F^{(\ell)},\{\theta_i\}_{i\in\mathscr{N}}}{\arg\min} \frac{1}{nT_\ell} \sum_{i=n_1}^{n_n} \left(Y_i^{(\ell)} - X_i^{(\ell)}\theta_i\right)' M_{F^{(\ell)}} \left(Y_i^{(\ell)} - X_i^{(\ell)}\theta_i\right), \tag{B.41}$$

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 (B.41) solve the following nonlinear system of equations:

$$\hat{\theta}_i^{(\ell)} = \left(X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} X_i^{(\ell)}\right)^{-1} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} Y_i^{(\ell)}, \tag{B.42}$$

$$\left[\frac{1}{nT_\ell} \sum_{i\in\mathscr{N}} \left(Y_i^{(\ell)} - X_i^{(\ell)}\hat{\theta}_i^{(\ell)}\right)' \left(Y_i^{(\ell)} - X_i^{(\ell)}\hat{\theta}_i^{(\ell)}\right)\right] \hat{F}^{(\ell)} = \hat{F}^{(\ell)} \hat{V}_{NT}^{(\ell)}, \tag{B.43}$$

where $\hat{V}_{NT}^{(\ell)}$ is a diagonal matrix that contains the r_0 largest eigenvalues of the matrix in the square brackets in (B.43). 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 $\hat{\Lambda}_n^{(\ell)} := (\hat{\lambda}_{n_1}^{(\ell)}, \cdots, \hat{\lambda}_{n_n}^{(\ell)})'$, 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 $\bar{\theta}^{0,(\ell)} = \frac{1}{n}\sum_{i\in\mathscr{N}} \theta_i^{0,(\ell)}$. Here, we consider testing the slope homogeneity for $i \in \mathscr{N}$. The null and alternative hypotheses are respectively given by

$$H_0 : c_i^{(\ell)} = 0 \ \forall i \in \mathscr{N} \text{ and } H_1 : c_i^{(\ell)} \neq 0 \text{ for some } i \in \mathscr{N}.$$

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

$$\hat{\Gamma}^{(\ell)} = \sqrt{n} \cdot \frac{\frac{1}{n}\sum_{i\in\{n_1\cdots,n_n\}} \hat{\mathbb{S}}_i^{(\ell)} - p}{\sqrt{2p}} \tag{B.44}$$

where

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

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

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

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

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.

B.4 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 3.1* where the error process is an m.d.s.. If we focus on Assumption 3.1, we can obtain similar results by using Davydov's inequality for strong mixing errors. Here we skip the analyses for static panels with serially correlated errors for brevity. Let M be a generic large positive constant and $\mathscr{F}^{(\ell)} := \left\{F^{(\ell)}\in\mathbb{R}^{T_\ell\times r_0} : \frac{F^{(\ell)\prime}F^{(\ell)}}{T_\ell} = I_{r_0}\right\}$.

Lemma B.13. *Under Assumptions 3.1*, 3.2 and 3.8, 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) $\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)\prime}P_{F^{(\ell)}}e_i^{(\ell)}\right| = o_p(1)$,

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

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

Proof (i) We notice that

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

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

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathscr{T}_\ell}f_t^0 e_{it}\right\|^2\middle|\mathscr{D}\right)\\
&\leq\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\sum_{s\in\mathscr{T}_\ell}\left\|\mathbb{E}\left(f_t^0 f_s^{0\prime}e_{it}e_{is}|\mathscr{D}\right)\right\|\\
&\leq\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\sum_{s\in\mathscr{T}_\ell}\|f_t^0\|\,\|f_s^{0\prime}\|\left|\mathbb{E}\left(e_{it}e_{is}|\mathscr{D}\right)\right|\\
&\lesssim\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\mathbb{E}(e_{it}^2|\mathscr{D})\leq M\ a.s.,
\end{aligned}\tag{B.45}$$

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 3.1*(i) and (iii). It follows that $\frac{1}{n}\sum_{i=n_1}^{n_n}||\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathscr{T}_\ell}f_t^0 e_{it}||^2=O_p(1)$ and $|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}e_i^{(\ell)\prime}P_{F^{0,(\ell)}}e_i^{(\ell)}|=O_p(T^{-1})$.

(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 $F^{(\ell)}\in\mathscr{F}^{(\ell)}$, we have $\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}e_i^{(\ell)\prime}P_{F^{(\ell)}}e_i^{(\ell)}|\leq\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}\frac{1}{n}\sum_{i=n_1}^{n_n}||\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}f_t e_{it}||^2$. Next,

$$\begin{aligned}
&\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}f_t e_{it}\right\|^2\\
&=\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}tr\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\frac{1}{T_\ell^2}\sum_{t\in\mathscr{T}_\ell}\sum_{s\in\mathscr{T}_\ell}f_t f_s' e_{it}e_{is}\right)\\
&=\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}tr\left\{\frac{1}{T_\ell^2}\sum_{t\in\mathscr{T}_\ell}\sum_{s\in\mathscr{T}_\ell}f_t f_s'\frac{1}{n}\sum_{i=n_1}^{n_n}\left[e_{it}e_{is}-\mathbb{E}\left(e_{it}e_{is}|\mathscr{D}\right)\right]\right\}\\
&+\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}tr\left\{\frac{1}{T_\ell^2}\sum_{t\in\mathscr{T}_\ell}\sum_{s\in\mathscr{T}_\ell}f_t f_s'\frac{1}{n}\sum_{i=n_1}^{n_n}\mathbb{E}\left(e_{it}e_{is}|\mathscr{D}\right)\right\}.
\end{aligned}\tag{B.46}$$

For the first term to the right the second equality of (B.46), we have $\max_{t,s}|\frac{1}{n}\sum_{i=n_1}^{n_n}[e_{it}e_{is}-\mathbb{E}(e_{it}e_{is}|\mathscr{D})]|=O_p(\sqrt{(\log T)/N})$ 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 3.1*(i). Then

$$\begin{aligned}
&\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}tr\left\{\frac{1}{T_\ell^2}\sum_{t\in\mathscr{T}_\ell}\sum_{s\in\mathscr{T}_\ell}f_t f_s'\frac{1}{n}\sum_{i=n_1}^{n_n}\left[e_{it}e_{is}-\mathbb{E}\left(e_{it}e_{is}|\mathscr{D}\right)\right]\right\}\\
&=O_p(\sqrt{(\log T)/N})\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}\left(\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}\|f_t\|\right)^2=O_p(\sqrt{(\log T)/N}).
\end{aligned}\tag{B.47}$$

For the second term on the second line of (B.46), we have

$$\begin{aligned}
&\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}} tr\left\{\frac{1}{T_\ell^2}\sum_{t\in\mathscr{T}_\ell}\sum_{s\in\mathscr{T}_\ell} f_t f_s' \frac{1}{n}\sum_{i=n_1}^{n_n}\mathbb{E}\left(e_{it}e_{is}\middle|\mathscr{D}\right)\right\}\\
&\leq \sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}} \frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\sum_{s\in\mathscr{T}_\ell}\|f_t\|\,\|f_s\|\,\mathbb{E}\left(e_{it}e_{is}\middle|\mathscr{D}\right)\\
&\lesssim \frac{1}{nT_\ell^2}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\left|\mathbb{E}\left(e_{it}^2\middle|\mathscr{D}\right)\right| = O_p(T^{-1}),
\end{aligned}\tag{B.48}$$

where the first inequality is by Cauchy's inequality, the third line is by the definition of $\mathscr{F}^{(\ell)}$ and similar arguments as in (B.45).

Combining (B.46)-(B.48), we have shown that $\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)\prime}P_{F^{(\ell)}}e_i^{(\ell)}| = O_p(\sqrt{(\log T)/N})$.

(iii) Note that

$$\begin{aligned}
\sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}M_{F^{(\ell)}}e_i^{(\ell)}\right| &\leq \left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}e_i^{(\ell)}\right|\\
&\quad + \sup_{F^{(\ell)}\in\mathscr{F}^{(\ell)}}\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}P_{F^{(\ell)}}e_i^{(\ell)}\right|.
\end{aligned}$$

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

$$\left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^{0\prime}F^{0,(\ell)\prime}e_i^{(\ell)}\right| = \left|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\lambda_i^{0\prime}f_t^0 e_{it}\right| = O_p((NT)^{-1/2})$$

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

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

where the third line is by Cauchy's inequality and the last line is by arguments in

(B.47) and (B.48). Combining the above results completes the proof.

(iv) We first observe that

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{n}\sum_{i=n_1}^{n_n}\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathscr{T}_\ell}X_{it}e_{it}\right\|^2\right)=\frac{1}{n}\sum_{i=n_1}^{n_n}\mathbb{E}\left(\left\|\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathscr{T}_\ell}X_{it}e_{it}\right\|^2\right)\\
&\leq\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\sum_{s\in\mathscr{T}_\ell}\left\|\mathbb{E}\left(X_{it}X_{is}'e_{it}e_{is}\right)\right\|=\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\left\|\mathbb{E}\left(X_{it}X_{it}'e_{it}^2\right)\right\|\\
&\leq M\ a.s.,
\end{aligned}\tag{B.49}$$

where the second equality is by Assumption 3.1*(ii) and the law of iterated expectations, and the last inequality is by Assumption 3.1*(v). It follows that

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

where the second inequality is by Cauchy's inequality, the sixth line holds by the fact that both θ_i and $\theta_i^{0,(\ell)}$ are bounded and $\frac{1}{n}\sum_{i=n_1}^{n_n}||\frac{1}{\sqrt{T_\ell}}\sum_{t\in\mathscr{T}_\ell}X_{it}e_{it}||^2=O_p(1)$ by (B.49), and the last line is due to (B.46) and the fact that

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

by Assumption 3.8(ii). ■

Lemma B.14. *Under Assumptions 3.1*, 3.2 and 3.8, we have* $\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\xrightarrow{p}0$ *and* $||P_{\hat{F}^{(\ell)}}-P_{F^{0,(\ell)}}||\xrightarrow{p}0$.

Proof Let

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

Recall from (B.41) that $(\{\hat{\theta}_i^{(\ell)}\},\hat{F}^{(\ell)})$ is the minimizer of $S_{NT}\left(\{\theta_i\},F^{(\ell)}\right)$. By (B.40) and Lemma B.13, we have

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

where

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

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 $(\{\theta_i^{0,(\ell)}\},F^{0,(\ell)}H^{(\ell)})$, where $H^{(\ell)}$ is a rotation matrix. Hence, we conclude that $\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\xrightarrow{p}0$. Following the proof of Proposition 1 of Bai (2009), we can show that $||P_{\hat{F}^{(\ell)}}-P_{F^{0,(\ell)}}||\xrightarrow{p}0$. ■

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

Lemma B.15. *Under Assumptions 3.1*, 3.2 and 3.8, we have* $\frac{1}{\sqrt{T_\ell}}\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|=O_p(B_N+\frac{1}{\sqrt{N\wedge T}})$, *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 $\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}(Y_i^{(\ell)}-X_i^{(\ell)}\theta_i)'(Y_i^{(\ell)}-X_i^{(\ell)}\theta_i)$ along its diagonal line. By inserting (B.40) into (B.43), we obtain that

$$\hat{F}^{(\ell)}V_{NT}^{(\ell)}-F^{0,(\ell)}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)\left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)=\sum_{m\in[8]}J_m^{(\ell)}, \tag{B.50}$$

where

$$J_1^{(\ell)} = \frac{1}{nT_\ell} \sum_{i=n_1}^{n_n} X_i^{(\ell)} (\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)})(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)})' X_i^{(\ell)\prime} \hat{F}^{(\ell)},$$

$$J_2^{(\ell)} = \frac{1}{nT_\ell} \sum_{i=n_1}^{n_n} X_i^{(\ell)} (\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}) \lambda_i^{0\prime} F^{0,(\ell)\prime} \hat{F}^{(\ell)},$$

$$J_3^{(\ell)} = \frac{1}{nT_\ell} \sum_{i=n_1}^{n_n} X_i^{(\ell)} (\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}) e_i^{(\ell)\prime} \hat{F}^{(\ell)},$$

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

$$J_5^{(\ell)} = \frac{1}{nT_\ell} \sum_{i=n_1}^{n_n} e_i^{(\ell)} (\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)})' X_i^{(\ell)\prime} \hat{F}^{(\ell)},$$

$$J_6^{(\ell)} = \frac{1}{nT_\ell} \sum_{i=n_1}^{n_n} F^{0,(\ell)} \lambda_i^0 e_i^{(\ell)\prime} \hat{F}^{(\ell)},$$

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

$$J_8^{(\ell)} = \frac{1}{nT_\ell} \sum_{i=n_1}^{n_n} e_i^{(\ell)} e_i^{(\ell)\prime} \hat{F}^{(\ell)}.$$

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

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

$$\frac{1}{\sqrt{T_\ell}} \left\| J_1^{(\ell)} \right\| \leq \max_{i \in \mathcal{N}} \left\| \hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)} \right\|^2 \frac{1}{\sqrt{T_\ell}} \left\| \hat{F}^{(\ell)} \right\| \frac{1}{nT_\ell} \sum_{i=n_1}^{n_n} \left\| X_i^{(\ell)} \right\|^2 = O_p(B_N^2), \tag{B.51}$$

where the equality holds by Assumption 3.8(ii) and normalization of the factor vector. Similarly, we have

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

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

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

$$\frac{1}{\sqrt{T_\ell}} \left\| J_8^{(\ell)} \right\| \leq \frac{\left\| \hat{F}^{(\ell)} \right\|}{\sqrt{T_\ell}} \frac{1}{nT_\ell} \left\| \sum_{i=n_1}^{n_n} e_i^{(\ell)} e_i^{(\ell)\prime} \right\| = O_p((N \wedge T)^{-1/2}),$$

where the third line is by the fact that $\frac{1}{T_\ell} \sum_{t \in \mathcal{T}_\ell} \left\| \frac{1}{\sqrt{n}} \sum_{i=n_1}^{n_n} \lambda_i^0 e_{it} \right\| = O_p(1)$ by similar arguments as in (B.45) and the last line is due to the fact that

$$\mathbb{E}\left(\left\|\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)} e_i^{(\ell)\prime}\right\|^2 \middle| \mathscr{D}\right) = \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{i^*=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\sum_{t^*\in\mathscr{T}_\ell}\mathbb{E}\left(e_{it}e_{it^*}e_{i^*t}e_{i^*t^*}\middle|\mathscr{D}\right)$$

$$= \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\sum_{t^*\in\mathscr{T}_\ell}\mathbb{E}\left(e_{it}^2 e_{it^*}^2\middle|\mathscr{D}\right) + \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{i^*\neq i}\sum_{t\in\mathscr{T}_\ell}\sum_{t^*\in\mathscr{T}_\ell}\mathbb{E}\left(e_{it}e_{it^*}\middle|\mathscr{D}\right)\mathbb{E}\left(e_{i^*t}e_{i^*t^*}\middle|\mathscr{D}\right)$$

$$= \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\mathbb{E}\left(e_{it}^4|\mathscr{D}\right) + \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{t\in\mathscr{T}_\ell}\sum_{t^*\in\mathscr{T}_\ell, t^*\neq t}\mathbb{E}\left(e_{it}^2|\mathscr{D}\right)\mathbb{E}\left(e_{it^*}^2|\mathscr{D}\right)$$

$$+ \frac{1}{(nT_\ell)^2}\sum_{i=n_1}^{n_n}\sum_{i^*\neq i}\sum_{t\in\mathscr{T}_\ell}\mathbb{E}\left(e_{it}^2|\mathscr{D}\right)\mathbb{E}\left(e_{i^*t}^2|\mathscr{D}\right)$$

$$= O\left((N)^{-1} + (T)^{-1}\right) \ a.s. \tag{B.52}$$

by Assumption 3.1*(i), (ii), (iii), and (v). Besides, we have $T_\ell^{-1/2}||_4^{(\ell)})|| = O_p(B_N)$, $T_\ell^{-1/2}||J_5^{(\ell)})|| = O_p(B_N)$, and $T_\ell^{-1/2}||J_4^{(\ell)}|| = O_p(N^{-1/2})$ 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 (B.50) by $\hat{F}^{(\ell)\prime}$ and using the fact that $\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)} = T_\ell I_r$, we have

$$\frac{1}{T_\ell}\left\|\hat{F}^{(\ell)}V_{NT}^{(\ell)} - F^{0,(\ell)}\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right\| = O_p(B_N) + O_p((N\wedge T)^{-1/2}), \tag{B.53}$$

and

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

Then $V_{NT}^{(\ell)}$ is invertible and $||V_{NT}^{(\ell)}|| = O_p(1)$. By the definition of $H^{(\ell)}$ and (B.53), we have $\frac{1}{\sqrt{T_\ell}}||\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}|| = O_p(B_N + (N\wedge T)^{-1/2})$. ■

Lemma B.16. *Under Assumptions 3.1*, 3.2 and 3.8, 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* $i, i^* \in \mathscr{N}$,

(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* $i \in \mathscr{N}$,

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

(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* $i \in \mathscr{N}$,

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

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

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

Proof (i) Let $\delta_1^{(\ell)} := (T_\ell^{-1}\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)})^{-1}-(T_\ell^{-1}F^{0,(\ell)\prime}F^{0,(\ell)})^{-1}$ and $\delta_2^{(\ell)} := T_\ell^{-1/2}(\hat{F}^{(\ell)\prime}-F^{0,(\ell)\prime}H^{(\ell)})$. Noting that $M_{\hat{F}^{(\ell)}} = I_{T_\ell}-\hat{F}^{(\ell)}(\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)})^{-1}\hat{F}^{(\ell)\prime}$ and $M_{F^{0,(\ell)}} = I_{T_\ell}-F^{0,(\ell)}(F^{0,(\ell)\prime}F^{0,(\ell)})^{-1}F^{0,(\ell)\prime}$, we can show that

$$\begin{aligned}
&M_{F^{0,(\ell)}}-M_{\hat{F}^{(\ell)}}\\
&=\frac{\hat{F}^{(\ell)}}{\sqrt{T_\ell}}\left(\frac{\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1}\frac{\hat{F}^{(\ell)\prime}}{\sqrt{T_\ell}}-\frac{F^{0,(\ell)}}{\sqrt{T_\ell}}\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\frac{F^{0,(\ell)\prime}}{\sqrt{T_\ell}}\\
&=\delta_2^{(\ell)}\delta_1^{(\ell)}\delta_2^{(\ell)\prime}+\delta_2^{(\ell)}\delta_1^{(\ell)}\left(\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\right)'+\delta_2^{(\ell)}\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\delta_2^{(\ell)\prime}\\
&+\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\delta_1^{(\ell)}\delta_2^{(\ell)\prime}+\delta_2^{(\ell)}\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\left(\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\right)'\\
&+\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\delta_1^{(\ell)}\left(\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\right)'+\frac{F^{0,(\ell)}H^{(\ell)}}{\sqrt{T_\ell}}\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\delta_2^{(\ell)\prime}.
\end{aligned}\tag{B.54}$$

By Lemma B.15, Assumption 3.8, the normalization for the factor space, and the fact that

$$\left\|\delta_1^{(\ell)}\right\| \le \frac{\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|^2}{T_\ell}+2\frac{\left\|F^{0,(\ell)}H^{(\ell)}\right\|\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|}{T_\ell} = O_p(B_N+(N\wedge T)^{-1/2}),\tag{B.55}$$

we can readily show that

$$\left\|M_{F^{0,(\ell)}}-M_{\hat{F}^{(\ell)}}\right\| = O_p(B_N+(N\wedge T)^{-1/2}).\tag{B.56}$$

Then $\max_{i,i^*\in\mathscr{N}}\left\|\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}X_{i^*}^{(\ell)}}{T_\ell}-\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_{i^*}^{(\ell)}}{T_\ell}\right\| \le \max_{i\in\mathscr{N}}\frac{1}{T_\ell}\left\|X_i^{(\ell)}\right\|^2\left\|M_{F^{0,(\ell)}}-M_{\hat{F}^{(\ell)}}\right\| = O_p(B_N+(N\wedge T)^{-1/2})$, where recall that $\mathscr{N} := \{n_1,\cdots,n_n\}$.

(ii) By (B.54), we notice that

$$\max_{i\in\mathcal{N}}\left\|\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_i^{(\ell)}}{\sqrt{T_\ell}}-\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}e_i^{(\ell)}}{\sqrt{T_\ell}}\right\|$$

$$=\sqrt{T_\ell}\max_{i\in\mathcal{N}}\frac{\left\|X_i^{(\ell)}\right\|}{\sqrt{T_\ell}}\frac{\left\|e_i^{(\ell)}\right\|}{\sqrt{T_\ell}}O_p\left[(B_N+(N\wedge T)^{-1/2})^3+(B_N+(N\wedge T)^{-1/2})^2\right]$$

$$+\sqrt{T_\ell}\max_{i\in\mathcal{N}}\frac{\left\|X_i^{(\ell)}\right\|}{\sqrt{T_\ell}}\frac{\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|}{\sqrt{T_\ell}}\left\|\left(\frac{F^{0,(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right)^{-1}\right\|\frac{\left\|\left(F^{0,(\ell)}H^{(\ell)}\right)'e_i^{(\ell)}\right\|}{T_\ell}$$

$$+\sqrt{T_\ell}\max_{i\in\mathcal{N}}\frac{\left\|X_i^{(\ell)}\right\|}{\sqrt{T_\ell}}\frac{\left\|F^{0,(\ell)}H^{(\ell)}\right\|}{\sqrt{T_\ell}}\left\|\delta_1^{(\ell)}\right\|\frac{\left\|\left(F^{0,(\ell)}H^{(\ell)}\right)'e_i^{(\ell)}\right\|}{T_\ell}$$

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

where the last line holds by combining Assumption 3.8(ii), (B.55), Lemma B.15 and the fact that

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

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

Note that

$$\max_{i\in\mathcal{N},t\in\mathcal{T}_\ell}\left\|Var\left(e_{it}f_t^0|\mathcal{D}\right)\right\|=\max_{i\in\mathcal{N},t\in\mathcal{T}_\ell}\left\|\mathbb{E}\left(e_{it}^2|\mathcal{D}\right)f_t^0f_t^{0\prime}\right\|=O_p(1)\quad\text{and}$$
$$\max_{i\in\mathcal{N},t\in\mathcal{T}_\ell}\left\|e_{it}f_t^0\right\|=O_p\left((NT)^{1/q}\right),\tag{B.57}$$

where the second line is by Assumption 3.1*(v) and the last line is by Assumption 3.1*(v). Define events $\mathscr{A}_{4,N}(M)=\left\{\max_{i\in\mathcal{N},t\in\mathcal{T}_\ell}\left\|e_{it}f_t^0\right\|\leq M(NT)^{1/q}\right\}$ and $\mathscr{A}_{4,N,i}(M)=\left\{\max_{t\in\mathcal{T}_\ell}\left\|e_{it}f_t^0\right\|\leq M(NT)^{1/q}\right\}$ 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

$$\mathbb{P}\left(\max_{i\in\mathcal{N}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}e_{it}f_t^0\right\|>c_6\sqrt{\frac{\log N}{T}}\right)$$
$$\leq\mathbb{P}\left(\max_{i\in\mathcal{N}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}e_{it}f_t^0\right\|>c_6\sqrt{\frac{\log N}{T}},\mathscr{A}_{4,N}(M)\right)+\mathbb{P}\left(\mathscr{A}_{4,N}^c(M)\right)$$
$$\leq\sum_{i=n_1}^{n_n}\mathbb{P}\left(\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}e_{it}f_t^0\right\|>c_6\sqrt{\frac{\log N}{T}},\mathscr{A}_{4,N}(M)\right)+\mathbb{P}\left(\mathscr{A}_{4,N}^c(M)\right)$$
$$\leq\sum_{i=n_1}^{n_n}\mathbb{P}\left(\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}e_{it}f_t^0\right\|>c_6\sqrt{\frac{\log N}{T}},\mathscr{A}_{4,N,i}(M)\right)+\mathbb{P}\left(\mathscr{A}_{4,N}^c(M)\right)$$

$$
\begin{aligned}
&\leq \sum_{i=n_1}^{n_n} \mathbb{E}\mathbb{P}\left(\left\| \frac{1}{T_\ell} \sum_{t\in\mathscr{T}_\ell} e_{it} f_t^0 \right\| > c_6 \sqrt{\frac{\log N}{T}} \middle| \mathscr{D} \right) \mathbf{1}\left\{ \mathscr{A}_{4,N,i}(M) \right\} + \mathbb{P}\left(\mathscr{A}_{4,N}^c(M) \right) \\
&\leq \sum_{i=n_1}^{n_n} \exp\left\{ -\frac{c_4 c_6^2 T \log N/2}{c_7 T + c_6 M \sqrt{T\log N}(NT)^{1/q}(\log T)^2/3} \right\} + o(1) \\
&= o(1), \qquad \text{(B.58)}
\end{aligned}
$$

where the last inequality holds by Lemma B.5(i), (B.57), and the definition of event $\mathscr{A}_{4,N,i}$, and the last line holds by Assumption 3.1*(vi) and the fact that $q > 8$.

(iii) By the fact that

$$
\begin{aligned}
&\frac{1}{nT_\ell^2} \sum_{i=n_1}^{n_n} \left\| e_i^{(\ell)\prime} \hat{F}^{(\ell)} \right\|^2 \\
&\leq \frac{1}{nT_\ell^2} \sum_{i=n_1}^{n_n} \left\| e_i^{(\ell)\prime} F^{0,(\ell)} H^{(\ell)} \right\|^2 + \frac{1}{nT_\ell^2} \sum_{i=n_1}^{n_n} \left\| e_i^{(\ell)\prime} \left(\hat{F}^{(\ell)} - F^{0,(\ell)} H^{(\ell)} \right) \right\|^2 \\
&\lesssim \frac{1}{nT_\ell^2} \sum_{i=n_1}^{n_n} \left\| e_i^{(\ell)\prime} F^{0,(\ell)} \right\|^2 + \frac{1}{nT_\ell} \sum_{i=n_1}^{n_n} \left\| e_i^{(\ell)\prime} \right\|^2 \frac{1}{T_\ell} \left\| \hat{F}^{(\ell)} - F^{0,(\ell)} H^{(\ell)} \right\|^2 \\
&\leq \frac{1}{nT_\ell^2} \sum_{i=n_1}^{n_n} \left\| e_i^{(\ell)\prime} F^{0,(\ell)} \right\|^2 + O_p(B_N^2 + (N\wedge T)^{-1}) \\
&= O_p(B_N^2 + (N\wedge T)^{-1}), \qquad \text{(B.59)}
\end{aligned}
$$

where the last inequality holds by Assumption 3.8(ii) and Lemma B.15, and last equality holds by the fact that

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

by (B.45).

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

$$
\begin{aligned}
&\max_{i\in\mathscr{N}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} F^{0,(\ell)} \right\| \\
&= \max_{i\in\mathscr{N}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \left(F^{0,(\ell)} - \hat{F}^{(\ell)} H^{(\ell)-1} \right) \right\| \\
&\leq \max_{i\in\mathscr{N}} \left\| \frac{1}{\sqrt{T_\ell}} X_i^{(\ell)} M_{\hat{F}^{(\ell)}} \right\| O_p(B_N + (N\wedge T)^{-1/2}) = O_p(B_N + (N\wedge T)^{-1/2}),
\end{aligned}
$$

where the last inequality is by Lemma B.15 and the last equality is by the fact that

$$
\max_{i\in\mathscr{N}} \left\| \frac{1}{\sqrt{T_\ell}} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \right\| \leq \max_{i\in\mathscr{N}} \left\| \frac{1}{\sqrt{T_\ell}} X_i^{(\ell)\prime} \right\| = O_p(1) \qquad \text{(B.60)}
$$

by Assumption 3.8(ii).

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

Under Assumptions 3.1* and Assumption 3.8(ii)

$$\max_{i,i^*\in\mathcal{N},t\in\mathcal{T}_\ell}\|X_{it}e_{i^*t}\| = O_p((NT)^{1/q}),$$

$$\max_{i,i^*\in\mathcal{N}}\left\|\sum_{t\in\mathcal{T}_\ell}\mathbb{E}(X_{it}X_{it}'e_{i^*t}^2|\mathcal{G}_{t-1})\right\| \leq \max_{i^*\in\mathcal{N},t}\mathbb{E}(e_{i^*t}^2|\mathcal{G}_{t-1})\max_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}_\ell}\|X_{it}\|^2 \leq c_8T\ a.s.$$

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

$$\begin{aligned}
&\mathbb{P}\left(\max_{i,i^*\in\mathcal{N}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\| > c_6\sqrt{\frac{\log N}{T}}\right)\\
&\leq \mathbb{P}\left(\max_{i,i^*\in\mathcal{N}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\| > c_6\sqrt{\frac{\log N}{T}},\mathscr{A}_{5,N}(M)\right) + \mathbb{P}\left(\mathscr{A}_{5,N}^c(M)\right)\\
&\leq \sum_{i=n_1}^{n_n}\sum_{i^*=n_1}^{n_n}\mathbb{P}\left(\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\| > c_6\sqrt{\frac{\log N}{T}},\mathscr{A}_{5,N}(M)\right) + \mathbb{P}\left(\mathscr{A}_{5,N}^c(M)\right)\\
&\leq \sum_{i=n_1}^{n_n}\sum_{i^*=n_1}^{n_n}\mathbb{P}\left(\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}X_{it}e_{i^*t}\right\| > c_6\sqrt{\frac{\log N}{T}},\mathscr{A}_{5,N,i,i^*}(M)\right) + \mathbb{P}\left(\mathscr{A}_{5,N}^c(M)\right)\\
&\leq \sum_{i=n_1}^{n_n}\sum_{i^*=n_1}^{n_n}\exp\left\{\frac{-c_6^2T\log N/2}{c_8T+Mc_6(NT)^{1/q}\sqrt{T\log N}/3}\right\} + o(1)\\
&= o(1),
\end{aligned}\tag{B.61}$$

where the last inequality holds by Lemma B.5(ii) and the last line is by Assumption 3.1*(vi).

(vi) Noted that

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

where the last line holds by Lemma B.7(i) and Assumption 3.1*. Similarly as above, we can also show that $\mathbb{E}\left[\left\|\frac{1}{\sqrt{nT_\ell}}\sum_{i=n_1}^{n_n}\lambda_i^0e_i^{(\ell)\prime}\right\|^2\middle|\mathscr{D}\right] = O_p(1)$. Then

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

Furthermore, we have

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

(vii) We first notice that

$$\max_{i \in \mathscr{N}} \left\| \frac{1}{nT_\ell} \sum_{i^*=n_1}^{n_n} X_i^{(\ell)\prime} e_{i^*}^{(\ell)} \lambda_{i^*}^{0\prime} \right\| = \max_{i \in \mathscr{N}} \left\| \frac{1}{nT_\ell} \sum_{i^*=n_1}^{n_n} \sum_{t \in \mathscr{T}_\ell} X_{it} e_{i^* t} \lambda_{i^*}^{0\prime} \right\| = O_p\left(\sqrt{\frac{\log N}{NT}} \right) \tag{B.62}$$

by similar arguments as used to obtain (B.61). This result, in conjunction with Lemma B.16(vi), implies that

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

(viii) By (B.61) and Lemma B.16(iii), we have

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

■

Define

$$\xi_i^{0,(\ell)} := \frac{X_i^{(\ell)\prime} M_{F^{0,(\ell)}} e_i^{(\ell)}}{T_\ell}, \quad S_{ii^*}^{0,(\ell)} := \frac{X_i^{(\ell)\prime} M_{F^{0,(\ell)}} X_{i^*}^{(\ell)}}{T_\ell}, \quad a_{ii^*}^0 := \lambda_i^{0\prime}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\lambda_{i^*}^0,$$

$G_{ii^*}^{0,(\ell)} := S_{ii^*}^{0,(\ell)} a_{ii^*}^0$, and $\Omega_i^{0,(\ell)} := Var(\xi_i^{0,(\ell)})$.

Lemma B.17. *Under Assumptions 3.1*, 3.2 and 3.8, we have*

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

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

(iii) $\max_{i\in\mathscr{N}} ||\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}|| = O_p(\sqrt{(\log N)/T})$.

Proof (i) Noting from (B.42) that $\hat{\theta}_i^{(\ell)} = (\hat{S}_i^{(\ell)})^{-1} T_\ell^{-1} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} Y_i^{(\ell)}$ with $\hat{S}_{ii}^{(\ell)} = T_\ell^{-1} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} X_i^{(\ell)}$, we have

$$\begin{aligned}
\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)} &= (\hat{S}_{ii}^{(\ell)})^{-1}\left[\frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} e_i^{(\ell)} + \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} F^{0,(\ell)} \lambda_i^0\right] \\
&= (\hat{S}_{ii}^{(\ell)})^{-1}\frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} e_i^{(\ell)}}{T_\ell} + (\hat{D}_i^{(\ell)})^{-1}\frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \cdot \\
&\quad \left[\hat{F}^{(\ell)} H^{(\ell)-1} - \sum_{m\in[8]} J_m^{(\ell)} \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\right]\lambda_i^0 \\
&= (\hat{S}_{ii}^{(\ell)})^{-1}\frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} e_i^{(\ell)}}{T_\ell} - (\hat{D}_i^{(\ell)})^{-1}\frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \sum_{m\in[8]} J_m^{(\ell)} \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\lambda_i^0,
\end{aligned} \tag{B.63}$$

where the second equality is from (B.50). Note that

$$\begin{aligned}
&\max_{i\in\mathscr{N}}\left\|\frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \sum_{m\in[8]} J_m^{(\ell)} \left(\frac{F^{0,(\ell)\prime}\hat{F}^{(\ell)}}{T_\ell}\right)^{-1}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\lambda_i^0\right\| \\
&\lesssim \sum_{m\in[8]} \max_{i\in\mathscr{N}}\left\|\frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} J_m^{(\ell)}\right\| =: \sum_{m\in[8]} II_m,
\end{aligned}$$

by Lemmas B.7(i) and B.15, and the normalization of the factor and factor loadings. Hence, it suffices to show the uniform convergence rate II_m for $m \in [8]\backslash\{2\}$. The term associated with II_2 needs to be kept.

For II, we have $II_1 \leq \max_{i\in\mathscr{N}}\left\|\frac{1}{\sqrt{T_\ell}} X_i^{(\ell)\prime}\right\|\left\|J_1^{(\ell)}\right\| = O_p(B_N^2)$ by (B.51) and (B.60).

Next, noting that

$$
\begin{aligned}
&II_{2,i}\\
&:= \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} J_2^{(\ell)} \left(\frac{F^{0,(\ell)\prime} \hat{F}^{(\ell)}}{T_\ell} \right)^{-1} \left(\frac{\Lambda_n^{0\prime} \Lambda_n^0}{n} \right)^{-1} \lambda_i^0\\
&= \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \frac{1}{nT_\ell} \sum_{i^*=n_1}^{n_n} X_{i^*}^{(\ell)} \left(\theta_{i^*}^{0,(\ell)} - \hat{\theta}_{i^*}^{(\ell)} \right) \lambda_{i^*}^{0\prime} F^{0,(\ell)\prime} \hat{F}^{(\ell)} \left(\frac{F^{0,(\ell)\prime} \hat{F}^{(\ell)}}{T_\ell} \right)^{-1} \left(\frac{\Lambda_n^{0\prime} \Lambda_n^0}{n} \right)^{-1} \lambda_i^0\\
&= \frac{1}{n} \sum_{i^*=n_1}^{n_n} \frac{X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} X_{i^*}^{(\ell)}}{T_\ell} \left(\theta_{i^*}^{0,(\ell)} - \hat{\theta}_{i^*}^{(\ell)} \right) \lambda_{i^*}^{0\prime} \left(\frac{\Lambda_n^{0\prime} \Lambda_n^0}{n} \right)^{-1} \lambda_i^0,
\end{aligned}
$$

we have $\max_{i \in \mathcal{N}} \left\| II_{2,i} \right\| = O_p(B_N)$, and this term will be kept in the linear expansion for $\hat{\theta}_i^{(\ell)}$. For II_3, we have

$$
\begin{aligned}
II_3 &= \max_{i \in \mathcal{N}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \frac{1}{nT_\ell} \sum_{i^*=n_1}^{n_n} X_{i^*}^{(\ell)} \left(\theta_{i^*}^{0,(\ell)} - \hat{\theta}_{i^*}^{(\ell)} \right) e_{i^*}^{(\ell)\prime} \hat{F}^{(\ell)} \right\|\\
&\leq \max_{i \in \mathcal{N}} \left\| \frac{1}{\sqrt{T_\ell}} X_i^{(\ell)\prime} \right\| \max_{i \in \mathcal{N}} \left\| \theta_{i^*}^{0,(\ell)} - \hat{\theta}_{i^*}^{(\ell)} \right\| \sqrt{\frac{1}{nT_\ell} \sum_{i^*=n_1}^{n_n} \left\| X_{i^*}^{(\ell)} \right\|^2} \sqrt{\frac{1}{nT_\ell^2} \sum_{i^*=n_1}^{n_n} \left\| e_{i^*}^{(\ell)\prime} \hat{F}^{(\ell)} \right\|}\\
&= O_p(B_N^2 + B_N (N \wedge T)^{-1/2})
\end{aligned}
$$

by (B.60), Assumption 3.8(ii) and Lemma B.16(iii). For II_4, we have

$$
\begin{aligned}
II_4 &= \max_{i \in \mathcal{N}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \frac{1}{nT_\ell} \sum_{i^*=n_1}^{n_n} F^{0,(\ell)} \lambda_{i^*}^0 \left(\theta_{i^*}^{0,(\ell)} - \hat{\theta}_{i^*}^{(\ell)} \right)' X_{i^*}^{(\ell)\prime} \hat{F}^{(\ell)} \right\|\\
&\leq \max_{i \in \mathcal{N}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} F^{0,(\ell)} \right\| \max_{i \in \mathcal{N}} \left\| \lambda_i^0 \right\| \max_{i \in \mathcal{N}} \left\| \theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)} \right\| \frac{1}{n\sqrt{T_\ell}} \sum_{i^*=n_1}^{n_n} \left\| X_{i^*}^{(\ell)} \right\| \frac{\left\| \hat{F}^{(\ell)} \right\|}{\sqrt{T_\ell}}\\
&= O_p(B_N + (N \wedge T)^{-1/2}) O_p(B_N) = O_p(B_N^2 + B_N (N \wedge T)^{-1/2}),
\end{aligned}
$$

where the last line holds by Lemma B.16(iv), the normalization of factors and the fact that $\frac{1}{n\sqrt{T_\ell}} \sum_{i=n_1}^{n_n} ||X_i^{(\ell)}|| = \frac{1}{n} \sum_{i=n_1}^{n_n} \sqrt{\frac{1}{T_\ell} \sum_{t \in \mathcal{T}_\ell} ||X_{it}||^2} = O_p(1)$ by Assumption 3.8(ii). For II_5, we have

$$
\begin{aligned}
II_5 &= \max_{i \in \mathcal{N}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} M_{\hat{F}^{(\ell)}} \frac{1}{nT_\ell} \sum_{i^*=n_1}^{n_n} e_{i^*}^{(\ell)} \left(\theta_{i^*}^{0,(\ell)} - \hat{\theta}_{i^*}^{(\ell)} \right)' X_{i^*}^{(\ell)\prime} \hat{F}^{(\ell)} \right\|\\
&\lesssim \max_{i \in \mathcal{N}} \left\| \frac{1}{nT_\ell \sqrt{T_\ell}} \sum_{i^*=n_1}^{n_n} X_i^{(\ell)\prime} e_{i^*}^{(\ell)} \left(\theta_{i^*}^{0,(\ell)} - \hat{\theta}_{i^*}^{(\ell)} \right)' X_{i^*}^{(\ell)\prime} \right\|\\
&\quad + \max_{i \in \mathcal{N}} \left\| \frac{1}{T_\ell} X_i^{(\ell)\prime} \hat{F}^{(\ell)} \frac{1}{n\sqrt{T_\ell}} \sum_{i^*=n_1}^{n_n} \frac{1}{T_\ell} \hat{F}^{(\ell)\prime} e_{i^*}^{(\ell)} \left(\theta_{i^*}^{0,(\ell)} - \hat{\theta}_{i^*}^{(\ell)} \right)' X_{i^*}^{(\ell)\prime} \right\|\\
&\leq \max_{i \in \mathcal{N}} \left\| \theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)} \right\|_2 \sqrt{\frac{1}{nT_\ell^2} \sum_{i^*=n_1}^{n_n} \left\| X_i^{(\ell)\prime} e_{i^*}^{(\ell)} \right\|^2} \sqrt{\frac{1}{nT_\ell} \sum_{i^*=n_1}^{n_n} \left\| X_{i^*}^{(\ell)\prime} \right\|^2}
\end{aligned}
$$

$$+\max_{i\in\mathcal{N}}\left\|\theta_i^{0,(\ell)}-\hat{\theta}_i^{(\ell)}\right\|_2\max_{i\in\mathcal{N}}\frac{\left\|X_i^{(\ell)}\right\|}{\sqrt{T_\ell}}\frac{\left\|\hat{F}^{(\ell)}\right\|}{\sqrt{T_\ell}}\sqrt{\frac{1}{nT_\ell^2}\sum_{i^*=n_1}^{n_n}\left\|e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|^2}\sqrt{\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}\left\|X_{i^*}^{(\ell)}\right\|^2}$$
$$=O_p(B_N)O_p(\sqrt{(\log N)/T})+O_p(B_N)O_p\left(B_N+(N\wedge T)^{-1/2}\right)$$
$$=O_p(B_N^2+B_N\sqrt{(\log N)/(N\wedge T)}),$$

where the last two lines hold by Lemma B.16(iii), B.16(v) and the fact that $\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}||X_{i^*}^{(\ell)}||^2=O_p(1)$ under Assumption 3.8(ii). Next,

$$II_6=\max_{i\in\mathcal{N}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}F^{0,(\ell)}\lambda_{i^*}^0e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|$$
$$\le\max_{i\in\mathcal{N}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}F^{0,(\ell)}\right\|\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}\lambda_{i^*}^0e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|$$
$$=O_p((B_N+(N\wedge T)^{-1/2})O_p(B_NN^{-1/2}+N^{-1}+(NT)^{-1/2})$$

by Lemma B.16(iv) and B.16(vi). By Lemma B.16(vii),

$$II_7=\max_{i\in\mathcal{N}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}F^{0,(\ell)\prime}\hat{F}^{(\ell)}\right\|$$
$$\lesssim\max_{i\in\mathcal{N}}\left\|\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_{i^*}^{(\ell)}\lambda_{i^*}^{0\prime}\right\|=O_p\left(B_NN^{-1/2}+N^{-1}+(NT/\log N)^{-1/2}\right). \tag{B.64}$$

By Lemma B.16(viii),

$$II_8=\max_{i\in\mathcal{N}}\left\|\frac{1}{T_\ell}X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}\frac{1}{nT_\ell}\sum_{i^*=n_1}^{n_n}e_{i^*}^{(\ell)}e_{i^*}^{(\ell)\prime}\hat{F}^{(\ell)}\right\|=O_p\left(B_N^2+B_N\sqrt{(\log N)/T}+\sqrt{\log N}(N\wedge T)^{-1}\right).$$

In sum, by (B.63) and the above analyses, we have

$$\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}=(\hat{S}_{ii}^{(\ell)})^{-1}\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}e_i^{(\ell)}}{T_\ell}$$
$$+(\hat{S}_{ii}^{(\ell)})^{-1}\frac{1}{n}\sum_{i^*=n_1}^{n_n}\frac{X_i^{(\ell)\prime}M_{\hat{F}^{(\ell)}}X_{i^*}^{(\ell)}}{T_\ell}\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)\lambda_{i^*}^{0\prime}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\lambda_i^0$$
$$+O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{N\wedge T}}+\frac{\sqrt{\log N}}{N\wedge T}\right) \tag{B.65}$$

uniformly in $i\in\mathcal{N}$. Then by (B.63) and Lemma B.16(i)-(ii), we have

$$\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}=\left(S_{ii}^{0,(\ell)}\right)^{-1}\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}e_i^{(\ell)}}{T_\ell}$$
$$+\left(S_{ii}^{0,(\ell)}\right)^{-1}\frac{1}{n}\sum_{i^*=n_1}^{n_n}\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_{i^*}^{(\ell)}}{T_\ell}\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)\lambda_{i^*}^{0\prime}\left(\frac{\Lambda_n^{0\prime}\Lambda_n^0}{n}\right)^{-1}\lambda_i^0$$

$$+O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{N\wedge T}}+\frac{\sqrt{\log N}}{N\wedge T}\right).$$

where recall that $S_{ii^*}^{0,(\ell)}:=T_\ell^{-1}X_i^{(\ell)\prime}M_{F^{0,(\ell)}}X_{i^*}^{(\ell)}$. 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 3.1*(i), (iii). Then we can show that

$$S_{ii^*}^{0,(\ell)}-\mathbb{E}(S_{ii^*}^{0,(\ell)}|\mathscr{D})=\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}[\mathfrak{x}_{it}^{0,(\ell)}\mathfrak{x}_{i^*t}^{0,(\ell)\prime}-\mathbb{E}(\mathfrak{x}_{it}^{0,(\ell)}\mathfrak{x}_{i^*t}^{0,(\ell)\prime}|\mathscr{D})]=O_p\left((T/\log N)^{-1/2}\right)$$

uniformly over $i,i^*\in\mathscr{N}$ by similar arguments in (B.58). Then by the fact that

$$\max_{i,i^*\in\mathscr{N}}\left\|\frac{X_i^{(\ell)\prime}M_{F^{0,(\ell)}}e_i^{(\ell)}}{T_\ell}\right\|=\max_{i,i^*\in\mathscr{N}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}X_{it}e_{it}\right\|+\max_{i,i^*\in\mathscr{N}}\left\|\frac{X_i^{(\ell)\prime}F^{0,(\ell)}}{T_\ell}\right\|\left\|\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}f_t^0e_{it}\right\|$$
$$=O_p\left((T/\log N)^{-1/2}\right)$$

by (B.58) and (B.61), we obtain that

$$\begin{aligned}&\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\\&=\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_i^{0,(\ell)}+\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\frac{1}{n}\sum_{i^*=n_1}^{n_n}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)\\&+O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{N\wedge T}}+\frac{\sqrt{\log N}}{N\wedge T}\right).\end{aligned}\tag{B.66}$$

For the second term on the right side of (B.66), we observe that

$$\begin{aligned}&\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\frac{1}{n}\sum_{i^*=n_1}^{n_n}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)\\&=\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\mathbb{E}\left(G_{ii}^{0,(\ell)}|\mathscr{D}\right)\frac{1}{n}\left(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right)\\&+\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)\\&=\frac{a_{ii}^0}{n}\left(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right)+\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right).\end{aligned}$$

By (B.66), it's clear that

$$\begin{aligned}&\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)\\&=\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left[\mathbb{E}\left(S_{i^*i^*}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_{i^*}^{0,(\ell)}\\&+\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left[\mathbb{E}\left(S_{i^*i^*}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\frac{1}{n}\sum_{j=n_1}^{n_n}\mathbb{E}\left(G_{i^*j}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_j^{(\ell)}-\theta_j^{0,(\ell)}\right)\end{aligned}$$

$$+O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{N\wedge T}}+\frac{\sqrt{\log N}}{N\wedge T}\right)$$

where the second term on the right side of the 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.,

$$\begin{aligned}
&\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left[\mathbb{E}\left(S_{i^*i^*}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_{i^*}^{0,(\ell)}\\
&=\frac{1}{nT_\ell}\sum_{i^*\neq i}\sum_{t\in\mathscr{T}_\ell}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left[\mathbb{E}\left(S_{i^*i^*}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\mathfrak{x}_{i^*t}e_{i^*t}\\
&=O_p\left(\sqrt{\frac{\log N}{NT}}\right)\text{ uniformly over } i\in\mathscr{N},
\end{aligned}$$

similarly to the result in (B.62). This yields

$$\max_{i\in\mathscr{N}}\left\|\frac{1}{n}\sum_{i^*\neq i}\mathbb{E}\left(G_{ii^*}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_{i^*}^{(\ell)}-\theta_{i^*}^{0,(\ell)}\right)\right\|=O_p\left(B_N^2+B_N\sqrt{\frac{\log N}{N\wedge T}}+\frac{\sqrt{\log N}}{N\wedge T}\right)$$

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 $B_N=\max_{i\in\mathscr{N}}||\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}||=O_p(\sqrt{(\log N)/T})$. Finally, we obtain that

$$\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\left(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right)\left(1-\frac{a_{ii}^0}{n}\right)=\xi_i^{0,(\ell)}+\mathscr{R}_i^{(\ell)}$$

such that $\max_{i\in\mathscr{N}}\left\|\mathscr{R}_i^{(\ell)}\right\|_2=O_p\left(\frac{\log N}{N\wedge T}\right)$.

(ii) Given the definition of $\Omega_i^{0,(\ell)}$ and by the 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 B.18. *Under Assumptions 3.1*, 3.2 and 3.8, we have*

(i) $\frac{1}{\sqrt{T_\ell}}\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|=O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right)$,

(ii) $\left\|M_{\hat{F}^{(\ell)}}-M_{F^{0,(\ell)}}\right\|=O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right)$,

(iii) $\max_{i\in\mathscr{N}}\left\|\hat{\lambda}_i^{(\ell)}-H^{(\ell)-1}\lambda_i^0\right\|=O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right)$,

(iv) $\max_t\left\|\hat{f}_t^{(\ell)}-H^{(\ell)\prime}f_t^0\right\|=O_p\left(\sqrt{\frac{\log(N\vee T)}{N\wedge T}}\right)$.

Proof (i) We obtain the result by combining Lemma B.15 and Lemma B.17(iii).

(ii) We obtain the result by combining (B.56) and Lemma B.17(iii).

(iii) Recall that

$$\begin{aligned}
\hat{\lambda}_i^{(\ell)} &= \left(\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}\right)^{-1}\hat{F}^{(\ell)\prime}\left(Y_i^{(\ell)} - X_i^{(\ell)\prime}\hat{\theta}_i^{(\ell)}\right) \\
&= \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left[Y_i^{(\ell)} - X_i^{(\ell)\prime}\theta_i^{0,(\ell)} - X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right)\right] \\
&= \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left[F^{0,(\ell)}\lambda_i^0 + e_i^{(\ell)} - X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right)\right] \\
&= \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left(F^{0,(\ell)} - \hat{F}^{(\ell)}H^{(\ell)-1}\right)\lambda_i^0 + \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)}H^{(\ell)-1}\lambda_i^0 + \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}e_i^{(\ell)} \\
&\quad - \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right) \\
&= H^{(\ell)-1}\lambda_i^0 + \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left(F^{0,(\ell)} - \hat{F}^{(\ell)}H^{(\ell)-1}\right)\lambda_i^0 + \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}e_i^{(\ell)} \\
&\quad - \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right)
\end{aligned}$$

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

$$\begin{aligned}
&\hat{\lambda}_i^{(\ell)} - H^{(\ell)-1}\lambda_i^0 \\
&= \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left(F^{0,(\ell)} - \hat{F}^{(\ell)}H^{(\ell)-1}\right)\lambda_i^0 + \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}e_i^{(\ell)} - \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right) \\
&= \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}\left(F^{0,(\ell)} - \hat{F}^{(\ell)}H^{(\ell)-1}\right)\lambda_i^0 + \frac{1}{T_\ell}\left(\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}\right)' e_i^{(\ell)} \\
&+ H^{(\ell)\prime}\frac{1}{T_\ell}F^{0,(\ell)}e_i^{(\ell)} - \frac{1}{T_\ell}\hat{F}^{(\ell)\prime}X_i^{(\ell)\prime}\left(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right) := I_{i,1}^{(\ell)} + I_{i,2}^{(\ell)} + I_{i,3}^{(\ell)} - I_{i,4}^{(\ell)}.
\end{aligned}$$

First, by Lemmas B.18(i) and B.7(i) and Assumption 3.8(ii),

$$\begin{aligned}
\max_{i\in\mathcal{N}}\left\|I_{i,1}^{(\ell)}\right\| &\leq \left\|\frac{\hat{F}^{(\ell)}}{\sqrt{T_\ell}}\right\|\frac{\left\|F^{0,(\ell)} - \hat{F}^{(\ell)}H^{(\ell)-1}\right\|}{\sqrt{T_\ell}}\max_{i\in\mathcal{N}}\|\lambda_i^0\| = O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right), \text{ and} \\
\max_{i\in\mathcal{N}}\left\|I_{i,2}^{(\ell)}\right\| &\leq \frac{1}{\sqrt{T_\ell}}\left\|\hat{F}^{(\ell)} - F^{0,(\ell)}H^{(\ell)}\right\|\frac{\left\|e_i^{(\ell)}\right\|}{\sqrt{T_\ell}} = O_p\left(\sqrt{\frac{\log N}{N\wedge T}}\right).
\end{aligned}$$

Similarly $\max_{i\in\mathcal{N}}||I_{i,3}^{(\ell)}|| = O_p(\sqrt{(\log N)/T})$ by (B.58). Now,

$$\max_{i\in\mathcal{N}}\left\|I_{i,4}^{(\ell)}\right\| \leq \left\|\frac{\hat{F}^{(\ell)}}{\sqrt{T_\ell}}\right\|\frac{\left\|X_i^{(\ell)}\right\|}{\sqrt{T_\ell}}\max_{i\in\mathcal{N}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\| = O_p\left(\sqrt{\frac{\log N}{T}}\right),$$

by Lemma B.17(iii). Combining the above results yields $\max_{i\in\mathcal{N}}||\hat{\lambda}_i - H^{(\ell)-1}\lambda_i||$ $= O_p(\sqrt{(\log N)/(N\wedge T)})$.

(iv) Recall from (B.50) that $\hat{F}^{(\ell)\prime} - H^{(\ell)\prime}F' = V_{NT}^{(\ell)-1}\sum_{m\in[8]} J_m^{(\ell)\prime}$, where $J_m^{(\ell)}$, $m \in [8]$, are defined in the proof of Lemma B.15. 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 $\hat{F}^{(\ell)\prime} - H^{(\ell)\prime}F'$. It remains to show the convergence rate for $J_{m,t}^{(\ell)}, m \in [8]$.

For $J_{1,t}^{(\ell)}$, we notice that

$$\begin{aligned}
\max_{t\in\mathscr{T}_\ell}\left\|J_{1,t}^{(\ell)}\right\| &= \max_{t\in\mathscr{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}\left[\frac{1}{NT_\ell}\sum_{i=n_1}^{n_n} X_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)' X_{it}\right]\right\| \\
&\leq \left\|V_{NT}^{(\ell)-1}\right\|\frac{\left\|\hat{F}^{(\ell)}\right\|}{\sqrt{T_\ell}}\max_{i\in\mathscr{N}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|^2 \max_{i\in\mathscr{N}}\frac{\left\|X_i^{(\ell)}\right\|}{\sqrt{T_\ell}}\max_{t\in\mathscr{T}_\ell}\frac{1}{n}\sum_{i=n_1}^{n_n}\|X_{it}\| \\
&\lesssim \max_{i\in\mathscr{N}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|^2 = O_p((\log N)/T),
\end{aligned}$$

by Lemma B.17(iii) and Assumption 3.8(ii). Similarly, by Lemma B.7(i),

$$\begin{aligned}
\max_{t\in\mathscr{T}_\ell}\left\|J_{2,t}^{(\ell)}\right\| &= \max_{t\in\mathscr{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}F^{0,(\ell)}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^0\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)' X_{it}\right]\right\| \\
&\lesssim \max_{i\in\mathscr{N}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\| = O_p(\sqrt{(\log N)/T}), \\
\max_{t\in\mathscr{T}_\ell}\left\|J_{3,t}^{(\ell)}\right\| &= \max_{t\in\mathscr{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)' X_{it}\right]\right\| \\
&\lesssim \max_{i\in\mathscr{N}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2 = O_p(\sqrt{(\log N)/T}) \\
\max_{t\in\mathscr{T}_\ell}\left\|J_{4,t}^{(\ell)}\right\| &= \max_{t\in\mathscr{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} X_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right)\lambda_i^{0\prime}\right] f_t^0\right\| \\
&\lesssim \max_{i\in\mathscr{N}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2 = O_p(\sqrt{(\log N)/T}), \\
\max_{t\in\mathscr{T}_\ell}\left\|J_{5,t}^{(\ell)}\right\|_2 &= \max_{t\in\mathscr{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} X_i^{(\ell)}\left(\theta_i^{0,(\ell)} - \hat{\theta}_i^{(\ell)}\right) e_{it}\right]\right\| \\
&\lesssim \max_{i\in\mathscr{N}}\left\|\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}\right\|_2 = O_p(\sqrt{(\log N)/T}).
\end{aligned}$$

Next,

$$\max_{t\in\mathscr{T}_\ell}\left\|J_{6,t}^{(\ell)}\right\| = \max_{t\in\mathscr{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n} e_i^{(\ell)}\lambda_i^{0\prime}\right] f_t^0\right\| \lesssim \frac{1}{n\sqrt{T_\ell}}\left\|\sum_{i=n_1}^{n_n} e_i^{(\ell)}\lambda_i^{0\prime}\right\| = O_p(N^{-1/2}),$$

by the fact that

$$\mathbb{E}\left(\frac{1}{nT_\ell}\left\|\sum_{i=n_1}^{n_n} e_i^{(\ell)}\lambda_i^{0\prime}\right\|^2\middle|\mathscr{D}\right) = \mathbb{E}\left(\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}\left\|\frac{1}{\sqrt{n}}\sum_{i\in[N]} e_{it}\lambda_i^0\right\|^2\middle|\mathscr{D}\right) = O_p(1)$$

with the same manner as (B.45). Similarly,

$$\max_{t\in\mathcal{T}_\ell}\left\|J_{7,t}^{(\ell)}\right\| = \max_{t\in\mathcal{T}_\ell}\left\|V_{NT}^{(\ell)-1}\hat{F}^{(\ell)\prime}F^{0,(\ell)}\left[\frac{1}{nT_\ell}\sum_{i=n_1}^{n_n}\lambda_i^0 e_{it}\right]\right\| \lesssim \max_{t\in\mathcal{T}_\ell}\left\|\frac{1}{n}\sum_{i=n_1}^{n_n}\lambda_i^0 e_{it}\right\|$$
$$= O_p(\sqrt{(\log T)/N}),$$

by using the Bernstein's inequality for the independent sequence in Lemma B.5(i). For $J_{8,t}^{(\ell)}$, we have

$$\max_{t\in\mathcal{T}_\ell}\left\|J_{8,t}^{(\ell)}\right\| \lesssim \frac{1}{\sqrt{n}}\max_{t\in\mathcal{T}_\ell}\left\|\frac{1}{\sqrt{nT_\ell}}\sum_{i=n_1}^{n_n}e_i^{(\ell)}e_{it}\right\| = \frac{1}{\sqrt{n}}\sqrt{\max_{t\in\mathcal{T}_\ell}\frac{1}{T_\ell}\sum_{s\in[T_\ell]}\left(\frac{1}{\sqrt{n}}\sum_{i=n_1}^{n_n}e_{is}e_{it}\right)^2}$$
$$\leq \frac{1}{\sqrt{n}}\sqrt{\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in[T_\ell]}\left(\frac{1}{\sqrt{n}}\sum_{i=n_1}^{n_n}e_{is}e_{it}\right)^2} = O_p(N^{-1/2}),$$

where the last equality holds by the fact that $\mathbb{E}[\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\sum_{s\in[T_\ell]}(\frac{1}{\sqrt{N}}\sum_{i=n_1}^{n_n}e_{is}e_{it})^2] = O(1)$ by (B.52). In sum, we have $\max_t ||\hat{f}_t - H^{(\ell)\prime}f_t^0|| = O_p(\sqrt{(\log N\vee T)/(N\wedge T)})$.

■

Lemma B.19. *Under Assumptions 3.1*, 3.2 and 3.8, we have*

$$\max_{i\in\mathcal{N}}\left\|\hat{S}_{ii}^{(\ell)} - \mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathcal{D}\right)\right\| = O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right) \text{ and } \max_{i\in\mathcal{N}}\left\|\hat{\Omega}_i^{(\ell)} - \Omega_i^{0,(\ell)}\right\| = o_p(1).$$

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 (B.64), Lemma B.16(i) and Lemma B.17(iii), we have

$$\max_{i\in\mathcal{N}}\left\|\hat{S}_{ii}^{(\ell)} - \mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathcal{D}\right)\right\| = O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right).$$

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 3.1*(iii), we have $\hat{\Omega}_i^{(\ell)} = \frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\hat{\mathfrak{x}}_{it}^{(\ell)}\hat{\mathfrak{x}}_{it}^{(\ell)\prime}\hat{e}_{it}^2$ and $\Omega_i^{0,(\ell)} = \frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\mathbb{E}(\mathfrak{x}_{it}^{(\ell)}\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2)$. It remains to show

$$\max_{i\in\mathcal{N}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left[\hat{\mathfrak{x}}_{it}^{(\ell)}\hat{\mathfrak{x}}_{it}^{(\ell)\prime}\hat{e}_{it}^2 - \mathbb{E}(\mathfrak{x}_{it}^{(\ell)}\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2)\right]\right\| = o_p(1). \tag{B.67}$$

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

$$\mathfrak{x}_{it}^{(\ell)} - \hat{\mathfrak{x}}_{it}^{(\ell)} = \frac{1}{T_\ell}X_i^{(\ell)\prime}(\hat{F}^{(\ell)}\hat{f}_t^0 - F^{0,(\ell)}f_t^0). \tag{B.68}$$

Note that

$$\max_{t\in\mathcal{T}_\ell}\frac{1}{\sqrt{T_\ell}}\left\|\hat{F}^{(\ell)}\hat{f}_t^0-F^{0,(\ell)}f_t^0\right\|_2$$
$$\leq\frac{\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|}{\sqrt{T_\ell}}\max_t\left\|\hat{f}_t^{(\ell)}-H^{(\ell)\prime}f_t^0\right\|+\frac{\left\|F^{0,(\ell)}H^{(\ell)}\right\|}{\sqrt{T_\ell}}\max_t\left\|\hat{f}_t^{(\ell)}-H^{(\ell)\prime}f_t^0\right\|$$
$$+\frac{\left\|\hat{F}^{(\ell)}-F^{0,(\ell)}H^{(\ell)}\right\|}{\sqrt{T_\ell}}\max_t\left\|H^{(\ell)\prime}f_t^0\right\|$$
$$=O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right),$$

by Lemma B.18(i) and (iv) and Lemma B.7(i). Then by (B.68) and Assumption 3.8(ii), we have

$$\max_{i\in\mathcal{N},t\in\mathcal{T}_\ell}\left\|\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right\|\leq\frac{1}{\sqrt{T_\ell}}\|X_i\|O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right)=O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right). \tag{B.69}$$

Next, for $i\in\mathcal{N},t\in\mathcal{T}_\ell$, note that

$$\hat{e}_{it}=Y_{it}-X_{it}'\hat{\theta}_i^{(\ell)}-\hat{\lambda}_i^{(\ell)\prime}\hat{f}_t^{(\ell)}$$
$$=e_{it}-\left[X_{it}'\left(\theta_i^{0,(\ell)}-\hat{\theta}_i^{(\ell)}\right)+\left(\hat{\lambda}_i^{(\ell)}-H^{(\ell)-1}\lambda_i^0\right)'\left(\hat{f}_t^{(\ell)}-H^{(\ell)\prime}f_t^0\right)\right.$$
$$\left.+\left(\hat{\lambda}_i^{(\ell)}-H^{(\ell)-1}\lambda_i^0\right)'H^{(\ell)\prime}f_t^0+\left(H^{(\ell)-1}\lambda_i^0\right)'\left(\hat{f}_t^{(\ell)}-H^{(\ell)\prime}f_t^0\right)\right].$$

Then

$$\max_{i\in\mathcal{N},t\in\mathcal{T}_\ell}|\hat{e}_{it}-e_{it}|=O_p\left(\sqrt{\frac{\log N}{T}}(NT)^{1/q}\right)+O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right),\quad\text{and}$$
$$\hat{e}_{it}^2-e_{it}^2=e_{it}\left(\hat{e}_{it}-e_{it}\right)+\left(\hat{e}_{it}-e_{it}\right)^2=e_{it}X_{it}'R_{1,it}+R_{2,it} \tag{B.70}$$
$$\text{s.t.}\max_{i\in\mathcal{N},t\in\mathcal{T}_\ell}\left\|R_{1,it}\right\|_2=O_p\left(\sqrt{\frac{\log N}{T}}\right),\max_{i\in[N],t\in\mathcal{T}_\ell}\left|R_{2,it}\right|=O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right)$$

by Lemmas B.17(iii), B.18(iii), and B.18(iv). It follows that

$$\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\hat{\mathfrak{x}}_{it}^{(\ell)}\hat{\mathfrak{x}}_{it}^{(\ell)\prime}\hat{e}_{it}^2$$
$$=\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\mathfrak{x}_{it}^{(\ell)}\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)'\left(\hat{e}_{it}^2-e_{it}^2\right)$$
$$+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)'e_{it}^2+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\mathfrak{x}_{it}^{(\ell)}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)'\left(\hat{e}_{it}^2-e_{it}^2\right)$$
$$+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)\mathfrak{x}_{it}^{(\ell)\prime}\left(\hat{e}_{it}^2-e_{it}^2\right)+\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell}\left(\mathfrak{x}_{it}^{(\ell)}-\hat{\mathfrak{x}}_{it}^{(\ell)}\right)\mathfrak{x}_{it}^{(\ell)\prime}e_{it}^2$$

$$
\begin{aligned}
&+\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}\mathfrak{r}_{it}^{(\ell)}\left(\mathfrak{r}_{it}^{(\ell)}-\hat{\mathfrak{r}}_{it}^{(\ell)}\right)'e_{it}^2+\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}\mathfrak{r}_{it}^{(\ell)}\mathfrak{r}_{it}^{(\ell)\prime}\left(\hat{e}_{it}^2-e_{it}^2\right)\\
&=\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}\mathfrak{r}_{it}^{(\ell)}\mathfrak{r}_{it}^{(\ell)\prime}e_{it}^2+O_p\left(\sqrt{\frac{\log N\vee T}{N\wedge T}}\right)\quad\text{uniformly over } i\in\mathscr{N}, \qquad \text{(B.71)}
\end{aligned}
$$

where the last line holds by (B.69), (B.70), and Assumptions 3.8(ii) and 3.1*(iv). Using similar arguments as used to derive (B.61) by the Bernstein's inequality for m.d.s., for a positive constant c_9, we have

$$
\mathbb{P}\left\{\max_{i\in\mathscr{N}}\left\|\frac{1}{T_\ell}\sum_{t\in\mathscr{T}_\ell}\left[\mathfrak{r}_{it}^{(\ell)}\mathfrak{r}_{it}^{(\ell)\prime}e_{it}^2-\mathbb{E}\left(\mathfrak{r}_{it}^{(\ell)}\mathfrak{r}_{it}^{(\ell)\prime}e_{it}^2\right)\right]\right\|>c_9\sqrt{\frac{\log N}{T}}\right\}=o(1). \qquad \text{(B.72)}
$$

Combining (B.71) and (B.72), we obtain (B.67). ■

Lemma B.20. *Under Assumptions 3.1*, 3.2 and 3.8, we have* $\hat{\Gamma}^{(\ell)}\rightsquigarrow\mathbb{N}(0,1)$ *under* H_0.

Proof Under the null that $\theta_i^{0,(\ell)}=\theta^{0,(\ell)}$ for $\forall i\in\mathscr{N}$, we use Lemma B.17(i) to obtain that

$$
\begin{aligned}
\hat{\hat{\theta}}^{(\ell)}-\theta^{0,(\ell)}=&\frac{1}{n}\sum_{i\in\mathscr{N}}\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_i^{(\ell)}\\
&+\frac{1}{n^2}\sum_{i\in\mathscr{N}}a_{ii}^0\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\left(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right)+O_p\left(\frac{\log N}{N\wedge T}\right),
\end{aligned}
\qquad \text{(B.73)}
$$

such that

$$
\left\|\frac{1}{n^2}\sum_{i\in\mathscr{N}}a_{ii}^0(\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)})\right\|\le\frac{1}{n}\max_{i\in\mathscr{N}}\left|a_{ii}^{(\ell)}\right|\left\|\hat{\theta}_i^{(\ell)}-\theta_i^{0,(\ell)}\right\|=O_p\left(\frac{\sqrt{\log N}}{N\sqrt{T}}\right)
$$

with Lemma B.17(iii) and the fact that $\max_{i\in\mathscr{N}}\left|a_{ii}^0\right|=O(1)$. For the first term on the right side of (B.73), we have

$$
\frac{1}{n}\sum_{i\in\mathscr{N}}\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\xi_i^{(\ell)}=\frac{1}{nT_\ell}\sum_{i\in\mathscr{N}}\sum_{t\in\mathscr{T}_\ell}\left[\mathbb{E}\left(S_{ii}^{0,(\ell)}|\mathscr{D}\right)\right]^{-1}\mathfrak{r}_{it}e_{it}=O_p(\frac{1}{\sqrt{nT}})
$$

by the central limit theorem for m.d.s., which yields that

$$
\left\|\hat{\hat{\theta}}^{0,(\ell)}-\theta^{0,(\ell)}\right\|=O_p\left(\frac{\log N}{N\wedge T}\right). \qquad \text{(B.74)}
$$

Recall from (B.44) that $\hat{\Gamma}^{(\ell)}=\sqrt{n}\cdot\frac{\frac{1}{n}\sum_{i\in\mathscr{N}}\hat{\mathbb{S}}_i^{(\ell)}-p}{\sqrt{2p}}$ such that

$$
\begin{aligned}
\hat{\mathbb{S}}_i^{(\ell)}&=T_\ell\left(\hat{\theta}_i^{(\ell)}-\hat{\hat{\theta}}^{(\ell)}\right)'\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\left(\hat{\theta}_i^{(\ell)}-\hat{\hat{\theta}}^{(\ell)}\right)\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2\\
&=T_\ell\left(\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)}\right)'\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\left(\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)}\right)\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2
\end{aligned}
$$

$$
\begin{aligned}
&+T_\ell\left(\hat{\bar{\theta}}^{(\ell)}-\theta^{0,(\ell)}\right)'\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\left(\hat{\bar{\theta}}^{(\ell)}-\theta^{0,(\ell)}\right)\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2\\
&-2T_\ell\left(\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)}\right)'\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\left(\hat{\bar{\theta}}^{(\ell)}-\theta^{0,(\ell)}\right)\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2\\
&:=\hat{\mathbb{S}}_{i,1}^{(\ell)}+\hat{\mathbb{S}}_{i,2}^{(\ell)}-\hat{\mathbb{S}}_{i,3}^{(\ell)}.
\end{aligned}
$$

Below we show that $\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\hat{\mathbb{S}}_{i,2}^{(\ell)}$ and $\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\hat{\mathbb{S}}_{i,3}^{(\ell)}$ are smaller terms and $\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\hat{\mathbb{S}}_{i,1}^{(\ell)}\rightsquigarrow\mathscr{N}(0,1)$.

First, noted that

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\hat{\mathbb{S}}_{i,2}^{(\ell)}\right|\leq\sqrt{n}T_\ell\left\|\hat{\bar{\theta}}^{(\ell)}-\theta^{0,(\ell)}\right\|^2\max_{i\in\mathscr{N}}\lambda_{\max}\left(\hat{S}_{ii}^{(\ell)}\left(\hat{\Omega}_i^{(\ell)}\right)^{-1}\hat{S}_{ii}^{(\ell)}\right)\max_{i\in\mathscr{N}}\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2\\
&=\sqrt{n}T_\ell O_p\left(\frac{(\log N)^2}{N^2\wedge T^2}\right)\max_{i\in\mathscr{N}}\left\|\hat{S}_{ii}^{(\ell)}\right\|^2\max_{i\in\mathscr{N}}\left\|\hat{\Omega}_i^{(\ell)}\right\|\max_{i\in\mathscr{N}}\left(1-\hat{a}_{ii}^{(\ell)}/n\right)^2\\
&=\sqrt{n}T_\ell O_p\left(\frac{(\log N)^2}{N^2\wedge T^2}\right)\left[\max_{i\in\mathscr{N}}\left\|S_{ii}^{0,(\ell)}\right\|^2\max_{i\in\mathscr{N}}\left\|\Omega_i^{0,(\ell)}\right\|\max_{i\in\mathscr{N}}\left(1-a_{ii}^0/N\right)^2+o_p(1)\right]\\
&=o_p(1),
\end{aligned}
$$

where the first equality is by (B.74), the second equality is by Lemma B.19 and the fact that $\max_{i\in\mathscr{N}}|\hat{a}_{ii}^{(\ell)}-a_{ii}^0|=o_p(1)$ owing to Lemma B.18(iii), and the last equality holds by the fact that $\max_{i\in\mathscr{N}}||S_{ii}^{0,(\ell)}||=O(1)$, $\max_{i\in\mathscr{N}}||\Omega_i^{0,(\ell)}||=O(1)$, and $\max_{i\in\mathscr{N}}|a_{ii}^{(\ell)}|=O(1)$ and Assumption 3.1*(vi).

Second, by analogous arguments as used above, we have

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\hat{\mathbb{S}}_{i,3}^{(\ell)}\right|\\
&\leq 2\sqrt{n}T_\ell\max_{i\in\mathscr{N}}\left\|\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)}\right\|\left\|\hat{\bar{\theta}}^{(\ell)}-\theta^{0,(\ell)}\right\|\max_{i\in\mathscr{N}}\lambda_{\max}\left(\hat{S}_{ii}^{(\ell)}(\hat{\Omega}_i^{(\ell)})^{-1}\hat{S}_{ii}^{(\ell)}\right)\max_{i\in\mathscr{N}}(1-\hat{a}_{ii}^{(\ell)}/n)^2\\
&=\sqrt{n}T_\ell O_p\left(\sqrt{\frac{\log N}{T}}\right)O_p\left(\frac{\log N}{N\wedge T}\right)=o_p(1).
\end{aligned}
$$

At last for $\hat{\mathbb{S}}_{i,1}^{(\ell)}$, it's clear that

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\hat{\mathbb{S}}_{i,1}^{(\ell)}&=\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}T_\ell(\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)})'\hat{S}_{ii}^{(\ell)}(\hat{\Omega}_i^{(\ell)})^{-1}\hat{S}_{ii}^{(\ell)}(\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)})(1-\hat{a}_{ii}^{(\ell)}/n)^2\\
&=\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\hat{z}_i^{(\ell)}+o_p(1),
\end{aligned}
$$

where $\hat{z}_i^{(\ell)}=T_\ell(\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)})'S_{ii}^{0,(\ell)}(\Omega_i^{0,(\ell)})^{-1}S_{ii}^{0,(\ell)}(\hat{\theta}_i^{(\ell)}-\theta^{0,(\ell)})\left(1-a_{ii}^0/n\right)^2$.

Then by the central limit theorem, we have $\hat{\Gamma}^{(\ell)}=\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\frac{\hat{z}_i^{(\ell)}-p}{\sqrt{2p}}+o_p(1)\rightsquigarrow\mathbb{N}(0,1)$.

■

Lemma B.21. *Under Assumptions 3.1*, 3.2 and 3.8, we have* $|\hat{\Gamma}^{(\ell)}| \to \infty$ *under* H_1 *if* $\frac{T_\ell}{\sqrt{n}} \sum_{i\in\mathscr{N}} ||c_i^{(\ell)}||^2 \to \infty$.

Proof Noting that $\theta_i^{0,(\ell)} = \theta^{0,(\ell)} + c_i^{(\ell)}$, we have

$$\begin{aligned}
\hat{\theta}_i^{(\ell)} - \hat{\bar{\theta}}^{(\ell)} &= (\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}) - (\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)}) + \theta_i^{0,(\ell)} - \theta^{0,(\ell)} \\
&= (\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)}) - (\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)}) + c_i^{(\ell)}.
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\mathbb{S}}_i^{(\ell)} &= T_\ell(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)})'\hat{S}_{ii}^{(\ell)}(\hat{\Omega}_i^{(\ell)})^{-1}\hat{S}_{ii}^{(\ell)}(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)})(1-\hat{a}_{ii}^{(\ell)}/n)^2 \\
&\quad + T_\ell(\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)})'\hat{S}_{ii}^{(\ell)}(\hat{\Omega}_i^{(\ell)})^{-1}\hat{S}_{ii}^{(\ell)}(\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)})(1-\hat{a}_{ii}^{(\ell)}/n)^2 \\
&\quad - 2T_\ell(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)})'\hat{S}_{ii}^{(\ell)}(\hat{\Omega}_i^{(\ell)})^{-1}\hat{S}_{ii}^{(\ell)}(\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)})(1-\hat{a}_{ii}^{(\ell)}/n)^2 \\
&\quad + T_\ell c_i^{(\ell)\prime}\hat{S}_{ii}^{(\ell)}(\hat{\Omega}_i^{(\ell)})^{-1}\hat{S}_{ii}^{(\ell)}c_i^{(\ell)}(1-\hat{a}_{ii}^{(\ell)}/n)^2 \\
&\quad + 2T_\ell(\hat{\theta}_i^{(\ell)} - \theta_i^{0,(\ell)})'\hat{S}_{ii}^{(\ell)}(\hat{\Omega}_i^{(\ell)})^{-1}\hat{S}_{ii}^{(\ell)}c_i^{(\ell)}(1-\hat{a}_{ii}^{(\ell)}/n)^2 \\
&\quad - 2T_\ell(\hat{\bar{\theta}}^{(\ell)} - \bar{\theta}^{0,(\ell)})'\hat{S}_{ii}^{(\ell)}(\hat{\Omega}_i^{(\ell)})^{-1}\hat{S}_{ii}^{(\ell)}c_i^{(\ell)}(1-\hat{a}_{ii}^{(\ell)}/n)^2 := \sum_{m=4}^{9}\hat{\mathbb{S}}_{i,m}^{(\ell)}.
\end{aligned}$$

In the proof of Lemma B.20, we have already shown that

$$\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\frac{\hat{\mathbb{S}}_{i,4}^{(\ell)} - p}{\sqrt{2p}} \rightsquigarrow \mathbb{N}(0,1) \quad \text{and} \quad \left|\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\hat{\mathbb{S}}_{i,m}^{(\ell)}\right| = o_p(1),\ m=5,6.$$

As for $\hat{\mathbb{S}}_{i,7}^{(\ell)}$, we can show that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{N}}\hat{\mathbb{S}}_{i,7}^{(\ell)} &= \frac{T_\ell}{\sqrt{n}}\sum_{i\in\mathscr{N}} c_i^{(\ell)\prime}\hat{S}_{ii}^{(\ell)}(\hat{\Omega}_i^{(\ell)})^{-1}\hat{S}_{ii}^{(\ell)}c_i^{(\ell)}(1-\hat{a}_{ii}^{(\ell)}/n)^2 \\
&= \frac{T_\ell}{\sqrt{n}}\sum_{i\in\mathscr{N}}\left[c_i^{(\ell)\prime}S_{ii}^{0,(\ell)}(\Omega_i^{0,(\ell)})^{-1}S_{ii}^{0,(\ell)}c_i^{(\ell)}\left(1-a_{ii}^0/n\right)^2 + o_p(1)\right] \\
&\geq \frac{T_\ell}{\sqrt{n}}\sum_{i\in\mathscr{N}}\left[[\lambda_{\max}(S_{ii}^{0,(\ell)-1}\Omega_i^{0,(\ell)}S_{ii}^{0,(\ell)-1})]^{-1}\left\|c_i^{(\ell)}\right\|_2^2\left(1-a_{ii}^0/n\right)^2 + o_p(1)\right] \\
&\geq \frac{T_\ell}{\sqrt{n}}\sum_{i\in\mathscr{N}}\left[\frac{\left\|c_i^{(\ell)}\right\|_2^2\left(1-a_{ii}^0/n\right)^2}{\left\|S_{ii}^{0,(\ell)-1}\right\|^2\left\|\Omega_i^{0,(\ell)}\right\|} + o_p(1)\right] \\
&= \frac{T_\ell}{\sqrt{n}}\sum_{i\in\mathscr{N}}\left[\frac{\left\|c_i^{(\ell)}\right\|_2^2}{\left\|S_{ii}^{0,(\ell)-1}\right\|^2\left\|\Omega_i^{0,(\ell)}\right\|} + o_p(1)\right] \\
&\geq \frac{1}{\max_{i\in\mathscr{N}}\left\|S_{ii}^{0,(\ell)-1}\right\|^2\left\|\Omega_i^{0,(\ell)}\right\|}\frac{T_\ell}{\sqrt{n}}\left[\sum_{i\in\mathscr{N}}\left\|c_i^{(\ell)}\right\|_2^2 + o_p(1)\right] \to \infty \text{ at a rate faster than (log}
\end{aligned}$$

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 $\max_{i\in\mathcal{N}}|\frac{a_{ii}^0}{n}| = o_p(1)$ and the last line draws from the assumption that $\frac{T_\ell}{\sqrt{n}}\sum_{i\in\mathcal{N}}||c_i^{(\ell)}||^2/(\log N)^{1/2} \to \infty$. Following this, by Cauchy's inequality, we also observe that

$$\left|\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{N}}\hat{\mathbb{S}}_{i,8}^{(\ell)}\right| \le 2\sqrt{\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{N}}\hat{\mathbb{S}}_{i,4}^{(\ell)}}\sqrt{\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{N}}\hat{\mathbb{S}}_{i,7}^{(\ell)}} = o_p\left(\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{N}}\hat{\mathbb{S}}_{i,7}^{(\ell)}\right),$$
$$\left|\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{N}}\hat{\mathbb{S}}_{i,9}^{(\ell)}\right| \le 2\sqrt{\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{N}}\hat{\mathbb{S}}_{i,5}^{(\ell)}}\sqrt{\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{N}}\hat{\mathbb{S}}_{i,7}^{(\ell)}} = o_p\left(\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{N}}\hat{\mathbb{S}}_{i,7}^{(\ell)}\right).$$

Combining arguments above, we obtain that $|\hat{\Gamma}^{(\ell)}|/(\log N)^{1/2} \to \infty$. ■

B.5 Algorithm for Nuclear Norm Regularization

To solve the optimization problem in (3.3), there are different algorithms in the literature such as the Alternating Direction Method of Multipliers (ADMM) algorithm and the singular value thresholding (SVT) procedure. Wang et al. (2022) provide the ADMM algorithm based on the quantile regression framework, which can be easily extended to the linear conditional mean regression framework. In this section, we focus on the SVT procedure for the case of low-rank estimation with two regressors. The case of more than two regressors is self-evident.

We can iteratively use SVT estimation to obtain the nuclear norm regularized regression estimates. Specifically, given Θ_1 and Θ_2, we solve for Θ_0 with

$$\Theta_0(\Theta_1,\Theta_2) = \arg\min_{\Theta_0} \|Y - X_1\odot\Theta_1 - X_2\odot\Theta_2 - \Theta_0\|_F + \nu_0 NT\,\|\Theta_0\|_*.$$

Given Θ_0 and Θ_2, we solve for Θ_1 with

$$\Theta_1(\Theta_0,\Theta_2) = \arg\min_{\Theta_1} \|Y - \Theta_0 - X_2\odot\Theta_2 - X_1\odot\Theta_1\|_F + \nu_1 NT\,\|\Theta_1\|_*.$$

Given Θ_0 and Θ_1, we solve for Θ_2 with

$$\Theta_2(\Theta_0,\Theta_1) = \arg\min_{\Theta_2} \|Y - \Theta_0 - X_1\odot\Theta_1 - X_2\odot\Theta_2\|_F + \nu_2 NT\,\|\Theta_2\|_*.$$

Specifically, the algorithm goes as follows:

Step 1: initialize Θ_0, Θ_1 and Θ_1 to be Θ_0^1, Θ_1^1 and Θ_1^1 and set $k = 1$.

Step 2: let

$$\Theta_0^{k+1} = S_{\frac{\nu_0 NT}{2}}\left(Y - X_1\odot\Theta_1^k - X_2\odot\Theta_2^k\right),$$

$$\Theta_1^{k+1} = S_{\frac{\tau v_1 NT}{2}} \left(\Theta_1^k - \tau X_1 \odot \left(X_1 \odot \Theta_1^k - Y + \Theta_0^{k+1} + X_2 \odot \Theta_2^k \right) \right),$$
$$\Theta_2^{k+1} = S_{\frac{\tau v_2 NT}{2}} \left(\Theta_2^k - \tau X_2 \odot \left(X_2 \odot \Theta_2^k - Y + \Theta_0^{k+1} + X_1 \odot \Theta_1^{k+1} \right) \right),$$
$$k = k+1,$$

where τ is the step size, and $S_\lambda(M)$ is the singular value operator for any matrix M and fixed parameter λ. By SVD, we have $M = U_M D_M V_M'$. Define $D_{M,\lambda}$ by replacing the diagonal entry $D_{M,ii}$ of D_M by $\max(D_{M,ii} - \lambda, 0)$, and then let $S_\lambda(M) = U_M D_{M,\lambda} V_M'$.

Step 3: repeat step 2 until convergence.

We can follow Chernozhukov et al. (2019), which gives the expression to pin down τ. Besides, Proposition 2.1 in Chernozhukov et al. (2019) also shows the convergence for the above algorithm.

Bibliography

Aktar, İ., Demirci, N., and Öztürk, L. (2009). Can unemployment be cured by economic growth and foreign direct investment? *Sosyal Ekonomik Araştırmalar Dergisi*, 9(17):452–467.

Ali, A., Kolter, J. Z., and Tibshirani, R. J. (2016). The multiple quantile graphical model. *Advances in Neural Information Processing Systems*, 29.

Alidaee, H., Auerbach, E., and Leung, M. P. (2020). Recovering network structure from aggregated relational data using penalized regression. *arXiv preprint arXiv:2001.06052*.

Almeida, H. and Campello, M. (2007). Financial constraints, asset tangibility, and corporate investment. *The Review of Financial Studies*, 20(5):1429–1460.

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.

Aquaro, M., Bailey, N., and Pesaran, M. H. (2021). Estimation and inference for spatial models with heterogeneous coefficients: An application to us house prices. *Journal of Applied Econometrics*, 36(1):18–44.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, pages 1–15.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

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. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

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

Bai, J. and Ng, S. (2020). Simpler proofs for approximate factor models of large dimensions. *arXiv preprint arXiv:2008.00254*.

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

Bai, J. and Perron, P. (2003). Critical values for multiple structural change tests. *The Econometrics Journal*, 6(1):72–78.

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.

Baltagi, B. H., Kao, C., and Wang, F. (2021). Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics*, 220(2):349–365.

Belloni, A., Chen, M., Madrid Padilla, O. H., and Wang, Z. (2023). High-dimensional latent panel quantile regression with an application to asset pricing. *The Annals of Statistics*, 51(1):96–121.

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.

Capozza, D. R., Hendershott, P. H., Mack, C., and Mayer, C. J. (2002). Determinants of real house price dynamics.

Castagnetti, C., Rossi, E., and Trapani, L. (2015). Inference on factor structures in heterogeneous panels. *Journal of Econometrics*, 184(1):145–157.

Chen, L. (2015). Estimating the common break date in large factor models. *Economics Letters*, 131:70–74.

Chen, L. (2022). Two-step estimation of quantile panel data models with interactive fixed effects. *Econometric Theory*, forthcoming.

Chen, L., Dolado, J. J., and Gonzalo, J. (2021). Quantile factor models. *Econometrica*, 89(2):875–910.

Chen, Y., Chi, Y., Fan, J., Ma, C., and Yan, Y. (2020). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM journal on optimization*, 30(4):3098–3121.

Cheng, X., Liao, Z., and Schorfheide, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies*, 83(4):1511–1543.

Chernozhukov, V., Hansen, C. B., Liao, Y., and Zhu, Y. (2019). 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.

Cleary, S. (1999). The relationship between firm investment and financial status. *The Journal of Finance*, 54(2):673–692.

Craigwell, R. (2006). Foreign direct investment and employment in the english and dutch-speaking caribbean. *ILO Subregional Office for the Caribbean*.

Devereux, M. and Schiantarelli, F. (1990). Investment, financial factors, and cash row: Evidence from uk panel data. *Asymmetric Information, Corporate Finance, and Investment*, pages 279–306.

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

Fan, J., Gong, W., and Zhu, Z. (2019). Generalized high-dimensional trace regression via nuclear norm regularization. *Journal of econometrics*, 212(1):177–202.

Fazzari, S., Hubbard, R. G., and Petersen, B. C. (1988). Financing constraints and corporate investment. *Brookings Papers on Economic Activity*, (1):141–206.

Feng, J. (2019). Regularized quantile regression with interactive fixed effects. *Econometric Theory*, forthcoming.

Gallin, J. (2006). The long-run relationship between house prices and income: evidence from local housing markets. *Real Estate Economics*, 34(3):417–438.

Galvao, A. F., Gu, J., and Volgushev, S. (2020). On the unbiased asymptotic normality of quantile regression with fixed effects. *Journal of Econometrics*, 218(1):178–215.

Galvao, A. F., Juhl, T., Montes-Rojas, G., and Olmo, J. (2018). Testing slope homogeneity in quantile regression panel data with an application to the cross-section of stock returns. *Journal of Financial Econometrics*, 16(2):211–243.

Galvao, A. F. and Kato, K. (2016). Smoothed quantile regression for panel data. *Journal of Econometrics*, 193(1):92–112.

Galvao, A. F. and Montes-Rojas, G. V. (2010). Penalized quantile regression for dynamic panel data. *Journal of Statistical Planning and Inference*, 140(11):3476–3497.

Galvao, A. F. and Wang, L. (2015). Efficient minimum distance estimator for quantile regression fixed effects panel data. *Journal of Multivariate Analysis*, 133:1–26.

Gilchrist, S. and Himmelberg, C. P. (1995). Evidence on the role of cash flow for investment. *Journal of Monetary Economics*, 36(3):541–572.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations*. JHU Press.

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

Hong, S., Su, L., and Jiang, T. (2023). 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.

Kaplan, S. N. and Zingales, L. (1995). Do financing constraints explain why investment is correlated with cash flow? *NBER Working paper*, (No. 5267).

Karlsson, S., Lundin, N., Sjöholm, F., and He, P. (2009). Foreign firms and chinese employment. *World Economy*, 32(1):178–201.

Kato, K., Galvao Jr, A. F., and Montes-Rojas, G. V. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170(1):76–91.

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.

Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *Annals of Statistics*, pages 755–770.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89.

Koltchinskii, V., Lounici, K., and Tsybakov, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5):2302 – 2329.

Leadbetter, M. and Rootzen, H. (1988). Extremal theory for stochastic processes. *The Annals of Probability*, pages 431–478.

Ledoux, M. and Talagrand, M. (1989). Comparison theorems, random geometry and some limit theorems for empirical processes. *The Annals of Probability*, pages 596–631.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: isoperimetry and processes*, volume 23. Springer Science & Business Media.

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. (2023). Uniform inference in linear panel data models with two-dimensional heterogeneity. *Journal of Econometrics*, forthcoming.

Lumsdaine, R. L., Okui, R., and Wang, W. (2023). Estimation of panel group structure models with structural breaks in group memberships and coefficients. *Journal of Econometrics*, 233(1):45–65.

Ma, S. and Su, L. (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics*, 207(1):1–29.

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

Machado, J. A. and Silva, J. S. (2019). Quantiles via moments. *Journal of Econometrics*, 213(1):145–173.

Malpezzi, S. (1999). A simple error correction model of house prices. *Journal of housing economics*, 8(1):27–62.

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. (2023). High-dimensional vars with common factors. *Journal of Econometrics*, 233(1):155–183.

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*.

Mucuk, M. and Demirsel, M. T. (2013). The effect of foreign direct investments on unemployment: Evidence from panel data for seven developing countries. *Journal of Business Economics and Finance*, 2(3):53–66.

Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 39(2):1069 – 1097.

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

Ortalo-Magne, F. and Rady, S. (2006). Housing market dynamics: On the contribution of income shocks and credit constraints. *The Review of Economic Studies*, 73(2):459–485.

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.

Rakhlin, A., Sridharan, K., and Tewari, A. (2015). Sequential complexities and uniform martingale laws of large numbers. *Probability Theory and Related Fields*, 161(1-2):111–153.

Rauh, J. D. (2006). Investment and financing constraints: Evidence from the funding of corporate pension plans. *The Journal of Finance*, 61(1):33–71.

Rohe, K., Chatterjee, S., and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878 – 1915.

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.

Strat, V. A., Davidescu, A., and Paul, A. M. (2015). Fdi and the unemployment-a causality analysis for the latest eu members. *Procedia Economics and Finance*, 23:635–643.

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

Su, L. and Chen, Q. (2013b). 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., Wang, W., and Xu, X. (2023). Identifying latent group structures in spatial dynamic panels. *Journal of Econometrics*.

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 Institute of Technology.

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*.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.

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.