10-2018

Identifying latent grouped patterns in panel data models with interactive fixed effects

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Gaosheng JU
Fudan University

Published in Journal of Econometrics, Available online 30 June 2018, pp. 1-20
https://doi.org/10.1016/j.jeconom.2018.06.014
Pre-print version

Identifying Latent Grouped Patterns in Panel Data Models with Interactive Fixed Effects*

Liangjun Su[a] and Gaosheng Ju[b]
[a] School of Economics, Singapore Management University
[b] School of Economics, Fudan University

August 8, 2016

Abstract

We consider the estimation of latent grouped patterns in dynamic panel data models with interactive fixed effects. We assume that the individual slope coefficients are homogeneous within a group and heterogeneous across groups but each individual's group membership is unknown to the researcher. We consider penalized principal component (PPC) estimation by extending the penalized-profile-likelihood-based C-Lasso of Su, Shi, and Phillips (2016) to panel data models with cross section dependence. Given the correct number of groups, we show that the C-Lasso can achieve simultaneous classification and estimation in a single step and exhibit the desirable property of uniform classification consistency. The C-Lasso-based PPC estimators of the group-specific parameters also have the oracle property. BIC-type information criteria are proposed to choose the numbers of factors and groups consistently and to select the data-driven tuning parameter. Simulations are conducted to demonstrate the finite-sample performance of the proposed method. We apply our C-Lasso to study the persistence of housing prices in China's large and medium-sized cities in the last decade and identify three groups.

JEL Classification: C33, C38, C51

Key Words: Classifier Lasso; Cross section dependence; Dynamic panel; High dimensionality; Latent structure; Parameter heterogeneity; Penalized method

*The authors thank the co-editors and referees for many constructive comments on the previous version of the paper. They also thank the participants at the 2016 Tsinghua International Conference on Econometrics in Beijing and the 2nd Dongbei Econometrics Workshop in Dalian, China. Su gratefully acknowledges the Singapore Ministry of Education for Tier-2 Academic Research Fund under grant number MOE2012-T2-2-021 and the funding support provided by the Lee Kong Chian Fund for Excellence. Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; E-mail: ljsu@smu.edu.sg, Phone: +65 6828 0386.

1 Introduction

Recently there has been a growing literature on large dimensional panel data models with interactive fixed effects (IFEs) where both the individual dimension N and the time dimension T pass to infinity. Most of the literature falls into two categories depending on whether the slope coefficients are allowed to be heterogeneous across individuals or not. The first category focuses on homogenous panel data models and includes Bai (2009), Greenaway-McGrevy et al. (2012), Bai and Li (2014), Moon and Weidner (2015a, 2015b), Lu and Su (2016), Su et al. (2015), and Su and Zhang (2016). The second category considers estimation and inference of heterogeneous panel data models; see Pesaran (2006), Kapetanios and Pesaran (2007), Chudik et al. (2011), Kapetanios et al. (2011), Pesaran and Tosetti (2011), Su and Jin (2012), Song (2013), Li and Lu (2014), Chudik and Pesaran (2015), among others. Although the assumption of slope homogeneity greatly facilitates the estimation and inference procedure for such models, inferences based on it could be misleading if the underlying models have heterogeneous slopes instead. On the other hand, if the models have homogeneous slopes, estimates based on slope heterogeneity could be inefficient and have slower rates of convergence. For this reason, Hsiao (2014, chapter 6) elaborates variable-coefficient models, Jin and Su (2013) propose a nonparametric test for poolability in nonparametric panel data models with IFEs, and Pesaran and Yamagata (2008) and Su and Chen (2013) propose various tests for slope homogeneity in linear panel data models with additive fixed effects (AFEs) and IFEs, respectively.

Since panel data usually cover individuals from different backgrounds over a period of time that frequently exhibit unobserved heterogeneity and neglecting it can lead to inconsistent estimation and misleading inference, it is of paramount importance to control unobserved heterogeneity in panel data models. The working paper version of Su, Shi, and Phillips (2016, SSP hereafter) documents a variety of empirical examples where cross-sectional slope homogeneity has been rejected in panel data models with AFEs. They also review the literature on the modeling of slope heterogeneity in such models and classify it into two broad categories. One assumes complete slope heterogeneity where the regression parameters are completely different for different individuals; see the survey by Baltagi et al. (2008) and Hsiao and Pesaran (2008). The other considers a panel structure model in which individuals form a number of homogeneous groups in a heterogeneous population, and the regression parameters are the same within each group but different across different groups. See, Bester and Hansen (2016), Sun (2005), Lin and Ng (2012), Bonhomme and Manresa (2015), and Sarafidis and Weber (2015), among others.

In this paper we follow the lead of SSP and extend their penalized estimation to the following panel data models with IFEs:

$$Y_{it} = \beta_i^{0\prime} X_{it} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it},\ \ i = 1, ..., N,\ \ t = 1, ..., T, \tag{1.1}$$

where X_{it} is a $p\times 1$ vector of predetermined variables that may include lagged dependent variables,

β_i^0 is a $p\times 1$ vector of unknown slope coefficients, λ_i^0 and F_t^0 denote an $R_0\times 1$ vector of unobservable factor loadings and common factors, respectively, both of which may be correlated with $\{X_{it}\}$, and ε_{it} is the idiosyncratic error term. Here, parameters with the superscript 0 denote the true values. We assume that the slope coefficients β_i^0's exhibit a certain grouped pattern so that they only take K_0 distinct values, say, $\left\{\alpha_k^0, k=1,...,K_0\right\}$ based on their group identities (see section 2.1 for detail).[1] Difficulty arises because the grouped pattern is unknown to the researcher. Due to the presence of IFEs, the penalized least squares (PLS) procedure of SSP is not applicable. We propose to extend the principal component (PC) approach to the current framework by considering penalized principal component (PPC) estimation of the unknown parameters in the model. Our PPC objective function is obtained by adding a novel penalty term to the usual form of the PC objective function that serves to shrink the individual slope parameter vectors β_i's to the unknown group-specific parameter vectors α_k's. Following the literature on large dimensional panel data models, we assume that both N and T pass to infinity.

Note that the parameters of interest in (1.1) include $\{\beta_i,\, i=1,...,N\}$, $\{\lambda_i,\, i=1,...,N\}$, and $\{F_t,\, t=1,...,T\}$. As both N and T pass to infinity, we have a divergent number of parameters to be estimated. For brevity we shall focus on the estimation and inference of β_i's as the asymptotics for the estimates of λ_i and F_t follow directly from Bai (2003). When we assume that β_i's exhibit certain grouped pattern, the *effective* number of unknown slope parameters in $\{\beta_i\}$ is not of order $O(N)$ but $O(K_0)$, where K_0 is typically a fixed constant in empirical applications. This motivates SSP and us to consider a variant of Lasso which can achieve simultaneous variable selection and estimation in a single step and is extremely useful when the set of parameters exhibit certain *sparsity* features. Unlike the typical Lasso or group-Lasso procedure that shrinks individual or groupwise coefficients to a fixed constant (zero), we have to shrink the individual coefficient vectors $\{\beta_i\}$ to certain *unknown* group-specific parameter vectors $\{\alpha_k\}$ through the use of a novel mixed *additive-multiplicative* penalty form. Our penalty has N *additive* terms, each of which takes a *multiplicative* expression as the product of K_0 penalty terms. Each of the K_0 penalty terms in the *multiplicative* expression shrinks the individual-level slope parameter vectors to a particular *unknown* group-level parameter vector. For easy reference, we also follow SSP and refer to our new Lasso method as the *classifier-Lasso* or *C-Lasso* method hereafter.

We first assume that K_0 and R_0 are known, demonstrate that our C-Lasso method can achieve simultaneous classification and estimation in a single step, and show that the PPC estimates of the slope parameters exhibit the desirable oracle property. Due to the presence of IFEs, the derivation of such results is much more difficult than that in SSP. First, we demonstrate the mean square convergence of the individual coefficient estimates and the factor estimates, based on which we can also obtain preliminary rates of consistency for both the individual-level and

[1] One can also consider the presence of grouped pattern in the factor loadings $\{\lambda_i\}$; see the remark at the end of Section 3.3.

group-level coefficient estimates. Second, we show that we can achieve the uniform classification consistency in the sense that all individuals belonging to a certain group can be classified into the same group correctly uniformly over both individuals and group identities with probability approaching one (w.p.a.1), and conversely, all individuals that are classified into a certain group belong to the same group uniformly over both individuals and group identities w.p.a.1. Third, based on the uniform classification consistency, we establish the *oracle* property of our PPC estimator that is asymptotically equivalent to the corresponding infeasible estimator of the group-specific parameter vector that is obtained by knowing all individuals' group identities. Fourth, the uniform classification consistency also allows us to study the asymptotic distributions of the post-Lasso estimators that are obtained by pooling all individuals in an estimated group to estimate the group-specific parameters. When R_0 and K_0 are unknown, we propose BIC-type information criteria to determine the number of factors and that of groups consistently.

We conduct a small set of Monte Carlo simulations to evaluate the finite sample performance of our PPC method. We consider two data generating processes (DGPs) that cover static and dynamic panels, respectively. In both settings, the classification errors, the root-mean-squared errors (RMSEs) and the biases of the estimates shrink quickly toward 0 as the time dimension T increases. Typically, the post-Lasso estimates outperform the C-Lasso estimates unless T is too small and the classification error is large. We also consider the bias-corrected PPC and post-Lasso estimates. We find that bias correction works remarkably well for the dynamic panel when the estimators are expected to exhibit large biases. Nevertheless, for static panels where the bias magnitude is small, the performance of the bias-corrected and non-bias-corrected estimators is similar. As to the proposed information criteria, we find that they work fairly well in determining the numbers of factors and groups.

As an illustration, we apply our method to study the persistency of housing prices in China's large and medium-sized cities in the last decade. Our findings show that the price persistence exhibits a group pattern across the 69 cities under study. Specifically, we identify 3 latent groups. Most cities in the first group are large cities and located in eastern China. The growth rates of housing prices in these cities are highly persistent. In the second and third groups, most are medium-sized and inland cities. Their persistence is small or close to zero.

The rest of the paper is organized as follows. We introduce the C-Lasso-based PPC estimation of panel structure models with IFEs in Section 2. We study the asymptotic properties of the C-Lasso procedure and the resulting PPC estimators in Section 3. We study the determination of the numbers of factors and groups in Section 4. Section 5 reports the Monte Carlo simulation results. In section 6 we apply the proposed method to an economic data set. Section 7 concludes. All proofs are relegated to the online supplemental appendix.

NOTATION. Throughout the paper we adopt the following notation. For an $m \times n$ real matrix A, we denote its transpose as A', its Frobenius norm as $\|A\|$, its spectral norm as $\|A\|_{\mathrm{sp}}$, and its Moore-Penrose generalized inverse as A^{+}. Let $P_A = A(A'A)^{+} A'$ and $M_A = I_m - P_A$, where I_m is an $m \times m$ identity matrix. When A is symmetric, we use $\mu_r(A)$ to denote its rth largest eigenvalue by counting multiple eigenvalues multiple times; we also use $\mu_{\max}(\cdot)$ and $\mu_{\min}(A)$ to denote A's largest and smallest eigenvalues, respectively. Let $\mathbf{0}_{p\times 1}$ denote a $p \times 1$ vector of zeros and $\mathbf{1}\{\cdot\}$ the usual indicator function. We use "p.d." and "p.s.d." to abbreviate "positive definite" and "positive semidefinite", respectively. The operator $\xrightarrow{P}$ denotes convergence in probability, $\xrightarrow{D}$ convergence in distribution, and plim probability limit. We use $(N, T) \to \infty$ to denote that N and T pass to infinity jointly.

2 Penalized principal component estimation of panel structure models with IFEs

In this section we consider a panel structure model with IFEs. We first assume that the number of groups is known and then consider the determination of the number of groups later on.

2.1 Panel structure models with IFEs

Let Y_{it} be the dependent variable for individual i measured at time t where $i = 1, ..., N$, and $t = 1, ..., T$. We consider the following panel structure model

$$Y_{it} = \beta_i^{0\prime} X_{it} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}, \tag{2.1}$$

where X_{it} is a $p \times 1$ vector of exogenous or predetermined variables, β_i^0 is a $p \times 1$ vector of unknown slope coefficients, λ_i^0 and F_t^0 denote an $R_0 \times 1$ vector of factor loadings and common factors, respectively, both of which may be correlated with $\{X_{it}\}$, ε_{it} is the idiosyncratic error term, and β_i^0 is a $p \times 1$ vector of slope parameters such that

$$\beta_i^0 = \begin{cases} \alpha_1^0 & \text{if } i \in G_1^0 \\ \vdots & \vdots \\ \alpha_{K_0}^0 & \text{if } i \in G_{K_0}^0 \end{cases} . \tag{2.2}$$

Here $\alpha_j^0 \neq \alpha_k^0$ for any $j \neq k$, $\cup_{k=1}^{K_0} G_k^0 = \{1, 2, ..., N\}$, and $G_k^0 \cap G_j^0 = \varnothing$ for any $j \neq k$. Let $N_k = \# G_k^0$, the cardinality of the set G_k^0. For the moment, we assume that both R_0 and K_0 are known and fixed but each individual's membership is unknown. In addition, following the lead of Sun (2005), Lin and Ng (2012), and SSP, we implicitly assume that the individual's membership does not vary over time. Let

$$\boldsymbol{\alpha} \equiv (\alpha_1, ..., \alpha_{K_0}), \ \boldsymbol{\beta} \equiv (\beta_1, ..., \beta_N), \ \Lambda = (\lambda_1, \lambda_2, ..., \lambda_N)', \text{ and } F = (F_1, F_2, ..., F_T)'. \tag{2.3}$$

The true values of $\boldsymbol{\alpha}, \boldsymbol{\beta}$, Λ, and F are denoted as $\boldsymbol{\alpha}^0$, $\boldsymbol{\beta}^0$, Λ^0, and F^0 respectively. We are interested in inferring each individual's group identity and estimating $(\boldsymbol{\alpha}^0, \Lambda^0, F^0)$ consistently.

2.2 Penalized principal component estimation

In this subsection we consider the PPC estimation of $\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, \Lambda^0, F^0\right)$ under the identification restrictions: $F'F/T = I_{R_0}$, and $\Lambda'\Lambda$ =diagonal with descending diagonal elements. Let $Y_i \equiv (Y_{i1}, \ldots, Y_{iT})'$ and $X_i \equiv (X_{i1}, \ldots, X_{iT})'$. The PPC objective function is given by

$$Q_{0NT,\kappa}^{(K_0)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \Lambda, F) = Q_{0NT}(\boldsymbol{\beta}, \Lambda, F) + \frac{\kappa}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \|\beta_i - \alpha_k\|, \tag{2.4}$$

where $Q_{0NT}(\boldsymbol{\beta}, \Lambda, F) = \frac{1}{NT} \sum_{i=1}^{N} \|Y_i - X_i\beta_i - F\lambda_i\|^2$ and $\kappa = \kappa_{NT} \geq 0$ is a tuning parameter. Noting that Λ and F only appears in the first term of the objective function, we can concentrate them out in turn. By concentrating Λ out, we can readily obtain the following profile PPC objective function

$$Q_{1NT,\kappa}^{(K_0)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, F) = Q_{1NT}(\boldsymbol{\beta}, F) + \frac{\kappa}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \|\beta_i - \alpha_k\|, \tag{2.5}$$

where $Q_{1NT}(\boldsymbol{\beta}, F) = \frac{1}{NT} \sum_{i=1}^{N} (Y_i - X_i\beta_i)' M_F (Y_i - X_i\beta_i)$. Following Moon and Weidner (2015a, 2015b), we can further concentrate F out and obtain the final profile PPC objective function

$$Q_{NT,\kappa}^{(K_0)}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = Q_{NT}(\boldsymbol{\beta}) + \frac{\kappa}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \|\beta_i - \alpha_k\|, \tag{2.6}$$

where $Q_{NT}(\boldsymbol{\beta}) = \frac{1}{T} \sum_{r=R_0+1}^{T} \mu_r \left[\frac{1}{N} \sum_{i=1}^{N} (Y_i - X_i\beta_i)(Y_i - X_i\beta_i)'\right]$.

Minimizing the profile PPC criterion function in (2.6) produces the *Classifier Lasso* (C-Lasso) estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, respectively. Let $\hat{\beta}_i$ and $\hat{\alpha}_k$ denote the ith and kth columns of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$, respectively, i.e., $\hat{\boldsymbol{\alpha}} \equiv (\hat{\alpha}_1, ..., \hat{\alpha}_{K_0})$ and $\hat{\boldsymbol{\beta}} \equiv (\hat{\beta}_1, ..., \hat{\beta}_N)$. We then obtain the estimate $(\hat{\Lambda}, \hat{F})$ of (Λ, F) via Bai and Ng's (2002) PC method under the identification restrictions: $F'F/T = I_{R_0}$ and $\Lambda'\Lambda$ is a diagonal matrix with descending diagonal elements. That is, $(\hat{\Lambda}, \hat{F})$ solves

$$\left[\frac{1}{NT} \sum_{i=1}^{N} \left(Y_i - X_i\hat{\beta}_i\right)\left(Y_i - X_i\hat{\beta}_i\right)'\right] \hat{F} = \hat{F} V_{NT} \text{ and } \hat{\Lambda} = (\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_N), \tag{2.7}$$

where V_{NT} is a diagonal matrix consisting of the R_0 largest eigenvalues of the above matrix in the square bracket, arranged in descending order, and $\hat{\lambda}_i = T^{-1}\hat{F}'(Y_i - X_i\hat{\beta}_i)$.

To proceed, it is worth mentioning that the penalty term in (2.4)-(2.6) takes a novel mixed *additive-multiplicative* form that does not appear in the literature. Traditionally a Lasso method adds a penalty term to the least-squares, GMM, or negative log-likelihood objective function additively and when multiple penalty terms are needed, they also enter the objective function additively. In sharp contrast, our C-Lasso method has N additive terms, each of which takes a multiplicative expression as the product of K_0 penalty terms. Each of the K_0 penalty terms in the multiplicative expression shrinks the individual-level slope parameter vector β_i to a particular *unknown* group-level parameter vector α_k, which also differs from the prototypical Lasso method of Tibshirani (1996) that shrinks a parameter to zero or the group Lasso method of Yuan and Lin (2006) that shrinks a parameter vector to a vector of zeros.

Note that the objective function in (2.6) is not convex in $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ even though it is convex in α_k when one fixes α_j for $j \neq k$. In the supplemental Appendix D, we propose an iterative algorithm to obtain the estimates $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, ..., \hat{\alpha}_{K_0})$ and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, ..., \hat{\beta}_N)$.

3 Asymptotic properties

In this section we study the asymptotic properties of the C-Lasso procedure and the resulting PPC estimates.

3.1 Preliminary rates of convergence

We first present sufficient conditions to ensure the consistency of $\hat{\boldsymbol{\beta}}$ and $P_{\hat{F}}$. Let $\varepsilon_i \equiv (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_N)'$. Let $\mathcal{F} = \{F \in \mathbb{R}^{T \times R_0} : T^{-1}F'F = I_{R_0}\}$ and $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$. For any $F \in \mathcal{F}$ and $\underline{c} > 0$, we define $\mathcal{N} = \{1, 2, ..., N\}$ and

$$\mathcal{S}_{N,F,\underline{c}} = \left\{i \in \mathcal{N}\colon\ P\left(\mu_{\min}\left(T^{-1}X_i'M_F X_i\right) \geq \underline{c}\right) = 1 - o\left(N^{-1}\right)\right\}.$$

Let $\mathcal{S}^*_{N,F,\underline{c}} = \mathcal{N} \backslash \mathcal{S}_{N,F,\underline{c}}$. Let $\#A$ denote the cardinality of the set A, and C a generic positive constant that may vary across lines. We make the following assumptions.

Assumption A.1 (i) $\max_{1\leq t\leq T} E\left\|F_t^0\right\|^8 \leq C$ and $T^{-1}F^{0\prime}F^0 \xrightarrow{P} \Sigma_{F^0} > 0$ as $T \to \infty$.

(ii) $\max_{1\leq i\leq N} E\left\|\lambda_i^0\right\|^8 \leq C$ and $N^{-1}\Lambda^{0\prime}\Lambda^0 \xrightarrow{P} \Sigma_{\Lambda^0} > 0$ as $N \to \infty$.

(iii) $\max_{1\leq i\leq N} T^{-1}E\left\|X_i\right\|^2 \leq C$ and $\max_{1\leq i\leq N} T^{-1/2}\left\|X_i\right\| = O_P(1)$.

(iv) $E(\varepsilon_{it}) = 0$, $\|\boldsymbol{\varepsilon}\|_{\rm sp} = O_P(\sqrt{N} + \sqrt{T})$, and $\max_{1\leq i\leq N}\max_{1\leq t\leq T} E(\varepsilon_{it}^8) \leq C$.

(v) Let $\sigma_{ij,ts} = E(\varepsilon_{it}\varepsilon_{js})$. $\max_{1\leq t\leq T} N^{-1}\sum_{i=1}^N \sigma_{ii,tt} \leq C$, $N^{-1}\sum_{i=1}^N\sum_{j=1}^N \max_{1\leq t\leq T}|\sigma_{ij,tt}| \leq C$, $T^{-1}\sum_{t=1}^T\sum_{s=1}^T \max_{1\leq i\leq N}|\sigma_{ii,ts}| \leq C$, and $(NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T|\sigma_{ij,ts}| \leq C$.

(vi) $\max_{1\leq t,s\leq T} E\left|N^{-1/2}\sum_{i=1}^N\left[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})\right]\right|^4 \leq C$.

(vii) $\max_{1\leq i\leq N} T^{-1}E\left\|F^{0\prime}\varepsilon_i\right\|^2 \leq C$, $\max_{1\leq i\leq N} T^{-1}E\left\|X_i'\varepsilon_i\right\|^2 \leq C$, and $(NT)^{-1/2}\sum_{i=1}^N X_i'\varepsilon_i = O_P(1)$.

(viii) $\max_{1\leq i\leq N}\frac{1}{NT}\sum_{j=1}^N E[\|\varepsilon_i'X_j\|^2(1 + \left\|\lambda_j^0\right\|^2)] \leq C$, $\max_{1\leq i\leq N}\frac{1}{NT^2}\sum_{j=1}^N E\left\|\varepsilon_j\varepsilon_i'X_j\right\|^2 \leq C$, $\max_{1\leq i\leq N}\frac{1}{NT}\sum_{j=1}^N E\left\|\varepsilon_i'\varepsilon_j\right\|^2 \leq C$, $\max_{1\leq i\leq N}\frac{1}{NT}E\left\|\sum_{j=1}^N \varepsilon_i'\varepsilon_j\lambda_j^{0\prime}\right\|^2 \leq C(1 + T/N)$, and $\max_{1\leq i\leq N}\frac{1}{NT}E\left\|\sum_{j=1}^N \varepsilon_i'\varepsilon_j\varepsilon_j'F^0\right\| \leq C$.

(ix) $\frac{1}{NT}E\left\|\lambda^{0\prime}\boldsymbol{\varepsilon}F^0\right\|^2 \leq C$, $\frac{1}{NT}\sum_{k=1}^{K_0} E\left\|\sum_{i\in G_k^0} X_i\varepsilon_i'F^0\right\| \leq C$, and $\frac{1}{N_k^2T^2}\left\|\sum_{i\in G_k^0}\sum_{j\in G_k^0}\sum_{t=1}^T\sum_{s=1}^T E(X_{it}X_{jt}'\varepsilon_{js}\varepsilon_{is})\right\| \leq C\delta_{NT}^{-2}$ for $k = 1, ..., K_0$.

(x) There exists a constant $\underline{c}_{XX} > 0$ such that $P\left(\min_{1\leq i\leq N}\mu_{\min}\left(T^{-1}X_i'M_{F^0}X_i\right)\geq \underline{c}_{XX}\right) = 1-o\left(N^{-1}\right)$.

(xi) There exists a constant $\underline{c} > 0$ such that $\sup_{F\in\mathcal{F}}\#\mathcal{S}^*_{N,F,\underline{c}}/N = o(1)$, $N^{-1}\sum_{i\in\mathcal{N}_s}\lambda_i^0\lambda_i^{0\prime}\xrightarrow{P}\Sigma_s > 0$ for any subset of $\mathcal{N}_s$ of $\mathcal{N}$ with $\#\mathcal{N}_s/N\to 1$, and $\max_{1\leq i\leq N}T^{-1}\varepsilon_i'M_{F^0}\varepsilon_i = O_P(1)$.

Assumption A.2 (i) $N_k/N\to\tau_k\in(0,1)$ for each $k=1,...,K_0$ as $N\to\infty$.

(ii) $T/N^2\to 0$ and $N/T^2\to 0$ as $(N,T)\to\infty$.

(iii) $T\kappa^4\to 0$ as $(N,T)\to\infty$.

A.1(i)-(iii) impose standard moment conditions on F_t^0, λ_i^0, and X_{it}; see, e.g., Bai and Ng (2002) and Bai (2003, 2009). Note that Bai and Ng (2002) assume only finite fourth moment for F_t^0 but require that λ_i^0 be uniformly bounded. A.1(iv)-(vi) impose conditions on the error processes $\{\varepsilon_{it}\}$. Except for the second part of A.1(iv) that is also assumed in Su and Chen (2013), they are adapted from Bai and Ng (2002) and Bai (2009) and allow for weak forms of cross sectional and serial dependence in the error processes. A.1(vii) requires weak exogeneity of the regressor X_{it} and the common factor F_t^0. A.1(viii) and (ix) can be satisfied under various primitive conditions in Section 3.4 below. In particular, the first part of A.1(ix) implies that $\left\|\lambda^{0\prime}\varepsilon F^0\right\| = O_P(N^{1/2}T^{1/2})$ by Chebyshev inequality, which further implies that $\left\|\lambda^{0\prime}\varepsilon\right\|^2 = O_P(NT)$ and $\left\|\varepsilon F^0\right\|^2 = \sum_{i=1}^N\left\|F^{0\prime}\varepsilon_i\right\|^2 = O_P(NT)$ under A.1(i)-(ii) by standard matrix operations (see Lu and Su, 2016). A.1(x) replaces Assumption A in Bai (2009) and it requires that the columns of X_i should not span the same space spanned by the columns of F^0. A.1(xi) is needed to demonstrate explicitly that the minimizer of our PPC objective function cannot be achieved at points $\{\beta_i\}$ such that $N^{-1}\sum_{i=1}^N\left\|\beta_i-\beta_i^0\right\|^2$ is explosive. Early literature on inferences with a diverging number of parameters (e.g., Fan and Peng (2004), Lam and Fan (2008), and Lu and Su (2016)) often assumed that the global solutions are achieved in the neighborhood of the true values directly. Note that not all of the conditions in Assumption A.1 are used in the proof of Theorems 3.1-3.2 below; some of them are used in the proofs of subsequent theorems instead.

A.2(i) implies that each group has an asymptotically non-negligible number of individuals as $N\to\infty$. This assumption can be relaxed at the cost of more lengthy arguments, in which case the estimates of α_k^0, $k=1,...,K_0$, will exhibit different convergence rates. A.2(ii) imposes conditions on the relative rates at which N and T can pass to infinity. A.2(iii) implies that κ has to shrink to zero sufficiently fast.

The following theorem establishes the mean square convergence of $\{\hat{\beta}_i\}$ and the consistency of the estimated factor space.

Theorem 3.1 *Suppose that Assumptions A.1, A.2(i), and A.2(iii) hold. Then*

(i) $N^{-1}\sum_{i=1}^N\left(\hat{\beta}_i-\beta_i^0\right)'\left(T^{-1}X_i'M_{\hat{F}}X_i\right)\left(\hat{\beta}_i-\beta_i^0\right) = o_P(1)$,

(ii) $T^{-1}F^{0\prime}\hat{F}$ *is invertible and* $\left\|P_{\hat{F}}-P_{F^0}\right\| = o_P(1)$,

(iii) $N^{-1}\sum_{i=1}^N\left\|\hat{\beta}_i-\beta_i^0\right\|^2 = o_P(1)$.

Theorems 3.1(i) and (iii) establish the weighted and nonweighted versions of the mean square consistency of $\{\hat{\beta}_i\}$, respectively; Theorem 3.1(ii) claims that the spaces spanned by the columns of $\hat{F}$ and F^0 are asymptotically the same. The above asymptotic results are based on the analysis of $Q_{1NT,\kappa}^{(K_0)}(\boldsymbol{\beta},\boldsymbol{\alpha},F)$ defined in (2.5). As in Bai (2009), the standard argument of consistency for extreme estimators (e.g., Amemiya (1985); Newey and McFadden (1994)) does not apply here because of the growing dimensions of both $\boldsymbol{\beta}$ and F. To overcome the difficulty of divergent parameter spaces, we follow the lead of Bai (2009) and adopt a proof strategy based on an auxiliary objective function that is uniformly close to the original objective function. This allows us to establish the weighted version of the mean square convergence of $\hat{\beta}_i$ to β_i^0, which is sufficient for the establishment of the consistency of estimated factor space. Then under Assumption A.1(x), we establish the nonweighted version of the mean square convergence of $\hat{\beta}_i$ to β_i^0.

Given consistency, we can further establish the rate of convergence for both the individual and group-specific parameter estimates.

Theorem 3.2 *Suppose that Assumptions A.1 and A.2(i)-(iii) hold. Then*

(i) $\frac{1}{N}\sum_{i=1}^N \left\|\hat{\beta}_i - \beta_i^0\right\|^2 = O_P\left(T^{-1}\right)$,

(ii) $\hat{\beta}_i - \beta_i^0 = O_P\left(T^{-1/2}+\kappa\right)$ *for* $i = 1,...,N$,

(iii) $\left(\hat{\alpha}_{(1)},...,\hat{\alpha}_{(K_0)}\right) - \left(\alpha_1^0,...,\alpha_{K_0}^0\right) = O_P\left(T^{-1/2}\right)$ *for some permutation* $\left(\hat{\alpha}_{(1)},...,\hat{\alpha}_{(K_0)}\right)$ *of* $(\hat{\alpha}_1,...,\hat{\alpha}_{K_0})$.

Theorems 3.2(i) and (ii) establish the mean square and pointwise convergence rates of $\{\hat{\beta}_i\}$, respectively. Theorem 3.2(iii) indicates that the group-specific parameters, $\alpha_1^0,...,\alpha_K^0$, can be estimated consistently by $\hat{\alpha}_1,...,\hat{\alpha}_{K_0}$ subject to suitable permutation. Our findings are similar to those in SSP who show that the pointwise convergence of $\hat{\beta}_i$ depends on κ while the mean square convergence of $\{\hat{\beta}_i\}$ and the convergence of $\hat{\alpha}_{(k)}$ do not.

For notational simplicity, hereafter we simply write $\hat{\alpha}_k$ for $\hat{\alpha}_{(k)}$ as the consistent estimator of α_k^0's, and define

$$\hat{G}_k = \left\{i \in \{1,2,...,N\} : \hat{\beta}_i = \hat{\alpha}_k\right\} \text{ for } k = 1,...,K_0. \tag{3.1}$$

Let $\hat{G}_0$ denote the group of individuals in $\{1,2,...,N\}$ that are not classified into any of the K_0 groups, i.e., $\hat{G}_0 = \{1,2,...,N\}\setminus\cup_{k=1}^{K_0}\hat{G}_k$.

To study the classification consistency, we need to establish the uniform consistency of $\hat{\beta}_i$. We add the following assumption.

Assumption A.3 (i) $P\left(\max_{1\le i\le N}\left\|T^{-1/2}\sum_{t=1}^T[\xi_{it}-E(\xi_{it})]\right\| \ge C(\ln T)^{\upsilon_1}\right) = o(N^{-1})\ \forall C>0$ for some $\upsilon_1 \ge 1$ and for $\xi_{it} = X_{it}X_{it}'$, ε_{it}^2, $X_{it}\varepsilon_{it}$, $X_{it}F_t^{0\prime}$,and $F_t^0\varepsilon_{it}$.

(ii) $P\left(\max_{1\le i,j\le N}\left\|T^{-1/2}\sum_{t=1}^T\xi_{it}\varepsilon_{jt}\right\| \ge C(\ln T)^{\upsilon_1}\right) = o(N^{-1})\ \forall C>0$ for $\xi_{it} = X_{it}$ and ε_{it}.

(iii) $P\left(\max_{1\le i\le N}\left\|(NT)^{-1/2}\sum_{j=1}^N\sum_{t=1}^T\varepsilon_{it}\varepsilon_{jt}\lambda_j^0\right\| \ge C(1+\sqrt{T/N})(\ln T)^{\upsilon_1}\right) = o(N^{-1})\ \forall C>0$.

(iv) $P\left(\max_{1\leq i\leq N}\left\|\lambda_i^0\right\|\geq c_{\lambda,N}\right)=o(N^{-1})$ for some $c_{\lambda,N}=o\left(N^{1/4}\right).$

A.3(i)-(iii) impose conditions to derive the uniform consistency of $\hat{\beta}_i$. Most of these conditions can be verified under some weak dependence conditions for $\left\{(X_{it},\varepsilon_{it},F_t^0,\lambda_i^0)\right\}$. See, e.g., Su et al. (2015) and Su and Wang (2016). In particular, Su et al. (2015) apply the concept of conditional strong mixing to study specification tests in dynamic panel data models with IFEs. It is well known that the strong mixing assumption generally does not hold for dynamic panel data model with IFEs because the random factor loadings introduce temporal dependence that does not vanish asymptotically. However, the process may still be conditional strong mixing given the sigma-field generated by the factor loadings (and factors). Some popular exponential inequalities for strong mixing processes also hold true for conditional strong mixing processes under some moment conditions; see Hahn and Kuersteiner (2011) and Su and Chen (2013). Under Assumptions A.1(i) and A.3(i), we can readily show that $P\left(\max_{1\leq i\leq N}T^{-1/2}X_i'M_{F^0}\varepsilon_i\geq C\left(\ln T\right)^{\upsilon_1}\right)=o\left(N^{-1}\right)$ for any $C>0$. A.3(iv) is implied by A.1(ii) if we only require $c_{\lambda,N}=o\left(N^{1/4}\right)$. For a better control on the classification, we will impose more stringent conditions on $c_{\lambda,N}$. If λ_i^0's are uniformly bounded, then A.3(iv) is satisfied with some $c_{\lambda,N}$ that does not depend on N. If $\max_{1\leq i\leq N}E[\exp\left(c_\lambda\left\|\lambda_i^0\right\|\right)]\leq C<\infty$ for some $c_\lambda>0$, A.2(iv) is satisfied by taking $c_{\lambda,N}=(\ln N)^a$ for $a>1/c_\lambda$.

The following theorem establishes the uniform consistency of $\hat{\beta}_i$.

Theorem 3.3 *Suppose that Assumptions A.1-A.2 and A.3(i)-(iii) hold. Then for any finite positive constant C that does not depend on (N,T) and any small $\upsilon_2>0$,*

$$P\left(\max_{1\leq i\leq N}\left\|\hat{\beta}_i-\beta_i^0\right\|\geq C[T^{-1/2}\left(\ln T\right)^{\upsilon_1}+\kappa\left(\ln T\right)^{\upsilon_2}]\right)=o\left(N^{-1}\right).$$

3.2 Classification consistency

To study the classification consistency, we follow SSP and define the following sequences of events:

$$\hat{E}_{kNT,i}=\left\{i\notin\hat{G}_k\mid i\in G_k^0\right\}\text{ and }\hat{F}_{kNT,i}=\left\{i\notin G_k^0\mid i\in\hat{G}_k\right\},\tag{3.2}$$

where $i=1,...,N$ and $k=1,...,K_0$. Let $\hat{E}_{kNT}=\cup_{i\in G_k^0}\hat{E}_{kNT,i}$ and $\hat{F}_{kNT}=\cup_{i\in\hat{G}_k}\hat{F}_{kNT,i}$. Apparently, $\hat{E}_{kNT}$ and $\hat{F}_{kNT}$ mimic Type I and II errors in statistical tests: $\hat{E}_{kNT}$ denotes the error event of not classifying an element of G_k^0 into the estimated group $\hat{G}_k$, and $\hat{F}_{kNT}$ denotes the error event of classifying an element that does not belong to G_k^0 into the estimated group $\hat{G}_k$. Our classification method is *uniformly consistent* if $P(\cup_{k=1}^{K_0}\hat{E}_{kNT})\to 0$ and $P(\cup_{k=1}^{K_0}\hat{F}_{kNT})\to 0$ as $(N,T)\to\infty$.

To study the consistency of our classification method, we add the following assumption.

Assumption A.2 (iv) $T\kappa^2/[(\ln T)^{2\upsilon_1}\,c_{\lambda,N}^2]\to\infty$ as $(N,T)\to\infty$.

Assumption A.2(iv) strengthens the conditions on the tuning parameter κ. If λ_i^0's are subgaussian, Assumptions A.2(iii)-(iv) indicate that we can specify κ such that

$$\kappa \propto T^{-a} \text{ for some } a \in (1/4, 1/2).$$

The following theorem establishes the uniform consistency for our classification method.

Theorem 3.4 *Suppose that Assumptions A.1, A.2(i)-(iv) and A.3 hold. Then*
(i) $P\left(\cup_{k=1}^{K_0} \hat{E}_{kNT}\right) \leq \sum_{k=1}^{K_0} P\left(\hat{E}_{kNT}\right) \to 0$ *as* $(N,T) \to \infty$,
(ii) $P\left(\cup_{k=1}^{K_0} \hat{F}_{kNT}\right) \leq \sum_{k=1}^{K_0} P\left(\hat{F}_{kNT}\right) \to 0$ *as* $(N,T) \to \infty$.

Theorem 3.4 implies that for all individuals within a group, say, G_k^0, they can be simultaneously correctly classified into the same group (denoted as $\hat{G}_k$) w.p.a.1. Conversely, for all individuals that are classified into the same group, say, $\hat{G}_k$, they simultaneously belong to the same group (G_k^0) w.p.a.1. Define the events $\hat{H}_{iNT} = \{i \in \hat{G}_0\}$. Theorem 3.4(i) implies that $P(\cup_{1\leq i\leq N}\hat{H}_{iNT}) \leq \sum_{k=1}^{K_0} P(\hat{E}_{kNT}) \to 0$. That is, all individuals can be correctly classified into one of the K_0 groups w.p.a.1. Nevertheless, when T is not large, it is possible for a small percentage of individuals to be left unclassified if we stick with the classification method defined in (3.1). To ensure that all individuals are classified into one of the K_0 groups in finite samples, SSP suggest that in practice one can modify the above classification rule a little bit: we classify $i \in \hat{G}_k$ if $\hat{\beta}_i = \hat{\alpha}_k$ for some $k = 1, ..., K_0$, and $i \in \hat{G}_l$ for some $l = 1, ..., K_0$ if

$$\left\|\hat{\beta}_i - \hat{\alpha}_l\right\| = \min\left\{\left\|\hat{\beta}_i - \hat{\alpha}_1\right\|, ..., \left\|\hat{\beta}_i - \hat{\alpha}_{K_0}\right\|\right\} \text{ and } \sum_{k=1}^{K_0} \mathbf{1}\left\{\hat{\beta}_i = \hat{\alpha}_k\right\} = 0.$$

Since the event $\sum_{k=1}^{K_0} \mathbf{1}\{\hat{\beta}_i = \hat{\alpha}_k\} = 0$ occurs with probability approaching zero uniformly in i, we ignore it in large samples in subsequent theoretical analysis and restrict our attention to the previous classification rule in (3.1) to avoid confusion. Let $\hat{N}_k = \#\hat{G}_k$ for $k = 0, 1, ..., K_0$. Based on Theorem 3.4, SSP also show that $\hat{N}_k = N_k + o_P(1)$ for $k = 1, ..., K_0$.

3.3 The oracle property and asymptotic properties of the post-Lasso estimators

Let $\nu_{ij} = \lambda_i^{0\prime}\left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\lambda_j^0$ and $\mu_{ts} = F_t^{0\prime}\left(T^{-1}F^{0\prime}F^0\right)^{-1}F_s^0$. Let $Q_{kNT} = \frac{1}{N_k T}\sum_{i\in G_k^0} X_i' M_{F^0} X_i$, $A_{k,l} = \frac{1}{N N_k T}\sum_{i\in G_k^0}\sum_{j\in G_l^0} \nu_{ij} X_i' M_{F^0} X_j$, and $\hat{V}_{kNT} = \frac{1}{N_k T}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}}(\varepsilon_i - \frac{1}{N}\sum_{j=1}^N \nu_{ji}\varepsilon_j)$ where k, $l = 1, ..., K_0$. Define

$$Q_{NT} = \begin{pmatrix} Q_{1NT} - A_{1,1} & -A_{1,2} & \cdots & -A_{1,K_0} \\ -A_{2,1} & Q_{2NT} - A_{2,2} & \cdots & -A_{2,K_0} \\ \vdots & \vdots & \ddots & \vdots \\ -A_{K_0,1} & -A_{K_0,2} & \cdots & Q_{K_0NT} - A_{K_0,K_0} \end{pmatrix} \text{ and } \hat{V}_{NT} = \begin{pmatrix} \hat{V}_{1NT} \\ \hat{V}_{2NT} \\ \vdots \\ \hat{V}_{K_0NT} \end{pmatrix}. \tag{3.3}$$

Let $\mathcal{D} \equiv \sigma\left(F^0, \lambda^0\right)$, the sigma-field generated by $\left(F^0, \lambda^0\right)$, and $E_{\mathcal{D}}(A) \equiv E(A|\mathcal{D})$. Let[2]

$$
\begin{aligned}
\mathcal{B}_{1,kNT} &= -\frac{1}{NN_kT^2}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}} \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\hat{F}(T^{-1}F^{0\prime}\hat{F})^{-1}(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0, \\
\mathcal{B}_{2,kNT} &= -\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j=1}^{N}(X_i - X_i^*)' F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_j^0\varepsilon_j'\varepsilon_i, \\
\mathcal{B}_{3,kNT} &= \frac{1}{N_kT}\sum_{i\in G_k^0} E_{\mathcal{D}}\left[X_i' P_{F^0}\varepsilon_i\right], \\
\mathcal{V}_{kNT} &= \frac{1}{NT}\sum_{i=1}^{N} Z_{k,i}'\varepsilon_i,
\end{aligned}
$$

where $X_i^* = \frac{1}{N}\sum_{j=1}^N \nu_{ij}X_j$ with the tth row given by $X_{it}^{*\prime}$ and $Z_{k,i} = \frac{N}{N_k}[X_i - P_{F^0}E_{\mathcal{D}}(X_i)]\mathbf{1}\left\{i \in G_k^0\right\}$ $-\frac{1}{N_k}M_{F^0}\sum_{j\in G_k^0}\nu_{ji}E_{\mathcal{D}}(X_j)$. Let Z_{it}' denote the tth row of $Z_i = (Z_{1,i}, ..., Z_{K_0,i})$. Let $\mathcal{B}_{s,NT} = (\mathcal{B}_{s,1NT}', ..., \mathcal{B}_{s,K_0NT}')'$ for $s = 1, 2, 3$ and $\mathcal{V}_{NT} = (\mathcal{V}_{1NT}', ..., \mathcal{V}_{K_0NT}')'$. Note that $\mathcal{V}_{NT} = \frac{1}{NT}\sum_{i=1}^N Z_i'\varepsilon_i$ $= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T Z_{it}\varepsilon_{it}$.

To obtain the oracle property of the C-Lasso estimators $\{\hat{\alpha}_k\}$, we add the following assumption.

Assumption A.4. (i) $Q_{NT} \xrightarrow{P} Q_0 > 0$ as $(N, T) \to \infty$.

(ii) For each $k = 1, ..., K_0$, $\frac{1}{NT}\sum_{i\in G_k^0} X_i' M_{F^0}\varepsilon_i = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$ and $\frac{1}{NT}\sum_{i\in G_k^0} X_i' M_{F^0}\varepsilon_i^* = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$ where $\varepsilon_i^* = \frac{1}{N}\sum_{j=1}^N \nu_{ij}\varepsilon_j$.

A.4(i) requires that the probability limit Q_0 be positive definite. In the special case where $K_0 = 1$, Q_{NT} reduces to $Q_{1NT} - A_{1,1}$ and Q_0 is identical to the probability limit matrix D_0 in Bai (2009). When $K_0 > 1$, Q_0 is typically not a block diagonal matrix, which reflects the fact the estimates of the group-specific parameters rely on the estimation of the common factors and thus are not asymptotically independent. A.4(ii) is a high level condition and it can be verified under various primitive conditions. In section 3.4 below, we follow Su and Chen (2013) and Lu and Su (2016) and assume that for each i, $\{(X_{it}, \varepsilon_{it})\}$ is a conditional strong mixing process given $\mathcal{D}$. Then we verify A.4(ii) in the supplemental appendix.

We are ready to state the next theorem.

Theorem 3.5 *Suppose that Assumptions A.1-A.4 hold. Then* $vec\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\right) = Q_{NT}^{-1}[\hat{V}_{NT} + \mathcal{B}_{1,NT}] + o_P((NT)^{-1/2}) = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$.

[2] Alternatively, one can define $\mathcal{B}_{1,kNT}$ as $\mathcal{B}_{1,kNT}^* = -\frac{1}{NN_kT^2}\sum_{i\in G_k^0} X_i' M_{\hat{F}}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}F^0(T^{-1}F^{0\prime}F^0)^{-1}(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0$. Following the proof of Lemma A.5 and using Lemma A.4 and Theorem 3.5, one can easily show that $\mathcal{B}_{1,kNT}^* = \mathcal{B}_{1,kNT} + o_P((NT)^{-1/2})$ and thus $\mathcal{B}_{1,kNT}^*$ and $\mathcal{B}_{1,kNT}$ play the same role in our asymptotic analysis. We use $\mathcal{B}_{1,kNT}$ as it naturally arises in the asymptotic expansion of our PPC estimator.

Theorem 3.5 gives the key Bahadur-type representation of $\text{vec}\left(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right)$. As we demonstrate in the proof of Theorem 3.5, $\mathcal{B}_{1,NT}$ contributes to the asymptotic bias of our group-specific estimators $\{\hat{\alpha}_k\}$ and $\hat{V}_{NT}$ contributes to both the asymptotic bias and variance. In the appendix, we prove the above theorem by a careful inspection of the Karush-Kuhn-Tucker (KKT) optimality conditions for minimizing the objective function in (2.5) based on subdifferential calculus (e.g., Bertsekas (1995, Appendix B.5)).

Let $V_{kNT}=\frac{1}{N_kT}\sum_{i\in G_k^0}X_i'M_{F^0}(\varepsilon_i-\frac{1}{N}\sum_{j=1}^N\nu_{ji}\varepsilon_j)$ and $V_{NT}=(V_{1NT}',...,V_{K_0NT}')'$. To study the asymptotic distribution of the PPC estimator $\{\hat{\alpha}_k\}$, we add the following assumption.

Assumption A.5. (i) For each $k=1,...,K_0$, $V_{kNT}=\mathcal{V}_{kNT}+\mathcal{B}_{3,kNT}+o_P((NT)^{-1/2})$.

(ii) For each $k=1,...,K_0$, $\mathcal{B}_{3,kNT}=O_P(T^{-1})$.

(iii) $\sqrt{NT}\mathcal{V}_{NT}\xrightarrow{D}N(0,\Theta_0)$ for some $\Theta_0>0$.

(iv) $\left\|\frac{1}{NT}\sum_{j=1}^N F^{0\prime}\varepsilon_j\varepsilon_j'\right\|=O_P\left(\delta_{NT}^{-1}\right)$.

A.5 is a high level assumption that parallels Assumption E in Bai (2009). It imposes conditions to ensure the $\sqrt{NT}$-consistency and asymptotic normality of our group-specific estimators $\{\hat{\alpha}_k\}$, subject to bias correction. In the special case where the idiosyncratic error term ε_{it} is independent of X_{js}, λ_j^0, and F_s^0 for all i, j, t, and s (i.e., Assumption D in Bai (2009) is maintained), it is easy to verify that A.5 holds with $V_{kNT}=\mathcal{V}_{kNT}$ and $\mathcal{B}_{3,kNT}=0$. If in addition $K_0=1$, noticing that $V_{NT}=\frac{1}{NT}\sum_{i=1}^N(X_i-X_i^*)'M_{F^0}\varepsilon_i$ in this case and $M_{F^0}(X_i-X_i^*)$ is defined as Z_i in Bai (2009), A.5(i)-(iii) become identical to Assumption E in Bai (2009). Nevertheless, we allow X_{it} to contain lagged dependent variables and ε_{it} to be dependent on F_t. In this case, $\mathcal{B}_{3,kNT}$ contributes to the asymptotic bias of our estimator and has to be corrected in finite samples. In the next subsection we specify a set of conditions such that A.5 can be verified in the presence of lagged dependent variables.

The following theorem establishes the oracle property of the PPC estimator $\{\hat{\alpha}_k\}$.

Theorem 3.6 *Suppose that Assumptions A.1-A.5 hold. Then* $\sqrt{NT}[vec\left(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right)-Q_{NT}^{-1}(\mathcal{B}_{1,NT}+\mathcal{B}_{2,NT}+\mathcal{B}_{3,NT})]\xrightarrow{D}N(0,\,Q_0^{-1}\Theta_0Q_0^{-1})$.

Clearly, Theorem 3.6 indicates that $\text{vec}(\hat{\boldsymbol{\alpha}})$ has three bias terms: $Q_{NT}^{-1}\mathcal{B}_{1,NT}$, $Q_{NT}^{-1}\mathcal{B}_{2,NT}$, and $Q_{NT}^{-1}\mathcal{B}_{3,NT}$. If Assumption D in Bai (2009) holds true, then $\mathcal{B}_{3,NT}=0$. Maintaining such an assumption, we can follow the analysis of Bai (2009) and show that $\mathcal{B}_{2,NT}$ is absent from the above expression in the absence of both *cross-sectional* correlation and heteroskedasticity in the error terms and that $\mathcal{B}_{1,NT}$ is absent in the absence of both *serial* correlation and heteroskedasticity in the error terms. See the remark after Theorem 3 in Bai (2009).

If we know each individual's group membership, we can obtain a (non-penalized) PC estimator $\bar{\alpha}_k$ of α_k^0 by utilizing such group information. Let $\bar{\boldsymbol{\alpha}}=(\bar{\alpha}_1,...,\bar{\alpha}_{K_0})$. Under Assumptions A.1, A.2(i) and A.4-A.5, we can show that $\sqrt{NT}[\text{vec}\left(\bar{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right)-Q_{NT}^{-1}(\mathcal{B}_{1,NT}+\mathcal{B}_{2,NT}+\mathcal{B}_{3,NT})]\xrightarrow{D}$

$N(0,\ Q_0^{-1}\Theta_0 Q_0^{-1})$. Theorem 3.6 indicates that the PPC estimators $\{\hat{\alpha}_k\}$ can achieve the same asymptotic distribution as an oracle would obtain by knowing the exact membership of each individual. In this sense, we say that the PPC estimators $\{\hat{\alpha}_k\}$ have the asymptotic oracle property.

Let $\{\tilde{\alpha}_{\hat{G}_k}\}$ denote the post-Lasso (non-penalized) PC estimators of $\{\alpha_k^0\}$ based on the estimated group structure in $\{\hat{G}_k\}$. Let $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_{\hat{G}_1}, ..., \tilde{\alpha}_{\hat{G}_{K_0}})$. The following theorem reports the asymptotic distribution of the post-Lasso PC estimators $\{\tilde{\alpha}_{\hat{G}_k}\}$.

Theorem 3.7 *Suppose that Assumptions A.1-A.5 hold. Then* $\sqrt{NT}[vec(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) - Q_{NT}^{-1}(\mathcal{B}_{1,NT} + \mathcal{B}_{2,NT} + \mathcal{B}_{3,NT})] \xrightarrow{D} N(0,\ Q_0^{-1}\Theta_0 Q_0^{-1})$.

That is, the post-Lasso PC estimators $\{\tilde{\alpha}_{\hat{G}_k}\}$ share the same first order asymptotic distribution as the PPC estimator $\{\hat{\alpha}_k\}$. We will compare the finite sample performance of these estimators through monte Carlo simulations.

Remark. If we also allow the factor loadings $\{\lambda_i^0\}$ to exhibit the same grouped pattern as $\{\beta_i^0\}$, we can let γ_k^0, $k = 1, ..., K_0$, denote the group-specific factor loadings. In this case, one can consider the following penalized PC objective function

$$\bar{Q}_{0NT,\kappa}^{(K_0)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \Lambda, F) = Q_{0NT}(\boldsymbol{\beta}, \Lambda, F) + \frac{\kappa}{N}\sum_{i=1}^{N}\Pi_{k=1}^{K_0}\left\|(\beta_i', \lambda_i') - (\alpha_k', \gamma_k')\right\|, \tag{3.4}$$

subject to certain identification constraints. Following the analysis in Section 2.1, we can concentrate F out to obtain

$$\bar{Q}_{1NT,\kappa}^{(K_0)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \Lambda) = \bar{Q}_{1NT}(\boldsymbol{\beta}, \Lambda) + \frac{\kappa}{N}\sum_{i=1}^{N}\Pi_{k=1}^{K_0}\left\|(\beta_i', \lambda_i') - (\alpha_k', \gamma_k')\right\|, \tag{3.5}$$

where $\bar{Q}_{1NT}(\boldsymbol{\beta}, \Lambda)$ is analogously defined as $Q_{1NT}(\boldsymbol{\beta}, F)$ by inverting the roles of F and Λ. We can also study the asymptotic properties of the estimates from the above minimization. But this is beyond the scope of the current paper.

3.4 Some primitive conditions and bias-correction

Now we present two assumptions that replace some high level conditions in Assumptions A.1, A.4, and A.5. They are also used for statistic inference based on the PPC estimators and the post-Lasso estimators. Let $\|A\|_{q,\mathcal{D}} \equiv [E_{\mathcal{D}}(\|A\|^q)]^{1/q}$.

Assumption B.1 (i) $\max_{1\le t\le T} E||F_t^0||^{8+4\sigma} \le C$ for some $\sigma > 0$ and $C < \infty$ and $T^{-1}F^{0\prime}F^0 \xrightarrow{P} \Sigma_{F^0} > 0$ as $T \to \infty$.

(ii) $\max_{1\le i\le N} E||\lambda_i^0||^{8+4\sigma} \le C$ and $N^{-1}\lambda^{0\prime}\lambda^0 \xrightarrow{P} \Sigma_{\lambda^0} > 0$ as $N \to \infty$.

(iii) $\max_{1\le i\le N, 1\le t\le T} E\left\|\zeta_{it}\right\|^{8+4\sigma} \le C$ for $\zeta_{it} = X_{it}$ and ε_{it}.

(iv) $\max_{1\leq i\leq N} T^{-1}\sum_{t=1}^{T} E_{\mathcal{D}}\left\Vert \zeta_{it}\right\Vert^{2}=O_{P}\left(1\right)$ and $\max_{1\leq t\leq T} N^{-1}\sum_{i=1}^{N} E_{\mathcal{D}}\left\Vert \zeta_{it}\right\Vert^{2}=O_{P}\left(1\right)$ for $\zeta_{it}=X_{it}$ and ε_{it}.

(v) $\max_{1\leq t\leq T} N^{-1}\sum_{i=1}^{N}\left\Vert \zeta_{it}\right\Vert_{8+4\sigma,\mathcal{D}}^{4}=O_{P}\left(1\right)$ for $\zeta_{it}=X_{it}$ and ε_{it}.

Note that Assumptions B.1(i)-(iii) strengthen the moment conditions in Assumptions A.1(i)-(iii) and require finite eighth plus moments for F_t^0, λ_i^0, X_{it}, and ε_{it} to derive the asymptotic distribution of our PPC estimator and to estimate the asymptotic bias and variance terms. Admittedly, our moment conditions are generally different and may sometimes be stronger than those assumed in the literature (e.g., Bai, 2009). For example, Bai (2009) only requires finite fourth moments for F_t^0, λ_i^0 and X_{it} and finite eighth moments for ε_{it}; but he assumes independence between ε_{it} and $(X_{js}, F_s^0, \lambda_j^0)$ for all i,j,t,s, and thus rules out dynamics in the model. Assumptions B.1(iv)-(v) are needed to show some uniform results.

To state the next assumption, we first provide the definition of conditional strong mixing processes.

Definition 1. (Conditional strong mixing) *Let $(\Omega,\mathcal{A},P)$ be a probability space and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let $P_{\mathcal{B}}\left(\cdot\right)\equiv P\left(\cdot|\mathcal{B}\right)$. Let $\{\xi_t, t\geq 1\}$ be a sequence of random variables defined on $(\Omega,\mathcal{A},P)$. The sequence $\{\xi_t, t\geq 1\}$ is said to be conditionally strong mixing given $\mathcal{B}$ (or $\mathcal{B}$-strong-mixing) if there exists a nonnegative $\mathcal{B}$-measurable random variable $\alpha^{\mathcal{B}}\left(t\right)$ converging to 0 a.s. as $t\to\infty$ such that*

$$\left|P_{\mathcal{B}}\left(A\cap B\right)-P_{\mathcal{B}}\left(A\right)P_{\mathcal{B}}\left(B\right)\right|\leq\alpha^{\mathcal{B}}\left(t\right)\ \text{ a.s.} \tag{3.6}$$

for all $A\in\sigma\left(\xi_1,...,\xi_k\right)$, $B\in\sigma\left(\xi_{k+t},\xi_{k+t+1},...\right)$ and $k\geq 1$, $t\geq 1$.

The above definition is due to Prakasa Rao (2009). When one takes $\alpha^{\mathcal{B}}\left(t\right)$ as the supremum of the left hand side object in (3.6) over the set $\{A\in\sigma\left(\xi_1,...,\xi_k\right), B\in\sigma\left(\xi_{k+t},\xi_{k+t+1},...\right), k\geq 1, t\geq 1\}$, we refer to it as the $\mathcal{B}$-strong-mixing coefficient.

Assumption B.2 (i) For each $i=1,...,N$, $\{(X_{it},\varepsilon_{it}): t=1,2,...\}$ is conditionally strong mixing given $\mathcal{D}$ with mixing coefficients $\left\{\alpha_{NT,i}^{\mathcal{D}}\left(\cdot\right)\right\}$. $\alpha_{\mathcal{D}}\left(\cdot\right)\equiv\alpha_{NT}^{\mathcal{D}}\left(\cdot\right)\equiv\max_{1\leq i\leq N}\alpha_{NT,i}^{\mathcal{D}}\left(\cdot\right)$ satisfies $\alpha_{\mathcal{D}}\left(s\right)=O_{a.s.}\left(s^{-\rho}\right)$ where $\rho=(2+\sigma)/(1+\sigma)+\epsilon$ for some arbitrarily small $\epsilon>0$ and σ is as defined in Assumption B.1(i). In addition, there exist integers τ_0, $\tau_*\in(1,T)$ such that $NT\alpha_{\mathcal{D}}\left(\tau_0\right)=o_{a.s.}\left(1\right)$, $T(T+N^{1/2})\alpha_{\mathcal{D}}\left(\tau_*\right)^{(1+\sigma)/(2+\sigma)}=o_{a.s.}\left(1\right)$, and $N^{1/2}T^{-1}\tau_*^2=o\left(1\right)$.

(ii) $\left(\varepsilon_i, X_i\right)$, $i=1,...,N$, are mutually independent of each other conditional on $\mathcal{D}$.

(iii) For each $i=1,...,N$, $E\left(\varepsilon_{it}|\mathcal{F}_{NT,t-1}\right)=0$ a.s., where $\mathcal{F}_{NT,t}\equiv\sigma(\mathcal{D},\{X_{i,t+1},X_{it},\varepsilon_{it},X_{i,t-1},\varepsilon_{i,t-1},...\}_{i=1}^{N})$.

(iv) As $(N,T)\to\infty$, $MN^{1/2}T^{-1/2}(\delta_{NT}^{-1}+\alpha_{NT})\to 0$ and $NT\alpha_{\mathcal{D}}\left(M+1\right)^{(3+2\sigma)/(2+\sigma)}=o_{a.s.}\left(1\right)$ where M is a positive integer defined below.

B.2(i) requires that each individual time series $\{(X_{it},\varepsilon_{it}): t=1,2,...\}$ be $\mathcal{D}$-strong-mixing. It is well known that a process may be conditionally strong mixing but not unconditionally strong mixing. This is true for the simple panel autoregressive models with additive factor structure under suitable conditions. For this reason, Hahn and Kuersteiner (2011), Su and Chen (2013), and Lu and Su (2016) consider the notion of conditional strong mixing where the mixing coefficient is defined by conditioning on either the fixed effects or both the common factors and factor loadings. The dependence of the mixing rate on σ defined in B.1 reflects the trade-off between the degree of dependence and the moment bounds of the process $\{(X_{it}, \varepsilon_{it}), t \geq 1\}$. As Su and Chen (2013) remark, Assumption B.2(ii) does not rule out cross sectional dependence among $(X_{it}, \varepsilon_{it})$. When $X_{it} = Y_{i,t-1}$ and ε_{it} exhibits conditional heteroskedasticity: $\varepsilon_{it} = \sigma_0 (Y_{i,t-1}) \epsilon_{it}$ where ϵ_{it} ~IID$(0,1)$ across both i and t and $\sigma_0(\cdot)$ is an unknown smooth function, $(X_{it}, \varepsilon_{it})$ are not independent across i because of the presence of common factors but independent across i conditional on $\mathcal{D}$ such that B.2(ii) is still satisfied. B.2(iii) requires that the error term ε_{it} be a martingale difference sequence (m.d.s.) with respect to the filter $\mathcal{F}_{NT,t}$ which allows for lagged dependent variables in X_{it}, and conditional heteroskedasticity, skewness, or kurtosis of an unknown form in ε_{it}. In contrast, both Bai (2009) and Pesaran (2006) assume that ε_{it} is independent of X_{js}, λ_j, and F_s for all i, t, j and s; Moon and Weidner (2015b) allow dynamics but assume that ε_{it}'s are independent conditional on $\mathcal{D}$ across both i and t. B.2(iv) requires that M should not grow too fast.

In the following proposition, we only verify the high-level conditions in Assumption A.5. In the supplemental appendix, we verify Assumption A.4(ii). The high-level conditions in A.1(viii)-(x) and A.3(i)-(iii) can be similarly verified.

Proposition 3.8 *Suppose that Assumptions A.1-A.4(i) and B.1-B.2 hold. Then Assumption A.5 holds with* $\Theta_0 = plim_{(N,T)\to\infty} \Theta_{NT}$ *and* $\Theta_{NT} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} Z_{it} Z_{it}' \varepsilon_{it}^2$.

To conduct inference, we need to estimate Q_{NT}, $\mathcal{B}_{1,NT}$, $\mathcal{B}_{2,NT}$, $\mathcal{B}_{3,NT}$, and Θ_{NT}. Let $\hat{\nu}_{ij} = \hat{\lambda}_i'(N^{-1}\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\lambda}_j$, $\hat{Q}_{kNT} = \frac{1}{N_k T}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}} X_i$, and $\hat{A}_{k,l} = \frac{1}{N\hat{N}_k T}\sum_{i\in\hat{G}_k}\sum_{j\in\hat{G}_l}\hat{\nu}_{ij} X_i' M_{\hat{F}} X_j$ where k, $l = 1, ..., K_0$. Define

$$\hat{Q}_{NT} = \begin{pmatrix} \hat{Q}_{1NT} - \hat{A}_{1,1} & -\hat{A}_{1,2} & \cdots & -\hat{A}_{1,K_0} \\ -\hat{A}_{2,1} & \hat{Q}_{2NT} - \hat{A}_{2,2} & \cdots & -\hat{A}_{2,K_0} \\ \vdots & \vdots & \ddots & \vdots \\ -\hat{A}_{K_0,1} & -\hat{A}_{K_0,2} & \cdots & \hat{Q}_{K_0NT} - \hat{A}_{K_0,K_0} \end{pmatrix}. \tag{3.7}$$

Let $\hat{\varepsilon}_{it} = Y_{it} - X_{it}'\hat{\alpha}_k - \hat{\lambda}_i'\hat{F}_t$ for $i \in \hat{G}_k$. Let $\hat{\Psi}_{NT} \equiv$diag$(\hat{\psi}_{1T}, ..., \hat{\psi}_{NT})$ and $\hat{\Phi}_{NT} \equiv$diag$(\hat{\varphi}_{1N}, ..., \hat{\varphi}_{TN})$ where $\hat{\psi}_{iT} \equiv T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$ and $\hat{\varphi}_{tN} \equiv N^{-1}\sum_{i=1}^{N}\hat{\varepsilon}_{it}^2$. For $k = 1, ..., K_0$, let $\hat{\Psi}_{kNT}$ denote the $N \times \hat{N}_k$ submatrix of $\hat{\Psi}_{NT}$; the column indices of $\hat{\Psi}_{kNT}$ correspond to $i \in \hat{G}_k$. Let $\hat{X}_i^* = \frac{1}{N}\sum_{j=1}^{N}\hat{\nu}_{ij} X_j$, and $\hat{Z}_{k,i} = \frac{N}{\hat{N}_k}[X_i - P_{\hat{F}} X_i]\mathbf{1}\{i \in \hat{G}_k\} - \frac{1}{\hat{N}_k} M_{\hat{F}} \sum_{j\in\hat{G}_k}\hat{\nu}_{ji} X_j$. Note that we can write the lth element of $\mathcal{B}_{2,kNT}$ and $\mathcal{B}_{3,NT}$ respectively as

$$\begin{aligned}
\mathcal{B}_{2,kNT,l} &= -\frac{1}{NN_kT^2}\mathrm{tr}\left\{F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\Lambda^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}(k)'\left[\mathbf{X}_l(k)-\mathbf{X}_l^*(k)\right]\right\}, \text{ and} \\
\mathcal{B}_{3,kNT,l} &= \frac{1}{N_kT}\mathrm{tr}\left\{P_{F^0}E_{\mathcal{D}}\left[\boldsymbol{\varepsilon}(k)'\mathbf{X}_l(k)\right]\right\},
\end{aligned}$$

where $\boldsymbol{\varepsilon}(k)$, $\mathbf{X}_l(k)$, and $\mathbf{X}_l^*(k)$ denote $N_k\times T$ matrices that have typical elements ε_{it}, $X_{it,l}$, and $X_{it,l}^*$, respectively, for $i\in G_k^0$ and $t=1,...,T$. Here, $X_{it,l}$ and $X_{it,l}^*$ denotes the lth element in X_{it} and X_{it}^*, respectively, for $l=1,...,p$. Let $\hat{\mathbf{X}}_l(k)$ and $\hat{\mathbf{X}}_l^*(k)$ denote $\hat{N}_k\times T$ matrix with typical elements $X_{it,l}$ and $\hat{X}_{it,l}^*$, respectively, for $i\in\hat{G}_k$ and $t=1,...,T$, where, e.g., $\hat{X}_{it,l}^*$ is the lth element of the tth row of $\hat{X}_i^*$. We propose to estimate $\mathcal{B}_{1,NT}$, $\mathcal{B}_{2,NT,l}$, $\mathcal{B}_{3,kNT,l}$, and Θ_{NT}, respectively, by

$$\begin{aligned}
\hat{\mathcal{B}}_{1,kNT} &= -\frac{1}{\hat{N}_kT^2}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}\hat{\Phi}_{NT}\hat{F}(N^{-1}\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\lambda}_i, \\
\hat{\mathcal{B}}_{2,kNT,l} &= -\frac{1}{N\hat{N}_kT}\mathrm{tr}\left\{\hat{F}(N^{-1}\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Psi}_{kNT}[\hat{\mathbf{X}}_l(k)-\hat{\mathbf{X}}_l^*(k)]\right\}, \\
\hat{\mathcal{B}}_{3,kNT,l} &= \frac{1}{\hat{N}_kT}\mathrm{tr}\left\{P_{\hat{F}}\left[\hat{\boldsymbol{\varepsilon}}(k)'\hat{\mathbf{X}}_l(k)\right]^{\mathrm{trunc}}\right\}, \\
\hat{\Theta}_{NT} &\equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2\hat{Z}_{it}\hat{Z}_{it}',
\end{aligned}$$

where $A^{\mathrm{trunc}}\equiv\sum_{t=1}^{T-M}\sum_{s=t+1}^{t+M}A_{ts}^*$ for any $T\times T$ matrix $A=(A_{ts})$ and A_{ts}^* is a $T\times T$ matrix with (t,s)th element given by A_{ts} and zeros elsewhere, $M=M(T)\to\infty$ as $T\to\infty$, $\hat{\boldsymbol{\varepsilon}}(k)$ denotes the $\hat{N}_k\times T$ matrix with typical elements $\hat{\varepsilon}_{it}$ for $i\in\hat{G}_k$ and $t=1,...,T$, and $\hat{Z}_{it}'$ denotes the tth row of $\hat{Z}_i$. Let $\hat{\mathcal{B}}_{s,kNT}=(\hat{\mathcal{B}}_{s,kNT,1},...,\hat{\mathcal{B}}_{s,kNT,p})'$ for $s=2,3$ and $k=1,...,K_0$. Let $\hat{\mathcal{B}}_{s,NT}=(\hat{\mathcal{B}}_{s,1NT}',...,\hat{\mathcal{B}}_{s,K_0NT}')'$ for $s=1,2,3$. Define the bias-corrected PPC estimator as

$$\mathrm{vec}\left(\hat{\boldsymbol{\alpha}}^{bc}\right)=\mathrm{vec}\left(\hat{\boldsymbol{\alpha}}\right)-\hat{Q}_{NT}^{-1}\left(\hat{\mathcal{B}}_{1,NT}+\hat{\mathcal{B}}_{2,NT}+\hat{\mathcal{B}}_{3,NT}\right),$$

where $\hat{\boldsymbol{\alpha}}^{bc}=(\hat{\alpha}_1^{bc},...,\hat{\alpha}_{K_0}^{bc})$ is a $p\times K_0$ matrix.

Corollary 3.9 *Suppose that Assumptions A.1-A.4(i) and B.1-B.2 hold. Then (i)*$\sqrt{NT}$*vec*$(\hat{\boldsymbol{\alpha}}^{bc}-\boldsymbol{\alpha}^0)\xrightarrow{D}N(0,\,Q_0^{-1}\Theta_0Q_0^{-1})$*, (ii)* $\hat{Q}_{NT}=Q_{NT}+O_P\left(\delta_{NT}^{-1}\right)$*, and (iii)* $\hat{\Theta}_{NT}=\Theta_{NT}+o_P(1)$.

Similarly, one can obtain a bias-corrected version for the post Lasso PC estimator $\tilde{\boldsymbol{\alpha}}$. The only difference is that now one relies on the post-Lasso estimates $(\tilde{\alpha}_k,\tilde{\lambda}_i,\tilde{F}_t)$ and residuals ($\tilde{\varepsilon}_{it}=Y_{it}-X_{it}'\tilde{\alpha}_k-\tilde{\lambda}_i'\tilde{F}_t$ for $i\in\hat{G}_k$) to construct estimates of Q_{NT}, $\mathcal{B}_{1,NT}$, $\mathcal{B}_{2,NT}$, $\mathcal{B}_{3,NT}$, and Θ_{NT}. The procedure is exactly the same as above and thus omitted.

4 Determination of the numbers of factors and groups

In the above analysis we assume that the number of factors R_0 and that of groups K_0 are known. In practice, one has to determine both R_0 and K_0 from data. Ideally we want to propose an information criterion that can be used to consistently estimate both simultaneously. Unfortunately, when both R_0 and K_0 deviate from the true values, we are unable to study the asymptotic properties of the PPC estimators. For this reason, in this section we propose to determine R_0 and K_0 sequentially.

4.1 Determination of the number of factors

Here we use R to denote a generic number of factors. We assume that the true value R_0 is bounded from above by a finite integer $R_{\max}$ and propose a BIC-type information criterion to determine R_0 before determining K_0.

Let $\dot{\beta}_{i,R}$, $\dot{F}_{t,R}$ and $\dot{\lambda}_{i,R}$ denote the PCA estimators (without the penalization device) of β_i, $F_{t,R}$ and $\lambda_{i,R}$ by assuming R factors in the model using the normalization rule: $T^{-1}F'_{(R)}F_{(R)} = I_R$ and $\Lambda'_{(R)}\Lambda_{(R)}$ is a diagonal matrix with descending diagonal elements. Note that we have made the dependence of the parameters and their estimators on R explicitly here, where, e.g., $F_{(R)} = (F_{1,R}, ..., F_{T,R})'$ and $F_{t,R}$ denotes an $R \times 1$ vector of factors when we assume there are R unobserved common factors in the model.

Let $\dot{\boldsymbol{\beta}}_{(R)} = \left(\dot{\beta}'_{1,R}, ..., \dot{\beta}'_{N,R}\right)'$. Define

$$V(R, \dot{\boldsymbol{\beta}}_{(R)}) = \frac{1}{T}\sum_{t=R+1}^{T}\mu_r\left[\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - X_i\dot{\beta}_{i,R}\right)\left(Y_i - X_i\dot{\beta}_{i,R}\right)'\right]. \tag{4.1}$$

Following Bai and Ng (2002), we consider the BIC-type information criterion defined by

$$IC_{1NT}(R) = \ln V(R, \dot{\boldsymbol{\beta}}_{(R)}) + \rho_{1NT}R, \tag{4.2}$$

where $\rho_{1,NT}$ is pre-determined which plays the role of $\ln(NT)/(NT)$ in the case of the conventional BIC criterion. Let $\hat{R} = \arg\min_{1\leq R\leq R_{\max}} IC_{1NT}(R)$, which estimates the number of factors.

To proceed, we add the following assumption.

Assumption A.6. As $(N,T) \to \infty$, $\rho_{1NT} \to 0$ and $\rho_{1NT}\delta_{NT}^2 \to \infty$.

Theorem 4.1 *Suppose that Assumptions A.1-A.6 hold. Then*

$$P\left(\hat{R} = R_0\right) \to 1 \;\; as \;\; (N,T) \to \infty.$$

The above theorem shows that the use of $IC_{1NT}(R)$ can consistently estimate R_0. To implement the above information criterion, one needs to choose the penalty coefficient ρ_{1NT}. Following Bai and Ng (2002), we can set

$$\rho_{1NT} = \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right) \quad \text{or } \rho_{1NT} = \frac{N+T}{NT} \ln\left(\delta_{NT}^2\right),$$

where $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$.

4.2 Determination of the number of groups

Here we use K to denote a generic number of groups. We assume that the true value K_0 is bounded from above by a finite integer $K_{\max}$ and study the determination of the number of groups via some information criterion after we pin down the correct number of factors via IC_{1NT}. Since IC_{1NT} identifies the correct number of factors in large samples, we assume that R_0 is known when we determine the number of groups.

Consider the following PPC criterion function

$$Q_{NT,\kappa}^{(K)}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = Q_{NT}(\boldsymbol{\beta}) + \frac{\kappa}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K} \|\beta_i - \alpha_k\|, \tag{4.3}$$

where $1 \leq K \leq K_{\max}$. By minimizing the above objective function, we can obtain the C-Lasso estimates $\{\hat{\beta}_i(K,\kappa), \hat{\alpha}_k(K,\kappa), \hat{\lambda}_i^{(K,\kappa)}, \hat{F}_t^{(K,\kappa)}\}$ of $\{\beta_i^0, \alpha_k^0, \lambda_i^0, F_t^0\}$, where we make the dependence of $\hat{\beta}_i$, $\hat{\alpha}_k$, $\hat{\lambda}_i$, and $\hat{F}_t$ on (K,κ) explicit. As above, we can classify individual i into group $\hat{G}_k(K,\kappa)$ if and only if $\hat{\beta}_i(K,\kappa) = \hat{\alpha}_k(K,\kappa)$, i.e.,

$$\hat{G}_k(K,\kappa) = \left\{i \in \{1,2,...,N\} : \hat{\beta}_i(K,\kappa) = \hat{\alpha}_k(K,\kappa)\right\} \text{ for } k = 1,...,K. \tag{4.4}$$

Let $\hat{G}(K,\kappa) = \{\hat{G}_1(K,\kappa), ..., \hat{G}_K(K,\kappa)\}$. Based on (4.4), we denote the post-Lasso estimate of $\{\alpha_k^0, \lambda_i^0, F_t^0\}$ as $\{\tilde{\alpha}_{\hat{G}_k(K,\kappa)}, \tilde{\lambda}_i^{(K,\kappa)}, \tilde{F}_t^{(K,\kappa)}\}$. Let $\hat{\sigma}^2_{\hat{G}(K,\kappa)} = \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k(K,\kappa)}\sum_{t=1}^{T}[Y_{it} - \tilde{\alpha}'_{\hat{G}_k(K,\kappa)}X_{it} - \tilde{\lambda}_i^{(K,\kappa)\prime}\tilde{F}_t^{(K,\kappa)}]^2$. We propose to select the number of groups by choosing K to minimize the following information criterion:

$$IC_{2NT}(K,\kappa) = \ln\left[\hat{\sigma}^2_{\hat{G}(K,\kappa)}\right] + \rho_{2NT}pK, \tag{4.5}$$

where ρ_{2NT} is a tuning parameter. Similar information criteria are used to choose the tuning parameter by Liao (2013), Lu and Su (2016), and SSP for shrinkage estimation in different contexts.

Let $\mathcal{K} = \{1,2,...,K_{\max}\}$. We divide $\mathcal{K}$ into three subsets $\mathcal{K}_0$, $\mathcal{K}_-$ and $\mathcal{K}_+$ as follows

$$\mathcal{K}_0 = \{K \in \mathcal{K} : K = K_0\}, \ \mathcal{K}_- = \{K \in \mathcal{K} : K < K_0\}, \text{ and } \mathcal{K}_+ = \{K \in \mathcal{K} : K > K_0\}.$$

Clearly, $\mathcal{K}_0$, $\mathcal{K}_-$ and $\mathcal{K}_+$ denote the three subsets of $\mathcal{K}$ in which the true, under-, and over-fitted models are produced, respectively. Let $G^{(K)} = (G_{K,1}, ..., G_{K,K})$ be any K-partition of the set of individual indices $\{1, 2, ..., N\}$. Let $\mathcal{G}_K$ denote the collection of such partitions. Let $\hat{\sigma}^2_{G^{(K)}} = \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in G_{K,k}}\sum_{t=1}^{T}[Y_{it} - \bar{\alpha}'_{G_{K,k}}X_{it} - \bar{\lambda}_i^{(K)\prime}\bar{F}_t^{(K)}]^2$, where $\{\bar{\alpha}_{G_{K,k}}, \bar{\lambda}_i^{(K)}, \bar{F}_t^{(K)}\}$ denote the estimates of $\{\alpha_k^0, \lambda_i^0, F_t^0\}$ given the group structure specified in $G^{(K)}$.

To proceed, we add the following two assumptions.

Assumption A.7. As $(N,T) \to \infty$, $\min_{1\le K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\hat{\sigma}^2_{G^{(K)}} \xrightarrow{P} \underline{\sigma}^2 > \sigma_0^2$, where $\sigma_0^2 = \text{plim}_{(N,T)\to\infty}\frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in G_k^0}\sum_{t=1}^{T}(Y_{it} - \alpha_k^{0\prime}X_{it} - \lambda_i^{0\prime}F_t^0)^2$.

Assumption A.8. As $(N,T) \to \infty$, $\rho_{2NT} \to 0$ and $\rho_{2NT}\delta_{NT}^2 \to \infty$.

A.7 is intuitively clear. It requires that all under-fitted models yield asymptotic mean square errors that are larger than σ_0^2, which is delivered by the true model. A.8 reflects the usual conditions for the consistency of model selection. The penalty coefficient ρ_{2NT} cannot shrink to zero either too fast or too slowly.

The following theorem justifies that one can choose K to minimize $IC_{2NT}(K, \kappa)$.

Theorem 4.2 *Suppose that Assumptions A.1-A.5 and A.7-A.8 hold. Then*

$$P\left(\inf_{K\in\mathcal{K}_-\cup\mathcal{K}_+} IC_{2NT}(K,\kappa) > IC_{2NT}(K_0,\kappa)\right) \to 1 \text{ as } (N,T)\to\infty.$$

Let $K(\kappa) = \arg\min_{1\le K\le K_{\max}} IC_{NT}(K,\kappa)$. As Theorem 4.2 indicates, as long as κ satisfies Assumptions A.2(iii)-(iv), we have $P(K(\kappa) = K_0) \to 1$ as $(N,T) \to \infty$. Consequently, the minimizer of $IC_{2NT}(K,\kappa)$ with respect to K is equal to K_0 w.p.a.1 for a variety of choices of κ. In practice, it is desirable to have a data-driven method to choose κ. For this purpose, define

$$IC^*_{2NT}(\kappa) = IC_{2NT}(K(\kappa), \kappa).$$

Then we can choose the tuning parameter to be $\hat{\kappa} = \arg\min_{\kappa\in\mathcal{K}} IC^*_{2NT}(\kappa)$, where $\mathcal{K} = \{\kappa : \kappa \propto T^{-a}$ for any $a \in (1/4, 1/2)\}$ provided some conditions on the moments of $\|X_{it}\varepsilon_{it}\|$ and on the relative rates at which N and T pass to infinity are satisfied. See the remark after Assumption A.3.

5 Monte Carlo simulation

In this section we conduct a small set of Monte Carlo simulations to evaluate the finite-sample performance of the C-Lasso and Post-Lasso estimates and that of information criteria in determining the number of groups and the number of common factors.

5.1 Data generating processes

We consider two data generating processes (DGPs) that are static and dynamic panels, respectively. The observations in each DGP are drawn from three groups, with the proportion of the number of observations $N_1 : N_2 : N_3 = 0.3 : 0.3 : 0.4$. We try six combinations of the sample sizes with $N = 100$, 200 and $T = 50, 100, 150$.

DGP1 (Static panel model). The observations (Y_{it}, X_{it}) are generated from the panel structure model

$$Y_{it} = \beta_{i1}^0 X_{it,1} + \beta_{i2}^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}.$$

The common factor vector F_t^0 has 2 dimensions and follows an AR(1) process: $F_t^0 = 0.5 + 0.5F_{t-1}^0 + \zeta_t$, where ζ_t, $t = 1, \ldots, T$, are IID $N[(0,0)', 0.5 \cdot I_2]$. To offset the starting-up effect, we drop the the first T observations of F_t^0. The factor loadings λ_i^0, $i = 1, \ldots, N$ are IID $N[(0.5, 0.5)', I_2]$. The regressors are generated according to

$$X_{it,1} = 0.25 \cdot \lambda_i^{0\prime} F_t^0 + \eta_{it,1},$$

$$X_{it,2} = 0.25 \cdot \lambda_i^{0\prime} F_t^0 + \eta_{it,2},$$

where $\eta_{it} = (\eta_{it,1}, \eta_{it,2})'$ are IID $N(0, I_2)$. Clearly, the regressors $X_{it,1}$ and $X_{it,2}$ are correlated with λ_i^0 and F_t^0. The true coefficients are set to

$$(\alpha_1^0, \alpha_2^0, \alpha_3^0) = \begin{pmatrix} 0.4 & 1.0 & 1.6 \\ 1.6 & 1.0 & 0.4 \end{pmatrix}.$$

The idiosyncratic errors are generated from a process with conditional heteroskedasticity

$$\varepsilon_{it} = \sigma_{it} e_{it}, \sigma_{it} = \sigma \left[0.25 + 0.05(X_{it,1}^2 + X_{it,2}^2)\right]^{1/2}, \; e_{it} \sim IID\ N(0,1),$$

where σ is set to 2.74 so that the signal-noise ratio is 4.

DGP2 (Dynamic panel model). The model is

$$Y_{it} = \beta_{i0}^0 Y_{i,t-1} + \beta_{i1}^0 X_{it,1} + \beta_{i2}^0 X_{it,2} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}.$$

The common factor vector F_t^0 has 2 dimensions and follows an AR(1) process: $F_t^0 = 0.2 + 0.5F_{t-1}^0 + \zeta_t$, where ζ_t, $t = 1, \ldots, T$, are IID $N[(0,0)', 0.5 \cdot I_2]$. The factor loadings $\lambda_i^0, i = 1, \ldots, N$, are IID $N[(0.1, 0.1)', 0.5 \cdot I_2]$. The exogenous regressors $X_{it,1}$ and $X_{it,2}$ are generated as in DGP 1. The true coefficients are

$$(\alpha_1^0, \alpha_2^0, \alpha_3^0) = \begin{pmatrix} 0.8 & 0.6 & 0.4 \\ 0.4 & 1.0 & 1.6 \\ 1.6 & 1.0 & 0.4 \end{pmatrix}.$$

As in DGP 1, the errors are generated from the process

$$\varepsilon_{it} = \sigma_{it} e_{it}, \sigma_{it} = \sigma \left[0.25 + 0.05(X_{it,1}^2 + X_{it,2}^2)\right]^{1/2}, \; e_{it} \sim IID\ N(0,1),$$

where σ is set to 2.84 so that the signal-noise ratio is about 4.

5.2 Determination of R_0 and K_0

Since the performance of classification and estimation may depend crucially on the specification of the number of common factors, R_0, and the number of groups, K_0, it is necessary to determine their values. We consider two methods to determine them. One is based on the sequential method studied in Section 4, and the other is a simultaneous method.

5.2.1 Method I: determining R_0 and K_0 sequentially

To apply the sequential method studied in Section 4, we need to specify the tuning parameter ρ_{1NT} and ρ_{2NT} in the information criteria. Following Bai and Ng (2002) and Li et al. (2016), we set $\rho_{1NT} = \frac{N+T}{NT} \ln\left(\delta_{NT}^2\right)$. Following SSP, we set $\rho_{2NT} = \frac{1}{25} \ln(NT)/\min(N,T)$.

Table 1 shows the performance of the information criterion IC_{1NT} in (4.2) for the determination of R_0 based on 500 replications. It reports the empirical probability that a particular choice of R minimizes IC_{1NT} when $R_0 = 2$. Apparently, IC_{1NT} works very well for the static model even for $T = 50$ and for the dynamic model when $T \geq 100$. In the latter case, we observe that the probability of correctly choosing R_0 exceeds 92%.

Table 2 shows the performance of the information criterion IC_{2NT} in (4.5) for the determination of K_0. To save time in computation and space for presentation, we conduct the simulation 500 times for DGP1 and DGP2 with 100 cross sectional units. The tuning parameter κ is set to be $0.2S_{\varepsilon}T^{-0.45}$, where S_{ε} denotes the sample variance of the residuals estimated from a model assuming that $K_0 = 1$. Table 2 presents the empirical probability that a particular choice of K minimizes IC_{2NT} when $K_0 = 3$. Similar to the findings in the first step, IC_{2NT} works very well for the static model even for $T = 50$ and for the dynamic model when $T \geq 100$. In the latter case, the probability of correctly choosing K_0 is over 99%.

Table 1: Frequency of selecting R ($R_0 = 2$; $R = 1, 2, \ldots, 6$)

	N	T	$R=1$	$R=2$	$R=3$	$R=4$	$R=5$	$R=6$
DGP1	100	50	0.004	0.996	0	0	0	0
	100	100	0	1.000	0	0	0	0
	100	150	0	1.000	0	0	0	0
	200	50	0	1.000	0	0	0	0
	200	100	0	1.000	0	0	0	0
	200	150	0	1.000	0	0	0	0
DGP2	100	50	0.334	0.666	0	0	0	0
	100	100	0.072	0.928	0	0	0	0
	100	150	0.002	0.998	0	0	0	0
	200	50	0.102	0.898	0	0	0	0
	200	100	0	1.000	0	0	0	0
	200	150	0	1.000	0	0	0	0

Table 2: Frequency of selecting K ($K_0 = 3$; $K = 1, 2, \ldots, 6$)

	N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$
DGP1	100	50	0	0	0.990	0.010	0	0
	100	100	0	0	1.000	0	0	0
	100	150	0	0	1.000	0	0	0
DGP2	100	50	0	0.164	0.828	0.006	0.002	0
	100	100	0	0	0.990	0.010	0	0
	100	150	0	0	1.000	0	0	0

5.2.2 Method II: determining R_0 and K_0 jointly

Our second method selects K and R jointly by minimizing the following information criterion:

$$IC_{NT}(K, \kappa, R) = \ln\left[\hat{\sigma}^2_{\hat{G}(K,\kappa,R)}\right] + \rho_{1NT} R + \rho_{2NT} pK, \tag{5.1}$$

where $\hat{\sigma}^2_{\hat{G}(K,\kappa,R)}$ is defined similar to the counterpart in (4.5) except that it is now dependent on R. Specifying κ, $\rho_{1,NT}$, and ρ_{NT} exactly the same as in Method I, we do simulation 500 times for DGP1 and DGP2 with $N = 100$. Table 3 shows that the above information criterion can correctly choose (R_0, K_0) with a probability over 98% for the static panel even when $T = 50$, and with a probability over 95% for the dynamic panel when $T \geq 100$.

Table 3: Frequency of selecting K and R ($R_0 = 2$, $K_0 = 3$; $K = 1, 2, \ldots, 6$; $R = 1, 2, 3, 4$)

DGP	N	T	R	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 5$	$K = 6$
1	100	50	1	0	0	0.004	0	0	0
			2	0	0	0.986	0.010	0	0
			3	0	0	0	0	0	0
			4	0	0	0	0	0	0
1	100	100	1	0	0	0	0	0	0
			2	0	0	1.000	0	0	0
			3	0	0	0	0	0	0
			4	0	0	0	0	0	0
1	100	150	1	0	0	0	0	0	0
			2	0	0	1.0000	0	0	0
			3	0	0	0	0	0	0
			4	0	0	0	0	0	0
2	100	50	1	0	0.052	0.224	0	0	0
			2	0	0.096	0.620	0.006	0.002	0
			3	0	0	0	0	0	0
			4	0	0	0	0	0	0
2	100	100	1	0	0	0.034	0	0	0
			2	0	0	0.956	0.010	0	0
			3	0	0	0	0	0	0
			4	0	0	0	0	0	0
2	100	150	1	0	0	0	0	0	0
			2	0	0	1.000	0	0	0
			3	0	0	0	0	0	0
			4	0	0	0	0	0	0

5.3 Classification and point estimation

In this section, we focus on the finite sample performance of classification and estimation by assuming that the number of groups, K_0, and the number of common factors, R_0, are known. The tuning parameter κ is set to be $C_\kappa S_\varepsilon T^{-0.45}$, where S_ε denotes the sample variance of ε_{it} and C_κ is a sequence of geometrically increasing constants. We try 5 values for C_κ, namely, $0.05, 0.1, 0.2, 0.4, 0.8$. Regarding the initial values of $\{\hat{\beta}_i^{(0)}\}_{i=1}^N$, we set them to be the within-group estimates calculated using the techniques of Pesaran (2006) and Chudik and Pesaran (2015) for the static and dynamic panels, respectively. The initial values of $\{\hat{\alpha}_k^{(0)}\}_{k=1}^{K_0}$ are all set to the average of $\{\hat{\beta}_i^{(0)}\}_{i=1}^N$. In addition, the truncation parameter M used for bias correction is set to be $\lfloor 2T^{1/6} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the largest integer that is not larger than $\cdot$.

We run $\mathcal{R}$ =250 replications for each scenario. Table 4 reports the classification results. We classify the observation with $\hat{\beta}_i$ into the k-th group whose $\hat{\alpha}_k$ is the closest to $\hat{\beta}_i$. The Type I classification error is summarized by averaging over all the observations and all the replications and is computed as

$$\bar{P}(E) = \frac{1}{N\mathcal{R}} \sum_{k=1}^{K_0} \sum_{i \in G_k^0} \sum_{r=1}^{\mathcal{R}} \mathbf{1}\left\{ i \notin \hat{G}_k^{(r)} \right\},$$

where $\hat{G}_k^{(r)}$ denotes the estimated kth group in the rth replication. The behavior of the Type II classification error $\bar{P}(F) = \frac{1}{N\mathcal{R}} \sum_{k=1}^{K_0} \sum_{i \in \hat{G}_k^{(r)}} \sum_{r=1}^{\mathcal{R}} \mathbf{1}\left\{ i \notin G_k^0 \right\}$ is similar and will not be reported to save space.

In Table 4, the classification errors shrink quickly to 0 as T increases. Particularly, when T =100, the classification error $\bar{P}(E)$ typically takes on values 2-5% for the static model and 5-10% for the dynamic model. The results are not very sensitive to the choice of the tuning parameter C_κ. However, the classification errors when T =50 are particularly large, ranging between 9% and 16%. We recommend a long panel of over 100 periods for empirical applications.

We now move on to the point estimation. Tables 5 and 6 report the root-mean-squared error (RMSE) and the bias of the estimates of the coefficient $\beta_{i,1}$ in each model. Since the estimates of coefficients are not directly comparable across groups, we follow SSP and define RMSE($\hat{\beta}_1$) and Bias($\hat{\beta}_1$) respectively as

$$\begin{aligned} \text{RMSE}(\hat{\beta}_1) &= \frac{1}{N} \sum_{k=1}^{K_0} N_k \sqrt{\frac{1}{\mathcal{R}} \sum_{r=1}^{\mathcal{R}} (\hat{\alpha}_{k1,r} - \alpha_{k1})^2}, \text{ and} \\ \text{Bias}(\hat{\beta}_1) &= \frac{1}{N} \sum_{k=1}^{K_0} N_k \frac{1}{\mathcal{R}} \sum_{r=1}^{\mathcal{R}} (\hat{\alpha}_{k1,r} - \alpha_{k1}), \end{aligned}$$

where $\mathcal{R} = 250$ is the number of replications and α_{k1} is the first element of the coefficient of X_{it} in the k-th group. In Tables 5 and 6, the bias-corrected estimates are denoted as C-Lasso BC and post-Lasso BC for the C-Lasso and post-Lasso estimates, respectively.

Table 4: Results of classification error (K_0 and R_0 are Known)

	N	T	C_κ= 0.05	C_κ= 0.1	C_κ= 0.2	C_κ= 0.4	C_κ= 0.8
DGP1	100	50	0.0940	0.0958	0.0996	0.1136	0.1390
	100	100	0.0242	0.0244	0.0251	0.0303	0.0479
	100	150	0.0059	0.0062	0.0068	0.0080	0.0140
	200	50	0.0904	0.0928	0.0989	0.1111	0.1337
	200	100	0.0223	0.0230	0.0252	0.0304	0.0489
	200	150	0.0053	0.0055	0.0061	0.0081	0.0147
DGP2	100	50	0.1528	0.1554	0.1585	0.1596	0.1522
	100	100	0.0539	0.0560	0.0604	0.0718	0.0901
	100	150	0.0208	0.0222	0.0248	0.0316	0.0493
	200	50	0.1492	0.1506	0.1539	0.1568	0.1474
	200	100	0.0533	0.0560	0.0608	0.0733	0.0934
	200	150	0.0200	0.0210	0.0237	0.0314	0.0506

The general pattern is clear from the findings in Tables 5 and 6. First, the RMSEs and the biases of the estimates of both static and dynamic models shrink toward 0 quickly as T increases. Second, the post-Lasso estimates tend to outperform the C-Lasso estimates in terms of RMSEs and is thus recommended for practical use. Third, the bias of post-Lasso estimates are smaller than that of C-Lasso estimates when T is sufficiently large, say over 100. When T is small, the classification error is large and the post-Lasso appears inferior to C-Lasso in terms of bias. This feature is salient in the dynamic model with $T = 50$. Fourth, Tables 5 and 6 suggest that bias correction is useful for dynamic models but not for static models as in the latter case the bias tends to be small. Fifth, as T increases the RMSEs of post-Lasso estimates tend to approach those of the oracle estimates.

5.4 Effects of misspecification of K and/or R

If $(R, K) \neq (R_0, K_0)$, we misspecify either the number of factors, or the number of groups, or both. To investigate the finite sample properties of the PPC estimator with misspecified K and R, we try six model specifications in which K takes on values 3 or 4 and R is set to 1, 2 or 4.[3] To save computing time, we focus on the samples with $N = 100$ and fix C_κ to be 0.2. In Tables 7 and 8, we present the RMSE and the bias of the estimates of the coefficient $\beta_{i,1}$ in each model specification. The main findings are as follows. First, if R is under-specified, the RMSEs and biases of the regression coefficients estimates are several times as large as those obtained from the benchmark model where both R and K are specified to be their true values. Second, if K is correctly specified but R is over-specified, the estimation shows little difference from the

[3]It is obvious that if one sets $K < K_0 = 3$, the PPC will produce inconsistent estimation of the model.

benchmark case. This is consistent with the prediction of Moon and Weidner (2015a) in the one-group case. Third, if K is over-specified and R is correctly specified or over-specified, the RMSEs increase significantly but the biases barely change. In sum, under-specification of K and/or R and over-specification of K are undesirable, because the former leads to inconsistency and the latter results in inefficiency.

Table 5: Estimation of $\beta_{i,1}$ in DGP 1 by PPC (K_0 and R_0 are Known)

		C_κ	0.05		0.1		0.2		0.4		0.8	
N	T		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
100	50	C-Lasso	0.0708	0.0282	0.0739	0.0303	0.0776	0.0346	0.0909	0.0431	0.1037	0.0551
		Post-Lasso	0.0646	0.0279	0.0661	0.0310	0.0689	0.0352	0.0788	0.0452	0.0966	0.0595
		C-Lasso BC	0.0709	0.0283	0.0740	0.0304	0.0776	0.0347	0.0910	0.0432	0.1037	0.0552
		Pos-Lasso BC	0.0647	0.0280	0.0662	0.0311	0.0690	0.0353	0.0789	0.0453	0.0967	0.0596
		Oracle	0.0502	0.0020	0.0502	0.0020	0.0502	0.0020	0.0502	0.0020	0.0502	0.0020
100	100	C-Lasso	0.0423	0.0165	0.0431	0.0176	0.0454	0.0199	0.0516	0.0251	0.0657	0.0403
		Post-Lasso	0.0360	0.0073	0.0363	0.0078	0.0368	0.0093	0.0391	0.0131	0.0476	0.0246
		C-Lasso BC	0.0424	0.0166	0.0432	0.0177	0.0455	0.0200	0.0517	0.0252	0.0658	0.0403
		Pos-Lasso BC	0.0361	0.0073	0.0364	0.0079	0.0369	0.0093	0.0393	0.0131	0.0477	0.0247
		Oracle	0.0334	0.0006	0.0334	0.0006	0.0334	0.0006	0.0334	0.0006	0.0334	0.0006
100	150	C-Lasso	0.0340	0.0139	0.0346	0.0146	0.0359	0.0160	0.0394	0.0197	0.0491	0.0310
		Post-Lasso	0.0274	0.0032	0.0277	0.0035	0.0278	0.0040	0.0279	0.0050	0.0295	0.0089
		C-Lasso BC	0.0341	0.0139	0.0347	0.0146	0.0360	0.0161	0.0394	0.0197	0.0491	0.0311
		Pos-Lasso BC	0.0275	0.0033	0.0277	0.0036	0.0279	0.0040	0.0280	0.0051	0.0296	0.0090
		Oracle	0.0266	0.0014	0.0266	0.0014	0.0266	0.0014	0.0266	0.0014	0.0266	0.0014
200	50	C-Lasso	0.0556	0.0295	0.0569	0.0313	0.0610	0.0352	0.0716	0.0420	0.0871	0.0497
		Post-Lasso	0.0471	0.0263	0.0492	0.0292	0.0542	0.0356	0.0651	0.0457	0.0850	0.0603
		C-Lasso BC	0.0556	0.0296	0.0569	0.0314	0.0611	0.0352	0.0717	0.0421	0.0872	0.0498
		Pos-Lasso BC	0.0471	0.0264	0.0493	0.0293	0.0543	0.0357	0.0652	0.0458	0.0850	0.0603
		Oracle	0.0340	0.0018	0.0340	0.0018	0.0340	0.0018	0.0340	0.0018	0.0340	0.0018
200	100	C-Lasso	0.0337	0.0161	0.0346	0.0173	0.0369	0.0198	0.0429	0.0257	0.0590	0.0432
		Post-Lasso	0.0249	0.0064	0.0252	0.0073	0.0261	0.0095	0.0285	0.0136	0.0382	0.0257
		C-Lasso BC	0.0337	0.0161	0.0346	0.0173	0.0369	0.0198	0.0429	0.0258	0.0590	0.0432
		Pos-Lasso BC	0.0249	0.0064	0.0252	0.0074	0.0261	0.0095	0.0285	0.0136	0.0382	0.0257
		Oracle	0.0225	-0.0002	0.0225	-0.0002	0.0225	-0.0002	0.0225	-0.0002	0.0225	-0.0002
200	150	C-Lasso	0.0275	0.0118	0.0280	0.0126	0.0292	0.0142	0.0344	0.0200	0.0441	0.0317
		Post-Lasso	0.0196	0.0013	0.0197	0.0015	0.0198	0.0021	0.0204	0.0037	0.0226	0.0079
		C-Lasso BC	0.0275	0.0118	0.0280	0.0126	0.0292	0.0142	0.0344	0.0200	0.0441	0.0317
		Pos-Lasso BC	0.0196	0.0013	0.0197	0.0016	0.0198	0.0021	0.0204	0.0037	0.0226	0.0079
		Oracle	0.0192	-0.0003	0.0192	-0.0003	0.0192	-0.0003	0.0192	-0.0003	0.0192	-0.0003

Table 6: Estimation of $\beta_{i,1}$ in DGP 2 by PPC (K_0 and R_0 are Known)

		C_κ	0.05		0.1		0.2		0.4		0.8	
N	T		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
100	50	C-Lasso	0.0924	0.0390	0.0934	0.0406	0.0956	0.0427	0.0952	0.0411	0.0908	0.0256
		Post-Lasso	0.0914	0.0517	0.0950	0.0553	0.1002	0.0591	0.1019	0.0561	0.0944	0.0417
		C-Lasso BC	0.0857	-0.0016	0.0871	0.0005	0.0895	0.0028	0.0907	0.0014	0.0925	-0.0138
		Post-Lasso BC	0.0805	0.0138	0.0829	0.0176	0.0872	0.0214	0.0914	0.0183	0.0888	0.0039
		Oracle	0.0650	0.0116	0.0650	0.0116	0.0650	0.0116	0.0650	0.0116	0.0650	0.0116
100	100	C-Lasso	0.0620	0.0249	0.0635	0.0269	0.0675	0.0311	0.0757	0.0385	0.0859	0.0445
		Post-Lasso	0.0523	0.0198	0.0529	0.0218	0.0563	0.0266	0.0645	0.0353	0.0771	0.0465
		C-Lasso BC	0.0580	0.0014	0.0596	0.0034	0.0636	0.0078	0.0715	0.0154	0.0815	0.0216
		Post-Lasso BC	0.0496	-0.0028	0.0494	-0.0008	0.0510	0.0041	0.0567	0.0128	0.0675	0.0241
		Oracle	0.0447	0.0041	0.0447	0.0041	0.0447	0.0041	0.0447	0.0041	0.0447	0.0041
100	150	C-Lasso	0.0462	0.0193	0.0477	0.0209	0.0510	0.0244	0.0585	0.0320	0.0699	0.0437
		Post-Lasso	0.0380	0.0093	0.0383	0.0105	0.0393	0.0129	0.0422	0.0181	0.0510	0.0291
		C-Lasso BC	0.0432	0.0038	0.0445	0.0055	0.0477	0.0090	0.0548	0.0167	0.0650	0.0285
		Post-Lasso BC	0.0379	-0.0058	0.0378	-0.0046	0.0379	-0.0021	0.0390	0.0030	0.0450	0.0141
		Oracle	0.0353	0.0022	0.0353	0.0022	0.0353	0.0022	0.0353	0.0022	0.0353	0.0022
200	50	C-Lasso	0.0736	0.0348	0.0729	0.0357	0.0745	0.0368	0.0738	0.0352	0.0671	0.0143
		Post-Lasso	0.0687	0.0450	0.0709	0.0469	0.0759	0.0507	0.0798	0.0511	0.0731	0.0370
		C-Lasso BC	0.0674	-0.0041	0.0671	-0.0031	0.0695	-0.0016	0.0702	-0.0027	0.0736	-0.0234
		Post-Lasso BC	0.0558	0.0082	0.0573	0.0101	0.0614	0.0140	0.0665	0.0144	0.0660	0.0003
		Oracle	0.0433	0.0042	0.0433	0.0042	0.0433	0.0042	0.0433	0.0042	0.0433	0.0042
200	100	C-Lasso	0.0486	0.0246	0.0502	0.0266	0.0541	0.0307	0.0633	0.0381	0.0763	0.0431
		Post-Lasso	0.0383	0.0189	0.0399	0.0213	0.0433	0.0262	0.0521	0.0358	0.0683	0.0475
		C-Lasso BC	0.0451	0.0012	0.0468	0.0033	0.0510	0.0075	0.0600	0.0151	0.0737	0.0203
		Post-Lasso BC	0.0348	-0.0038	0.0351	-0.0013	0.0361	0.0036	0.0415	0.0132	0.0569	0.0249
		Oracle	0.0299	0.0013	0.0299	0.0013	0.0299	0.0013	0.0299	0.0013	0.0299	0.0013
200	150	C-Lasso	0.0375	0.0191	0.0391	0.0210	0.0428	0.0250	0.0511	0.0335	0.0646	0.0477
		Post-Lasso	0.0276	0.0077	0.0281	0.0092	0.0293	0.0116	0.0327	0.0171	0.0427	0.0292
		C-Lasso BC	0.0342	0.0036	0.0357	0.0055	0.0394	0.0096	0.0472	0.0181	0.0591	0.0324
		Post-Lasso BC	0.0281	-0.0075	0.0279	-0.0059	0.0278	-0.0035	0.0288	0.0020	0.0352	0.0141
		Oracle	0.0249	0.0012	0.0249	0.0012	0.0249	0.0012	0.0249	0.0012	0.0249	0.0012

Table 7: Estimation of $\beta_{i,1}$ in DGP 1
(Misspecification of K and/or R; $C_\kappa = 0.2$; $(R_0, K_0) = (2, 3)$)

			$K = 3$					
			$R = 1$		$R = 2$		$R = 4$	
N	T		RMSE	Bias	RMSE	Bias	RMSE	Bias
100	50	C-Lasso	0.1876	0.1692	0.0776	0.0346	0.0825	0.0372
		Post-Lasso	0.2040	0.1858	0.0689	0.0352	0.0767	0.0396
		C-Lasso BC	0.1877	0.1693	0.0776	0.0347	0.0821	0.0372
		Post-Lasso BC	0.2041	0.1859	0.0690	0.0353	0.0766	0.0396
		Oracle	0.1649	0.1480	0.0502	0.0020	0.0538	0.0025
100	100	C-Lasso	0.1628	0.1527	0.0454	0.0199	0.0477	0.0216
		Post-Lasso	0.1723	0.1617	0.0368	0.0093	0.0394	0.0113
		C-Lasso BC	0.1629	0.1527	0.0455	0.0200	0.0473	0.0216
		Post-Lasso BC	0.1724	0.1617	0.0369	0.0093	0.0394	0.0114
		Oracle	0.1588	0.1491	0.0334	0.0006	0.0358	0.0008
100	150	C-Lasso	0.1491	0.1422	0.0359	0.0160	0.0372	0.0171
		Post-Lasso	0.1605	0.1523	0.0278	0.0040	0.0301	0.0051
		C-Lasso BC	0.1491	0.1422	0.0360	0.0161	0.0370	0.0171
		Post-Lasso BC	0.1605	0.1523	0.0279	0.0040	0.0300	0.0052
		Oracle	0.1556	0.1479	0.0266	0.0014	0.0290	0.0019

			$K = 4$					
			$R = 1$		$R = 2$		$R = 4$	
N	T		RMSE	Bias	RMSE	Bias	RMSE	Bias
100	50	C-Lasso	0.1843	0.1071	0.1378	-0.0195	0.1418	-0.0031
		Post-Lasso	0.1853	0.1093	0.1370	-0.0047	0.1393	0.0048
		C-Lasso BC	0.1843	0.1072	0.1377	-0.0194	0.1407	-0.0030
		Post-Lasso BC	0.1854	0.1093	0.1370	-0.0046	0.1383	0.0048
		Oracle	0.1649	0.1480	0.0502	0.0020	0.0538	0.0025
100	100	C-Lasso	0.1368	0.0808	0.0938	-0.0464	0.0981	-0.0443
		Post-Lasso	0.1408	0.1040	0.0738	-0.0104	0.0782	-0.0106
		C-Lasso BC	0.1368	0.0808	0.0938	-0.0463	0.0972	-0.0441
		Post-Lasso BC	0.1409	0.1040	0.0738	-0.0104	0.0777	-0.0107
		Oracle	0.1588	0.1491	0.0334	0.0006	0.0358	0.0008
100	150	C-Lasso	0.1257	0.0900	0.0714	-0.0303	0.0740	-0.0337
		Post-Lasso	0.1351	0.1198	0.0486	0.0019	0.0512	-0.0000
		C-Lasso BC	0.1257	0.0900	0.0714	-0.0303	0.0731	-0.0335
		Post-Lasso BC	0.1351	0.1198	0.0486	0.0019	0.0508	-0.0001
		Oracle	0.1556	0.1479	0.0266	0.0014	0.0290	0.0019

Table 8: Estimation of $\beta_{i,1}$ in DGP 2
(Misspecification of K and/or R; $C_\kappa = 0.2$; $(R_0, K_0) = (2, 3)$)

			$K=3$					
			$R=1$		$R=2$		$R=4$	
N	T		RMSE	Bias	RMSE	Bias	RMSE	Bias
100	50	C-Lasso	0.1856	0.1634	0.0956	0.0427	0.0995	0.0431
		Post-Lasso	0.2007	0.1803	0.1002	0.0591	0.1043	0.0562
		C-Lasso BC	0.1720	0.1470	0.0895	0.0028	0.0943	-0.0046
		Post-Lasso BC	0.1883	0.1655	0.0872	0.0214	0.0921	0.0122
		Oracle	0.1526	0.1354	0.0650	0.0116	0.0693	0.0096
100	100	C-Lasso	0.1617	0.1487	0.0675	0.0311	0.0707	0.0339
		Post-Lasso	0.1670	0.1555	0.0563	0.0266	0.0602	0.0281
		C-Lasso BC	0.1530	0.1388	0.0636	0.0078	0.0654	0.0095
		Post-Lasso BC	0.1588	0.1463	0.0510	0.0041	0.0545	0.0048
		Oracle	0.1418	0.1309	0.0447	0.0041	0.0480	0.0041
100	150	C-Lasso	0.1532	0.1446	0.0510	0.0244	0.0536	0.0260
		Post-Lasso	0.1544	0.1465	0.0393	0.0129	0.0425	0.0136
		C-Lasso BC	0.1474	0.1382	0.0477	0.0090	0.0495	0.0100
		Post-Lasso BC	0.1489	0.1405	0.0379	-0.0021	0.0409	-0.0018
		Oracle	0.1399	0.1324	0.0353	0.0022	0.0377	0.0018

			$K=4$					
			$R=1$		$R=2$		$R=4$	
N	T		RMSE	Bias	RMSE	Bias	RMSE	Bias
100	50	C-Lasso	0.1680	0.1038	0.1306	-0.0089	0.1352	0.0005
		Post-Lasso	0.1801	0.1043	0.1529	0.0094	0.1597	0.0047
		C-Lasso BC	0.1683	0.1035	0.1311	-0.0097	0.1345	-0.0005
		Post-Lasso BC	0.1805	0.1040	0.1531	0.0088	0.1588	0.0040
		Oracle	0.1526	0.1354	0.0650	0.0116	0.0693	0.0096
100	100	C-Lasso	0.1574	0.0903	0.1123	-0.0395	0.1156	-0.0301
		Post-Lasso	0.1580	0.1099	0.1030	-0.0034	0.1090	-0.0004
		C-Lasso BC	0.1576	0.0904	0.1123	-0.0392	0.1147	-0.0299
		Post-Lasso BC	0.1583	0.1099	0.1031	-0.0034	0.1084	-0.0006
		Oracle	0.1418	0.1309	0.0447	0.0041	0.0480	0.0041
100	150	C-Lasso	0.1288	0.0713	0.0956	-0.0513	0.0991	-0.0520
		Post-Lasso	0.1332	0.1045	0.0747	-0.0072	0.0766	-0.0078
		C-Lasso BC	0.1288	0.0714	0.0956	-0.0511	0.0982	-0.0515
		Post-Lasso BC	0.1333	0.1044	0.0749	-0.0072	0.0762	-0.0080
		Oracle	0.1399	0.1324	0.0353	0.0022	0.0377	0.0018

6 An application to China's housing price data

In this application, we study the persistence of housing prices in China's large and medium-sized cities. Applying our methodology, we document that the persistence in the housing price growth from 2005 to 2014 is heterogeneous across cities and exhibits grouped patterns.

In the last decade, China's housing markets experienced a remarkably rapid price appreciation. Among the 70 large and medium-sized cities, 37 cities had an annual growth rate over 4%, and ten of them even reached over 7% per year. The observed price growth rates in most cities are very persistent. Meanwhile, the persistence also exhibits differences across cities. We seek to identify the persistence heterogeneity by assuming the existence of a grouped pattern.

Monthly data of growth rates of second-hand house prices between 2005m7 and 2014m4 are available from China Economic Information Network. After excluding the records with missing values, a panel of 69 cities and 104 periods can be used for our estimation. The econometric model is specified as

$$Y_{i,t} = Y_{i,t-1}\beta_i + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it},$$

where $Y_{i,t}$ is the annualized growth rate of the average second-hand housing prices in city i from month t-1 to t, and β_i is the key coefficient measuring the persistence of housing prices in city i. The F_t^0 stands for common factors. As suggested by Bai et al. (2014) and Zhang et al. (2012), the housing prices are mainly determined by a couple of common factors such as inflation, credit policies, taxation, and public expectation. The idiosyncratic term ε_{it} represents heterogeneous policies of local governments and various city-specific attributes such as Hukou population and income levels (Wang and Zhang, 2014).

The estimation requires specifications of the tuning parameter κ, the number of groups K_0, and the number of common factors R_0. As in the simulations, we set $\kappa = C_\kappa S_\varepsilon(R) T^{-0.45}$ and let C_k take on values 0.05, 0.1, 0.2, 0.4, and 0.8. The variance $S_\varepsilon(R)$ is obtained by running the panel regression (assuming that $K_0 = 1$) and computing the sample variance of the fitted residuals. We use two methods proposed in the simulation section to determine R_0 and K_0. Both methods suggest that $R_0 = 5$ and $K_0 = 3$ (see Tables 9, 10, and 11).

When $R_0 = 5$, $K_0 = 3$ and $C_\kappa = 0.2$,[4] the C-Lasso method groups the 69 cities as follows:

- Group 1 (27 cities): Beijing, Shanghai, Guangzhou, Shenzhen, Tianjin, Chengdu, Shenyang, Dalian, Nanjing, Hangzhou, Ningbo, Fuzhou, Xiamen, Quanzhou, Zhanjiang, Jinan, Taiyuan, Yinchuan, Urumqi, Jilin, Jinhua, Anqing, Jiujiang, Pingdingshan, Yichang, Xiangfan, Changde;

- Group 2 (31 cities): Shijiazhuang, Hohhot, Changchun, Harbin, Hefei, Nanchang, Qingdao, Wuhan, Changsha, Chongqing, Guiyang, Kunming, Xining, Tangshan, Qinhungdao, Dandong, Jinzhou, Wuxi, Xuzhou, Wenzhou, Ganzhou, Yantai, Jining, Luoyang, Yueyang, Shaoguan, Beihai, Sanya, Luzhou, Nanchong, Dali;

[4] The classification is not sensitive to the choice of C_κ.

- Group 3 (11 cities): Zhengzhou, Nanning, Guilin, Haikou, Xi'an, Lanzhou, Baotou, Mudanjiang, Bengbu, Huizhou, Zunyi.

Table 12 reports C-Lasso and post-Lasso bias-corrected estimates of the persistence parameter β_1. The housing prices in the first group are most persistent, with β_1 being between 0.51 and 0.56. Interestingly, those cities are mostly large cities located in eastern China. The prices in the second group are less persistent, while the persistence in the third group is close to zero. Those are mainly medium-sized or inland cities.

Table 9: Information criteria for R_0(Method I)

	R=2	R=3	R=4	R=5	R=6	R=7	R=8
$IC_{1NT}(R)$	4.8295	4.8012	4.7895	$\underline{4.7809}$	4.7859	4.7949	4.8002

Table 10: Information criteria for K_0 (Method I)

C_κ	K=1	K=2	K=3	K=4	K=5
0.05	4.3342	4.3098	$\underline{4.2977}$	4.3100	4.3082
0.1	4.4146	4.3098	$\underline{4.2977}$	4.3100	4.3082
0.2	4.4146	4.3098	$\underline{4.2977}$	4.3100	4.3082
0.4	4.4146	4.3098	$\underline{4.2977}$	4.3100	4.3082
0.8	4.4146	4.3098	$\underline{4.2966}$	4.3100	4.3082

Table 11: Information criteria for (R_0, K_0) (Method II)

	C_κ= 0.05					C_κ= 0.1				
K=	R=2	R=3	R=4	R=5	R=6	R=2	R=3	R=4	R=5	R=6
1	4.9128	4.8664	4.8541	4.8446	4.8532	4.9498	4.9227	4.9263	4.9250	4.9266
2	4.8771	4.8376	4.8315	4.8202	4.8251	4.8771	4.8377	4.8315	4.8202	4.8251
3	4.8588	4.8257	4.8160	<u>4.8080</u>	4.8160	4.8588	4.8257	4.8160	<u>4.8080</u>	4.8160
4	4.8689	4.8308	4.8209	4.8204	4.8143	4.8689	4.8308	4.8209	4.8204	4.8143
5	4.8663	4.8313	4.8226	4.8186	4.8223	4.8653	4.8312	4.8230	4.8186	4.8223
	C_κ= 0.2					C_κ= 0.4				
K=	R=2	R=3	R=4	R=5	R=6	R=2	R=3	R=4	R=5	R=6
1	4.9498	4.9227	4.9263	4.9250	4.9266	4.9498	4.9227	4.9263	4.9250	4.9266
2	4.8771	4.8377	4.8315	4.8202	4.8255	4.8771	4.8377	4.8315	4.8202	4.8251
3	4.8588	4.8257	4.8160	<u>4.8080</u>	4.8160	4.8588	4.8257	4.8160	<u>4.8081</u>	4.8160
4	4.8689	4.8308	4.8209	4.8204	4.8143	4.8689	4.8308	4.8209	4.8204	4.8142
5	4.8664	4.8312	4.8230	4.8185	4.8223	4.8658	4.8312	4.8228	4.8185	4.8223
	C_κ= 0.8									
K=	R=2	R=3	R=4	R=5	R=6					
1	4.9498	4.9227	4.9263	4.9250	4.9266					
2	4.8771	4.8376	4.8315	4.8202	4.8251					
3	4.8588	4.8257	4.8160	<u>4.8070</u>	4.8160					
4	4.8689	4.8308	4.8209	4.8204	4.8142					
5	4.8657	4.8312	4.8225	4.8185	4.8223					

Table 12: C-Lasso and Post-Lasso bias-corrected estimates

	$C_\kappa = 0.05$		$C_\kappa = 0.1$		$C_\kappa = 0.2$		$C_\kappa = 0.4$		$C_\kappa = 0.8$	
Group#	β_1	t-value	β_1	t-value	β_1	t-value	β_1	t-value	β_1	t-value
C-Lasso bias-corrected estimates										
1	0.5115	13.9621	0.5116	13.9658	0.5118	13.9711	0.5108	14.8863	0.5114	15.3179
2	0.3134	8.6061	0.3132	8.5996	0.3128	8.5866	0.3052	8.0482	0.3039	8.0280
3	-0.0065	-0.1558	-0.0061	-0.1478	-0.0065	-0.1569	-0.0064	-0.1538	-0.0064	-0.1455
Post-Lasso bias-corrected estimates										
1	0.5563	15.0382	0.5563	15.0382	0.5563	15.0382	0.5464	15.7838	0.5314	15.8928
2	0.2697	7.2744	0.2697	7.2744	0.2697	7.2744	0.2625	6.8102	0.2778	7.2771
3	0.0074	0.1793	0.0074	0.1793	0.0074	0.1793	0.0075	0.1804	-0.0371	-0.8577

7 Conclusion

In this paper we study a panel structure model with interactive fixed effects. We propose penalized principal component (PPC) estimation of the latent group structure and unknown parameters simultaneously. Our PPC method achieves the uniform classification consistency and oracle property. We also propose information criteria to determine the number of factors and the number of groups in the model. Simulations suggest that the good performance of our method typically requires larger amount of time series observations than in SSP. We apply our method to study the persistence of housing prices in China's large and medium-sized cities in the period 2005m7–2014m4 and identify three groups among 69 cities.

References

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge.

Bai, C., Li, Q., and Ouyang M., 2014. Property taxes and home prices: A tale of two cities. Journal of Econometrics 180, 1-15.

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

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

Bai, J., Li, K., 2014. Theory and methods of panel data models with interactive effects. Annals of Statistics 42, 142-170.

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

Baltagi, B. H., Bresson, G., Pirotte, A., 2008. To pool or not to pool? In L. Mátyás and P. Sevestre (Eds.), The Econometrics of Panel Data:Fundamentals and Recent Developments in Theory and Practice, pp. 517-546, 3rd Edition. Springer-Verlag, Berlin.

Bertsekas, D.P., 1995. Nonlinear Programming. Athena Scientific, Belmont, MA.

Bester, C. A., Hansen, C. B., 2016. Grouped effects estimators in fixed effects models. Journal of Econometrics 190, 197-208.

Bonhomme, S., Manresa, E., 2015. Grouped patterns of heterogeneity in panel data. Econometrica 83, 1147-1187.

Chudik, A., Pesaran M. H., 2015. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics 188, 393-420.

Chudik, A., Pesaran M. H., Tosetti, E., 2011. Weak and strong cross section dependence and estimation of large panels. Econometrics Journal 14, C45-C90.

Fan, J., Peng, H., 2004. Nonconcave penalized likelihood with a diverging number of parameters. Annals of Statistics 32, 928-961.

Greenaway-McGrevy, R., C. Han, D. Sul, 2012. Asymptotic distribution of factor augmented estimators for panel regression. Journal of Econometrics 169, 48-53.

Hahn, J., Kuersteiner, G., 2011. Bias reduction for dynamic nonlinear panel models with fixed effects. Econometric Theory 27, 1152-1191.

Hsiao, C., 2014. Analysis of Panel Data. Cambridge University Press, Cambridge.

Hsiao, C., Pesaran, M. H., 2008. Random coefficient panel data models. In L. Matyas and P. Sevestre (Eds.), The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice, pp. 187-216, 3rd Edition. Springer-Verlag, Berlin.

Jin, S., Su, L., 2013. A Nonparametric poolability test for panel data models with cross section dependence. Econometric Reviews 32, 469-512.

Kapetanios, G., Pesaran, M. H., 2007. Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modeling of asset returns. In G. Phillips & E. Tzavalis (eds.), The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis, pp. 239–281, Cambridge University Press, New York.

Kapetanios, G., Pesaran, M. H., Yagamata, T., 2011. Panels with non-stationary multifactor error structures. Journal of Econometrics 160, 326-348.

Lam, C., Fan, J. 2008. Profile-kernel likelihood inference with diverging number of parameters. Annals of Statistics 36, 2232-2260.

Li, D., J. Qian, and L. Su, 2016. Panel data models with interactive fixed effects and multiple structural breaks. Journal of the American Statistical Association, forthcoming.

Li, K., Lu, L., 2014. Efficient estimation of heterogeneous coefficients in panel data models with common shock. Working paper, Capital University of Economics and Business.

Liao, Z., 2013. Adaptive GMM shrinkage estimation with consistent moment selection. Econometric Theory 29, 857-904.

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

Lu, X., Su, L., 2016. Shrinkage estimation of dynamic panel data models with interactive fixed effects. Journal of Econometrics 190, 148-175.

Moon, H. R. Weidner, M., 2015a. Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83, 1543-1579.

Moon, H. R. Weidner, M., 2015b. Dynamic linear panel data regression models with interactive fixed effects. Econometric Theory, forthcoming.

Newey, W., McFadden, D., 1994. Large sample estimation and hypothesis testing. In R.F. Engle and D. McFadden (eds), Handbook of Econometrics, 2111-2245. Amsterdam: North Holland.

Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967-1012.

Pesaran, M. H. Tosetti, E., 2011. Large panels with common factors and spatial correlation. Journal of Econometrics 161, 182–202.

Pesaran, M. H., Yamagata, T., 2008. Testing slope homogeneity in large panels. Journal of Econometrics 142, 50-93.

Prakasa Rao, B. L. S., 2009. Conditional independence, conditional mixing and conditional association. Annals of the Institute of Statistical Mathematics 61, 441-460.

Sarafidis, V., Weber, N., 2015. A Partially heterogeneous framework for analyzing panel data. Oxford Bulletin of Economics and Statistics 77, 274-296.

Song, M., 2013. Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Working paper, Columbia University.

Su, L., Chen, Q., 2013. Testing homogeneity in panel data models with interactive fixed effects. Econometric Theory, 29, 1079-1135.

Su, L., Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. Journal of Econometrics 169, 34-47.

Su, L., Jin, S, Zhang, Y., 2015. Specification test for panel data models with interactive fixed effects. Journal of Econometrics 186, 222-244.

Su, L., Wang, X., 2016. On time-varying factor models: estimation and testing. Working paper, Singapore Management University.

Su, L., Shi, Z., Phillips, P.C.B., 2016. Identifying latent structures in panel data. Econometrica, forthcoming.

Su, L., Zhang, Y., 2016. Nonparametric dynamic panel data models with interactive fixed effects: sieve estimation and specification testing. Working paper, Singapore Management University..

Sun, Y., 2005. Estimation and inference in panel structure models. Working paper, UCSD.

Tibshirani, R. J., 1996. Regression Shrinkage and Selection via the LASSO. Journal of the Royal Statistical Society, Series B. 58, 267-288.

Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society, Series B. 68, 49-67.

Zhang, Y., Hua, X., and Zhao, L., 2012. Exploring determinants of housing prices: A case study of Chinese experience in 1999-2010. Economic Modeling 29, 2349-2361.

Wang, Z., Zhang, Q., 2014. Fundamental factors in the housing markets of China. Journal of Housing Economics 25, 53-61.

Supplemental Material for
"Identifying Latent Grouped Patterns in Panel Data Models with Interactive Fixed Effects"
(Not for Publication)

Liangjun Su[a] and Gaosheng Ju[b]

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

This supplemental material consists of four parts. Appendix A contains proofs of the main results in the paper and the statement of some technical lemmas. Appendix B contains the proofs of these technical lemmas. Appendix C verifies Assumptions A.4(ii) under Assumptions B.1-B.2 and Appendix D contains numerical algorithm for our PPC estimation.

A Proof of the results in Section 3

In this appendix we prove the main results in the paper. The proof relies on some technical lemmas whose proofs are given in Appendix B. Throughout the appendix, we frequently use the decomposition

$$\begin{aligned} P_{\hat{F}} - P_{F^0 H} &= T^{-1}\hat{F}\hat{F}' - F^0 H \left(H' F^{0\prime} F^0 H\right)^{-1} H' F^{0\prime} \\ &= T^{-1}(\hat{F} - F^0 H)(\hat{F} - F^0 H)' - T^{-1}(\hat{F} - F^0 H) H' F^{0\prime} \\ &\quad + T^{-1} F^0 H(\hat{F} - F^0 H)' + T^{-1} F^0 H (I_{R_0} - \left(T^{-1} H' F^{0\prime} F^0 H\right)^{-1})' H' F^{0\prime} \\ &\equiv d_1 + d_2 + d_3 + d_4, \text{ say,} \end{aligned} \tag{A.1}$$

where $H = (N^{-1}\Lambda^{0\prime}\Lambda^0)(T^{-1}F^{0\prime}\hat{F})V_{NT}^{-1}$. Let $G = V_{NT}^{-1}H^{-1} = (T^{-1}F^{0\prime}\hat{F})^{-1}(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}$, $\mu_{ts} = F_t^{0\prime} \times \left(T^{-1}F^{0\prime}F^0\right)^{-1} F_s^0$ and $\nu_{ij} = \lambda_i^{0\prime}\left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\lambda_j^0$. Let $\eta_{NT}^2 = \frac{1}{N}\sum_{i=1}^N ||\hat{b}_i||^2$. Let ω_p, $\bar{\omega}_p \in \mathbb{R}^p$ be two non-random vectors with $\|\omega_p\| = 1$ and $\|\bar{\omega}_p\| = 1$. The next lemma is used in the proof of Theorem 3.1.

Lemma A.1 *Suppose Assumption A.1 hold. Then*

(i) $\sup_{N^{-1/2}\|\mathbf{b}\| \le L} \sup_{F\in\mathcal{F}} \left| \frac{1}{NT}\sum_{i=1}^N b_i' X_i' M_F \varepsilon_i \right| = o_P(1)$,

(ii) $\sup_{F\in\mathcal{F}} \left\| \frac{1}{NT}\sum_{i=1}^N \lambda_i^{0\prime} F^{0\prime} M_F \varepsilon_i \right\| = o_P(1)$,

(iii) $\sup_{F\in\mathcal{F}} \left\| \frac{1}{NT}\sum_{i=1}^N \varepsilon_i' P_F \varepsilon_i \right\| = o_P(1)$,

(iv) $\frac{1}{NT}\sum_{i=1}^N \varepsilon_i' P_{F^0} \varepsilon_i = O_P\left(T^{-1}\right)$,

where $\mathbf{b} = (b_1, ..., b_N)$, *$L$ is a large positive constant that does not depend on N and T and* $\mathcal{F} = \left\{F \in \mathbb{R}^{T\times R_0} : T^{-1}F'F = I_{R_0}\right\}$.[2]

[2] From the proof of Lemma A.1(i) we can tell that the result in (i) also holds if we allow L to diverge at a rate $o(\min(N^{1/4}, T^{1/4}))$.

Proof of Theorem 3.1. Let $Q_{1NT,i}\left(\beta_i, F\right)=\frac{1}{T}\left(Y_i-X_i\beta_i\right)^{\prime} M_F\left(Y_i-X_i\beta_i\right)$ and $Q_{1iNT,\kappa}^{(K_0)}\left(\beta_i, \boldsymbol{\alpha}, F\right)=Q_{1NT,i}\left(\beta_i, F\right)+\kappa\Pi_{k=1}^{K_0}\left\|\beta_i-\alpha_k\right\|$. Note that $Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}, F\right)=\frac{1}{N}\sum_{i=1}^N Q_{1iNT,\kappa}^{(K_0)}\left(\beta_i, \boldsymbol{\alpha}, F\right)$. We prove the theorem under Assumptions A.1(i)-(x) by assuming that the minimizer $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ can only occur in the L-neighborhood of $\boldsymbol{\beta}^0=(\beta_1^0,...,\beta_N^0)$: $\mathcal{N}_L\left(\boldsymbol{\beta}^0\right)=\left\{\boldsymbol{\beta}\in\mathbb{R}^{p\times N}: N^{-1/2}\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\right\|\leq L\right\}$, where L is a large positive constant that does not grow with N and T. Then we demonstrate that Assumption A.1(xi) rules out the possibility that $N^{-1/2}\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right\|$ is divergent.

(i) Let $\mathbf{b}=(b_1,...,b_N)$ and $b=$vec$(\mathbf{b})$ where $b_i=\beta_i-\beta_i^0$ for $i=1,...,N$. Let $\hat{\mathbf{b}}=(\hat{b}_1,...,\hat{b}_N)$ and $\hat{b}=$vec$(\hat{\mathbf{b}})$ where $\hat{b}_i=\hat{\beta}_i-\beta_i^0$. Note that

$$\begin{aligned}&Q_{1NT,i}\left(\beta_i, F\right)-Q_{1NT,i}\left(\beta_i^0, F^0\right)\\=\;&S_{NT,i}\left(\beta_i, F\right)-\frac{2}{T}b_i^{\prime}X_i^{\prime}M_F\varepsilon_i+\frac{2}{T}\lambda_i^{0\prime}F^{0\prime}M_F\varepsilon_i-\frac{1}{T}\varepsilon_i^{\prime}\left(P_F-P_{F^0}\right)\varepsilon_i,\end{aligned}\tag{A.2}$$

where $S_{NT,i}\left(\beta_i, F\right)=\frac{1}{T}b_i^{\prime}X_i^{\prime}M_FX_ib_i+\frac{1}{T}\lambda_i^{0\prime}F^{0\prime}M_FF^0\lambda_i^0-\frac{2}{T}b_i^{\prime}X_i^{\prime}M_FF^0\lambda_i^0$. By Lemma A.1, the last three terms on the right hand side of (A.2), after averaging over i, are $o_P(1)$ uniformly in $\boldsymbol{\beta}\in\mathcal{N}_L\left(\boldsymbol{\beta}^0\right)$ and $F\in\mathcal{F}$. It follows that uniformly in $\boldsymbol{\beta}\in\mathcal{N}_L\left(\boldsymbol{\beta}^0\right)$ and $F\in\mathcal{F}$ we have

$$\begin{aligned}Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}, F\right)-Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, F^0\right)&=\frac{1}{N}\sum_{i=1}^N\left[Q_{1NT,i}\left(\beta_i, F\right)-Q_{1NT,i}\left(\beta_i^0, F^0\right)\right]\\&\quad+\frac{\kappa}{N}\sum_{i=1}^N\Pi_{k=1}^{K_0}\left\|\beta_i-\hat{\alpha}_k\right\|\\&\geq\frac{1}{N}\sum_{i=1}^N S_{NT,i}\left(\beta_i, F\right)+o_P(1).\end{aligned}\tag{A.3}$$

Let $A=T^{-1}\mathrm{diag}(X_1^{\prime}M_FX_1,...,X_N^{\prime}M_FX_N)$, $B=\left(\Lambda^{0\prime}\Lambda^0\right)\otimes I_T$, and $C=T^{-1/2}((\lambda_1^0\otimes M_FX_1),...,(\lambda_N^0\otimes M_FX_N))$. Let $\eta=\frac{1}{T^{1/2}}\mathrm{vec}(M_FF^0)$. Then

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N S_{NT,i}\left(\beta_i, F\right)&=\frac{1}{N}b^{\prime}Ab+\mathrm{tr}\left[\left(\frac{1}{T^{1/2}}M_FF^0\right)\left(\frac{1}{N}\Lambda^{0\prime}\Lambda^0\right)\left(\frac{1}{T^{1/2}}F^{0\prime}M_F\right)\right]\\&\quad-\frac{2}{NT}\sum_{i=1}^N\mathrm{tr}\left(M_FF^0\lambda_i^0b_i^{\prime}X_i^{\prime}M_F\right)\\&=N^{-1}b^{\prime}Ab+N^{-1}\eta^{\prime}\left[\left(\Lambda^{0\prime}\Lambda^0\right)\otimes I_T\right]\eta-\frac{2}{NT^{1/2}}\eta^{\prime}\sum_{i=1}^N\left(\lambda_i^0\otimes M_FX_i\right)b_i\\&=N^{-1}b^{\prime}Ab+N^{-1}\eta^{\prime}B\eta-2N^{-1}b^{\prime}C^{\prime}\eta\\&=N^{-1}b^{\prime}Ab+N^{-1}\vartheta^{\prime}B\vartheta-N^{-1}b^{\prime}C^{\prime}B^{-1}Cb\end{aligned}\tag{A.4}$$

where $\vartheta=\eta-B^{-1}Cb$, the second equality follows from the fact that

$$\mathrm{tr}\left(B_1B_2B_3\right)=\mathrm{vec}\left(B_1\right)^{\prime}\left(B_2\otimes I\right)\mathrm{vec}\left(B_3^{\prime}\right)\text{ and }\mathrm{tr}\left(B_1B_2B_3B_4\right)=\mathrm{vec}\left(B_1\right)^{\prime}\left(B_2\otimes B_4^{\prime}\right)\mathrm{vec}\left(B_3^{\prime}\right)$$

for any conformable matrices B_1, B_2, B_3, B_4 and an identity matrix I (see, e.g., Bernstein (2005, p.253)). We now argue that the last term in (A.4) is $o_P(1)$ uniformly in b such that $\boldsymbol{\beta}\in\mathcal{N}_L\left(\boldsymbol{\beta}^0\right)$ and $N^{-1}b^{\prime}b=O(1)$. Observe that $N^{-1}b^{\prime}C^{\prime}B^{-1}Cb\leq\mu_{\max}\left(C^{\prime}B^{-1}C\right)N^{-1}b^{\prime}b=o_P(1)$ for any $N^{-1}b^{\prime}b=O(1)$ provided

$\mu_{\max}\left(C'B^{-1}C\right)=o_P\left(1\right).$ Note that $\mu_{\max}\left(C'B^{-1}C\right)\leq\left[\mu_{\min}\left(N^{-1}B\right)\right]^{-1}\mu_{\max}\left(N^{-1}C'C\right)=\underline{c}_{\Lambda}^{-1}\mu_{\max}(N^{-1}\times C'C)$, where $\underline{c}_{\Lambda}\equiv\mu_{\min}\left(\Lambda^{0\prime}\Lambda^0/N\right).$ Define the upper block-triangular matrix

$$C_1=T^{-1}\begin{pmatrix}\lambda_1^{0\prime}\lambda_1^0X_1'M_FX_1 & \lambda_1^{0\prime}\lambda_2^0X_1'M_FX_2 & \cdots & \lambda_1^{0\prime}\lambda_N^0X_1'M_FX_N\\ 0 & \lambda_2^{0\prime}\lambda_2^0X_2'M_FX_2 & \cdots & \lambda_2^{0\prime}\lambda_N^0X_2'M_FX_N\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_N^{0\prime}\lambda_N^0X_N'M_FX_N\end{pmatrix}.$$

Noting that the $Np\times Np$ matrix $C'C$ has a typical $p\times p$ block submatrix $T^{-1}\lambda_i^{0\prime}\lambda_j^0X_i'M_FX_j$, we have $C'C=C_1+C_1'-C_d$ where $C_d=T^{-1}\text{diag}(\lambda_1^{0\prime}\lambda_1^0X_1'M_FX_1,...,\lambda_N^{0\prime}\lambda_N^0X_N'M_FX_N)$. By the fact that the eigenvalues of a block upper/lower triangular matrix are the combined eigenvalues of its diagonal block matrices, Weyl's inequality, and Assumptions A.1(ii)-(iii), we have

$$\begin{aligned}N^{-1}\mu_{\max}\left(C'C\right) &\leq N^{-1}\left\{2\mu_{\max}\left(C_1\right)-\mu_{\min}(C_d)\right\}\\ &\leq 2N^{-1}\max_{1\leq i\leq N}\left\|\lambda_i^0\right\|^2\mu_{\max}\left(T^{-1}X_i'M_FX_i\right)\\ &\leq 2N^{-1}\max_{1\leq i\leq N}\left\|\lambda_i^0\right\|^2\max_{1\leq i\leq N}T^{-1}\left\|X_i\right\|^2\\ &= O_P(N^{-1})o_P(N^{1/4})O_P(1)=o_P(N^{-3/4}),\end{aligned}\tag{A.5}$$

where the third inequality follows from the fact that $\mu_{\max}\left(T^{-1}X_i'M_FX_i\right)\leq\mu_{\max}\left(T^{-1}X_i'X_i\right)=T^{-1}\left\|X_i\right\|_{\text{sp}}^2\leq T^{-1}\left\|X_i\right\|^2$ as $\mu_{\max}\left(M_F\right)=1$ and the first equality follows from the fact that $\max_{1\leq i\leq N}\left\|\lambda_i^0\right\|^2=o_P\left(N^{1/4}\right)$ under Assumption A.1(ii) by the Markov inequality. It follows that $\mu_{\max}\left(C'B^{-1}C\right)=o_P(N^{-3/4})$ and $N^{-1}b'C'B^{-1}Cb=o_P\left(1\right)$ uniformly in b such that $N^{-1}b'b=O\left(1\right).$ This, in conjunction with (A.2)-(A.4) and the fact that $Q_{1NT,\kappa}^{(K_0)}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\hat{F})-Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,F^0\right)\leq 0$ implies that $N^{-1}\hat{b}'\hat{A}\hat{b}=o_P\left(1\right),$ where $\hat{A}$ is defined analogously to A with F replaced by $\hat{F}$ which satisfies $T^{-1}\hat{F}'\hat{F}=I_{R_0}$. Then we have

$$N^{-1}\hat{b}'\hat{A}\hat{b}=N^{-1}\sum_{i=1}^N\left(\hat{\beta}_i-\beta_i^0\right)'\left(T^{-1}X_i'M_{\hat{F}}X_i\right)\left(\hat{\beta}_i-\beta_i^0\right)=o_P\left(1\right).$$

(ii) By (A.3)-(A.4), the Cauchy-Schwarz inequality and (i),

$$\begin{aligned}0 &\geq Q_{1NT,\kappa}^{(K_0)}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\hat{F})-Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,F^0\right)=\frac{1}{N}\sum_{i=1}^N S_{NT,i}(\hat{\beta}_i,\hat{F})+o_P\left(1\right)\\ &= \frac{1}{N}\hat{b}'\hat{A}\hat{b}+\frac{1}{NT}\text{tr}\left[\left(F^{0\prime}M_{\hat{F}}F^0\right)\left(\Lambda^{0\prime}\Lambda^0\right)\right]-\frac{2}{NT}\sum_{i=1}^N\hat{b}_i'X_i'M_{\hat{F}}F^0\lambda_i^0+o_P\left(1\right)\\ &\geq \frac{1}{N}\hat{b}'\hat{A}\hat{b}+\frac{1}{NT}\text{tr}\left[\left(F^{0\prime}M_{\hat{F}}F^0\right)\left(\Lambda^{0\prime}\Lambda^0\right)\right]-2\left\{\frac{1}{N}\hat{b}'\hat{A}\hat{b}\right\}^{1/2}\left\{\frac{1}{NT}\text{tr}\left[\left(F^{0\prime}M_{\hat{F}}F^0\right)\left(\Lambda^{0\prime}\Lambda^0\right)\right]\right\}^{1/2}+o_P\left(1\right)\\ &= o_P\left(1\right)+\frac{1}{NT}\text{tr}\left[\left(F^{0\prime}M_{\hat{F}}F^0\right)\left(\Lambda^{0\prime}\Lambda^0\right)\right]-2o_P\left(1\right)\left\{\frac{1}{NT}\text{tr}\left[\left(F^{0\prime}M_{\hat{F}}F^0\right)\left(\Lambda^{0\prime}\Lambda^0\right)\right]\right\}^{1/2}.\end{aligned}$$

It follows that $\frac{1}{NT}\text{tr}\left[\left(F^{0\prime}M_{\hat{F}}F^0\right)\left(\Lambda^{0\prime}\Lambda^0\right)\right]=o_P\left(1\right).$ As in Bai (2009, p.1265), this further implies that $\frac{1}{T}\text{tr}\left(F^{0\prime}M_{\hat{F}}F^0\right)=o_P\left(1\right)$ under Assumption A.1(ii), $\frac{1}{T}\hat{F}'F^0$ is invertible, and $\left\|P_{\hat{F}}-P_{F^0}\right\|=o_P\left(1\right).$

(iii) By Assumption A.1(iii) and the result in part (ii),

$$
\begin{aligned}
& \left| N^{-1} \sum_{i=1}^{N} \left(\hat{\beta}_i - \beta_i^0\right)' \left(T^{-1} X_i' \left(M_{F^0} - M_{\hat{F}}\right) X_i\right) \left(\hat{\beta}_i - \beta_i^0\right) \right| \\
\leq & \max_{1 \leq i \leq N} T^{-1} \left\| X_i' \left(P_{\hat{F}} - P_{F^0}\right) X_i \right\| N^{-1} \sum_{i=1}^{N} \left(\hat{\beta}_i - \beta_i^0\right)' \left(\hat{\beta}_i - \beta_i^0\right) \\
\leq & \max_{1 \leq i \leq N} T^{-1} \left\| X_i \right\|^2 \left\| P_{\hat{F}} - P_{F^0} \right\| N^{-1} \sum_{i=1}^{N} \left\| \hat{\beta}_i - \beta_i^0 \right\|^2 \\
= & \; o_P(1) \, N^{-1} \sum_{i=1}^{N} \left\| \hat{\beta}_i - \beta_i^0 \right\|^2 .
\end{aligned}
$$

This result, in conjunction with the result in part (i), implies that

$$
\begin{aligned}
o_P(1) &= N^{-1} \sum_{i=1}^{N} \left(\hat{\beta}_i - \beta_i^0\right)' \left(T^{-1} X_i' M_{\hat{F}} X_i\right) \left(\hat{\beta}_i - \beta_i^0\right) \\
&= N^{-1} \sum_{i=1}^{N} \left(\hat{\beta}_i - \beta_i^0\right)' \left(T^{-1} X_i' M_{F^0} X_i\right) \left(\hat{\beta}_i - \beta_i^0\right) \\
&\quad - N^{-1} \sum_{i=1}^{N} \left(\hat{\beta}_i - \beta_i^0\right)' \left(T^{-1} X_i' \left(M_{F^0} - M_{\hat{F}}\right) X_i\right) \left(\hat{\beta}_i - \beta_i^0\right) \\
&\geq \left[\min_{1 \leq i \leq N} \mu_{\min} \left(T^{-1} X_i' M_{F^0} X_i\right) - o_P(1) \right] N^{-1} \sum_{i=1}^{N} \left\| \hat{\beta}_i - \beta_i^0 \right\|^2 .
\end{aligned}
$$

Then by Assumption A.1(x), $N^{-1} \sum_{i=1}^{N} \left\| \hat{\beta}_i - \beta_i^0 \right\|^2 = o_P(1)$.

Now, we argue that Assumption A.1(xi) rules out the possibility that $N^{-1/2} \left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 \right\|$ is divergent so that it is sufficient to consider minimization of $Q_{1NT,\kappa}^{(K_0)}$ over $\boldsymbol{\beta} \in \mathcal{N}_L\left(\boldsymbol{\beta}^0\right)$. In the following analysis, we will show that $N^{-1} \left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 \right\|^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{b}_i' \hat{b}_i = O_P(1)$ without restricting $\boldsymbol{\beta} \in \mathcal{N}_L\left(\boldsymbol{\beta}^0\right)$ when we solve the minimization problem. Once this is shown, the results in (i)-(iii) can be used to conclude the proof of the theorem.

Suppose $\hat{F} \in \mathcal{F}$ and $\hat{\boldsymbol{\alpha}}$ are the estimates of F^0 and $\boldsymbol{\alpha}^0$, respectively. By (A.2)-(A.3) and Lemmas A.1(ii)-(iv)

$$
\begin{aligned}
& Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}, \hat{F}\right) - Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, F^0\right) \\
\geq & \; \frac{1}{N} \sum_{i \in \mathcal{S}_{N,\hat{F},\underline{c}}} \left[Q_{1NT,i}\left(\beta_i, \hat{F}\right) - Q_{1NT,i}\left(\beta_i^0, F^0\right) \right] + \frac{1}{N} \sum_{i \in \mathcal{S}_{N,\hat{F},\underline{c}}^*} \left[Q_{1NT,i}\left(\beta_i, \hat{F}\right) - Q_{1NT,i}\left(\beta_i^0, F^0\right) \right] \\
\geq & \; \frac{1}{N} \sum_{i \in \mathcal{S}_{N,\hat{F},\underline{c}}} \left[Q_{1NT,i}\left(\beta_i, \hat{F}\right) - Q_{1NT,i}\left(\beta_i^0, F^0\right) \right] - \frac{1}{N} \sum_{i \in \mathcal{S}_{N,\hat{F},\underline{c}}^*} Q_{1NT,i}\left(\beta_i^0, F^0\right) \\
\geq & \; \frac{1}{N} \sum_{i \in \mathcal{S}_{N,\hat{F},\underline{c}}} \left(S_{NT,i}(\beta_i, \hat{F}) - \frac{2}{T} b_i' X_i' M_{\hat{F}} \varepsilon_i \right) + o_P(1),
\end{aligned}
$$

where the last inequality follows because by Assumption A.1(xi)

$$\frac{1}{N}\sum_{i\in\mathcal{S}^*_{N,\hat{F},\underline{c}}} Q_{1NT,i}\left(\beta_i^0, F^0\right) = \frac{1}{N}\sum_{i\in\mathcal{S}^*_{N,\hat{F},\underline{c}}} \frac{\varepsilon_i' M_{F^0}\varepsilon_i}{T} \leq \frac{\#\mathcal{S}^*_{N,\hat{F},\underline{c}}}{N}\max_{1\leq i\leq N}\frac{\varepsilon_i' M_{F^0}\varepsilon_i}{T} = o_P(1).$$

Then following the proof in (i) and using the fact that $0 \geq Q_{1NT,\kappa}^{(K_0)}\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{F}\right) - Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, F^0\right)$, we can readily show that

$$\begin{aligned}
0 &\geq \frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i' X_i' M_{\hat{F}} X_i \hat{b}_i - \frac{2}{NT}\sum_{i\in\mathcal{S}^*_{N,\hat{F},\underline{c}}} \hat{b}_i' X_i' M_{\hat{F}}\left(F^0\lambda_i^0 + \varepsilon_i\right) + o_P(1) \\
&\quad + \frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \lambda_i^{0\prime} F^{0\prime} M_{\hat{F}} F^0 \lambda_i^0 \\
&\geq \frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i' X_i' M_{\hat{F}} X_i \hat{b}_i - \frac{2}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i' X_i' M_{\hat{F}}\left(F^0\lambda_i^0 + \varepsilon_i\right) + o_P(1) \\
&\geq \frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i' X_i' M_{\hat{F}} X_i \hat{b}_i - 2\left\{\frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i' X_i' M_{\hat{F}} X_i \hat{b}_i\right\}^{1/2}\left\{\Gamma_{NT}\right\}^{1/2} + o_P(1), \qquad \text{(A.6)}
\end{aligned}$$

where $\Gamma_{NT} = \frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}}\left(F^0\lambda_i^0 + \varepsilon_i\right)'\left(F^0\lambda_i^0 + \varepsilon_i\right) = O_P(1)$. It follows that $\frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i' X_i' M_{\hat{F}} X_i \hat{b}_i$ $= O_P(1)$, which further implies that $\frac{1}{N}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i'\hat{b}_i = O_P(1)$ under Assumption A.1(xi). But when $\frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i'\hat{b}_i = O_P(1)$, we have $\frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i' X_i' M_{\hat{F}}\varepsilon_i = o_P(1)$ by the same argument as used in the proof of Lemma A.1(i). This, in conjunction with the first inequality in (A.6), implies that

$$\begin{aligned}
0 &\geq \frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i' X_i' M_{\hat{F}} X_i \hat{b}_i - \frac{2}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \hat{b}_i' X_i' M_{\hat{F}} F^0\lambda_i^0 + \frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \lambda_i^{0\prime} F^{0\prime} M_{\hat{F}} F^0\lambda_i^0 + o_P(1) \\
&= N^{-1}\tilde{b}'\tilde{A}b + N^{-1}\hat{\eta}'\tilde{B}\hat{\eta} - 2N^{-1}\tilde{b}'\tilde{C}'\hat{\eta} \qquad \text{(A.7)}
\end{aligned}$$

where $\hat{\eta} = \frac{1}{T^{1/2}}\text{vec}(M_{\hat{F}}F^0)$, $\tilde{b} = (\hat{b}_i, i\in\mathcal{S}_{N,\hat{F},\underline{c}})$ is a subvector of $\hat{b}$ associated with $i\in\mathcal{S}_{N,\hat{F},\underline{c}}$, $\tilde{A} = \text{diag}(T^{-1} X_i' M_{\hat{F}} X_i,\ i\in\mathcal{S}_{N,\hat{F},\underline{c}})$, and $\tilde{C} = T^{-1/2}((\lambda_i^0 \otimes M_{\hat{F}} X_1),\ i\in\mathcal{S}_{N,\hat{F},\underline{c}})$. Following the above arguments in (i), we can readily show that $N^{-1}\tilde{b}'\tilde{A}b = o_p(1)$. Then using the same argument as used in (ii) but restricting our attention to the summation over $i\in\mathcal{S}_{N,\hat{F},\underline{c}}$, we can show that $\frac{1}{NT}\sum_{i\in\mathcal{S}_{N,\hat{F},\underline{c}}} \lambda_i^{0\prime} F^{0\prime} M_{\hat{F}} F^0\lambda_i^0 = o_P(1)$, implying that $\frac{1}{T}\text{tr}\left(F^{0\prime} M_{\hat{F}} F^0\right) = o_P(1)$ under Assumptions A.1(ii) and (xi), $\frac{1}{T}\hat{F}'F^0$ is invertible, and $\left\|P_{\hat{F}} - P_{F^0}\right\| = o_P(1)$.

Again, by (A.2) and Lemmas A.1(ii)-(iv) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
0 &\geq Q_{1NT,\kappa}^{(K_0)}\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{F}\right) - Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, F^0\right) \\
&\geq \frac{1}{N}\sum_{i=1}^N\left[Q_{1NT,i}\left(\hat{\beta}_i, \hat{F}\right) - Q_{1NT,i}\left(\beta_i^0, F^0\right)\right] \\
&= \frac{1}{N}\sum_{i=1}^N\left[S_{NT,i}\left(\hat{\beta}_i, \hat{F}\right) - \frac{2}{T}\hat{b}_i' X_i' M_{\hat{F}}\varepsilon_i\right] + o_P(1) \\
&= \frac{1}{N}\sum_{i=1}^N\left[\frac{1}{T}\hat{b}_i' X_i' M_{\hat{F}} X_i \hat{b}_i + \frac{1}{T}\lambda_i^{0\prime} F^{0\prime} M_{\hat{F}} F^0\lambda_i^0 - \frac{2}{T}\hat{b}_i' X_i' M_{\hat{F}} F^0\lambda_i^0 - \frac{2}{T}\hat{b}_i' X_i' M_{\hat{F}}\varepsilon_i\right] + o_P(1)
\end{aligned}$$

$$
\begin{aligned}
&\geq \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\hat{b}_i'X_i'M_{\hat{F}}X_i\hat{b}_i-\frac{2}{T}\hat{b}_i'X_i'M_{\hat{F}}F^0\lambda_i^0-\frac{2}{T}\hat{b}_i'X_i'M_{\hat{F}}\varepsilon_i\right]+o_P(1)\\
&\geq \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\hat{b}_i'X_i'M_{\hat{F}}X_i\hat{b}_i-2\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\hat{b}_i'X_i'M_{\hat{F}}X_i\hat{b}_i\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}M_{\hat{F}}F^0\lambda_i^0\right\}^{1/2}\\
&\quad -2\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\hat{b}_i'X_i'M_{\hat{F}}X_i\hat{b}_i\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i'\varepsilon_i\right\}^{1/2}+o_P(1).
\end{aligned}
$$

This, in conjunction with the fact that $\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}M_{\hat{F}}F^0\lambda_i^0=o_P(1)$ and $\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i'\varepsilon_i=O_P(1)$, implies that $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\hat{b}_i'X_i'M_{\hat{F}}X_i\hat{b}_i=O_P(1)$. Noting that

$$
\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\hat{b}_i'X_i'M_{\hat{F}}X_i\hat{b}_i\geq \min_{1\leq i\leq N}\mu_{\min}\left(\frac{1}{T}X_i'M_{\hat{F}}X_i\right)\frac{1}{N}\sum_{i=1}^{N}\hat{b}_i'\hat{b}_i
$$

and

$$
\begin{aligned}
\mu_{\min}\left(\frac{1}{T}X_i'M_{\hat{F}}X_i\right) &= \mu_{\min}\left(\frac{1}{T}X_i'M_{F^0}X_i+\frac{1}{T}X_i'\left(M_{\hat{F}}-M_{F^0}\right)X_i\right)\\
&\geq \mu_{\min}\left(\frac{1}{T}X_i'M_{F^0}X_i\right)-\frac{1}{T}\left\|X_i'\left(M_{\hat{F}}-M_{F^0}\right)X_i\right\|\\
&\geq \mu_{\min}\left(\frac{1}{T}X_i'M_{F^0}X_i\right)-\frac{1}{T}\left\|X_i\right\|^2\left\|P_{\hat{F}}-P_{F^0}\right\|\\
&\geq \underline{c}_{XX}-o_P(1)\ \text{ w.p.a.1,}
\end{aligned}
$$

we have $\frac{1}{N}\sum_{i=1}^{N}\hat{b}_i'\hat{b}_i=O_P(1)$. This implies that for any $\epsilon>0$, there exists an $L=L(\epsilon)$ such that the minimizer $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ satisfies $\hat{\beta}\in\mathcal{N}_L\left(\boldsymbol{\beta}^0\right)$ with probability at least $1-\epsilon$. By the the above analyses in (i)-(iii), we have completed the proof of Theorem 3.1. ■

To prove Theorem 3.2, we need the following two lemmas.

Lemma A.2 *Suppose Assumptions A.1-A.2 hold. Then*

(i) $T^{-1}\left\|\hat{F}-F^0H\right\|^2=O_P\left(\eta_{NT}^2\right)+O_P\left(\delta_{NT}^{-2}\right)$,

(ii) $T^{-1}(\hat{F}-F^0H)'F^0H=O_P\left(\eta_{NT}\right)+O_P\left(\delta_{NT}^{-2}\right)$,

(iii) $T^{-1}(\hat{F}-F^0H)'\hat{F}=O_P\left(\eta_{NT}\right)+O_P\left(\delta_{NT}^{-2}\right)$,

(iv) $T^{-1}(\hat{F}'\hat{F}-H'F^{0\prime}F^0H)=I_{R_0}-T^{-1}H'F^{0\prime}F^0H=O_P\left(\eta_{NT}\right)+O_P\left(\delta_{NT}^{-2}\right)$,

(v) $\left\|P_{\hat{F}}-P_{F^0H}\right\|=O_P\left(\eta_{NT}\right)+O_P\left(\delta_{NT}^{-1}\right)$.

Lemma A.3 *Let* $R_{1i}\equiv\frac{1}{T}X_i'(M_{\hat{F}}-M_{F^0})\varepsilon_i$ *and* $R_{2i}\equiv\frac{1}{T}X_i'M_{\hat{F}}(\hat{F}-F^0H)H^{-1}\lambda_i^0+\frac{1}{NT}\sum_{j=1}^{N}X_i'M_{\hat{F}}X_j\nu_{ji}\hat{b}_j$. *Suppose Assumptions A.1-A.2 hold. Then*

(i) $R_{1i}=O_P\left(\varsigma_{NT}\right)$ *for each* $i=1,...,N,$ *and* $N^{-1}\sum_{i=1}^{N}\left\|R_{1i}\right\|^2=O_P\left(\varsigma_{NT}^2\right)$,

(ii) $R_{2i}=O_P\left(\varsigma_{NT}\right)$ *for each* $i=1,...,N,$ *and* $N^{-1}\sum_{i=1}^{N}\left\|R_{2i}\right\|^2=O_P\left(\varsigma_{NT}^2\right)$,

where $\varsigma_{NT}=O_P\left(\eta_{NT}^2\right)+O_P\left(T^{-1/2}\eta_{NT}\right)+O_P\left(\delta_{NT}^{-2}\right)$.

Proof of Theorem 3.2. (i) We invoke subdifferential calculus (e.g., Bertsekas (1995, Appendix B.5)). A necessary condition for $\{\hat{\beta}_i\}$, $\hat{F}$, and $\{\hat{\alpha}_k\}$ to minimize the objective function in (2.5) is that for each

$i = 1, ..., N$ (resp. $k = 1, ..., K_0$), $\mathbf{0}_{p\times 1}$ belongs to the subdifferential of $Q^{(K_0)}_{1NT,\kappa}(\boldsymbol{\beta}, F, \boldsymbol{\alpha})$ with respect to β_i (resp. α_k) evaluated at $\{\hat{\beta}_i\}$, $\hat{F}$, and $\{\hat{\alpha}_k\}$. That is, for each $i = 1, ..., N$ and $k = 1, ..., K_0$, we have

$$\mathbf{0}_{p\times 1} = -\frac{2}{T} X_i' M_{\hat{F}} (Y_i - X_i \hat{\beta}_i) + \kappa \sum_{j=1}^{K_0} \hat{e}_{ij} \Pi^{K_0}_{l=1, l\neq j} \left\| \hat{\beta}_i - \hat{\alpha}_l \right\|, \tag{A.8}$$

where $\hat{e}_{ij} = \frac{\hat{\beta}_i - \hat{\alpha}_j}{\|\hat{\beta}_i - \hat{\alpha}_j\|}$ if $\left\|\hat{\beta}_i - \hat{\alpha}_j\right\| \neq 0$ and $\|\hat{e}_{ij}\| \leq 1$ if $\left\|\hat{\beta}_i - \hat{\alpha}_j\right\| = 0$. Let $\hat{Q}_{X_i X_i} = \frac{1}{T} X_i' M_{\hat{F}} X_i$. Noting that $Y_i = X_i \beta^0_i + \hat{F} H^{-1} \lambda^0_i + \varepsilon_i + (F^0 - \hat{F} H^{-1}) \lambda^0_i$, (A.8) implies that

$$\hat{Q}_{X_i X_i} (\hat{\beta}_i - \beta^0_i) = \frac{1}{T} X_i' M_{\hat{F}} \varepsilon_i + \frac{1}{T} X_i' M_{\hat{F}} (F^0 - \hat{F} H^{-1}) \lambda^0_i - \frac{\kappa}{2} \sum_{j=1}^{K_0} \hat{e}_{ij} \Pi^{K_0}_{l=1, l\neq j} \left\| \hat{\beta}_i - \hat{\alpha}_l \right\|. \tag{A.9}$$

Using the definition of $\hat{b}_i (= \hat{\beta}_i - \beta^0_i)$, (A.9) can be rewritten as

$$\hat{Q}_{X_i X_i} \hat{b}_i = \frac{1}{NT} \sum_{j=1}^{N} X_i' M_{\hat{F}} X_j \nu_{ij} \hat{b}_j + R_i, \tag{A.10}$$

where $R_i = \frac{1}{T} X_i' M_{F^0} \varepsilon_i + R_{1i} - R_{2i} + R_{3i}$, where R_{1i} and R_{2i} are defined in Lemma A.3, and $R_{3i} = -\frac{\kappa}{2} \sum_{j=1}^{K_0} \hat{e}_{ij} \Pi^{K_0}_{l=1, l\neq j} \left\| \hat{\beta}_i - \hat{\alpha}_l \right\|$. Noting that $\|X_i' M_{F^0} \varepsilon_i\|^2 \leq 2 \|X_i' \varepsilon_i\|^2 + 2 \|X_i' P_{F^0} \varepsilon_i\|^2$ and $\|X_i' P_{F^0} \varepsilon_i\|^2 = \mathrm{tr}(X_i X_i' P_{F^0} \varepsilon_i \varepsilon_i' P_{F^0}) \leq \mathrm{tr}(X_i X_i') \mathrm{tr}(P_{F^0} \varepsilon_i \varepsilon_i' P_{F^0}) = \mathrm{tr}(X_i X_i') \mathrm{tr}(\varepsilon_i' P_{F^0} \varepsilon_i) \leq \left[\mu_{\min}\left(T^{-1} F^{0\prime} F^0\right)\right]^{-1} T^{-1} \|X_i\|^2 \times \left\|F^0 \varepsilon_i\right\|^2$, we can show that $\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} X_i' M_{F^0} \varepsilon_i \right\|^2 = O_P\left(T^{-1}\right)$ by Assumptions A.1(i), (iii), and (vii), and

$$\frac{1}{N} \sum_{i=1}^{N} \|R_{3i}\|^2 \leq \frac{\kappa^2 K_0}{4N} \sum_{j=1}^{K_0} \sum_{i=1}^{N} \left\{ \Pi^{K_0}_{l=1, l\neq j} \left\| \hat{\beta}_i - \hat{\alpha}_l \right\| \right\}^2 = O_P\left(\kappa^2\right).$$

Then $N^{-1} \sum_{i=1}^{N} \|R_i\|^2 = O_P\left(\eta^4_{NT}\right) + O_P\left(T^{-1} + \delta^{-4}_{NT} + \kappa^2\right)$ by Lemma A.3. Let $\hat{Q}_1 = \mathrm{diag}(\hat{Q}_{X_1 X_1}, ..., \hat{Q}_{X_N X_N})$. Define $\hat{Q}_2$ as an $Np \times Np$ matrix with typical blocks $\frac{1}{NT} X_i' M_{\hat{F}} X_j \nu_{ij}$. Let $\mathbf{R} = (R_1', ..., R_N')'$. Then (A.10) implies that

$$\left(\hat{Q}_1 - \hat{Q}_2\right) \hat{b} = \mathbf{R}. \tag{A.11}$$

It follows that

$$\|\mathbf{R}\|^2 = \mathrm{tr}\left[\hat{b}' (\hat{Q}_1 - \hat{Q}_2)(\hat{Q}_1 - \hat{Q}_2) \hat{b}\right] \geq \left\|\hat{b}\right\|^2 \left[\mu_{\min}(\hat{Q}_1 - \hat{Q}_2)\right]^2.$$

By Weyl's inequality and analogous arguments as used to deduce (A.5), we have w.p.a.1

$$\begin{aligned} \mu_{\min}(\hat{Q}_1 - \hat{Q}_2) &\geq \mu_{\min}(\hat{Q}_1) - \mu_{\max}(\hat{Q}_2) = \mu_{\min}(\hat{Q}_1) - o_P\left(N^{-3/4}\right) \\ &\geq \min_{1\leq i\leq N} \mu_{\min}(\hat{Q}_{X_i X_i})/2 \geq \underline{c}_{XX}/2 > 0, \end{aligned}$$

where we use the fact that

$$\begin{aligned} \mu_{\min}(\hat{Q}_{X_i X_i}) &= \mu_{\min}\left(\frac{1}{T} X_i' M_{F^0} X_i + \frac{1}{T} X_i' \left(M_{\hat{F}} - M_{F^0}\right) X_i\right) \\ &\geq \mu_{\min}\left(\frac{1}{T} X_i' M_{F^0} X_i\right) - \frac{1}{T} \left\| X_i' \left(M_{\hat{F}} - M_{F^0}\right) X_i \right\| \\ &\geq \underline{c}_{XX} - \frac{1}{T} \|X_i\|^2 \left\|P_{\hat{F}} - P_{F^0}\right\| \\ &= \underline{c}_{XX} - o_P(1) \ \text{ w.p.a.1} \end{aligned}$$

by Theorem 3.1(ii) and Assumptions A.1(iii) and (x). It follows that $\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{b}_i\right\|^2=\frac{1}{N}\left\|\hat{b}\right\|^2\leq 4\underline{c}_{XX}^{-2}\frac{1}{N}\sum_{i=1}^{N}\left\|R_i\right\|^2=O_P\left(\eta_{NT}^4\right)+O_P\left(T^{-1}+\delta_{NT}^{-4}+\kappa^2\right)=O_P(T^{-1}+\kappa^2)$ by Assumptions A.2(ii).

Next, we strengthen the above result to obtain $\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{b}_i\right\|^2=O_P\left(T^{-1}\right)$. Let $\boldsymbol{\beta}=\boldsymbol{\beta}^0+T^{-1/2}\mathbf{v}$, where $\mathbf{v}=(v_1,...,v_N)$ is a $p\times N$ matrix. We want to show that for any given $\epsilon^*>0$, there exists a large constant $L=L\left(\epsilon^*\right)$ such that for sufficiently large N and T we have

$$P\left\{\inf_{N^{-1}\sum_{i=1}^{N}\|v_i\|^2=L}Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0+T^{-1/2}\mathbf{v},\hat{\boldsymbol{\alpha}},\hat{F}\right)>Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\hat{F}\right)\right\}\geq 1-\epsilon^*, \tag{A.12}$$

where $\hat{\boldsymbol{\alpha}}\equiv\hat{\boldsymbol{\alpha}}\left(\mathbf{v}\right)$ and $\hat{F}\equiv\hat{F}\left(\mathbf{v}\right)$ are chosen such that $(\boldsymbol{\beta}^0+T^{-1/2}\mathbf{v},\hat{\boldsymbol{\alpha}},\hat{F})$ minimizes $Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta},\boldsymbol{\alpha},F\right)$ for some given $\mathbf{v}$. This implies that w.p.a.1 there is a local minimum $\{\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}}\}$ such that $N^{-1}\sum_{i=1}^{N}||\hat{b}_i||^2=O_P\left(T^{-1}\right)$. By Lemmas A.2(v) and A.3(ii)

$$\begin{aligned}
& T\left[Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0+T^{-1/2}\mathbf{v},\hat{\boldsymbol{\alpha}},\hat{F}\right)-Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\hat{F}\right)\right]\\
\geq\ & T\left[Q_{1NT}\left(\boldsymbol{\beta}^0+T^{-1/2}\mathbf{v},\hat{F}\right)-Q_{1NT}\left(\boldsymbol{\beta}^0,\hat{F}\right)\right]\\
=\ & \frac{1}{N}\sum_{i=1}^{N}\left\{\left[Y_i-X_i\left(\beta_i^0+T^{-1/2}v_i\right)\right]'M_{\hat{F}}\left[Y_i-X_i\left(\beta_i^0+T^{-1/2}v_i\right)\right]-\left[Y_i-X_i\beta_i^0\right]'M_{\hat{F}}\left[Y_i-X_i\beta_i^0\right]\right\}\\
=\ & \frac{1}{NT}\sum_{i=1}^{N}v_i'X_i'M_{\hat{F}}X_iv_i-\frac{2}{N\sqrt{T}}\sum_{i=1}^{N}v_i'X_i'M_{\hat{F}}\left[Y_i-X_i\beta_i^0\right]\\
=\ & \frac{1}{NT}\sum_{i=1}^{N}v_i'X_i'M_{\hat{F}}X_iv_i+\frac{2\sqrt{T}}{N}\sum_{i=1}^{N}\frac{1}{T}v_i'X_i'M_{\hat{F}}\left(\hat{F}-F^0H\right)H^{-1}\lambda_i^0-\frac{2}{N\sqrt{T}}\sum_{i=1}^{N}v_i'X_i'M_{\hat{F}}\varepsilon_i\\
=\ & \frac{1}{NT}\sum_{i=1}^{N}v_i'X_i'M_{\hat{F}}X_iv_i-\frac{2}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}v_i'\left(\frac{1}{NT}X_i'M_{\hat{F}}X_j\nu_{ji}\right)v_j+\frac{2\sqrt{T}}{N}\sum_{i=1}^{N}v_i'R_{2i}-\frac{2}{N\sqrt{T}}\sum_{i=1}^{N}v_i'X_i'M_{\hat{F}}\varepsilon_i\\
=\ & \left\{\frac{1}{NT}\sum_{i=1}^{N}v_i'X_i'M_{F^0}X_iv_i-\frac{2}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}v_i'X_i'M_{F^0}X_j\nu_{ji}v_j\right\}\left\{1+o_P\left(1\right)\right\}+\frac{2\sqrt{T}}{N}\sum_{i=1}^{N}v_i'R_{2i}\\
& -\frac{2}{N\sqrt{T}}\sum_{i=1}^{N}v_i'X_i'M_{\hat{F}}\varepsilon_i\\
=\ & \frac{1}{N}v'\left(A^0-J^0\right)v\left\{1+o_P\left(1\right)\right\}+\frac{2\sqrt{T}}{N}\sum_{i=1}^{N}v_i'R_{2i}-\frac{2}{N\sqrt{T}}\sum_{i=1}^{N}\varepsilon_i'M_{\hat{F}}X_iv_i,
\end{aligned}$$

where $v=\text{vec}(\mathbf{v})$, A^0 is defined analogously to A with F replaced by F^0, J^0 is an $Np\times Np$ matrix with a typical $p\times p$ block submatrix $2\left(NT\right)^{-1}X_i'M_{F^0}X_j\nu_{ji}$. As in the proof of Theorem 3.1, we can show that w.p.a.1

$$\begin{aligned}
& \mu_{\min}\left(A^0-J^0\right)\\
\geq\ & \mu_{\min}\left(A^0\right)-\mu_{\max}\left(J^0\right)\\
\geq\ & \min_{1\leq i\leq N}\mu_{\min}\left(T^{-1}X_i'M_{F^0}X_i\right)-4\max_{1\leq i\leq N}\mu_{\max}\left(\left(NT\right)^{-1}X_i'M_{F^0}X_i\nu_{ii}\right)\\
\geq\ & \min_{1\leq i\leq N}\mu_{\min}\left(T^{-1}X_i'M_{F^0}X_i\right)-4N^{-1}\max_{1\leq i\leq N}\left\|\lambda_i^0\right\|^2\left[\mu_{\min}\left(N^{-1}\Lambda^0\Lambda^0\right)\right]^{-1}\max_{1\leq i\leq N}\mu_{\max}\left(T^{-1}X_i'M_{F^0}X_i\right)\\
\geq\ & \underline{c}_{XX}/2,
\end{aligned}$$

where we use the fact that $\nu_{ii}=\lambda_i^{0\prime}\left(N^{-1}\Lambda^0\Lambda^0\right)^{-1}\lambda_i^0\leq\left[\mu_{\min}\left(N^{-1}\Lambda^0\Lambda^0\right)\right]^{-1}\left\|\lambda_i^0\right\|^2$ and that $\max_{1\leq i\leq N}\left\|\lambda_i^0\right\|^2 = o\left(N^{1/4}\right)$ under Assumption A.1(ii). Then $\frac{1}{N}v'\left(A^0-J^0\right)v\geq\frac{\underline{c}_{XX}}{2N}\|v\|^2$ w.p.a.1. In addition, by the Cauchy-Schwarz inequality,

$$\frac{\sqrt{T}}{N}\sum_{i=1}^N v_i'R_{2i}\leq\left\{\frac{1}{N}\sum_{i=1}^N\|v_i\|^2\right\}^{1/2}\left\{\frac{T}{N}\sum_{i=1}^N\|R_{2i}\|^2\right\}^{1/2}=N^{-1/2}\|v\|\,o_P(1),$$

and

$$\begin{aligned}\frac{1}{N\sqrt{T}}\sum_{i=1}^N v_i'X_i'M_{\hat{F}}\varepsilon_i &\leq \left\{\frac{1}{N}\sum_{i=1}^N\|v_i\|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^N\varepsilon_i'M_{\hat{F}}X_iX_i'M_{\hat{F}}\varepsilon_i\right\}^{1/2}\\ &= N^{-1/2}\|v\|\,C_1^{1/2}\left\{1+o_P(1)\right\},\end{aligned}$$

where $C_1=\text{plim}_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^N\varepsilon_i'M_{F^0}X_iX_i'M_{F^0}\varepsilon_i$, and we use the fact that $\frac{T}{N}\sum_{i=1}^N\|R_{2i}\|^2=o_P(1)$ by Lemma A.3(ii), the above analysis, and Assumptions A.2(ii)-(iii). It follows that

$$\begin{aligned}&T\left[Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0+T^{-1/2}\mathbf{v},\hat{\boldsymbol{\alpha}},\hat{F}\right)-Q_{1NT,\kappa}^{(K_0)}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\hat{F}\right)\right]\\ \geq\;&\frac{\mu_{\min}\left(A^0-J^0\right)}{N}\|v\|^2\left\{1+o_P(1)\right\}-2N^{-1/2}\|v\|\,C_1^{1/2}\left\{1+o_P(1)\right\}+o_P(1),\end{aligned}$$

where $\mu_{\min}\left(A^0-J^0\right)\geq\underline{c}_{XX}/2>0$ w.p.a.1. The first term dominates the second term in the last display for sufficiently large L. That is $T[Q_{1NT,\kappa}^{(K_0)}(\boldsymbol{\beta}^0+T^{-1/2}\mathbf{v},\hat{\boldsymbol{\alpha}},\hat{F})-Q_{1NT,\kappa}^{(K_0)}(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\hat{F})]>0$ for sufficiently large L. Consequently, the minimizer $\hat{\boldsymbol{\beta}}$ must satisfy $N^{-1}\sum_{i=1}^N||\hat{b}_i||^2=O_P\left(T^{-1}\right)$.

(ii) By Lemma A.3(i), the fact that $\frac{1}{T}X_i'M_{F^0}\varepsilon_i=O_P\left(T^{-1/2}\right)$ under Assumptions A.1(i) and (vii), and the result in (i), we have

$$\frac{1}{T}X_i'M_{\hat{F}}\varepsilon_i=O_P\left(\eta_{NT}^2\right)+O_P\left(T^{-1/2}\eta_{NT}\right)+O_P\left(\delta_{NT}^{-2}\right)+O_P\left(T^{-1/2}\right)=O_P\left(T^{-1/2}\right).$$

By Lemma A.3(ii), the fact that $\frac{1}{NT}\sum_{j=1}^N X_i'M_{\hat{F}}X_j\nu_{ji}\hat{b}_j=O_P\left(\eta_{NT}\right)$, and (i),

$$\frac{1}{T}X_i'M_{\hat{F}}(F^0-\hat{F}H^{-1})\lambda_i^0=O_P\left(\eta_{NT}\right)+O_P\left(T^{-1/2}\eta_{NT}\right)+O_P\left(\delta_{NT}^{-2}\right)=O_P\left(T^{-1/2}\right).$$

The last term on the right hand side of (A.9) is $O_P(\kappa)$. In addition, $\hat{Q}_{X_iX_i}$ is positive definite w.p.a.1. It follows from the above results and (A.9) that $\hat{\beta}_i-\beta_i^0=O_P\left(T^{-1/2}+\kappa\right)$.

(iii) Let $P_{NT}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)=\frac{1}{N}\sum_{i=1}^N\Pi_{k=1}^{K_0}\|\beta_i-\alpha_k\|$ and $\hat{c}_{iNT}\left(\boldsymbol{\alpha}\right)=\Pi_{k=1}^{K_0-1}\left\|\hat{\beta}_i-\alpha_k\right\|+\Pi_{k=1}^{K_0-2}\left\|\hat{\beta}_i-\alpha_k\right\|\times\left\|\beta_i^0-\alpha_{K_0}\right\|+...+\Pi_{k=2}^{K_0}\left\|\beta_i^0-\alpha_k\right\|$. By the repeated use of Minkowski's inequality (see, e.g., Su, Shi, and Phillips (2016, SSP hereafter)) and (ii), we have that as $(N,T)\to\infty$,

$$\left|\Pi_{k=1}^{K_0}\left\|\hat{\beta}_i-\alpha_k\right\|-\Pi_{k=1}^{K_0}\left\|\hat{\beta}_i^0-\alpha_k\right\|\right|\leq\hat{c}_{iNT}\left(\boldsymbol{\alpha}\right)\left\|\beta_i-\beta_i^0\right\|\text{ and }\hat{c}_{iNT}\left(\boldsymbol{\alpha}\right)\leq C_{K_0NT}\left(\boldsymbol{\alpha}\right)\left(1+2\left\|\hat{\beta}_i-\beta_i^0\right\|\right),$$

where $C_{K_0NT}\left(\boldsymbol{\alpha}\right)=\max_{1\leq i\leq N}\max_{1\leq s\leq k\leq K_0-1}\Pi_{k=1}^s a_{ks}\left\|\beta_i^0-\alpha_k\right\|^{K_0-1-s}=\max_{1\leq l\leq K_0}\max_{1\leq s\leq k\leq K_0-1}\Pi_{k=1}^s a_{ks}\left\|\alpha_l^0-\alpha_k\right\|^{K_0-1-s}=O(1)$ and a_{ks}'s are finite integers. It follows that as $(N,T)\to\infty$,

$$\begin{aligned}\left|P_{NT}(\hat{\boldsymbol{\beta}},\boldsymbol{\alpha})-P_{NT}(\boldsymbol{\beta}^0,\boldsymbol{\alpha})\right| &\leq C_{K_0NT}\left(\boldsymbol{\alpha}\right)\frac{1}{N}\sum_{i=1}^N\left\|\hat{b}_i\right\|+2C_{K_0NT}\left(\boldsymbol{\alpha}\right)\frac{1}{N}\sum_{i=1}^N\left\|\hat{b}_i\right\|^2\\ &\leq C_{K_0NT}\left(\boldsymbol{\alpha}\right)\left\{\frac{1}{N}\sum_{i=1}^N\left\|\hat{b}_i\right\|^2\right\}^{1/2}+O_P\left(T^{-1}\right)=O_P(T^{-1/2}).\qquad\text{(A.13)}\end{aligned}$$

By (A.13) and the fact that $P_{NT}\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) = 0$, we have

$$
\begin{aligned}
0 &\geq P_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - P_{NT}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}^0) = P_{NT}\left(\boldsymbol{\beta}^0, \hat{\boldsymbol{\alpha}}\right) - P_{NT}\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) + O_P(T^{-1/2}) \\
&= \frac{1}{N}\sum_{i=1}^{N} \Pi_{k=1}^{K_0} \left\|\beta_i^0 - \hat{\alpha}_k\right\| + O_P(T^{-1/2}) \\
&= \frac{N_1}{N}\Pi_{k=1}^{K_0}\left\|\hat{\alpha}_k - \alpha_1^0\right\| + ... + \frac{N_{K_0}}{N}\Pi_{k=1}^{K_0}\left\|\hat{\alpha}_k - \alpha_{K_0}^0\right\| + O_P(T^{-1/2}). \qquad \text{(A.14)}
\end{aligned}
$$

Then by Assumption A.2(i), we have $\Pi_{k=1}^{K_0}\left\|\hat{\alpha}_k - \alpha_l^0\right\| = O_P\left(T^{-1/2}\right)$ for $l = 1, ..., K_0$. It follows that $\left(\hat{\alpha}_{(1)}, ..., \hat{\alpha}_{(K_0)}\right) - (\alpha_1^0, ..., \alpha_{K_0}^0) = O_P\left(T^{-1/2}\right)$ for some suitable permutation $\left(\hat{\alpha}_{(1)}, ..., \hat{\alpha}_{(K_0)}\right)$ of $(\hat{\alpha}_1, ..., \hat{\alpha}_{K_0})$. ■

To prove Theorems 3.3 and 3.4, we need the following lemma.

Lemma A.4 *Suppose Assumptions A.1-A.3 hold. Then*

(i) $T^{-1}\varepsilon_i'(\hat{F} - F^0H) = O_P\left(\delta_{NT}^{-2}\right)$ *and* $P\left(\max_{1\leq i\leq N}\left\|T^{-1}\varepsilon_i'(\hat{F} - F^0H)\right\| \geq C\delta_{NT}^{-2}(\ln T)^{\upsilon_1}\right) = o\left(N^{-1}\right)$,

(ii) $T^{-1/2}X_i'(\hat{F} - F^0H) = O_P(1)$ *and* $P\left(\max_{1\leq i\leq N}\left\|T^{-1/2}X_i'(\hat{F} - F^0H)\right\| \geq C(\ln T)^{\upsilon_1}\right) = o\left(N^{-1}\right)$,

(iii) $T^{-1}X_i'M_{\hat{F}}\varepsilon_i = T^{-1}X_i'M_{F^0}\varepsilon_i + O_P\left(\delta_{NT}^{-2}\right)$ *and* $P(\max_{1\leq i\leq N}\|R_{1i}\| \geq C\delta_{NT}^{-2}(\ln T)^{\upsilon_1}) = o\left(N^{-1}\right)$,

(iv) $T^{-1/2}X_i'M_{\hat{F}}F^0\lambda_i^0 = O_P(1)$ *and* $P(\max_{1\leq i\leq N}\left\|T^{-1/2}X_i'M_{\hat{F}}F^0\lambda_i^0\right\| \geq C\left(\ln T\right)^{\upsilon_1}c_{\lambda,N}) = o\left(N^{-1}\right)$,

(v) $P\left(\max_{1\leq i\leq N}\left\|T^{-1}X_i'\left(M_{\hat{F}} - M_{F^0}\right)X_i\right\| \geq C\delta_{NT}^{-1}(\ln T)^{\upsilon_2}\right) = o\left(N^{-1}\right)$,

(vi) $P\left(\max_{1\leq i\leq N}\|R_{2i}\| \geq C\delta_{NT}^{-2}(\ln T)^{\upsilon_1}\right) = o\left(N^{-1}\right)$,

where C is any positive finite constant that does not depend on N and T and R_{1i} and R_{2i} are as defined in Lemma A.3.

Proof of Theorem 3.3 Let $\hat{b}$ and $\hat{b}_i$ be as defined in the proof of Theorem 3.1. Let $\hat{Q}_1$, $\hat{Q}_2$, $\mathbf{R}$, and R_i be as defined in the proof of Theorem 3.2. Let $S_i = (0_{p\times p}, ..., 0_{p\times p}, I_p, 0_{p\times p}, ..., 0_{p\times p})$ be a $p \times Np$ selection matrix such that $S_i\hat{b} = \hat{b}_i$ and $S_i\mathbf{R} = R_i$. Then $\hat{b}_i = S_i\left(\hat{Q}_1 - \hat{Q}_2\right)^{-1}\mathbf{R}$ by (A.11). It follows that

$$
\begin{aligned}
\left\|\hat{b}_i\right\|^2 &= \operatorname{tr}\left(S_i'S_i\left(\hat{Q}_1 - \hat{Q}_2\right)^{-1}RR'\left(\hat{Q}_1 - \hat{Q}_2\right)^{-1}\right) \\
&= \left(\operatorname{vec}\left(S_i'S_i\right)\right)'\left(\left(\hat{Q}_1 - \hat{Q}_2\right)^{-1} \otimes \left(\hat{Q}_1 - \hat{Q}_2\right)^{-1}\right)\operatorname{vec}\left(\mathbf{RR}'\right) \\
&\leq \mu_{\max}\left(\left(\hat{Q}_1 - \hat{Q}_2\right)^{-1} \otimes \left(\hat{Q}_1 - \hat{Q}_2\right)^{-1}\right)\left(\operatorname{vec}\left(S_i'S_i\right)\right)'\operatorname{vec}\left(\mathbf{RR}'\right) \\
&= \left[\mu_{\min}\left(\hat{Q}_1 - \hat{Q}_2\right)\right]^{-2}\operatorname{tr}\left(S_i'S_i\mathbf{RR}'\right) \\
&= \left[\mu_{\min}\left(\hat{Q}_1 - \hat{Q}_2\right)\right]^{-2}\|R_i\|^2,
\end{aligned}
$$

where we use the fact that $\operatorname{tr}(B_1B_2B_3B_4) = \operatorname{vec}(B_1)'(B_2 \otimes B_4')\operatorname{vec}(B_3')$ for conformable matrices B_1, B_2, B_3, and B_4 (e.g., Bernstein (2005, p.253)) and $\operatorname{tr}(B_1B_3) = \operatorname{vec}(B_1)'\operatorname{vec}(B_3')$ (e.g., Bernstein (2005, p.247)). Following the arguments used in the proof of Theorem 3.1, we can show that

$$
\begin{aligned}
\mu_{\min}(\hat{Q}_1 - \hat{Q}_2) &\geq \mu_{\min}(\hat{Q}_1) - \mu_{\max}(\hat{Q}_2) \\
&\geq \min_{1\leq i\leq N}\mu_{\min}(\hat{Q}_{X_iX_i}) - 2\underline{c}_{\Lambda}^{-1}\max_{1\leq i\leq N}N^{-1}\left\|\lambda_i^0\right\|^2\max_{1\leq i\leq N}T^{-1}\left\|X_i\right\|^2.
\end{aligned}
$$

Under Assumptions A.3(i) and (iv), we have $P(\max_{1\le i\le N} N^{-1}\left\|\lambda_i^0\right\|^2 \max_{1\le i\le N} T^{-1}\left\|X_i\right\|^2 \geq C) = o\left(N^{-1}\right)$ for any small $C > 0$. This, in conjunction with Lemma A.4(v), and Assumption A.1(x), implies that $P(\mu_{\min}(\hat{Q}_1 - \hat{Q}_2) \geq \underline{c}_{XX}/2) = 1 - o\left(N^{-1}\right)$. On the other hand,

$$\left\|R_i\right\|^2 = \left\|\frac{1}{T}X_i' M_{F^0}\varepsilon_i + R_{1i} - R_{2i} + R_{3i}\right\|^2 \leq 2\left\|\frac{1}{T}X_i' M_{F^0}\varepsilon_i + R_{1i} - R_{2i}\right\|^2 + 2\left\|R_{3i}\right\|^2.$$

By Assumption A.3, Lemmas A.4(iii) and (vi), and the triangle inequality, we can readily show that

$$P\left(\max_{1\le i\le N}\left\|\frac{1}{T}X_i' M_{F^0}\varepsilon_i + R_{1i} - R_{2i}\right\| \geq CT^{-1/2}\left(\ln T\right)^{\upsilon_1}\right) = o\left(N^{-1}\right) \text{ for any } C > 0.$$

By the triangle inequality and the fact that $\Pi_{l=1,l\neq j}^{K_0}\left\|\hat{\beta}_i - \hat{\alpha}_l\right\| - \Pi_{l=1,l\neq j}^{K_0}\left\|\beta_i^0 - \hat{\alpha}_l\right\| \leq C_{K_0}\left(\hat{\boldsymbol{\alpha}}\right)(1 + 2\left\|\hat{\beta}_i - \beta_i^0\right\|)$ for some $C_{K_0}\left(\hat{\boldsymbol{\alpha}}\right) = O_P\left(1\right)$ (see, e.g., the argument used in the proof of Theorem 3.2), we have

$$\begin{aligned}
\left\|R_{3i}\right\| &\leq \frac{\kappa}{2}\sum_{j=1}^{K_0}\Pi_{l=1,l\neq j}^{K_0}\left\|\hat{\beta}_i - \hat{\alpha}_l\right\| \\
&\leq \frac{\kappa}{2}\sum_{j=1}^{K_0}\left|\Pi_{l=1,l\neq j}^{K_0}\left\|\hat{\beta}_i - \hat{\alpha}_l\right\| - \Pi_{l=1,l\neq j}^{K_0}\left\|\beta_i^0 - \hat{\alpha}_l\right\|\right| + \frac{\kappa}{2}\sum_{j=1}^{K_0}\Pi_{l=1,l\neq j}^{K_0}\left\|\beta_i^0 - \hat{\alpha}_l\right\| \\
&\leq \kappa K_0 C\left(\hat{\boldsymbol{\alpha}}\right)\left\|\hat{b}_i\right\| + r_{NT}.
\end{aligned}$$

where $r_{NT} = \frac{\kappa}{2}[K_0 C_{K_0}\left(\hat{\boldsymbol{\alpha}}\right) + \max_{1\le i\le N}\sum_{j=1}^{K_0}\Pi_{l=1,l\neq j}^{K_0}\left\|\beta_i^0 - \hat{\alpha}_l\right\|] = O\left(\kappa\right)$ as β_i^0's can only take finite K_0 values. It follows that

$$\begin{aligned}
\left\|\hat{b}_i\right\|^2 &\leq \left[\mu_{\min}\left(\hat{Q}_1 - \hat{Q}_2\right)\right]^{-2}\left\|R_i\right\|^2 \\
&\leq \left[\mu_{\min}\left(\hat{Q}_1 - \hat{Q}_2\right)\right]^{-2}\left\{2\left\|\frac{1}{T}X_i' M_{F^0}\varepsilon_i + R_{1i} - R_{2i}\right\|^2 + 2\left\|R_{3i}\right\|^2\right\} \\
&\leq \left[\mu_{\min}\left(\hat{Q}_1 - \hat{Q}_2\right)\right]^{-2}\left\{2\left\|\frac{1}{T}X_i' M_{F^0}\varepsilon_i + R_{1i} - R_{2i}\right\|^2 + 4\left[\kappa K_0 C\left(\hat{\boldsymbol{\alpha}}\right)\right]^2\left\|\hat{b}_i\right\|^2 + 4r_{NT}^2\right\}.
\end{aligned}$$

That is, $\left\|\hat{b}_i\right\|^2 \leq [\mu_{\min}(\hat{Q}_1 - \hat{Q}_2)]^{-2}\{2\left\|\frac{1}{T}X_i' M_{F^0}\varepsilon_i + R_{1i} - R_{2i}\right\|^2 + 4r_{NT}^2\}/(1 - 4\left[\kappa K_0 C\left(\hat{\boldsymbol{\alpha}}\right)\right]^2)$. It follows that for sufficiently small κ (as $\kappa \to 0$), we have

$$\begin{aligned}
& P\left(\max_{1\le i\le N}\left\|\hat{b}_i\right\|^2 \geq C\left[T^{-1}\left(\ln T\right)^{2\upsilon_1} + \kappa^2\left(\ln T\right)^{2\upsilon_2}\right]\right) \\
\leq\ & P\left(\max_{1\le i\le N}\left\|\hat{b}_i\right\|^2 \geq C\left[T^{-1}\left(\ln T\right)^{2\upsilon_1} + \kappa^2\left(\ln T\right)^{2\upsilon_2}\right],\ \mu_{\min}(\hat{Q}_1 - \hat{Q}_2) \geq \underline{c}_{XX}/2\right) + o\left(N^{-1}\right) \\
\leq\ & P\left(\max_{1\le i\le N} 8\underline{c}_{XX}^{-2}\left\{2\left\|\frac{1}{T}X_i' M_{F^0}\varepsilon_i + R_{1i} - R_{2i}\right\|^2 + 4r_{NT}^2\right\} \geq C\left[T^{-1}\left(\ln T\right)^{2\upsilon_1} + \kappa^2\left(\ln T\right)^{2\upsilon_2}\right]\right) + o\left(N^{-1}\right) \\
\leq\ & P\left(\max_{1\le i\le N} 2\left\|\frac{1}{T}X_i' M_{F^0}\varepsilon_i + R_{1i} - R_{2i}\right\|^2 \geq CT^{-1}\left(\ln T\right)^{2\upsilon_1}\underline{c}_{XX}^2/8\right) \\
& + P\left(4r_{NT}^2 \geq C\kappa^2\left(\ln T\right)^{2\upsilon_2}\underline{c}_{XX}^2/8\right) + o\left(N^{-1}\right) \\
=\ & o(N^{-1}) + o\left(N^{-1}\right) + o(N^{-1}) = o(N^{-1}),
\end{aligned}$$

where we use the fact that $P\left(\max_{1\le i\le N}\left\|T^{-1/2}X_i'M_{F^0}\varepsilon_i\right\|\ge C(\ln T)^{\upsilon_1}\right)=o\left(N^{-1}\right)$ $\forall C>0$ by Assumptions A.1(i) and A.3(i). That is, $P\left(\max_{1\le i\le N}\left\|\hat{b}_i\right\|\ge C\left[T^{-1/2}(\ln T)^{\upsilon_1}+\kappa(\ln T)^{\upsilon_2}\right]\right)=o(N^{-1})$ for any $C>0$. ■

Proof of Theorem 3.4 (i) Fix $k\in\{1,...,K_0\}$. By the consistency of $\hat{\alpha}_k$ and $\hat{\beta}_i$ in Theorem 3.2, we have $\hat{\beta}_i-\hat{\alpha}_l\xrightarrow{P}\alpha_k^0-\alpha_l^0\neq 0$ for all $i\in G_k^0$ and $l\neq k$ and $\hat{c}_{ki}=\Pi_{l=1,l\neq k}^{K_0}\left\|\hat{\beta}_i-\hat{\alpha}_l\right\|\xrightarrow{P}c_k^0\equiv\Pi_{l=1,l\neq k}^{K_0}\left\|\alpha_k^0-\alpha_l^0\right\|>0$ for $i\in G_k^0$. Now, suppose that $\left\|\hat{\beta}_i-\hat{\alpha}_k\right\|\neq 0$ for some $i\in G_k^0$. Then the first order condition (with respect to β_i) for the minimization problem in (2.5) implies that

$$\begin{aligned}
\mathbf{0}_{p\times 1} &= -\frac{2}{\sqrt{T}}X_i'M_{\hat{F}}(Y_i-X_i\hat{\beta}_i)+\sqrt{T}\kappa\sum_{j=1}^{K_0}\hat{e}_{ij}\Pi_{l=1,l\neq j}^{K_0}\left\|\hat{\beta}_i-\hat{\alpha}_l\right\| \\
&= -\frac{2}{\sqrt{T}}X_i'M_{F^0}\varepsilon_i+\left(\frac{2}{T}X_i'M_{\hat{F}}X_i+\frac{\kappa\hat{c}_{ki}}{\left\|\hat{\beta}_i-\hat{\alpha}_k\right\|}I_p\right)\sqrt{T}(\hat{\beta}_i-\hat{\alpha}_k) \\
&\quad +\frac{2}{\sqrt{T}}X_i'(M_{F^0}-M_{\hat{F}})\varepsilon_i-\frac{2}{\sqrt{T}}X_i'M_{\hat{F}}F^0\lambda_i^0 \\
&\quad +\frac{2}{T}X_i'M_{\hat{F}}X_i\sqrt{T}\left(\hat{\alpha}_k-\alpha_k^0\right)+\sqrt{T}\kappa\sum_{j=1}^{K_0}\hat{e}_{ij}\Pi_{l=1,l\neq j}^{K_0}\left\|\hat{\beta}_i-\hat{\alpha}_l\right\| \\
&\equiv \hat{B}_{i1}+\hat{B}_{i2}+\hat{B}_{i3}+\hat{B}_{i4}+\hat{B}_{i5}+\hat{B}_{i6},\ \text{say,}
\end{aligned}\tag{A.15}$$

where $\hat{e}_{ij}=\frac{\hat{\beta}_i-\hat{\alpha}_j}{\left\|\hat{\beta}_i-\hat{\alpha}_j\right\|}$ if $\left\|\hat{\beta}_i-\hat{\alpha}_j\right\|\neq 0$ and $\left\|\hat{e}_{ij}\right\|\le 1$ otherwise.

Let $\chi_{NT}=T^{-1/2}(\ln T)^{\upsilon_1}+\kappa(\ln T)^{\upsilon_2}$ where $0<\upsilon_2\le\upsilon_1$. By Theorem 3.3,

$$P\left(\max_{1\le i\le N}\left\|\hat{b}_i\right\|\ge C\chi_{NT}\right)=o(N^{-1})\text{ for any }C>0.\tag{A.16}$$

This, in conjunction with the proof of Theorem 3.2(iii), implies that

$$P\left(\left\|\hat{\alpha}_k-\alpha_k^0\right\|\ge CT^{-1/2}(\ln T)^{\upsilon_1}\right)=o\left(N^{-1}\right)\text{ and }P\left(\max_{i\in G_k^0}\left|\hat{c}_{ki}-c_k^0\right|\ge c_k^0/2\right)=o\left(N^{-1}\right).\tag{A.17}$$

By (A.16)-(A.17), Lemma A.4(v) and Assumption A.1(x),

$$P\left(\max_{i\in G_k^0}\left\|\hat{B}_{i5}\right\|>C(\ln T)^{\upsilon_1}\right)=o\left(N^{-1}\right)\text{ and }P\left(\max_{i\in G_k^0}\left\|\hat{B}_{i6}\right\|>C\sqrt{T}\kappa\chi_{NT}\right)=o\left(N^{-1}\right).$$

By Lemmas A.4(iii)-(iv),

$$P\left(\max_{i\in G_k^0}\left\|\hat{B}_{i3}\right\|>CT^{1/2}\delta_{NT}^{-2}(\ln T)^{\upsilon_1}\right)=o\left(N^{-1}\right)\text{ and }P\left(\max_{i\in G_k^0}\left\|\hat{B}_{i4}\right\|>C(\ln T)^{\upsilon_1}c_{\lambda,N}\right)=o\left(N^{-1}\right).$$

It follows that $P(\Xi_{kNT})=1-o\left(N^{-1}\right)$, where

$$\begin{aligned}
\Xi_{kNT} &\equiv \left\{\max_{i\in G_k^0}\left|\hat{c}_{ki}-c_k^0\right|\le c_k^0/2\right\}\cap\left\{\max_{i\in G_k^0}\left\|\hat{B}_{i3}\right\|\le CT^{1/2}\delta_{NT}^{-2}(\ln T)^{\upsilon_1}\right\} \\
&\quad\cap\left\{\max_{i\in G_k^0}\left\|\hat{B}_{i4}\right\|\le C(\ln T)^{\upsilon_1}c_{\lambda,N}\right\}\cap\left\{\max_{i\in G_k^0}\left\|\hat{B}_{i5}\right\|\le C(\ln T)^{\upsilon_1}\right\} \\
&\quad\cap\left\{\max_{i\in G_k^0}\left\|\hat{B}_{i6}\right\|\le C\sqrt{T}\kappa\chi_{NT}\right\}.
\end{aligned}$$

Then conditional on Ξ_{kNT}, we have that uniformly in $i \in G_k^0$,

$$
\begin{aligned}
& \left\| \left(\hat{\beta}_i - \hat{\alpha}_k\right)' \left(\hat{B}_{i2} + \hat{B}_{i3} + \hat{B}_{i4} + \hat{B}_{i5} + \hat{B}_{i6}\right) \right\| \\
\geq & \left\| \left(\hat{\beta}_i - \hat{\alpha}_k\right)' \hat{B}_{i2} \right\| - \left\| \left(\hat{\beta}_i - \hat{\alpha}_k\right)' \left(\hat{B}_{i3} + \hat{B}_{i4} + \hat{B}_{i5} + \hat{B}_{i6}\right) \right\| \\
\geq & \sqrt{T}\kappa \hat{c}_{ki} \left\| \hat{\beta}_i - \hat{\alpha}_k \right\| - C \left\| \hat{\beta}_i - \hat{\alpha}_k \right\| \{3 (\ln T)^{\upsilon_1} c_{\lambda,N} + \sqrt{T}\kappa\chi_{NT}\} \\
\geq & \sqrt{T}\kappa c_k^0 \left\| \hat{\beta}_i - \hat{\alpha}_k \right\| /4 \text{ for sufficiently large } (N,T),
\end{aligned}
$$

where the last inequality follows because $\hat{c}_{ki} \geq c_k^0/2$ on Ξ_{kNT}, $\sqrt{T}\kappa \gg 3 (\ln T)^{\upsilon_1} c_{\lambda,N} + \sqrt{T}\kappa\chi_{NT}$ by Assumptions A.2(iii)-(iv). It follows that for all $i \in G_k^0$

$$
\begin{aligned}
P(\hat{E}_{kNT,i}) &= P\left(i \notin \hat{G}_k \mid i \in G_k^0\right) \\
&= P\left(-\hat{B}_{i1} = \hat{B}_{i2} + \hat{B}_{i3} + \hat{B}_{i4} + \hat{B}_{i5} + \hat{B}_{i6}\right) \\
&\leq P\left(\left|(\hat{\beta}_i - \hat{\alpha}_k)'\hat{B}_{i1}\right| \geq \left|(\hat{\beta}_i - \hat{\alpha}_k)' \left(\hat{B}_{i2} + \hat{B}_{i3} + \hat{B}_{i4} + \hat{B}_{i5} + \hat{B}_{i6}\right)\right|\right) \\
&\leq P\left(\left\|\hat{\beta}_i - \hat{\alpha}_k\right\| \left\|\hat{B}_{i1}\right\| \geq \sqrt{T}\kappa c_k^0 \left\|\hat{\beta}_i - \hat{\alpha}_k\right\| /4, \Xi_{kNT}\right) + P\left(\Xi_{kNT}^*\right) \\
&\leq P\left(\left\|\hat{B}_{i1}\right\| \geq \sqrt{T}\kappa c_k^0/4\right) + P\left(\Xi_{kNT}^*\right) \\
&\to 0 \text{ as } (N,T) \to \infty,
\end{aligned}
$$

where Ξ_{kNT}^* denotes the complement of Ξ_{kNT} and the convergence follows by Assumptions A.1(i), A.2(iv) and A.3(i) (see the remark after Assumption A.3), and the fact that $P\left(\Xi_{kNT}^*\right) = o\left(N^{-1}\right)$. Consequently, we can conclude that with probability $1 - o\left(N^{-1}\right)$, $\hat{\beta}_i - \hat{\alpha}_k$ must be in a position where $\|\beta_i - \alpha_k\|$ is not differentiable with respect to β_i for any $i \in G_k^0$. That is, $P\left(\left\|\hat{\beta}_i - \hat{\alpha}_k\right\| = 0 \mid i \in G_k^0\right) = 1 - o\left(N^{-1}\right)$ as $(N,T) \to \infty$. Then the rest of the proof follows SSP. ■

The following lemma improves the results in Lemma A.2.

Lemma A.5 *Suppose Assumption A.1-A.3 hold. Then*

(i) $T^{-1}\left\|\hat{F} - F^0H\right\|^2 = O_P\left(\sum_{k=1}^{K_0} \left\|\hat{\alpha}_k - \alpha_k^0\right\|^2\right) + O_P\left(\delta_{NT}^{-2}\right)$,

(ii) $T^{-1}(\hat{F} - F^0H)'F^0H = O_P\left(\sum_{k=1}^{K_0} \left\|\hat{\alpha}_k - \alpha_k^0\right\|\right) + O_P\left(\delta_{NT}^{-2}\right)$,

(iii) $T^{-1}(\hat{F} - F^0H)'\hat{F} = O_P\left(\sum_{k=1}^{K_0} \left\|\hat{\alpha}_k - \alpha_k^0\right\|\right) + O_P\left(\delta_{NT}^{-2}\right)$,

(iv) $T^{-1}(\hat{F}'\hat{F} - H'F^{0\prime}F^0H) = I_{R_0} - T^{-1}H'F^{0\prime}F^0H = O_P\left(\sum_{k=1}^{K_0} \left\|\hat{\alpha}_k - \alpha_k^0\right\|\right) + O_P\left(\delta_{NT}^{-2}\right)$,

(v) $\left\|P_{\hat{F}} - P_{F^0H}\right\| = O_P\left(\sum_{k=1}^{K_0} \left\|\hat{\alpha}_k - \alpha_k^0\right\|\right) + O_P\left(\delta_{NT}^{-1}\right)$.

To state the next lemma, for k, $l = 1, ..., K_0$, define

$$
\begin{aligned}
\bar{A}_{1k,l} &= -\frac{1}{NN_kT^2} \sum_{j\in\hat{G}_l} \sum_{i\in\hat{G}_k} X_i'M_{\hat{F}}X_j \left(\hat{\alpha}_l - \alpha_l^0\right) \lambda_i^{0\prime}G'\hat{F}'X_j, \\
\bar{A}_{2k,l} &= \frac{1}{NN_kT} \sum_{j\in\hat{G}_l} \sum_{i\in\hat{G}_k} \nu_{ji}X_i'M_{\hat{F}}X_j, \\
\bar{A}_{3k,l} &= \frac{1}{NN_kT^2} \sum_{j\in\hat{G}_l} \sum_{i\in\hat{G}_k} X_i'M_{\hat{F}}X_j\varepsilon_j'\hat{F}G\lambda_i^0,
\end{aligned}
$$

$$\bar{A}_{4k,l} = \frac{1}{NN_kT^2}\sum_{j\in\hat{G}_l}\sum_{i\in\hat{G}_k} X_i'M_{\hat{F}}F^0\lambda_j^0\lambda_i^{0\prime}G'\hat{F}'X_j, \text{ and}$$

$$\bar{A}_{5k,l} = \frac{1}{NN_kT^2}\sum_{j\in\hat{G}_l}\sum_{i\in\hat{G}_k} X_i'M_{\hat{F}}\varepsilon_j\lambda_i^{0\prime}G'\hat{F}'X_j.$$

Let $\bar{A}_{k,l}=\sum_{s=1}^{5}\bar{A}_{sk,l}$ and $A_{k,l}=\frac{1}{NN_kT}\sum_{j\in G_l^0}\sum_{i\in G_k^0}\nu_{ji}X_i'M_{F^0}X_j$.

Lemma A.6 *Suppose Assumptions A.1-A.3 hold. Then for each k, $l=1,2,...,K_0$, we have*

(i) $\bar{A}_{1k,l}=O_P\left(T^{-1/2}\right)$,

(ii) $\bar{A}_{2k,l}=A_{k,l}+O_P\left(\delta_{NT}^{-1}\right)$,

(iii) $\bar{A}_{3k,l}=O_P\left(\delta_{NT}^{-1}\right)$,

(iv) $\bar{A}_{4k,l}=O_P\left(\delta_{NT}^{-1}\right)$,

(v) $\bar{A}_{5k,l}=O_P\left(\delta_{NT}^{-1}\right)$,

(vi) $\bar{A}_{k,l}=\bar{A}_{2k,l}+O_P\left(\delta_{NT}^{-1}\right)=A_{k,l}+O_P\left(\delta_{NT}^{-1}\right)$.

Lemma A.7 *Suppose Assumptions A.1-A.3 hold. Then*

(i) $\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i'(\hat{F}-F^0H)=O_P\left(\delta_{NT}^{-2}\right)$,

(ii) $\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^0\varepsilon_i'(\hat{F}-F^0H)=O_P\left(\delta_{NT}^{-2}\right)$,

(iii) $\frac{1}{NT^2}\sum_{i=1}^{N}\varepsilon_i'(\hat{F}-F^0H)(\hat{F}-F^0H)'\varepsilon_i=O_P\left(\delta_{NT}^{-4}\right)$,

(iv) $\frac{1}{N_kT^2}\sum_{i\in G_k^0}X_i'(\hat{F}-F^0H)(\hat{F}-F^0H)'\varepsilon_i=O_P\left(\delta_{NT}^{-3}\right)$,

(v) $\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j=1}^{N}X_i'\varepsilon_j\varepsilon_j'(\hat{F}-F^0H)G\lambda_i^0=O_P\left(T^{-1}\delta_{NT}^{-2}\right)$,

(vi) $\frac{1}{N_kT}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}F^0\lambda_i^0=\sum_{l=1}^{K_0}\bar{A}_{k,l}\left(\hat{\alpha}_l-\alpha_l^0\right)-\frac{1}{NN_kT}\sum_{i\in\hat{G}_k}\sum_{j=1}^{N}\nu_{ji}X_i'M_{\hat{F}}\varepsilon_j+\mathcal{B}_{1,kNT}+o_P((NT)^{-1/2})$.

Let $Q_{kNT}=\frac{1}{N_kT}\sum_{i\in G_k^0}X_i'M_{F^0}X_i$ and $V_{kNT}=\frac{1}{N_kT}\sum_{i\in G_k^0}X_i'M_{F^0}\ (\varepsilon_i-\frac{1}{N}\sum_{j=1}^{N}\nu_{ji}\varepsilon_j)$.

Lemma A.8 *Suppose Assumptions A.1-A.3 hold. Then for $k=1,...,K_0$,*

(i) $\hat{Q}_{kNT}=Q_{kNT}+O_P\left(\delta_{NT}^{-1}\right)$,

(ii) $\mathcal{B}_{1,kNT}=O_P\left(T^{-1}\right)$,

(iii) $\hat{V}_{kNT}=V_{kNT}+O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$,

(iv) $V_{kNT}=O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$.

Proof of Theorem 3.5. To study the oracle property of the Lasso estimator, we invoke subdifferential calculus (e.g., Bertsekas (1995, Appendix B.5)). A necessary and sufficient condition for $\{\hat{\beta}_i\}$ and $\{\hat{\alpha}_k\}$ to minimize the objective function in (2.5) is that for each $i=1,...,N$ (resp. $k=1,...,K_0$), $\mathbf{0}_{p\times1}$ belongs to the subdifferential of $Q_{1NT,\kappa}^{(K_0)}(\boldsymbol{\beta},\boldsymbol{\alpha},\hat{F})$ with respect to β_i (resp. α_k) evaluated at $\{\hat{\beta}_i\}$ and $\{\hat{\alpha}_k\}$. That is, for each $i=1,...,N$ and $k=1,...,K_0$, we have

$$\mathbf{0}_{p\times1} = -\frac{2}{NT}X_i'M_{\hat{F}}(Y_i-X_i\hat{\beta}_i)+\frac{\kappa}{N}\sum_{j=1}^{K_0}\hat{e}_{ij}\Pi_{l=1,l\neq j}^{K_0}\left\|\hat{\beta}_i-\hat{\alpha}_l\right\|, \text{ and} \tag{A.18}$$

$$\mathbf{0}_{p\times1} = \frac{\kappa}{N}\sum_{i=1}^{N}\hat{e}_{ik}\Pi_{l=1,l\neq k}^{K_0}\left\|\hat{\beta}_i-\hat{\alpha}_l\right\|, \tag{A.19}$$

where $\hat{e}_{ij} = \frac{\hat{\beta}_i - \hat{\alpha}_j}{\|\hat{\beta}_i - \hat{\alpha}_j\|}$ if $\left\|\hat{\beta}_i - \hat{\alpha}_j\right\| \neq 0$ and $\|\hat{e}_{ij}\| \leq 1$ if $\left\|\hat{\beta}_i - \hat{\alpha}_j\right\| = 0$. Fix $k \in \{1, ..., K_0\}$. Observe that (a) $\left\|\hat{\beta}_i - \hat{\alpha}_k\right\| = 0$ for any $i \in \hat{G}_k$ by the definition of $\hat{G}_k$, and (b) $\hat{\beta}_i - \hat{\alpha}_l \xrightarrow{P} \alpha_k^0 - \alpha_l^0 \neq 0$ for any $i \in \hat{G}_k$ and $l \neq k$. It follows that $\|\hat{e}_{ik}\| \leq 1$ for any $i \in \hat{G}_k$ and $\hat{e}_{ij} = \frac{\hat{\beta}_i - \hat{\alpha}_j}{\|\hat{\beta}_i - \hat{\alpha}_j\|} = \frac{\hat{\alpha}_k - \hat{\alpha}_j}{\|\hat{\alpha}_k - \hat{\alpha}_j\|}$ w.p.a.1 for any $i \in \hat{G}_k$ and $j \neq k$. This further implies that w.p.a.1

$$\sum_{i \in \hat{G}_k} \sum_{j=1, j \neq k}^{K_0} \hat{e}_{ij} \Pi_{l=1, l \neq j}^{K_0} \left\|\hat{\beta}_i - \hat{\alpha}_l\right\| = \sum_{i \in \hat{G}_k} \sum_{j=1, j \neq k}^{K_0} \frac{\hat{\alpha}_k - \hat{\alpha}_j}{\|\hat{\alpha}_k - \hat{\alpha}_j\|} \Pi_{l=1, l \neq j}^{K_0} \|\hat{\alpha}_k - \hat{\alpha}_l\| = \mathbf{0}_{p \times 1}$$

and

$$\begin{aligned} \mathbf{0}_{p \times 1} &= \sum_{i=1}^{N} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \left\|\hat{\beta}_i - \hat{\alpha}_l\right\| \\ &= \sum_{i \in \hat{G}_k} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \|\hat{\alpha}_k - \hat{\alpha}_l\| + \sum_{i \in \hat{G}_0} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \left\|\hat{\beta}_i - \hat{\alpha}_l\right\| + \sum_{j=1, j \neq k}^{K_0} \sum_{i \in \hat{G}_j} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \|\hat{\alpha}_j - \hat{\alpha}_l\| \\ &= \sum_{i \in \hat{G}_k} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \|\hat{\alpha}_k - \hat{\alpha}_l\| + \sum_{i \in \hat{G}_0} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \left\|\hat{\beta}_i - \hat{\alpha}_l\right\|. \end{aligned}$$

Then by (A.18) we have

$$\frac{2}{NT} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} (Y_i - X_i \hat{\alpha}_k) + \frac{\kappa}{N} \sum_{i \in \hat{G}_0} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \left\|\hat{\beta}_i - \hat{\alpha}_l\right\| = \mathbf{0}_{p \times 1}. \tag{A.20}$$

Noting that $\mathbf{1}\{i \in \hat{G}_k\} = \mathbf{1}\{i \in G_k^0\} + \mathbf{1}\{i \in \hat{G}_k \backslash G_k^0\} - \mathbf{1}\{i \in G_k^0 \backslash \hat{G}_k\}$ and $Y_i = X_i \alpha_k^0 + F^0 \lambda_i^0 + \varepsilon_i$ when $i \in G_k^0$, we have

$$\begin{aligned} \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} Y_i &= \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} X_i \beta_i^0 + \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} F^0 \lambda_i^0 + \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} \varepsilon_i \\ &= \frac{1}{NT} \sum_{i \in G_k^0} X_i' M_{\hat{F}} X_i \alpha_k^0 + \frac{1}{NT} \sum_{i \in \hat{G}_k \backslash G_k^0} X_i' M_{\hat{F}} X_i \beta_i^0 - \frac{1}{NT} \sum_{i \in G_k^0 \backslash \hat{G}_k} X_i' M_{\hat{F}} X_i \beta_i^0 \\ &\quad + \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} F^0 \lambda_i^0 + \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} \varepsilon_i. \end{aligned} \tag{A.21}$$

Combining (A.20)-(A.21) yields

$$\frac{1}{N_k T} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} X_i \left(\hat{\alpha}_k - \alpha_k^0\right) = \frac{1}{N_k T} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} F^0 \lambda_i^0 + \frac{1}{N_k T} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} \varepsilon_i + \hat{\mathcal{R}}_{1k} - \hat{\mathcal{R}}_{2k} + \hat{\mathcal{R}}_{3k}, \tag{A.22}$$

where $\hat{\mathcal{R}}_{1k} = \frac{1}{NT} \sum_{i \in \hat{G}_k \backslash G_k^0} X_i' M_{\hat{F}} X_i \beta_i^0$, $\hat{\mathcal{R}}_{2k} = \frac{1}{NT} \sum_{i \in G_k^0 \backslash \hat{G}_k} X_i' M_{\hat{F}} X_i \beta_i^0$, and $\hat{\mathcal{R}}_{3k} = \frac{\kappa}{2N} \sum_{i \in \hat{G}_0} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0}$ $||\hat{\beta}_i - \hat{\alpha}_l||$. Noting that for any $\epsilon > 0$, by Theorem 3.4 and the proof of Theorem 2.2 in SSP, we have $P(||\hat{\mathcal{R}}_{1k}|| \geq \epsilon/(NT)^{1/2}) \leq P(\hat{F}_{kNT}) \to 0$, $P(||\hat{\mathcal{R}}_{2k}|| \geq \epsilon/(NT)^{1/2}) \leq P(\hat{E}_{kNT}) \to 0$, and $P(||\hat{\mathcal{R}}_{3k}|| \geq \epsilon/(NT)^{1/2}) \leq \sum_{k=1}^{K_0} \sum_{i \in G_k^0} P(i \in \hat{G}_0 | i \in G_k^0) \leq \sum_{k=1}^{K_0} \sum_{i \in G_k^0} P(\hat{E}_{kNT,i}) = o(1)$. It follows that

$$\left\|\hat{\mathcal{R}}_{1k} - \hat{\mathcal{R}}_{2k} + \hat{\mathcal{R}}_{3k}\right\| = o_P((NT)^{-1/2}). \tag{A.23}$$

Next, by Lemma A.7(vi),

$$\frac{1}{N_k T} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} F^0 \lambda_i^0 = \sum_{l=1}^{K_0} \bar{A}_{k,l} \left(\hat{\alpha}_l - \alpha_l^0\right) - \frac{1}{N N_k T} \sum_{i \in \hat{G}_k} \sum_{j=1}^{N} \nu_{ji} X_i' M_{\hat{F}} \varepsilon_j + \mathcal{B}_{1,kNT} + o_P((NT)^{-1/2}). \tag{A.24}$$

Combining (A.22)-(A.24) yields

$$\hat{Q}_{kNT}\left(\hat{\alpha}_k-\alpha_k^0\right)=\sum_{l=1}^{K_0}\bar{A}_{k,l}\left(\hat{\alpha}_l-\alpha_l^0\right)+\hat{V}_{kNT}+\mathcal{B}_{1,kNT}+o_P((NT)^{-1/2}),$$

where $\hat{Q}_{kNT}=\frac{1}{N_kT}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}X_i$ and $\hat{V}_{kNT}=\frac{1}{N_kT}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}(\varepsilon_i-\frac{1}{N}\sum_{j=1}^N\nu_{ji}\varepsilon_j)$. The above result holds for $k=1,...,K_0$. It follows that

$$\operatorname{vec}\left(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right)=\bar{Q}_{NT}^{-1}\left[\hat{V}_{NT}+\mathcal{B}_{1,NT}+o_P((NT)^{-1/2})\right], \tag{A.25}$$

where

$$\bar{Q}_{NT}=\begin{pmatrix}\hat{Q}_{1NT}-\bar{A}_{1,1} & -\bar{A}_{1,2} & \cdots & -\bar{A}_{1,K_0}\\ -\bar{A}_{2,1} & \hat{Q}_{2NT}-\bar{A}_{2,2} & \cdots & -\bar{A}_{2,K_0}\\ \vdots & \vdots & \ddots & \vdots\\ -\bar{A}_{K_0,1} & -\bar{A}_{K_0,2} & \cdots & \hat{Q}_{K_0NT}-\bar{A}_{K_0,K_0}\end{pmatrix},\ \hat{V}_{NT}=\begin{pmatrix}\hat{V}_{1NT}\\ \hat{V}_{2NT}\\ \vdots\\ \hat{V}_{K_0NT}\end{pmatrix}, \tag{A.26}$$

and $\mathcal{B}_{1,NT}=(\mathcal{B}_{1,1NT}',...,\mathcal{B}_{1,K_0NT}')'$. By Lemmas A.7(vi) and A.8(i), $\bar{Q}_{NT}=Q_{NT}+O_P\left(\delta_{NT}^{-1}\right)$ where Q_{NT} is defined in (3.3). By Lemmas A.8(ii)-(iv), $\mathcal{B}_{1,NT}=O_P(T^{-1})$ and $\hat{V}_{NT}=O_P(T^{-1/2}\delta_{NT}^{-1})$. Then by (A.25), $\operatorname{vec}\left(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right)=\left[Q_{NT}^{-1}+O_P\left(\delta_{NT}^{-1}\right)\right]\left[\hat{V}_{NT}+\mathcal{B}_{1,NT}\right]+o_P((NT)^{-1/2})=Q_{NT}^{-1}[O_P(T^{-1/2}\delta_{NT}^{-1})+O_P(T^{-1})]+o_P((NT)^{-1/2})=O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$. ■

Lemma A.9 *Suppose Assumptions A.1-A.5 hold. Let $\omega_p=(\omega_{1p},...,\omega_{pp})'$ be an arbitrary $p\times1$ nonrandom vector with $\|\omega_p\|=1$. Then*

(i) $\left\|\frac{1}{NN_kT^2}\sum_{j=1}^N\sum_{i\in G_k^0}\hat{F}'\varepsilon_j\varepsilon_j'\varepsilon_i\omega_p'X_i'F^0\right\|=O_P\left(T^{1/2}\delta_{NT}^{-3}\right)$,

(ii) $\frac{1}{N_kT^2}\sum_{i\in G_k^0}X_i'F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}(\hat{F}H^{-1}-F^0)'\varepsilon_i=\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j=1}^N X_i'F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}\times(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_j^0\varepsilon_j'\varepsilon_i+o_P((NT)^{-1/2})$,

(iii) $\frac{1}{N_kT^2}\sum_{i\in G_k^0}X_i^{*\prime}F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}(\hat{F}H^{-1}-F^0)'\varepsilon_i=\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j=1}^N X_i^{*\prime}F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}\times(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_j^0\varepsilon_j'\varepsilon_i+o_P((NT)^{-1/2})$.

Proof of Theorem 3.6. By (A.25), the fact that $\bar{Q}_{NT}=Q_{NT}+O_P\left(\delta_{NT}^{-1}\right)$, and Assumptions A.4-A.5, it suffices to prove the theorem by showing that (i) $\hat{V}_{NT}=V_{NT}+\mathcal{B}_{2,NT}+o_P((NT)^{-1/2})$, and (ii) $\mathcal{B}_{NT}\equiv\mathcal{B}_{1,NT}+\mathcal{B}_{2,NT}+\mathcal{B}_{3,NT}=O_P(T^{-1/2}\delta_{NT}^{-1})$. This follows because (i)-(ii), in conjunction with (A.25) and Assumptions A.4(i) and A.5(i)-(iii), imply that

$$\begin{aligned}\sqrt{NT}\left[\operatorname{vec}\left(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right)-Q_{NT}^{-1}\mathcal{B}_{NT}\right] &= \bar{Q}_{NT}^{-1}\sqrt{NT}\left[\hat{V}_{NT}+\mathcal{B}_{1NT}+o_P((NT)^{-1/2})\right]-\sqrt{NT}Q_{NT}^{-1}\mathcal{B}_{NT}\\ &= \bar{Q}_{NT}^{-1}\sqrt{NT}\left[\mathcal{V}_{NT}+\mathcal{B}_{NT}+o_P((NT)^{-1/2})\right]-\sqrt{NT}Q_{NT}^{-1}\mathcal{B}_{NT}\\ &= \bar{Q}_{NT}^{-1}\sqrt{NT}\mathcal{V}_{NT}+\left(\bar{Q}_{NT}^{-1}-Q_{NT}^{-1}\right)\sqrt{NT}\mathcal{B}_{NT}+o_P(1)\\ &= Q_{NT}^{-1}\sqrt{NT}\mathcal{V}_{NT}+O_P\left(\delta_{NT}^{-1}\right)\sqrt{NT}O_P(T^{-1/2}\delta_{NT}^{-1})+o_P(1)\\ &= Q_{NT}^{-1}\sqrt{NT}\mathcal{V}_{NT}+o_P(1)\xrightarrow{d}N\left(0,Q_0^{-1}\Theta_0Q_0^{-1}\right).\end{aligned}$$

We first show (i). Let $\tilde{V}_{kNT}=\frac{1}{N_kT}\sum_{i\in G_k^0}X_i'M_{\hat{F}}(\varepsilon_i-\frac{1}{N}\sum_{j=1}^N\nu_{ji}\varepsilon_j)$. Following the proof of Lemma A.6(ii), we can show that $\hat{V}_{kNT}-\tilde{V}_{kNT}=o_P((NT)^{-1/2})$. Next,

$$\tilde{V}_{kNT}-V_{kNT}=\frac{1}{N_kT}\sum_{i\in G_k^0}X_i'\left(M_{\hat{F}}-M_{F^0}\right)\varepsilon_i-\frac{1}{NN_kT}\sum_{i\in G_k^0}\sum_{j=1}^N X_i'\left(M_{\hat{F}}-M_{F^0}\right)\nu_{ji}\varepsilon_j\equiv v_{1kNT}-v_{2kNT},\ \text{say.}$$

Using (A.1), we have $v_{1kNT} = \sum_{l=1}^{4} \frac{-1}{N_k T} \sum_{i \in G_k^0} X_i' d_l \varepsilon_i \equiv \sum_{l=1}^{4} v_{1kNT,l}$, and $v_{2kNT} = \sum_{l=1}^{4} \frac{-1}{N_k T} \sum_{i \in G_k^0} X_i^{*\prime} d_l \varepsilon_i \equiv \sum_{l=1}^{4} v_{2kNT,l}$, where $X_i^* = \frac{1}{N} \sum_{k=1}^{N} \nu_{ki} X_k$. We have shown in the proof of Lemma A.8(iv) that $v_{1kNT,1} = O_P\left(\delta_{NT}^{-3}\right)$ and $v_{2kNT,1} = O_P\left(\delta_{NT}^{-3}\right)$ under the preliminary convergence rate $\text{vec}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) = O_P\left(T^{-1/2}\right)$. With the faster convergence rate, $\text{vec}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$, obtained in Theorem 3.5, we have $\frac{1}{T}\left\|(\hat{F} - F^0 H)H' F^{0\prime}\right\| = O_P\left(T^{-1/2}\delta_{NT}^{-1} + \delta_{NT}^{-2}\right) = O_P\left(\delta_{NT}^{-2}\right)$ and $\left\|I_{R^0} - \left(T^{-1}H' F^{0\prime} F^0 H\right)^{-1}\right\| = O_P(T^{-1/2}\ \delta_{NT}^{-1} + \delta_{NT}^{-2}) = O_P\left(\delta_{NT}^{-2}\right)$ by Lemma A.5. With these improved rates, we can obtain an improved stochastic order for $v_{1kNT,2}$, $v_{1kNT,4}$, $v_{2kNT,2}$, and $v_{2kNT,4}$: $v_{1kNT,2} = O_P\left(\delta_{NT}^{-3}\right)$, $v_{1kNT,4} = O_P\left(T^{-1/2}\delta_{NT}^{-2}\right)$, $v_{2kNT,2} = O_P\left(\delta_{NT}^{-3}\right)$, and $v_{2kNT,4} = O_P\left(T^{-1/2}\delta_{NT}^{-2}\right)$. We now study $v_{1kNT,3}$ and $v_{2kNT,3}$. For $v_{1kNT,3}$, we make the following decomposition

$$
\begin{aligned}
v_{1kNT,3} &= \frac{-1}{N_k T} \sum_{i \in G_k^0} X_i' d_3 \varepsilon_i = \frac{-1}{N_k T^2} \sum_{i \in G_k^0} X_i' F^0 H H' (\hat{F} H^{-1} - F^0)' \varepsilon_i \\
&= \frac{-1}{N_k T^2} \sum_{i \in G_k^0} X_i' F^0 \left(T^{-1} F^{0\prime} F^0\right)^{-1} (\hat{F} H^{-1} - F^0)' \varepsilon_i \\
&\quad + \frac{-1}{N_k T^2} \sum_{i \in G_k^0} X_i' F^0 \left[\left(T^{-1} F^{0\prime} F^0\right)^{-1} - H H'\right] (\hat{F} H^{-1} - F^0)' \varepsilon_i \\
&\equiv v_{1kNT,3}(1) + v_{1kNT,3}(2).
\end{aligned}
$$

$v_{1kNT,3}(1)$ is studied in Lemma A.9(ii). By Lemma A.5 and Theorem 3.5 and following the proof of Lemma A.7 in Bai (2009), we can readily show that $\left(T^{-1}F^{0\prime}F^0\right)^{-1} - HH' = O_P\left(T^{-1/2}\delta_{NT}^{-1} + \delta_{NT}^{-2}\right) = O_P\left(\delta_{NT}^{-2}\right)$. Then by Lemma A.7(iii)

$$
\begin{aligned}
\left|\omega_p' v_{1kNT,3}(2)\right| &\leq \frac{1}{N_k T^2} \left| \sum_{i \in G_k^0} \omega_p' X_i' F^0 \left[\left(T^{-1} F^{0\prime} F^0\right)^{-1} - H H'\right] (\hat{F} H^{-1} - F^0)' \varepsilon_i \right| \\
&\leq \left\{ \frac{1}{N_k T^2} \sum_{i \in G_k^0} \omega_p' X_i' F^0 \left[\left(T^{-1} F^{0\prime} F^0\right)^{-1} - H H'\right] \left[\left(T^{-1} F^{0\prime} F^0\right)^{-1} - H H'\right] F^0 X_i \omega_p \right\}^{1/2} \\
&\quad \times \left\{ \frac{1}{N_k T^2} \sum_{i \in G_k^0} \varepsilon_i (\hat{F} H^{-1} - F^0)(\hat{F} H^{-1} - F^0)' \varepsilon_i \right\}^{1/2} \\
&\leq \left\| \left(T^{-1} F^{0\prime} F^0\right)^{-1} - H H' \right\| \left\{ \frac{1}{N_k T^2} \sum_{i \in G_k^0} \left\| X_i' F^0 \right\|^2 \right\}^{1/2} O_P\left(\delta_{NT}^{-2}\right) \\
&= O_P\left(\delta_{NT}^{-2}\right) O_P\left(\delta_{NT}^{-2}\right) = O_P\left(\delta_{NT}^{-4}\right).
\end{aligned}
$$

It follows that $v_{1kNT,3} = -\frac{1}{N N_k T^2} \sum_{i \in G_k^0} \sum_{j=1}^{N} X_i' F^0 \left(T^{-1} F^{0\prime} F^0\right)^{-1} (N^{-1} \Lambda^{0\prime} \Lambda^0)^{-1} \lambda_j^0 \varepsilon_j' \varepsilon_i + o_P((NT)^{-1/2})$. Analogously, by Lemma A.9 and using $\left(T^{-1}F^{0\prime}F^0\right)^{-1} - HH' = O_P\left(\delta_{NT}^{-2}\right)$, we can show that

$$
\begin{aligned}
v_{2kNT,3} &= \frac{-1}{N_k T} \sum_{i \in G_k^0} X_i^{*\prime} d_3 \varepsilon_i = \frac{-1}{N_k T^2} \sum_{i \in G_k^0} X_i^{*\prime} F^0 H (\hat{F} - F^0 H)' \varepsilon_i \\
&= \frac{-1}{N N_k T^2} \sum_{i \in G_k^0} \sum_{j=1}^{N} X_i^{*\prime} F^0 \left(T^{-1} F^{0\prime} F^0\right)^{-1} (N^{-1} \Lambda^{0\prime} \Lambda^0)^{-1} \lambda_j^0 \varepsilon_j' \varepsilon_i + o_P\left((NT)^{-1/2}\right).
\end{aligned}
$$

It follows that $\hat{V}_{kNT} - V_{kNT} = \mathcal{B}_{2,kNT} + o_P((NT)^{-1/2})$ and (i) follows. Now, by Lemma A.8(ii), $\mathcal{B}_{1NT} = O_P(T^{-1})$. Lemmas A.8(iii)-(iv), in conjunction with the result in (i), imply that $\mathcal{B}_{2,NT} = O_P(T^{-1/2}\delta_{NT}^{-1})$. By Assumption A.5(ii), $\mathcal{B}_{3,NT} = O_P(T^{-1/2}\delta_{NT}^{-1})$. It follows that $\mathcal{B}_{NT} = O_P(T^{-1/2}\delta_{NT}^{-1})$. That is, (ii) follows. ■

Proof of Theorem 3.7. The post Lasso estimates $\tilde{F}$ and $\tilde{\alpha}_k$ based on the group identity estimates $\{\hat{G}_k, k = 1, ..., K_0\}$ are also obtained via PCA. This implies that for $k = 1, ..., K_0$,

$$\frac{1}{\hat{N}_k T}\sum_{i\in\hat{G}_k} X_i' M_{\tilde{F}} X_i \tilde{\alpha}_k^0 = \frac{1}{\hat{N}_k T}\sum_{i\in\hat{G}_k} X_i' M_{\tilde{F}} Y_i, \tag{A.27}$$

$$\left[\frac{1}{\hat{N}T}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k}\left(Y_i - X_i\tilde{\alpha}_k\right)\left(Y_i - X_i\tilde{\alpha}_k\right)'\right]\tilde{F} = \tilde{F}\tilde{V}_{NT} \text{ and } \tilde{\Lambda} = (\tilde{\lambda}_1, \tilde{\lambda}_2, ..., \tilde{\lambda}_{\hat{N}}), \tag{A.28}$$

where $\tilde{V}_{NT}$ is a diagonal matrix consisting of the R largest eigenvalues of the above matrix in the square bracket, arranged in descending order, and $\tilde{\lambda}_i = T^{-1}\tilde{F}'\left(Y_i - X_i\tilde{\alpha}_k\right)$ for $i \in \hat{G}_k$ and $k = 1, ..., K_0$. Following the derivation of (A.22), we can derive from (A.27) to obtain that

$$\frac{1}{N_k T}\sum_{i\in\hat{G}_k} X_i' M_{\tilde{F}} X_i\left(\tilde{\alpha}_k - \alpha_k^0\right) = \frac{1}{N_k T}\sum_{i\in\hat{G}_k} X_i' M_{\tilde{F}} F^0\lambda_i^0 + \frac{1}{N_k T}\sum_{i\in\hat{G}_k} X_i' M_{\tilde{F}}\varepsilon_i + \tilde{\mathcal{R}}_{1k} - \tilde{\mathcal{R}}_{2k}. \tag{A.29}$$

where $\tilde{\mathcal{R}}_{1k} = \frac{1}{NT}\sum_{i\in\hat{G}_k\backslash G_k^0} X_i' M_{\tilde{F}} X_i\beta_i^0$ and $\tilde{\mathcal{R}}_{2k} = \frac{1}{NT}\sum_{i\in G_k^0\backslash\hat{G}_k} X_i' M_{\tilde{F}} X_i\beta_i^0$. By arguments like those used in the derivation of (A.23), we can show that $\tilde{\mathcal{R}}_{sk} = o_P((NT)^{-1/2})$ for $s = 1, 2$, and $k = 1, ..., K_0$. By the uniform selection consistency obtained in Theorem 3.4, we can show that $\tilde{F}$ shares the same first order asymptotic properties as $\hat{F}$ in that the results in Lemmas A.2-A.9 continue to hold with $\hat{F}$ replaced by $\tilde{F}$. Consequently, following the proof of Theorem 3.5 we can show that

$$\text{vec}\left(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0\right) = \tilde{Q}_{NT}^{-1}\left[\tilde{V}_{NT} + \tilde{\mathcal{B}}_{1NT} + o_P((NT)^{-1/2})\right],$$

where $\tilde{Q}_{NT}$, $\tilde{V}_{NT}$, and $\tilde{\mathcal{B}}_{1NT}$ are analogously defined as $\hat{Q}_{NT}$, $\hat{V}_{NT}$, and $\mathcal{B}_{1NT}$ with $\hat{F}$ replaced by $\tilde{F}$ everywhere. The results then follow by arguments as used in the proof of Theorem 3.6. ■

Proof of Proposition 3.8. (i) We verify that $V_{kNT} = \mathcal{V}_{kNT} + \mathcal{B}_{3,kNT} + o_P((NT)^{-1/2})$. Observe that

$$\begin{aligned}
V_{kNT} &= \frac{1}{N_k T}\sum_{i\in G_k^0} X_i' M_{F^0}\varepsilon_i - \frac{1}{NN_k T}\sum_{i\in G_k^0} X_i' M_{F^0}\sum_{j=1}^{N}\nu_{ji}\varepsilon_j \\
&= \frac{1}{N_k T}\sum_{i\in G_k^0} X_i'\varepsilon_i - \frac{1}{N_k T}\sum_{i\in G_k^0} X_i' P_{F^0}\varepsilon_i - \frac{1}{NT}\sum_{i=1}^{N}\left[\frac{1}{N_k}\sum_{j\in G_k^0}\nu_{ji} X_j' M_{F^0}\right]\varepsilon_i \\
&= \left\{\frac{1}{N_k T}\sum_{i\in G_k^0} X_i'\varepsilon_i - \frac{1}{N_k T}\sum_{i\in G_k^0} E_{\mathcal{D}}\left(X_i'\right) P_{F^0}\varepsilon_i - \frac{1}{NT}\sum_{i=1}^{N}\frac{1}{N_k}\sum_{j\in G_k^0}\nu_{ji} E_{\mathcal{D}}(X_j') M_{F^0}\varepsilon_i\right\} \\
&\quad + \frac{1}{N_k T}\sum_{i\in G_k^0} E_{\mathcal{D}}\left(X_i' P_{F^0}\varepsilon_i\right) + \frac{1}{N_k T}\sum_{i\in G_k^0}\left\{\left[X_i - E_{\mathcal{D}}\left(X_i\right)\right]' P_{F^0}\varepsilon_i - E_{\mathcal{D}}\left(X_i' P_{F^0}\varepsilon_i\right)\right\} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\left\{\frac{1}{N_k}\sum_{j\in G_k^0}\nu_{ji}\left[X_j' - E_{\mathcal{D}}(X_j')\right] M_{F^0}\right\}\varepsilon_i \\
&= \mathcal{V}_{kNT} + \mathcal{B}_{3,kNT} + \mathcal{R}_{1,kNT} + \mathcal{R}_{2,kNT},
\end{aligned}$$

where $\mathcal{V}_{kNT}$ and $\mathcal{B}_{3,kNT}$ are defined in Section 3.3, $Z_{ki} = \frac{N}{N_k}[X_i - P_{F^0} E_{\mathcal{D}}(X_i)] \mathbf{1}\{i \in G_k^0\} - \frac{1}{N_k} M_{F^0} \sum_{j \in G_k^0} \nu_{ji} E_{\mathcal{D}}(X_j)$, $\mathcal{R}_{1,kNT} = \frac{1}{N_k T} \sum_{i \in G_k^0} \{[X_i - E_{\mathcal{D}}(X_i)]' P_{F^0} \varepsilon_i - E_{\mathcal{D}}(X_i' P_{F^0} \varepsilon_i)\}$, and $\mathcal{R}_{2,kNT} = \frac{1}{NT} \sum_{i=1}^{N} \{\frac{1}{N_k} \sum_{j \in G_k^0} \nu_{ji} [X_j' - E_{\mathcal{D}}(X_j')] M_{F^0}\} \varepsilon_i$. It suffices to prove that (i1) $\mathcal{R}_{1,kNT} = o_P((NT)^{-1/2})$ and (i2) $\mathcal{R}_{2,kNT} = o_P((NT)^{-1/2})$. For the proof of (i1), see the proof of Corollary 3.4 in Lu and Su (2016). To prove (i2), write

$$\mathcal{R}_{2,kNT} = \frac{1}{NT} \sum_{i=1}^{N} \left[\frac{1}{N_k} \sum_{j \in G_k^0} \nu_{ji} \bar{X}_j' \right] \varepsilon_i - \frac{1}{NT} \sum_{i=1}^{N} \left[\frac{1}{N_k} \sum_{j \in G_k^0} \nu_{ji} \bar{X}_j' P_{F^0} \right] \varepsilon_i \equiv \mathcal{R}_{2,kNT,1} - \mathcal{R}_{2,kNT,2}, \text{ say.}$$

Let $\bar{X}_{jt}'$ denote the tth row of $X_j - E_{\mathcal{D}}(X_j)$. Then

$$\begin{aligned} ER &\equiv E_{\mathcal{D}} \left[\omega_p' \mathcal{R}_{2,kNT,1} \right]^2 = \frac{1}{N_k^2 N^2 T^2} E_{\mathcal{D}} \left[\sum_{i=1}^{N} \sum_{j \in G_k^0} \sum_{t=1}^{T} \nu_{ji} \omega_p' \bar{X}_{jt} \varepsilon_{it} \right]^2 \\ &= \frac{1}{N_k^2 N^2 T^2} \sum_{i=1}^{N} \sum_{j \in G_k^0} \sum_{l=1}^{N} \sum_{m \in G_k^0} \sum_{t=1}^{T} \sum_{s=1}^{T} \nu_{ji} \nu_{ml} E_{\mathcal{D}} \left[\omega_p' \bar{X}_{jt} \varepsilon_{it} \omega_p' \bar{X}_{ms} \varepsilon_{ls} \right]. \end{aligned}$$

Let $S = \{i, j, l, m\}$. We consider two cases: (a) $\#S \geq 3$, (b) $\#S = 2$, and (c) $\#S = 1$. We use $ER(a)$, $ER(b)$, and $ER(c)$ to denote ER when the individual indices in S are restricted to cases (a), (b), and (c), respectively. Note that $ER(a) = 0$ under Assumption B.2. In view of the fact that $E_{\mathcal{D}} \left[\omega_p' \bar{X}_{jt} \varepsilon_{it} \right] = 0$ for all i, j, and t, we have

$$\begin{aligned} ER(b) &= \frac{1}{N_k^2 N^2 T^2} \sum_{i=1}^{N} \sum_{j \in G_k^0} \sum_{t=1}^{T} \sum_{s=1}^{T} \nu_{ji} \nu_{ji} E_{\mathcal{D}} \left[\omega_p' \bar{X}_{jt} \omega_p' \bar{X}_{js} \right] E_{\mathcal{D}} \left[\varepsilon_{it} \varepsilon_{is} \right] \\ &\quad + \frac{1}{N_k^2 N^2 T^2} \sum_{i \in G_k^0} \sum_{j \in G_k^0} \sum_{t=1}^{T} \sum_{s=1}^{T} \nu_{ji} \nu_{ij} E_{\mathcal{D}} \left[\varepsilon_{it} \omega_p' \bar{X}_{is} \right] E_{\mathcal{D}} \left[\omega_p' \bar{X}_{jt} \varepsilon_{js} \right] \\ &= \frac{1}{N_k^2 N^2 T^2} \sum_{i=1}^{N} \sum_{j \in G_k^0} \sum_{t=1}^{T} \nu_{ji}^2 E_{\mathcal{D}} \left[\left(\omega_p' \bar{X}_{jt} \right)^2 \right] E_{\mathcal{D}} \left[\varepsilon_{it}^2 \right] + 0 = O_P \left(N^{-2} T^{-1} \right), \end{aligned}$$

where the second inequality holds because $E_{\mathcal{D}} [\varepsilon_{it} \varepsilon_{is}] = 0$ if $t \neq s$, $E_{\mathcal{D}} \left[\varepsilon_{it} \omega_p' \bar{X}_{is} \right] = 0$ if $t \geq s$, and $E_{\mathcal{D}} \left[\omega_p' \bar{X}_{jt} \varepsilon_{js} \right] = 0$ if $s \geq t$. Next,

$$\begin{aligned} ER(c) &= \frac{1}{N_k^2 N^2 T^2} \sum_{i \in G_k^0} \sum_{t=1}^{T} \sum_{s=1}^{T} \nu_{ii}^2 E_{\mathcal{D}} \left[\omega_p' \bar{X}_{it} \varepsilon_{it} \omega_p' \bar{X}_{is} \varepsilon_{is} \right] \\ &= \frac{1}{N_k^2 N^2 T^2} \sum_{i \in G_k^0} \sum_{t=1}^{T} \nu_{ii}^2 E_{\mathcal{D}} \left[\left(\omega_p' \bar{X}_{it} \right)^2 \varepsilon_{it}^2 \right] = O_P \left(N^{-3} T^{-1} \right). \end{aligned}$$

It follows that $ER = O_P \left(N^{-2} T^{-1} \right)$ and hence $\mathcal{R}_{2,kNT,1} = O_P \left(N^{-1} T^{-1/2} \right) = o_P((NT)^{-1/2})$. Analogously, we can show that $\mathcal{R}_{2,kNT,2} = o_P((NT)^{-1/2})$. Then $\mathcal{R}_{2,kNT} = o_P((NT)^{-1/2})$.

(ii) We verify that $\mathcal{B}_{3,kNT} = O_P(T^{-1})$. Observe that $\mathcal{B}_{3,kNT} \equiv \frac{1}{N_k T} \sum_{i \in G_k^0} E_{\mathcal{D}} (X_i' P_{F^0} \varepsilon_i) = \frac{1}{N_k T^2} \sum_{i \in G_k^0} \sum_{t=2}^{T} \sum_{s=1}^{t} \mu_{ts} E_{\mathcal{D}} (X_{it} \varepsilon_{is})$. By Davydov inequality for conditional strong mixing processes and Cauchy-

Schwarz inequality,

$$
\begin{aligned}
\left|\omega_p' \mathcal{B}_{3,kNT}\right| &\le \frac{1}{N_k T^2} \sum_{i\in G_k^0} \left| \sum_{t=2}^T \sum_{s=1}^t \mu_{ts} E_{\mathcal{D}} \left[\omega_p' X_{it} \varepsilon_{is}\right] \right| \\
&\le \frac{8}{N_k T^2} \sum_{i\in G_k^0} \sum_{t=2}^T \sum_{s=1}^t \left\|F_t^0\right\| \left\|F_s^0\right\| \left\|\bar{X}_{it}\right\|_{8+4\sigma,\mathcal{D}} \left\|\varepsilon_{is}\right\|_{8+4\sigma,\mathcal{D}} \alpha_{NT}^{\mathcal{D}} (t-s)^{(4+2\sigma)/(3+2\sigma)} \\
&\le \frac{4}{N_k T^2} \sum_{i\in G_k^0} \sum_{t=2}^T \left\|F_t^0\right\|^2 \left\|\bar{X}_{it}\right\|^2_{8+4\sigma,\mathcal{D}} \sum_{\tau=1}^\infty \alpha_{NT}^{\mathcal{D}} (\tau)^{(4+2\sigma)/(3+2\sigma)} \\
&\quad + \frac{4}{N_k T^2} \sum_{i\in G_k^0} \sum_{s=1}^{T-1} \left\|F_s^0\right\|^2 \left\|\varepsilon_{is}\right\|^2_{8+4\sigma,\mathcal{D}} \sum_{\tau=1}^\infty \alpha_{NT}^{\mathcal{D}} (\tau)^{(4+2\sigma)/(3+2\sigma)} \\
&= O_P\left(T^{-1}\right) + O_P\left(T^{-1}\right) = O_P\left(T^{-1}\right).
\end{aligned}
$$

Thus $\mathcal{B}_{3,kNT} = O_P\left(T^{-1}\right)$.

(iii) We show that $\sqrt{NT}\mathcal{V}_{NT} \xrightarrow{d} N(0,\Omega_0)$. Let $Z_{k,it}'$ denote the tth row of $Z_{k,i} \equiv \frac{N}{N_k}[X_i - P_{F^0} E_{\mathcal{D}}(X_i)] \times \mathbf{1}\left\{i \in G_k^0\right\} - \frac{1}{N_k} M_{F^0} \sum_{j\in G_k^0} \nu_{ji} E_{\mathcal{D}}(X_j)$. That is $Z_{k,it} = \frac{N}{N_k}[X_{it} - \frac{1}{T}\sum_{s=1}^T \mu_{st} E_{\mathcal{D}}(X_{is})]\mathbf{1}\left\{i\in G_k^0\right\} - \frac{1}{N_k}\frac{1}{T}\sum_{s=1}^T \left[\mathbf{1}\left\{t=s\right\} - \mu_{st}\right] \sum_{j\in G_k^0} \nu_{ji} E_{\mathcal{D}}(X_{jt})$. Let $Z_{it} = (Z_{1,i1}', ..., Z_{K_0,iT}')'$ and $Z_i = (Z_{1,i}, ..., Z_{K_0,i})$. Then $\sqrt{NT}\mathcal{V}_{NT} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N Z_i'\varepsilon_i = \frac{1}{\sqrt{NT}}\sum_{t=1}^T \sum_{i=1}^N Z_{it}\varepsilon_{it} = \sum_{t=1}^T \xi_{NT,t}$, where $\xi_{NT,t} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N Z_{it}\varepsilon_{it}$. By Assumption B.2(iii), $E\left(\xi_{NT,t}|\mathcal{F}_{NT,t-1}\right) = \frac{1}{\sqrt{NT}}\sum_{i=1}^N Z_{it} E\left(\varepsilon_{it}|\mathcal{F}_{NT,t-1}\right) = 0$. That is, $\left\{\xi_{NT,t}, \mathcal{F}_{NT,t}\right\}$ is an m.d.s. By the martingale CLT (e.g., Pollard, 1984, p. 171) and Cramér-Wold device, it suffices to show that

$$
\mathcal{Z}_1 \equiv \sum_{t=1}^T E\left[\left|\omega'\xi_{NT,t}\right|^4 |\mathcal{F}_{NT,t-1}\right] = o_P(1) \text{ and } \mathcal{Z}_2 \equiv \sum_{t=1}^T \left|\omega'\xi_{NT,t}\right|^2 - \omega'\Omega_{NT}\omega = o_P(1) \tag{A.30}
$$

for any nonrandom $K_0 p \times 1$ real vector ω with $\|\omega\| = 1$, where $E_{\mathcal{F}_{NT,t-1}}$ denote expectation conditional on $\mathcal{F}_{NT,t-1}$. Observing that $\mathcal{Z} \ge 0$, it suffices to show the first part of (A.30) by showing that $E_{\mathcal{D}}(\mathcal{Z}_1) = o_P(1)$ by conditional Markov inequality. Noting that $\{Z_{it}, \varepsilon_{it}\}$ are independent across i given $\mathcal{D}$ by Assumption B.2(iii) and $E_{\mathcal{D}}(Z_{it}\varepsilon_{it}) = 0$, we have

$$
\begin{aligned}
E_{\mathcal{D}}(\mathcal{Z}_1) &= \frac{1}{N^2T^2}\sum_{t=1}^T E_{\mathcal{D}}\left[\left|\sum_{i=1}^N \omega' Z_{it}\varepsilon_{it}\right|^4\right] \\
&= \frac{1}{N^2T^2}\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N E_{\mathcal{D}}\left[\omega' Z_{it}\omega' Z_{jt}\omega' Z_{kt}\omega' Z_{lt}\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}\right] \\
&= \frac{3}{N^2T^2}\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1,j\ne i}^N E_{\mathcal{D}}[(\omega' Z_{it})^2 \varepsilon_{jt}^2] E_{\mathcal{D}}[(\omega' Z_{jt})^2 \varepsilon_{jt}^2] + \frac{1}{N^2T^2}\sum_{t=1}^T\sum_{i=1}^N E_{\mathcal{D}}[(\omega' Z_{it})^4 \varepsilon_{it}^4] \\
&= O_P(T^{-1}) + O_P(N^{-1}T^{-1}) = o_P(1).
\end{aligned}
$$

Note that

$$
\begin{aligned}
\mathcal{Z}_2 &= \omega'\sum_{t=1}^T \xi_{NT,t}\xi_{NT,t}'\omega - \omega'\Omega_{NT}\omega = \frac{1}{NT}\omega'\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1}^N Z_{it}Z_{jt}'\varepsilon_{it}\varepsilon_{jt}\omega - \omega'\Omega_{NT}\omega \\
&= \frac{1}{NT}\omega'\sum_{t=1}^T\sum_{1\le i<j\le N} Z_{it}Z_{jt}'\varepsilon_{it}\varepsilon_{jt}\omega + \frac{1}{NT}\omega'\sum_{t=1}^T\sum_{1\le j<i\le N} Z_{it}Z_{jt}'\varepsilon_{it}\varepsilon_{jt}\omega \equiv \mathcal{Z}_{21} + \mathcal{Z}_{22}, \text{ say,}
\end{aligned}
$$

where $\sum_{1\leq i<j\leq N}=\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}$.

$$
\begin{aligned}
E_{\mathcal{D}}\left(\mathcal{Z}_{21}^{2}\right) &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{1\leq i<j\leq N}\sum_{1\leq k<l\leq N}E_{\mathcal{D}}\left(\omega' Z_{it}Z_{jt}'\omega\varepsilon_{it}\varepsilon_{jt}\ \omega' Z_{ks}Z_{ls}'\omega\varepsilon_{ks}\varepsilon_{ls}\right) \\
&= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{1\leq i<j\leq N}E_{\mathcal{D}}\left(\omega' Z_{it}\ \omega' Z_{is}\varepsilon_{it}\varepsilon_{is}\ \omega' Z_{jt}\omega' Z_{js}\varepsilon_{jt}\varepsilon_{js}\right) \\
&= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{1\leq i<j\leq N}E_{\mathcal{D}}[(\omega' Z_{it})^2\,\varepsilon_{it}^2]E_{\mathcal{D}}[(\omega' Z_{jt})^2\,\varepsilon_{jt}^2]=O_P(T^{-1}).
\end{aligned}
$$

Similarly, $E_{\mathcal{D}}\left(\mathcal{Z}_{22}^{2}\right)=O_P\left(T^{-1}\right)$. Thus $\mathcal{Z}_2=\omega'\sum_{t=1}^{T}\xi_{NT,t}\xi_{NT,t}'\omega-\omega'\Omega_{NT}\omega=o_P(1)$ and (iii) follows.

(iv) We verify that $\left\|\frac{1}{NT}\sum_{i=1}^{N}F^{0\prime}\varepsilon_i\varepsilon_i'\right\|=O_P\left(\delta_{NT}^{-1}\right)$. Note that

$$
\begin{aligned}
E_{\mathcal{D}}\left\|\frac{1}{NT}\sum_{i=1}^{N}F^{0\prime}\varepsilon_i\varepsilon_i'\right\|^2 &= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E_{\mathcal{D}}\left[\operatorname{tr}(F^{0\prime}\varepsilon_i\varepsilon_i'\varepsilon_j\varepsilon_j'F^0)\right] \\
&= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}E_{\mathcal{D}}\left(\varepsilon_{it}\varepsilon_{is}\varepsilon_{js}\varepsilon_{jr}\right)F_r^{0\prime}F_t^0 \\
&= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}E_{\mathcal{D}}\left(\varepsilon_{it}\varepsilon_{is}^2\varepsilon_{ir}\right)F_r^{0\prime}F_t^0 \\
&\quad +\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{s=1}^{T}E_{\mathcal{D}}\left(\varepsilon_{is}^2\right)E_{\mathcal{D}}(\varepsilon_{js}^2)F_s^{0\prime}F_s^0.
\end{aligned}
$$

By Davydov inequality for conditional strong mixing processes, we can show the first term in the last expression is $O_P\left(N^{-1}\right)$. The second term is $O_P\left(T^{-1}\right)$. It follows that $E_{\mathcal{D}}\left\|\frac{1}{NT}\sum_{i=1}^{N}F^{0\prime}\varepsilon_i\varepsilon_i'\right\|^2=O_P\left(\delta_{NT}^{-2}\right)$ and the result follows by conditional Chebyshev inequality. ■

Proof of Corollary 3.9. In view of the fact that

$$
\begin{aligned}
\sqrt{NT}\operatorname{vec}\left(\hat{\boldsymbol{\alpha}}^{bc}-\boldsymbol{\alpha}^0\right) &= \sqrt{NT}\left[\operatorname{vec}\left(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0\right)-Q_{NT}^{-1}\left(\mathcal{B}_{1,NT}+\mathcal{B}_{2,NT}+\mathcal{B}_{3,NT}\right)\right] \\
&\quad -\sqrt{NT}\left[\hat{Q}_{NT}^{-1}(\hat{\mathcal{B}}_{1,NT}+\hat{\mathcal{B}}_{2,NT}+\hat{\mathcal{B}}_{3,NT})-Q_{NT}^{-1}\left(\mathcal{B}_{1,NT}+\mathcal{B}_{2,NT}+\mathcal{B}_{3,NT}\right)\right],
\end{aligned}
$$

it suffices to show that (i) $\hat{\mathcal{B}}_{1NT}-\mathcal{B}_{1NT}=o_P\left((NT)^{-1/2}\right)$, (ii) $\hat{\mathcal{B}}_{2NT}-\mathcal{B}_{2NT}=o_P\left((NT)^{-1/2}\right)$, (iii) $\hat{\mathcal{B}}_{3NT}-\mathcal{B}_{3NT}=o_P\left((NT)^{-1/2}\right)$, (iv) $\hat{\Theta}_{NT}=\Theta_{NT}+o_P(1)$, and (v) $\hat{Q}_{NT}-Q_{NT}=O_P\left(\delta_{NT}^{-1}\right)$. (v) holds by Lemmas A.6(vi) and A.8(i). We now show (i)-(v) in order.

First, we prove (i) $\hat{\mathcal{B}}_{1NT}-\mathcal{B}_{1NT}=o_P\left((NT)^{-1/2}\right)$. Let $\hat{\mathcal{B}}_{1,kNT}$ and $\mathcal{B}_{1,kNT}$ denote the kth block of $\hat{\mathcal{B}}_{1NT}$ and $\mathcal{B}_{1NT}$, respectively. Let $\bar{\mathcal{B}}_{1,kNT}=-\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}\Phi_{NT}\hat{F}G\lambda_i^0$. We prove (i) by showing that (i1) $\sqrt{N_k T}(\bar{\mathcal{B}}_{1,kNT}-\mathcal{B}_{1,kNT})=o_P(1)$ and (i2) $\sqrt{N_k T}(\bar{\mathcal{B}}_{1,kNT}-\hat{\mathcal{B}}_{1,kNT})=o_P(1)$ for $k=1,...,K_0$.

We make the following decomposition:

$$
\begin{aligned}
& \sqrt{N_k T}(\bar{\mathcal{B}}_{1,kNT} - \mathcal{B}_{1,kNT}) \\
= \; & \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in\hat{G}_k} X_i' M_{\hat{F}} \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) \hat{F} G \lambda_i^0 \\
= \; & \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in\hat{G}_k} X_i' \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) F^0 H G \lambda_i^0 + \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in\hat{G}_k} X_i' \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) (\hat{F} - F^0 H) G \lambda_i^0 \\
& - \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in\hat{G}_k} X_i' F^0 (F^{0\prime}F^0)^{-1} F^{0\prime} \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) F^0 H G \lambda_i^0 \\
& - \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in\hat{G}_k} X_i' F^0 (F^{0\prime}F^0)^{-1} F^{0\prime} \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) (\hat{F} - F^0 H) G \lambda_i^0 \\
& + \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in\hat{G}_k} X_i' \left(M_{\hat{F}} - M_{F^0}\right) \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) F^0 H G \lambda_i^0 \\
\equiv \; & b_{11} + b_{12} - b_{13} - b_{14} + b_{15}, \text{ say.}
\end{aligned}
$$

Define $\bar{b}_{1s}$ analogously as b_{1s} for $s = 1, ..., 5$, with the summation $\sum_{i\in\hat{G}_k}$ replaced by $\sum_{i\in G_k^0}$. Following the proof of Lemma A.5(i), we can readily show that $\bar{b}_{1s} = b_{1s} + o_P(1)$ for $s = 1, ..., 5$. Let $\omega_p = (\omega_{1p}, ..., \omega_{pp})'$ be an arbitrary $p \times 1$ nonrandom vector with $\|\omega_p\| = 1$. Let $\mathbf{X}_{lk}$ denote an $N_k \times T$ matrix with a typical element $X_{it,l}$ for $i \in G_k^0$, $t = 1, ..., T$, and $l = 1, ..., p$, where $X_{it,l}$ denotes the lth element in X_{it}. Let Λ_k^0 denote an $N_k \times R$ matrix with a typical row $\lambda_i^{0\prime}$ for $i \in G_k^0$. Then

$$
\begin{aligned}
\left|\omega_p' \bar{b}_{11}\right| &= \left| \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in G_k^0} \omega_p' X_i' \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) F^0 H G \lambda_i^0 + o_P(1) \right| \\
&= \left| \frac{1}{N_k^{1/2}T^{3/2}} \sum_{l=1}^{p} \omega_{lp} \mathrm{tr}\left[\mathbf{X}_{lk}' \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) F^0 H G \Lambda_k^{0\prime}\right] + o_P(1) \right| \\
&\leq \left\{ \sum_{l=1}^{p} \frac{1}{T^{3/2}} \left\| \mathbf{X}_{lk}' \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) F^0 \right\| \right\} \|HG\| \frac{1}{N_k^{1/2}} \left\|\Lambda_k^0\right\| + o_P(1) \\
&= O(1) \left\{ O_P(N_k^{-1/2}) + o_P\left(T^{-(7+4\sigma)/(16+8\sigma)}\right) \right\} O_P(1) O_P(1) + o_P(1) = o_P(1),
\end{aligned}
$$

where we use the fact that $\frac{1}{T^{3/2}} \left\| \mathbf{X}_{lk} \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) F^0 \right\| = O_P(N_k^{-1/2}) + o_P\left(T^{-(7+4\sigma)/(16+8\sigma)}\right)$ by Lemma D.4(iii) in Lu and Su (2016). It follows that $b_{11} = o_P(1)$. Similarly, using $\frac{1}{T^{3/2}} \left\| \mathbf{X}_{lk} \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) \right\|$ $= O_P(N_k^{-1/2}) + o_P\left(T^{-(7+4\sigma)/(16+8\sigma)}\right)$, Lemma A.2(i), and Theorem 3.2,

$$
\begin{aligned}
\left|\omega_p' \bar{b}_{12}\right| &= \frac{1}{N_k^{1/2}T^{3/2}} \left| \sum_{l=1}^{p} \omega_{lp} \mathrm{tr}\left[\mathbf{X}_{lk} \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) (\hat{F} - F^0 H) G \Lambda_k^{0\prime}\right] \right| + o_P(1) \\
&\leq p T^{1/2} \left\{ \frac{1}{T^{3/2}} \left\| \mathbf{X}_{lk} \left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right) \right\| \right\} \frac{1}{T^{1/2}} \left\| \hat{F} - F^0 H \right\| \frac{1}{N_k^{1/2}} \left\|\Lambda_k^0\right\| \|G\| \\
&= T^{1/2} O_P(1) \left\{ O_P(N_k^{-1/2}) + o_P\left(T^{-(7+4\sigma)/(16+8\sigma)}\right) \right\} O_P\left(\delta_{NT}^{-1}\right) O_P(1) = o_P(1).
\end{aligned}
$$

Using the fact that $T^{-1}\left\|F^{0\prime}\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}-\Phi_{NT}\right)F^{0}\right\|=O_P(\delta_{NT}^{-1})$ by Lemma D.4(i) in Lu and Su (2016), we have

$$
\begin{aligned}
\left|\omega_p'\bar{b}_{13}\right| &= \frac{1}{N_k^{1/2}T^{5/2}}\sum_{l=1}^{p}\omega_{lp}\mathrm{tr}\left[\mathbf{X}_{lk}F^0(T^{-1}F^{0\prime}F^0)^{-1}F^{0\prime}\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}-\Phi_{NT}\right)F^0HG\Lambda_k^{0\prime}\right]+o_P(1)\\
&\leq \frac{N_k^{1/2}}{T^{1/2}}\frac{1}{N_k^{1/2}T}\left\|\mathbf{X}_{lk}F^0\right\|\left\|(T^{-1}F^{0\prime}F^0)^{-1}\right\|\frac{1}{T}\left\|F^{0\prime}\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}-\Phi_{NT}\right)F^0\right\|\frac{1}{N_k^{1/2}}\left\|HG\Lambda_k^{0\prime}\right\|+o_P(1)\\
&= N_k^{1/2}T^{-1/2}O_P(1)O_P(1)O_P(\delta_{NT}^{-1})O_P(1)+o_P(1)=o_P(1).
\end{aligned}
$$

Similarly, using the fact that $T^{-1}\left\|F^{0\prime}\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}-\Phi_{NT}\right)\right\|=O_P(\delta_{NT}^{-1})$, $\frac{1}{T^{1/2}}\left\|\hat{F}-F^0H\right\|=O_P(\delta_{NT}^{-1})$, and $\left\|M_{\hat{F}}-P_{F^0}\right\|=O_P(\delta_{NT}^{-1})$ we can readily show that $\bar{b}_{14}=o_P(1)$ and $\bar{b}_{15}=o_P(1)$. It follows that $\sqrt{N_kT}(\bar{\mathcal{B}}_{1kNT}-\mathcal{B}_{1kNT})=o_P(1)$. This proves (i1).

To show (i2), notice that

$$
\begin{aligned}
&\sqrt{N_kT}(\bar{\mathcal{B}}_{1,kNT}-\hat{\mathcal{B}}_{1,kNT})\\
&= \frac{1}{N_k^{1/2}T^{3/2}}\sum_{i\in\hat{G}_k}\left[X_i'M_{\hat{F}}\hat{\Phi}_{NT}\hat{F}\hat{G}\hat{\lambda}_i-X_i'M_{\hat{F}}\Phi_{NT}\hat{F}G\lambda_i^0\right]+\frac{\hat{N}_k^{-1/2}-N_k^{-1/2}}{T^{3/2}}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}\hat{\Phi}_{NT}\hat{F}\hat{G}\hat{\lambda}_i\\
&= \frac{1}{N_k^{1/2}T^{3/2}}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}(\hat{\Phi}_{NT}-\Phi_{NT})\hat{F}\hat{G}\hat{\lambda}_i+\frac{1}{N_k^{1/2}T^{3/2}}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}\Phi_{NT}\hat{F}(\hat{G}\hat{\lambda}_i-G\lambda_i^0)\\
&\quad+\frac{\hat{N}_k^{-1/2}-N_k^{-1/2}}{T^{3/2}}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}\hat{\Phi}_{NT}\hat{F}\hat{G}\hat{\lambda}_i\\
&\equiv b_{16}+b_{17}+b_{18},\ \text{say,}
\end{aligned}
$$

where $\hat{G}=(T^{-1}\hat{F}'\hat{F})^{-1}(N^{-1}\hat{\Lambda}'\hat{\Lambda})^{-1}=(N^{-1}\hat{\Lambda}'\hat{\Lambda})^{-1}$. Using the fact that $\left\|\hat{\Phi}_{NT}-\Phi_{NT}\right\|_{\mathrm{sp}}=O_P\left(\delta_{NT}^{-1}T^{1/(8+4\sigma)}\right)$ by Lemma D.4(ii) in Lu and Su (2016), we have

$$
\begin{aligned}
\left\|b_{16}\right\| &\leq \frac{1}{N_k^{1/2}T^{3/2}}\sum_{i\in\hat{G}_k}\left\|X_i'M_{\hat{F}}(\hat{\Phi}_{NT}-\Phi_{NT})\hat{F}\hat{G}\hat{\lambda}_i\right\|_{\mathrm{sp}}\\
&\leq \frac{N_k^{1/2}}{T^{1/2}}\left\|\hat{\Phi}_{NT}-\Phi_{NT}\right\|_{\mathrm{sp}}\frac{1}{N_kT^{1/2}}\sum_{i\in\hat{G}_k}\left\|X_i\right\|\left\|\hat{\lambda}_i\right\|\frac{1}{T^{1/2}}\left\|\hat{F}\hat{G}\right\|\\
&= N_k^{1/2}T^{-1/2}O_P(\delta_{NT}^{-1}T^{1/(8+4\sigma)})O_P(1)O_P(1)=o_P(1),
\end{aligned}
$$

where we also use the fact that $\frac{1}{N_kT^{1/2}}\sum_{i\in\hat{G}_k}\left\|X_i\right\|||\hat{\lambda}_i||\leq\{\frac{1}{N_kT}\sum_{i\in\hat{G}_k}\left\|X_i\right\|^2\}^{1/2}\ \{\frac{1}{N_k}\sum_{i\in\hat{G}_k}||\hat{\lambda}_i||^2\}^{1/2}$ $+o_P(1)=O_P(1)$ by the mean-square convergence of $\{\hat{\lambda}_i\}$. Note that

$$
\begin{aligned}
\hat{G}\hat{\lambda}_i-G\lambda_i^0 &= (T^{-1}\hat{F}'\hat{F})^{-1}(N^{-1}\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\lambda}_i-(T^{-1}H'F^{0\prime}\hat{F})^{-1}(N^{-1}H^{-1}\Lambda^{0\prime}\Lambda^0H'^{-1})^{-1}H^{-1}\lambda_i^0\\
&= \varsigma_1\hat{\lambda}_i+\varsigma_2(\hat{\lambda}_i-H^{-1}\lambda_i^0),
\end{aligned}
$$

where $\varsigma_1=(T^{-1}\hat{F}'\hat{F})^{-1}(N^{-1}\hat{\Lambda}'\hat{\Lambda})^{-1}-(T^{-1}H'F^{0\prime}\hat{F})^{-1}(N^{-1}H^{-1}\Lambda^{0\prime}\Lambda^0H'^{-1})^{-1}=O_P(\delta_{NT}^{-1})$ and $\varsigma_2=(T^{-1}H'F^{0\prime}\hat{F})^{-1}(N^{-1}H^{-1}\Lambda^{0\prime}\Lambda^0H'^{-1})^{-1}=O_P(1)$. By the triangle and Jensen inequalities, the fact that

$\frac{1}{N_k}\sum_{i\in\hat{G}_k}\left\|\hat{\lambda}_i - H^{-1}\lambda_i^0\right\|^2 = O_P\left(T^{-1}\right)$ and $\|\Phi_{NT}\|_{\text{sp}} = O_P\left(1\right),$ we have

$$\begin{aligned}
\|b_{17}\| &= \frac{1}{N_k^{1/2}T^{3/2}}\left\|\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}}\Phi_{NT}\hat{F}(\hat{G}\hat{\lambda}_i - G\lambda_i^0)\right\|_{\text{sp}} \\
&\le \frac{1}{N_k^{1/2}T^{3/2}}\left\|\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}}\Phi_{NT}\hat{F}\varsigma_1\hat{\lambda}_i\right\|_{\text{sp}} + \frac{1}{N_k^{1/2}T^{3/2}}\left\|\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}}\Phi_{NT}\hat{F}\varsigma_2\left(\hat{\lambda}_i - H^{-1}\lambda_i^0\right)\right\|_{\text{sp}} \\
&\le \frac{N_k^{1/2}}{T^{1/2}}\|\Phi_{NT}\|_{\text{sp}}\frac{1}{T^{1/2}}\left\|\hat{F}\right\|\left\{\|\varsigma_1\|\left[\frac{1}{N_kT}\sum_{i\in\hat{G}_k}\|X_i\|^2\right]^{1/2}\left[\frac{1}{N_k}\sum_{i\in\hat{G}_k}\left\|\hat{\lambda}_i\right\|^2\right]^{1/2}\right. \\
&\quad \left. + \|\varsigma_2\|\left[\frac{1}{N_kT}\sum_{i\in\hat{G}_k}\|X_i\|^2\right]^{1/2}\left[\frac{1}{N_k}\sum_{i\in\hat{G}_k}\left\|\hat{\lambda}_i - H^{-1}\lambda_i^0\right\|^2\right]^{1/2}\right\} \\
&= \frac{N_k^{1/2}}{T^{1/2}}O_P\left(1\right)O_P\left(1\right)\left[O_P\left(\delta_{NT}^{-1}\right)O_P\left(1\right) + O_P\left(1\right)O_P(T^{-1/2})\right] = o_P\left(1\right).
\end{aligned}$$

For b_{18}, we use the fact that $\hat{N}_k - N_k = o_P\left(1\right)$ as shown by SSP and that $||\hat{\Phi}_{NT}||_{\text{sp}} \le \|\Phi_{NT}\|_{\text{sp}} + ||\hat{\Phi}_{NT} - \Phi_{NT}||_{\text{sp}} = O_P\left(1\right) + o_P\left(1\right) = O_P\left(1\right)$ to obtain that

$$\begin{aligned}
\|b_{18}\| &= \frac{\left|\hat{N}_k - N_k\right|}{\hat{N}_k^{1/2}N_k^{1/2}(\hat{N}_k^{1/2} + N_k^{1/2})T^{3/2}}\left\|\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}}\hat{\Phi}_{NT}\hat{F}\hat{G}\hat{\lambda}_i\right\| \\
&\le o_P(N_k^{-3/2}T^{-3/2})\left\{\sum_{i\in\hat{G}_k}\|X_i\|\left\|\hat{\lambda}_i\right\|\right\}\left\|\hat{F}\hat{G}\right\|\left\|\hat{\Phi}_{NT}\right\|_{\text{sp}} \\
&= o_P(N_k^{-3/2}T^{-3/2})O_P(N_kT^{1/2})O_P(T^{1/2})O_P\left(1\right) = o_P(N_k^{-1/2}T^{-1/2}).
\end{aligned}$$

Consequently, we have $\sqrt{N_kT}(\bar{\mathcal{B}}_{1,kNT} - \hat{\mathcal{B}}_{1,kNT}) = o_P\left(1\right).$

Next, we prove (ii) $\hat{\mathcal{B}}_{2,NT} - \mathcal{B}_{2,NT} = o_P\left((NT)^{-1/2}\right)$. Let $\bar{\mathcal{B}}_{2,kNT,l} = -\frac{1}{N\hat{N}_kT}\text{tr}\{F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}$ $(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\Lambda^{0\prime}\Psi_{kNT}[\mathbf{X}_l\left(k\right) - \hat{\mathbf{X}}_l^*\left(k\right)]\}$ for $l = 1, ..., p$. We prove (ii) by showing that (ii1) $\sqrt{N_kT}(\bar{\mathcal{B}}_{2,kNT,l} - \mathcal{B}_{2,kNT,l}) = o_P\left(1\right)$ and (ii2) $\sqrt{N_kT}(\bar{\mathcal{B}}_{2,kNT,l} - \hat{\mathcal{B}}_{2,kNT,l}) = o_P\left(1\right)$ for $k = 1, ..., K_0$ and $l = 1, ..., p$. Using arguments as used in the proof of Lemma D.3(vi) in Lu and Su (2016), we can show that

$$N^{-3/2}\left\|\Lambda^{0\prime}\left(T^{-1}\boldsymbol{\varepsilon\varepsilon}\left(k\right)' - \Psi_{kNT}\right)\left[\mathbf{X}_l\left(k\right) - \mathbf{X}_l^*\left(k\right)\right]\right\| = O_P\left(\delta_{NT}^{-1}\right).$$

It follows that

$$\begin{aligned}
&\sqrt{N_kT}\left|\bar{\mathcal{B}}_{2,kNT,l} - \mathcal{B}_{2,kNT,l}\right| \\
&= \frac{1}{NN_k^{1/2}T^{1/2}}\left|\text{tr}\{F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\Lambda^{0\prime}\left(T^{-1}\boldsymbol{\varepsilon\varepsilon}\left(k\right)' - \Psi_{kNT}\right)\left[\mathbf{X}_l\left(k\right) - \mathbf{X}_l^*\left(k\right)\right]\}\right| \\
&\le \frac{N^{1/2}}{N_k^{1/2}T^{1/2}}\left\|F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\right\|\frac{1}{N^{3/2}}\left\|\Lambda^{0\prime}\left(T^{-1}\boldsymbol{\varepsilon\varepsilon}\left(k\right)' - \Psi_{kNT}\right)\left[\mathbf{X}_l\left(k\right) - \mathbf{X}_l^*\left(k\right)\right]\}\right\| \\
&= O_P\left(1\right)O_P(\delta_{NT}^{-1}) = o_P\left(1\right).
\end{aligned}$$

This proves (ii1). The proof of (ii2) is analogous to that of (i2) except that now $\left\|\hat{\Psi}_{kNT} - \Psi_{kNT}\right\|_{\mathrm{sp}} = o_P\left(\delta_{NT}^{-1} N^{1/(8+4\sigma)}\right)$ by Lemma G.3(iii) in Lu and Su (2016).

The proof of (iii) and (iv) are analogous to that of the fourth and fifth parts in the proof of Corollary 3.4 in Lu and Su (2016) and thus omitted. ■

To prove Theorem 4.1, we need the following lemma.

Lemma A.10 *Suppose that the conditions in Theorem 4.1 hold. Then*

(i) there exists a $c_R > 0$ *such that plim inf*$_{(N,T)\to\infty}[V(R, \dot{\boldsymbol{\beta}}_{(R)}) - V(R_0, \dot{\boldsymbol{\beta}}_{(R_0)})] \geq c_R$ *for each* R *with* $1 \leq R < R_0$,

(ii) $V(R, \dot{\boldsymbol{\beta}}_{(R)}) - V(R_0, \dot{\boldsymbol{\beta}}_{(R_0)}) = O_P(\delta_{NT}^{-2})$ *for each* R *with* $R \geq R_0$.

Proof of Theorem 4.1. The proof is similar to that of Corollary 1 in Bai and Ng (2002). For notational simplicity, let $V(R) = V(R, \dot{\boldsymbol{\beta}}_{(R)})$ for all R. Note that

$$IC_{1NT}(R) - IC_{1NT}(R_0) = \ln\left[V(R)/V(R_0)\right] + (R - R_0)\,\rho_{1NT}.$$

We discuss the following two cases: (a) $R < R_0$, and (b) $R_0 < R \leq R_{\max}$.

For case (a), by Lemma A.10(i), $V(R)/V(R_0) > 1 + \epsilon_0$ and thus $\ln\left[V(R)/V(R_0)\right] \geq \epsilon_0/2$ for some $\epsilon_0 > 0$ w.p.a.1. This, in conjunction with the fact that $(R_0 - R)\,\rho_{1NT} \to 0$ under Assumption A.6, implies that $IC_{1NT}(R) - IC_{1NT}(R_0) \geq \epsilon_0/4$ w.p.a.1. It follows that $P\left(IC_{1NT}(R) - IC_{1NT}(R_0) > 0\right) \to 1$ as $(N, T) \to \infty$ for any $R < R_0$.

For case (b), by Lemma A.10(ii) and Assumption A.6, we have

$$\begin{aligned} P\left(IC_{1NT}(R) - IC_{1NT}(R_0) > 0\right) &= P\left(\ln\left[V(R)/V(R_0)\right] + (R - R_0)\,\rho_{1NT} > 0\right) \\ &= P\left(O_P(1) + (R - R_0)\,\rho_{1NT}\delta_{NT}^2 > 0\right) \to 1 \end{aligned}$$

as $(N, T) \to \infty$ for any $R_0 < R \leq R_{\max}$. Consequently, the minimizer of $IC_{1NT}(R)$ can only be achieved at $R = R_0$ w.p.a.1. That is, $P(\hat{R} = R_0) \to 1$ for any $R \in [1, R_{\max}]$ as $(N, T) \to \infty$. ■

To prove Theorem 4.2, we need the following lemma.

Lemma A.11 *Suppose that the conditions in Theorem 4.2 hold. Let* $\bar{\sigma}_{G^0}^2 = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it}^2$. *Then* $\max_{K_0 \leq K \leq K_{\max}} \left|\hat{\sigma}_{\hat{G}(K,\kappa)}^2 - \bar{\sigma}_{G^0}^2\right| = O_P\left(\delta_{NT}^{-2}\right)$.

Proof of Theorem 4.2. The proof is similar to that of Theorem 2.6 in SSP. Using Theorems 3.4 and 3.7 and Assumption A.8, we can readily show that

$$\begin{aligned} IC_{2NT}(K_0, \kappa) &= \ln\left[\hat{\sigma}_{\hat{G}(K_0,\kappa)}^2\right] + \rho_{2NT}\, p K_0 \\ &= \ln\left\{\frac{1}{NT}\sum_{k=1}^{K_0} \sum_{i \in \hat{G}_k(K_0,\kappa)} \sum_{t=1}^{T}\left[Y_{it} - \tilde{\alpha}_{\hat{G}_k(K_0,\kappa)}' X_{it} - \tilde{\lambda}_i^{(K_0,\kappa)\prime} \tilde{F}_t^{(K_0,\kappa)}\right]^2\right\} + o(1) \\ &\xrightarrow{P} \ln\left(\sigma_0^2\right). \end{aligned}$$

We consider the cases of under- and over-fitted models separately. *Case 1: Under-fitted model* ($K < K_0$). Noting that

$$\begin{aligned}\hat{\sigma}^2_{\hat{G}_k(K,\kappa)} &= \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k(K,\kappa)}\sum_{t=1}^{T}\left[Y_{it}-\tilde{\alpha}'_{\hat{G}_k(K,\kappa)}X_{it}-\tilde{\lambda}_i^{(K,\kappa)\prime}\tilde{F}_t^{(K,\kappa)}\right]^2 \\ &\geq \min_{1\leq K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in G_{K,k}}\sum_{t=1}^{T}\left[Y_{it}-\bar{\alpha}'_{G_{K,k}}X_{it}-\bar{\lambda}_i^{(K)\prime}\bar{F}_t^{(K)}\right]^2=\min_{1\leq K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\hat{\sigma}^2_{G^{(K)}},\end{aligned}$$

we have by Assumptions A.7-A.8 and the Slutsky Lemma

$$\min_{1\leq K<K_0} IC_{2NT}(K,\kappa)\geq\min_{1\leq K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\ln(\hat{\sigma}^2_{G^{(K)}})+\rho_{2NT}pK\overset{P}{\to}\ln(\underline{\sigma}^2)>\ln(\sigma_0^2).$$

It follows that $P\left(\min_{K\in\Omega_-} IC_{2NT}(K,\kappa)>IC_{2NT}(K_0,\kappa)\right)\to 1$.

Case 2: Over-fitted model ($K > K_0$). Let $K\in\Omega_+$. By Lemma A.11 and the fact that $\delta^2_{NT}\rho_{2NT}\to\infty$ under Assumption A.8, we have

$$\begin{aligned}& P\left(\min_{K\in\Omega_+} IC_{2NT}(K,\kappa)>IC_{2NT}(K_0,\kappa)\right)\\ =\;& P\left(\min_{K\in\Omega_+}\left[\delta^2_{NT}\ln\left(\hat{\sigma}^2_{\hat{G}(K,\kappa)}/\hat{\sigma}^2_{\hat{G}(K_0,\kappa)}\right)+\delta^2_{NT}\rho_{2NT}(K-K_0)\right]>0\right)\\ =\;& P\left(\min_{K\in\Omega_+}\delta^2_{NT}\left(\hat{\sigma}^2_{\hat{G}(K,\kappa)}-\hat{\sigma}^2_{\hat{G}(K_0,\kappa)}\right)/\hat{\sigma}^2_{\hat{G}(K_0,\kappa)}+\delta^2_{NT}\rho_{2NT}(K-K_0)+o_P(1)>0\right)\\ \to\;& 1 \text{ as } (N,T)\to\infty.\end{aligned}$$

Combining the results in Cases 1 and 2, we complete the proof. ■

B Proof of the technical lemmas in Appendix A

Proof of Lemma A.1. The results in (ii) and (iii) follow from Lemma A.1 in Bai (2009). We only sketch the proofs of (i) and (iv). Note that $\frac{1}{NT}\sum_{i=1}^N b_i'X_i'M_F\varepsilon_i=\frac{1}{NT}\sum_{i=1}^N b_i'X_i'\varepsilon_i-\frac{1}{NT}\sum_{i=1}^N b_i'X_i'P_F\varepsilon_i\equiv A_1(b)-A_2(b,F)$, say. By the Cauchy-Schwarz inequality, Assumption A.1(vii), and Markov inequality

$$\begin{aligned}|A_1(b)| &= \left|\frac{1}{NT}\sum_{i=1}^N b_i'X_i'\varepsilon_i\right|\leq\frac{1}{\sqrt{T}}\left\{\frac{1}{N}\sum_{i=1}^N\|b_i\|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^N\|X_i'\varepsilon_i\|^2\right\}^{1/2}\\ &= T^{-1/2}O(1)O_P(1)=O_P(T^{-1/2}) \text{ uniformly in } \|b\|\leq N^{1/2}L.\end{aligned}$$

Noting that $T^{-1}\|b_i'X_i'F\|\leq\sqrt{R_0}T^{-1/2}\|b_i\|\|X_i\|$ as $T^{-1}F'F=I_{R_0}$, we have by Assumption A.1(iii),

$$\begin{aligned}|A_2(b,F)| &= \left|\frac{1}{NT^2}\sum_{i=1}^N b_i'X_i'FF'\varepsilon_i\right|\leq\frac{1}{NT^2}\sum_{i=1}^N\|b_i'X_i'F\|\|F'\varepsilon_i\|\\ &\leq \sqrt{R_0}\max_{1\leq i\leq N}T^{-1/2}\|X_i\|\frac{1}{NT^3}\sum_{i=1}^N\|b_i\|\|F'\varepsilon_i\|\\ &\leq O_P(1)\left\{\frac{1}{N}\sum_{i=1}^N\|b_i\|^2\right\}^{1/2}\left\{\frac{1}{NT^2}\sum_{i=1}^N\|F'\varepsilon_i\|^2\right\}^{1/2}\\ &= O_P(1)O(1)O_P\left(N^{-1/4}+T^{-1/4}\right)=o_P(1) \text{ uniformly in } (b,F)\end{aligned}$$

where we have used the fact that $\frac{1}{NT^2}\sum_{i=1}^N \|F'\varepsilon_i\|^2 = O_P\left(N^{-1/2}+T^{-1/2}\right)$ uniformly in F by the proof of Lemma A.1(i) in Bai (2009). Then (i) follows.

Using $P_{F^0} = F^0(F^{0\prime}F^0)^{-1}F^{0\prime}$ and the fact that $\frac{1}{NT}\sum_{i=1}^N E\left\|F^{0\prime}\varepsilon_i\right\|^2 = O(1)$, we have by Markov inequality and Assumptions A.1 (i) and (vii)

$$\frac{1}{NT}\sum_{i=1}^N \varepsilon_i' P_{F^0}\varepsilon_i \le \frac{1}{T}\left\|(T^{-1}F^{0\prime}F^0)^{-1}\right\|\frac{1}{NT}\sum_{i=1}^N \left\|F^{0\prime}\varepsilon_i\right\|^2 = O_P\left(T^{-1}\right).$$

That is, (iv) follows. ■

Proof of Lemma A.2. (i) By (2.7) and using $Y_i - X_i\hat{\beta}_i = -X_i\hat{b}_i + F^0\lambda_i^0 + \varepsilon_i$ with $\hat{b}_i = \hat{\beta}_i - \beta_i^0$, we have

$$\begin{aligned}
\hat{F}V_{NT} - F^0HV_{NT} &= \left[\frac{1}{NT}\sum_{i=1}^N\left(-X_i\hat{b}_i + F^0\lambda_i^0+\varepsilon_i\right)\left(-X_i\hat{b}_i + F^0\lambda_i^0+\varepsilon_i\right)'\right]\hat{F} - F^0HV_{NT} \\
&= \frac{1}{NT}\sum_{i=1}^N X_i\hat{b}_i\hat{b}_i'X_i'\hat{F} - \frac{1}{NT}\sum_{i=1}^N X_i\hat{b}_i\lambda_i^{0\prime}F^{0\prime}\hat{F} - \frac{1}{NT}\sum_{i=1}^N X_i\hat{b}_i\varepsilon_i'\hat{F} \\
&\quad -\frac{1}{NT}\sum_{i=1}^N F^0\lambda_i^0\hat{b}_i'X_i'\hat{F} - \frac{1}{NT}\sum_{i=1}^N \varepsilon_i\hat{b}_i'X_i'\hat{F} + \frac{1}{NT}\sum_{i=1}^N F^0\lambda_i^0\varepsilon_i'\hat{F} \\
&\quad +\frac{1}{NT}\sum_{i=1}^N \varepsilon_i\lambda_i^{0\prime}F^{0\prime}\hat{F} + \frac{1}{NT}\sum_{i=1}^N\varepsilon_i\varepsilon_i'\hat{F} \\
&\equiv a_1+a_2+a_3+a_4+a_5+a_6+a_7+a_8, \text{ say.}
\end{aligned} \tag{B.1}$$

Noting that $T^{-1/2}\left\|\hat{F}\right\| = \sqrt{R_0}$, we have by Assumptions A.1(i)-(iii), $T^{-1/2}\|a_1\| \le \sqrt{R_0}\max_{1\le i\le N}T^{-1}\|X_i\|^2 \times\frac{1}{N}\sum_{i=1}^N\left\|\hat{b}_i\right\|^2 = O_P\left(\eta_{NT}^2\right)$, and

$$\begin{aligned}
T^{-1/2}\|a_2\| &\le \sqrt{R_0}\max_{1\le i\le N}T^{-1/2}\|X_i\|\,T^{-1/2}\left\|F^0\right\|\frac{1}{N}\sum_{i=1}^N\left\|\hat{b}_i\right\|\left\|\lambda_i^0\right\| \\
&\le O_P(1)\left\{\frac{1}{N}\sum_{i=1}^N\left\|\lambda_i^0\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\left\|\hat{b}_i\right\|^2\right\}^{1/2} = O_P\left(\eta_{NT}\right).
\end{aligned}$$

Analogously, we can show that $T^{-1/2}\|a_s\| = O_P(\eta_{NT})$ for $s=3,4,5$. By Bai and Ng (2002), $T^{-1/2}\|a_s\| = O_P\left(\delta_{NT}^{-1}\right)$ for $s=6,7,8$. It follows that $T^{-1/2}\left\|\hat{F}V_{NT} - F^0HV_{NT}\right\| = O_P\left(\eta_{NT}+\delta_{NT}^{-1}\right)$. Then following the arguments as used in the proof of Proposition A.1 in Bai (2009), we can readily show that V_{NT} and H are invertible, and $T^{-1}\left\|\hat{F}-F^0H\right\|^2 = O_P\left(\eta_{NT}^2+\delta_{NT}^{-2}\right)$.

(ii) Writing $T^{-1}(\hat{F}-F^0H)'F^0H = \sum_{l=1}^8 T^{-1}V_{NT}^{-1}a_l'F^0H \equiv \sum_{l=1}^8 A_l$, it suffices to show that $\|A_l\| = O_P(\eta_{NT}) + O_P\left(\delta_{NT}^{-2}\right)$ for $l=1,2,...,8$. By the definitions of a_l's in (B.1), we have $\|A_1\| \le \left\{T^{-1/2}\|a_1\|\right\} \times\left\|V_{NT}^{-1}\right\|T^{-1/2}\left\|F^0H\right\| = O_P\left(\eta_{NT}^2\right)$ and $\|A_2\|\le\left\{T^{-1/2}\|a_2\|\right\}\left\|V_{NT}^{-1}\right\|T^{-1/2}\left\|F^0H\right\| = O_P(\eta_{NT})\,O_P(1) = O_P(\eta_{NT})$. For A_3, we have

$$\begin{aligned}
\|A_3\| &= N^{-1}T^{-2}\left\|V_{NT}^{-1}\hat{F}'\sum_{i=1}^N\varepsilon_i\hat{b}_i'X_i'F^0H\right\| \\
&\le N^{-1}T^{-2}\left\|V_{NT}^{-1}\left(\hat{F}-F^0H\right)'\sum_{i=1}^N\varepsilon_i\hat{b}_i'X_i'F^0H\right\| + N^{-1}T^{-2}\left\|V_{NT}^{-1}H'\sum_{i=1}^N F^{0\prime}\varepsilon_i\hat{b}_i'X_i'F^0H\right\| \\
&\equiv A_{3,1}+A_{3,2}, \text{ say.}
\end{aligned}$$

Noting that $N^{-1}\sum_{i=1}^{N}\|\varepsilon_i\|\left\|\hat{b}_i\right\|\leq\left\{N^{-1}\sum_{i=1}^{N}\|\varepsilon_i\|^2\right\}^{1/2}\left\{N^{-1}\sum_{i=1}^{N}\left\|\hat{b}_i\right\|^2\right\}^{1/2}=O_P\left(\eta_{NT}\right)$, we have

$$\begin{aligned}A_{3,1} &\leq \left\|V_{NT}^{-1}\right\|T^{-1/2}\left\|\hat{F}-F^0H\right\|\max_{1\leq i\leq N}T^{-1/2}\|X_i\|T^{-1/2}\left\|F^0H\right\|N^{-1}T^{-1/2}\sum_{i=1}^{N}\|\varepsilon_i\|\left\|\hat{b}_i\right\| \\ &\leq O_P\left(\eta_{NT}+\delta_{NT}^{-1}\right)O_P\left(\eta_{NT}\right)\ \text{by (i) and Assumptions A.1(i), (iii) and (vii).}\end{aligned}$$

Similarly,

$$\begin{aligned}A_{3,2} &\leq \left\|V_{NT}^{-1}\right\|\|H\|\max_{1\leq i\leq N}T^{-1/2}\|X_i\|T^{-1/2}\left\|F^0H\right\|N^{-1}T^{-1}\sum_{i=1}^{N}\left\|F'\varepsilon_i\right\|\left\|\hat{b}_i\right\| \\ &\leq T^{-1/2}O_P(1)\left\{N^{-1}T^{-1}\sum_{i=1}^{N}\left\|F^{0\prime}\varepsilon_i\right\|^2\right\}^{1/2}\left\{N^{-1}\sum_{i=1}^{N}\left\|\hat{b}_i\right\|^2\right\}^{1/2} \\ &= O_P(T^{-1/2}\eta_{NT})\ \text{by Assumptions A.1(i), (iii) and(vii).}\end{aligned}$$

Hence $\|A_3\|=O_P\left(\eta_{NT}+\delta_{NT}^{-1}\right)O_P\left(\eta_{NT}\right)$. As in the analysis of A_2 and $A_{3,2}$, we can readily show that $\|A_4\|=N^{-1}T^{-2}\left\|V_{NT}^{-1}\hat{F}'\sum_{i=1}^{N}X_i\hat{b}_i\lambda_i^{0\prime}F^{0\prime}F^0H\right\|=O_P\left(\eta_{NT}\right)$ and $\|A_5\|=N^{-1}T^{-2}\left\|V_{NT}^{-1}\hat{F}'\sum_{i=1}^{N}X_i\hat{b}_i\varepsilon_iF^0H\right\|$ $=O_P\left(T^{-1/2}\eta_{NT}\right)$. In addition,

$$\begin{aligned}\|A_6\| &= \frac{1}{NT^2}\left\|V_{NT}^{-1}\hat{F}'\sum_{i=1}^{N}\varepsilon_i\lambda_i^{0\prime}F^{0\prime}F^0H\right\|=N^{-1}T^{-2}\left\|V_{NT}^{-1}\hat{F}'\boldsymbol{\varepsilon}'\Lambda^0F^{0\prime}F^0H\right\| \\ &\leq N^{-1}T^{-2}\left\{\left\|V_{NT}^{-1}(\hat{F}-F^0H)'\boldsymbol{\varepsilon}'\Lambda^0F^{0\prime}F^0H\right\|+\left\|V_{NT}^{-1}H'F^{0\prime}\boldsymbol{\varepsilon}'\Lambda^0F^{0\prime}F^0H\right\|\right\} \\ &\leq N^{-1}T^{-2}\left\{\left\|V_{NT}^{-1}\right\|\left\|\hat{F}-F^0H\right\|\left\|\boldsymbol{\varepsilon}'\Lambda^0\right\|\left\|F^{0\prime}F^0H\right\|+\left\|V_{NT}^{-1}H'\right\|\left\|F^{0\prime}\boldsymbol{\varepsilon}'\Lambda^0\right\|\right\}\left\|F^{0\prime}F^0H\right\| \\ &= N^{-1}T^{-2}\left\{O_P(1)\,O_P(T^{1/2}(\eta_{NT}+\delta_{NT}^{-1}))O_P(N^{1/2}T^{1/2})O_P(T)+O_P(1)\,O_P(N^{1/2}T^{1/2})O_P(T)\right\} \\ &= O_P(\eta_{NT}N^{-1/2})+O_P(\delta_{NT}^{-2}), \\ \|A_7\| &= N^{-1}T^{-2}\left\|V_{NT}^{-1}\hat{F}'F^0\sum_{i=1}^{N}\lambda_i^0\varepsilon_i'F^0H\right\|=N^{-1}T^{-2}\left\|V_{NT}^{-1}\hat{F}'F^0\Lambda^{0\prime}\boldsymbol{\varepsilon}F^0H\right\| \\ &\leq N^{-1}T^{-2}\left\|V_{NT}^{-1}\right\|\left\|\hat{F}'F^0\right\|\left\|\Lambda^{0\prime}\boldsymbol{\varepsilon}F^0\right\|\|H\| \\ &= N^{-1}T^{-2}O_P(1)\,O_P(T)\,O_P(N^{1/2}T^{1/2})O_P(1)=O_P(\delta_{NT}^{-2}),\ \text{and} \\ \|A_8\| &= N^{-1}T^{-2}\left\|V_{NT}^{-1}\hat{F}'\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'F^0H\right\|=N^{-1}T^{-2}\left\|V_{NT}^{-1}\hat{F}'\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}F^0H\right\| \\ &\leq N^{-1}T^{-3/2}\sqrt{r}\left\|V_{NT}^{-1}\right\|\|H\|\|\boldsymbol{\varepsilon}\|_{\text{sp}}\left\|\boldsymbol{\varepsilon}F^0\right\| \\ &= N^{-1}T^{-3/2}O_P(1)\,O_P(N^{1/2}+T^{1/2})O_P(N^{1/2}T^{1/2})=O_P\left(\delta_{NT}^{-2}\right).\end{aligned}$$

It follows that $T^{-1}(\hat{F}-F^0H_{(1)})'F^0H=O_P\left(\eta_{NT}\right)+O_P\left(\delta_{NT}^{-2}\right)$.

(iii) Noting that $T^{-1}(\hat{F}-F^0H)'\hat{F}=T^{-1}(\hat{F}-F^0H)'(\hat{F}-F^0H)+T^{-1}(\hat{F}-F^0H)'F^0H$, the result follows from (i) and (ii).

(iv) In view of the fact that $T^{-1}(\hat{F}'\hat{F}-H'F^{0\prime}F^0H)=T^{-1}(\hat{F}-F^0H)'(\hat{F}-F^0H)+T^{-1}(\hat{F}-F^0H)'F^0H$ $+T^{-1}\left(F^0H\right)'(\hat{F}-F^0H)$, the result follows from (i) and (ii).

(v) By (A.1), the result follows from (i) and the fact that $T^{-1/2}\left\|F^0H\right\|=O_P(1)$. ■

Proof of Lemma A.3. (i) We only prove the second part of the claim in (i) and the first part can be shown analogously. By (A.1), we have $-R_{1i} = \sum_{l=1}^{4} \frac{1}{T} X_i' d_l \varepsilon_i \equiv \sum_{l=1}^{4} R_{1li}$, say. It suffices to show that $\mathcal{R}_{1l} \equiv N^{-1} \sum_{i=1}^{N} \|R_{1li}\|^2 = O_P\left(\eta_{NT}^4\right) + O_P\left(T^{-1}\eta_{NT}^2\right) + O_P\left(\delta_{NT}^{-4}\right)$ for $l = 1, 2, 3, 4$. By Lemma A.2 and Assumption A.1

$$
\begin{aligned}
\mathcal{R}_{11} &\leq \frac{1}{NT^4} \sum_{i=1}^{N} \left\| X_i'(\hat{F} - F^0 H)(\hat{F} - F^0 H)' \varepsilon_i \right\|^2 \\
&\leq T^{-2} \left\| \hat{F} - F^0 H \right\|^4 \frac{1}{NT^2} \sum_{i=1}^{N} \|X_i\|^2 \|\varepsilon_i\|^2 = O_P\left(\eta_{NT}^4 + \delta_{NT}^{-4}\right), \text{ and} \\
\mathcal{R}_{12} &\leq \frac{1}{NT^4} \sum_{i=1}^{N} \left\| X_i'(\hat{F} - F^0 H) H' F^{0\prime} \varepsilon_i \right\| \\
&\leq T^{-2} \left\| \hat{F} - F^0 H \right\|^2 \|H\|^2 \frac{1}{NT^2} \sum_{i=1}^{N} \|X_i\|^2 \left\| F^{0\prime} \varepsilon_i \right\|^2 = T^{-1} O_P\left(\eta_{NT}^2 + \delta_{NT}^{-2}\right).
\end{aligned}
$$

For $\mathcal{R}_{13}$, we apply (B.1) and the C_r inequality to obtain

$$
\mathcal{R}_{13} = \frac{1}{NT^4} \sum_{i=1}^{N} \left\| X_i' F^0 H (\hat{F} - F^0 H)' \varepsilon_i \right\|^2 \leq 8 \sum_{l=1}^{8} \frac{1}{NT^4} \sum_{i=1}^{N} \left\| X_i' F^0 H V_{NT}^{-1} a_l' \varepsilon_i \right\|^2 \equiv 8 \sum_{l=1}^{8} \mathcal{R}_{13,l}, \text{ say.}
$$

For $\mathcal{R}_{13,1}$, we use the fact that $\|a_1\| = O_P\left(\eta_{NT}^2\right)$ to obtain the rough bound $\mathcal{R}_{12,1} = O_P\left(\eta_{NT}^4\right)$. Next by Assumption A.1(viii)

$$
\begin{aligned}
\mathcal{R}_{13,2} &= \frac{1}{N^3 T^6} \sum_{i=1}^{N} \left\| X_i' F^0 H V_{NT}^{-1} \hat{F}' F^0 \sum_{j=1}^{N} \lambda_j^0 \hat{b}_j' X_j' \varepsilon_i \right\|^2 \\
&\leq T^{-3} \left\| F^0 H V_{NT}^{-1} \hat{F}' F^0 \right\|^2 \max_{1 \leq i \leq N} T^{-1} \|X_i\|^2 \frac{1}{NT^2} \sum_{i=1}^{N} \left\{ \frac{1}{N} \sum_{j=1}^{N} \left\| \lambda_j^0 \varepsilon_i' X_j \right\| \left\| \hat{b}_j \right\| \right\}^2 \\
&\leq O_P(1) \left\{ \frac{1}{N} \sum_{j=1}^{N} \left\| \hat{b}_j \right\|^2 \right\} \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\| \lambda_j^0 \varepsilon_i' X_j \right\|^2 = \eta_{NT}^2 O_P\left(T^{-1}\right).
\end{aligned}
$$

Similarly, we can show that $\mathcal{R}_{13,l} = O_P\left(T^{-1}\eta_{NT}^2\right)$ for $l = 3, 4, 5$. For $\mathcal{R}_{12,6}$, we have

$$
\begin{aligned}
\mathcal{R}_{13,6} &= \frac{1}{N^3 T^6} \sum_{i=1}^{N} \left\| X_i' F^0 H V_{NT}^{-1} \hat{F}' \sum_{j=1}^{N} \varepsilon_j \lambda_j^{0\prime} F^{0\prime} \varepsilon_i \right\|^2 \\
&\leq \frac{2}{N^3 T^6} \sum_{i=1}^{N} \left\| X_i' F^0 H V_{NT}^{-1} H' \sum_{j=1}^{N} F^{0\prime} \varepsilon_j \lambda_j^{0\prime} F^{0\prime} \varepsilon_i \right\|^2 \\
&\quad + \frac{2}{N^3 T^6} \sum_{i=1}^{N} \left\| X_i' F^0 H V_{NT}^{-1} (\hat{F} - F^0 H)' \sum_{j=1}^{N} \varepsilon_j \lambda_j^{0\prime} F^{0\prime} \varepsilon_i \right\|^2.
\end{aligned}
$$

The first term is bounded by

$$
N^{-1} T^{-2} \left\{ T^{-1} \left\| F^0 H V_{NT}^{-1} H' \right\|^2 \right\} \left\{ \frac{1}{NT^2} \sum_{i=1}^{N} \|X_i\|^2 \left\| F^{0\prime} \varepsilon_i \right\|^2 \right\} \left\{ \frac{1}{NT} \left\| F^{0\prime} \boldsymbol{\varepsilon}' \Lambda^0 \right\|^2 \right\} = O_P\left(N^{-1} T^{-2}\right),
$$

and the second term is bounded by

$$
\begin{aligned}
& N^{-1}T^{-1}T^{-1}\left\|F^0HV_{NT}^{-1}\right\|^2 T^{-1}\left\|\hat{F}-F^0H\right\|^2\left\{\frac{1}{NT^2}\sum_{i=1}^N\|X_i\|^2\left\|F^{0\prime}\varepsilon_i\right\|^2\right\}\frac{1}{NT}\left\|\sum_{j=1}^N\varepsilon_j\lambda_j^{0\prime}\right\|^2 \\
= \;& N^{-1}T^{-1}\left[O_P\left(\eta_{NT}^2\right)+O_P\left(\delta_{NT}^{-2}\right)\right].
\end{aligned}
$$

So $\mathcal{R}_{13,6}=N^{-1}T^{-1}\left[O_P\left(\eta_{NT}^2\right)+O_P\left(\delta_{NT}^{-2}\right)\right]$. Similarly,

$$
\begin{aligned}
\mathcal{R}_{13,7} &= \frac{1}{N^3T^6}\sum_{i=1}^N\left\|X_i'F^0HV_{NT}^{-1}\hat{F}'F^0\sum_{j=1}^N\lambda_j^0\varepsilon_j'\varepsilon_i\right\|^2 \\
&\leq N^{-1}T^{-1}\max_{1\leq i\leq N}T^{-1}\|X_i\|^2\left\{T^{-3}\left\|F^0HV_{NT}^{-1}\hat{F}'F^0\right\|^2\right\}\left\{\frac{1}{N^2T}\sum_{i=1}^N\left\|\sum_{j=1}^N\lambda_j^0\varepsilon_j'\varepsilon_i\right\|^2\right\} \\
&= N^{-1}T^{-1}O_P(1)\,O_P(1)\,O_P(1)=O_P\left(N^{-1}T^{-1}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{R}_{13,8} &= \frac{1}{N^3T^6}\sum_{i=1}^N\left\|X_i'F^0HV_{NT}^{-1}\hat{F}'\sum_{j=1}^N\varepsilon_j\varepsilon_j'\varepsilon_i\right\|^2=\frac{1}{N^3T^6}\sum_{i=1}^N\left\|X_i'F^0HV_{NT}^{-1}\hat{F}'\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\varepsilon_i\right\|^2 \\
&\leq \left\{\frac{1}{T^2}\left\|F^0HV_{NT}^{-1}H'\hat{F}\right\|^2\right\}\max_{1\leq i\leq N}T^{-1}\|X_i\|^2\frac{1}{N^2T^2}\|\boldsymbol{\varepsilon}\|_{\mathrm{sp}}^4\left\{\frac{1}{NT}\sum_{i=1}^N\|\varepsilon_i\|^2\right\} \\
&= O_P(1)\,O_P(1)\,O_P\left(N^{-2}+T^{-2}\right)O_P(1)=O_P\left(\delta_{NT}^{-4}\right).
\end{aligned}
$$

In addition, using $\hat{F}'\hat{F}/T=I_R$ and Lemma A.2(i),

$$
\begin{aligned}
\mathcal{R}_{14} &\leq \frac{1}{NT^4}\sum_{i=1}^N\left\|X_i'F^0H\left(I_r-\left(T^{-1}H'F^{0\prime}F^0H\right)^{-1}\right)'H'F^{0\prime}\varepsilon_i\right\|^2 \\
&\leq T^{-1}\left\|I_r-\left(T^{-1}H'F^{0\prime}F^0H\right)^{-1}\right\|^2T^{-1}\left\|F^0H\right\|^2\|H\|^2\left\{\frac{1}{NT^2}\sum_{i=1}^N\|X_i\|^2\left\|F^{0\prime}\varepsilon_i\right\|^2\right\} \\
&= T^{-1}\left[O_P\left(\eta_{NT}^2\right)+O_P\left(\delta_{NT}^{-4}\right)\right].
\end{aligned}
$$

It follows that $\mathcal{R}_{14}=T^{-1}\left[O_P\left(\eta_{NT}^2\right)+O_P\left(\delta_{NT}^{-4}\right)\right]$. Then (i) follows.

(ii) We only prove the second part of the claim in (ii) and the first part can be shown analogously. By (B.1), $\frac{1}{T}X_i'M_{\hat{F}}(\hat{F}-F^0H)H^{-1}\lambda_i^0=\sum_{l=1}^8\frac{1}{T}X_i'M_{\hat{F}}a_lV_{NT}^{-1}H^{-1}\lambda_i^0=\sum_{l=1}^8R_{2li}$, say. Noting that $R_{22i}=-\frac{1}{NT^2}\sum_{j=1}^NX_i'M_{\hat{F}}X_j\hat{b}_j\lambda_j^{0\prime}F^{0\prime}\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0=-\frac{1}{NT}\sum_{j=1}^NX_i'M_{\hat{F}}X_j\nu_{ji}\hat{b}_j$, we have $\mathcal{R}_2\equiv N^{-1}\sum_{i=1}^N\|R_{2i}\|^2\leq 7\sum_{l=1,l\neq 2}^8\mathcal{R}_{2l}$, where $\mathcal{R}_{2l}=N^{-1}\sum_{t=1}^N\|R_{2li}\|^2$. Following the analyses of $\mathcal{R}_{11}$ and $\mathcal{R}_{13}$, we can readily show that

$$
\begin{aligned}
\mathcal{R}_{21} &= \frac{1}{N^3T^4}\sum_{i=1}^N\left\|X_i'M_{\hat{F}}\sum_{j=1}^NX_j\hat{b}_j\hat{b}_j'X_j'\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&\leq T^{-1}\left\|\hat{F}V_{NT}^{-1}H^{-1}\right\|^2\left\{\max_{1\leq i\leq N}T^{-3}\|X_i\|^6\right\}\left\{\frac{1}{N}\sum_{i=1}^N\left\|\lambda_i^0\right\|^2\right\}\left\{\frac{1}{N}\sum_{j=1}^N\left\|\hat{b}_j\right\|^2\right\}^2=O_P\left(\eta_{NT}^4\right),
\end{aligned}
$$

and

$$\begin{aligned}
\mathcal{R}_{23} &= \frac{1}{N^3T^4}\sum_{i=1}^{N}\left\|X_i'M_{\hat{F}}\sum_{j=1}^{N}X_j\hat{b}_j\lambda_j^{0\prime}F^{0\prime}\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&\leq T^{-1}\left\{\max_{1\leq i\leq N}T^{-1}\left\|X_i\right\|^2\right\}T^{-2}\left\|F^{0\prime}\hat{F}V_{NT}^{-1}H^{-1}\right\|^2\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2\right\}\left\{\frac{1}{N}\sum_{j=1}^{N}\left\|\hat{b}_j\right\|\left\|\lambda_j^0\right\|\right\}^2 \\
&\leq T^{-1}O_P(1)\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2\right\}^2\frac{1}{N}\sum_{j=1}^{N}\left\|\hat{b}_j\right\|^2 = T^{-1}O_P\left(\eta_{NT}^2\right).
\end{aligned}$$

For $\mathcal{R}_{24}$, by the fact that $M_{\hat{F}}\hat{F}H^{-1}=0$ and Minkowski inequality, we have

$$\begin{aligned}
\mathcal{R}_{24} &\leq \frac{1}{N^3T^4}\sum_{i=1}^{N}\left\|X_i'M_{\hat{F}}\left(F^0-\hat{F}H^{-1}\right)\sum_{j=1}^{N}\lambda_j^0\hat{b}_j'X_j'\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&\leq T^{-1}\left\|F^0-\hat{F}H^{-1}\right\|^2T^{-1}\left\|\hat{F}'V_{NT}^{-1}H^{-1}\right\|^2\frac{1}{NT}\sum_{i=1}^{N}\left\|X_i\right\|^2\left\|\lambda_i^0\right\|^2\left\{\frac{1}{NT^{1/2}}\sum_{j=1}^{N}\left\|\lambda_j^0\right\|\left\|\hat{b}_j\right\|\left\|X_j\right\|\right\}^2 \\
&\leq O_P\left(\eta_{NT}^2+\delta_{NT}^{-4}\right)O_P(1)\left\{\frac{1}{NT}\sum_{j=1}^{N}\left\|\lambda_j^0\right\|^2\left\|X_j\right\|^2\right\}^2\frac{1}{N}\sum_{j=1}^{N}\left\|\hat{b}_j\right\|^2 = O_P\left(\eta_{NT}^2+\delta_{NT}^{-4}\right)O_P\left(\eta_{NT}^2\right).
\end{aligned}$$

By the fact that $M_{\hat{F}}=\left(M_{\hat{F}}-M_{F^0}\right)+I_T-P_{F^0}$ and Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathcal{R}_{25} &\leq \frac{1}{N^3T^4}\sum_{i=1}^{N}\left\|X_i'M_{\hat{F}}\sum_{j=1}^{N}\varepsilon_j\hat{b}_j'X_j'\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&\leq \frac{3}{N^3T^4}\sum_{i=1}^{N}\left\|X_i'\left(M_{\hat{F}}-M_{F^0}\right)\sum_{j=1}^{N}\varepsilon_j\hat{b}_j'X_j'\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2+\frac{3}{N^3T^4}\sum_{i=1}^{N}\left\|\sum_{j=1}^{N}X_i'\varepsilon_j\hat{b}_j'X_j'\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&\quad+\frac{3}{N^3T^4}\sum_{i=1}^{N}\left\|X_i'F^0\left(F^{0\prime}F^0\right)^{-1}\sum_{j=1}^{N}F^{0\prime}\varepsilon_j\hat{b}_j'X_j'\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&= \eta_{NT}^2O_P\left(\eta_{NT}^2+\delta_{NT}^{-4}\right)+O_P\left(T^{-1}\eta_{NT}^2\right)+O_P\left(T^{-1}\eta_{NT}^2\right)=O_P\left(T^{-1}\eta_{NT}^2\right)+O_P\left(\eta_{NT}^4+\delta_{NT}^{-4}\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{R}_{26} &\leq \frac{1}{N^3T^4}\sum_{t=1}^{N}\left\|X_i'M_{\hat{F}}\left(F^0-\hat{F}H^{-1}\right)\sum_{j=1}^{N}\lambda_j^0\varepsilon_j'\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&\leq \frac{2}{N^3T^4}\sum_{t=1}^{N}\left\|X_i'M_{\hat{F}}\left(F^0-\hat{F}H^{-1}\right)\Lambda^{0\prime}\boldsymbol{\varepsilon}F^0HV_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&\quad+\frac{2}{N^3T^4}\sum_{t=1}^{N}\left\|X_i'M_{\hat{F}}\left(F^0-\hat{F}H^{-1}\right)\Lambda^{0\prime}\boldsymbol{\varepsilon}\left(\hat{F}-F^0H\right)V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&\leq T^{-1}O_P\left(\eta_{NT}^2+\delta_{NT}^{-2}\right)+O_P\left(\eta_{NT}^4+\delta_{NT}^{-4}\right).
\end{aligned}$$

Using $M_{\hat{F}} = \left(M_{\hat{F}} - M_{F^0}\right) + I_T - P_{F^0}$ and Cauchy-Schwarz inequality,

$$
\begin{aligned}
\mathcal{R}_{27} &\leq \frac{1}{N^3T^4}\sum_{i=1}^{N}\left\|X_i' M_{\hat{F}}\sum_{j=1}^{N}\varepsilon_j\lambda_j^{0\prime}F^{0\prime}\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&\leq \frac{3}{N^3T^4}\sum_{i=1}^{N}\left\|X_i'\left(M_{\hat{F}} - M_{F^0}\right)\boldsymbol{\varepsilon}'\Lambda^0F^{0\prime}\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 + \frac{3}{N^3T^4}\sum_{i=1}^{N}\left\|X_i'\boldsymbol{\varepsilon}'\Lambda^0F^{0\prime}\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&\quad + \frac{3}{N^3T^4}\sum_{i=1}^{N}\left\|X_i'F^0\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\boldsymbol{\varepsilon}'\Lambda^0F^{0\prime}\hat{F}V_{NT}^{-1}H^{-1}\lambda_i^0\right\|^2 \\
&= T^{-1}O_P\left(\eta_{NT}^2 + \delta_{NT}^{-4}\right) + O_P\left(N^{-1}T^{-1}\right) + O_P\left(N^{-1}T^{-1}\right) = T^{-1}O_P\left(\eta_{NT}^2\right) + O_P\left(\delta_{NT}^{-4}\right)
\end{aligned}
$$

where, e.g., the third term is bounded by

$$
\begin{aligned}
&3N^{-1}T^{-1}\left\{T\left\|F^0\left(F^{0\prime}F^0\right)^{-1}\right\|^2\right\}T^{-2}\left\|F^{0\prime}\hat{F}V_{NT}^{-1}H^{-1}\right\|^2\max_{1\leq i\leq N}T^{-1}\left\|X_i\right\|^2 \\
&\times\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2\right\}\left\{\frac{1}{NT}\left\|F^{0\prime}\boldsymbol{\varepsilon}'\Lambda^0\right\|^2\right\} = O_P\left(N^{-1}T^{-1}\right).
\end{aligned}
$$

Similarly, we can show that $\mathcal{R}_{28} = T^{-1}O_P\left(\eta_{NT}^2\right) + O_P\left(\delta_{NT}^{-4}\right)$. So (ii) follows. ■

Proof of Lemma A.4. (i) By (B.1), $T^{-1}\varepsilon_i'(\hat{F} - F^0H) = \sum_{l=1}^{8}\frac{1}{T}\varepsilon_i'a_lV_{NT}^{-1} \equiv \sum_{l=1}^{8}A_{il}$, say. For A_{i1}, using $\|a_1\| = O_P\left(T^{1/2}\eta_{NT}^2\right) = O_P\left(T^{-1/2}\right)$ by Theorem 3.2, we can readily show that $\|A_{i1}\| = T^{-1/2}\|\varepsilon_i\|T^{-1/2}\|a_1\|\left\|V_{NT}^{-1}\right\| = O_P\left(T^{-1}\right)$. Noting that $T^{-1}\|\varepsilon_i\|^2 = T^{-1}\sum_{t=1}^{T}E(\varepsilon_{it}^2) + T^{-1}\sum_{t=1}^{T}[\varepsilon_{it}^2 - E(\varepsilon_{it}^2)] \leq c_\varepsilon + T^{-1}\sum_{t=1}^{T}[\varepsilon_{it}^2 - E(\varepsilon_{it}^2)]$, where $c_\varepsilon = \max_{1\leq i\leq N}T^{-1}\sum_{t=1}^{T}E(\varepsilon_{it}^2) \leq C < \infty$ under Assumption A.1(iv), we can readily show that $P(\max_{1\leq i\leq N}T^{-1}\|\varepsilon_i\|^2 \geq 3c_\varepsilon/2) = o\left(N^{-1}\right)$ by Assumption A.3(i). Then $P\left(\max_{1\leq i\leq N}\|A_{i1}\| \geq CT^{-1}(\ln T)^{\upsilon_2}\right) = o\left(N^{-1}\right)$ for any $C > 0$ and $\upsilon_2 > 0$.

For A_{i2}, we have by the submultiplicative property of Frobenius norm, Cauchy-Schwarz inequality, Assumption A.1(viii), and Theorem 3.2

$$
\begin{aligned}
\|A_{i2}\| &= \frac{1}{NT^2}\left\|\sum_{j=1}^{N}\varepsilon_i'X_j\hat{b}_j\lambda_j^{0\prime}F^{0\prime}\hat{F}V_{NT}^{-1}\right\| \leq \frac{1}{NT^2}\sum_{j=1}^{N}\left\|\varepsilon_i'X_j\right\|\left\|\lambda_j^{0\prime}\right\|\left\|\hat{b}_j\right\|\left\|F^{0\prime}\hat{F}V_{NT}^{-1}\right\| \\
&\leq \frac{\eta_{nT}}{T^{1/2}}\left\{\frac{1}{NT}\sum_{j=1}^{N}\left\|\varepsilon_i'X_j\right\|^2\left\|\lambda_j^0\right\|^2\right\}^{1/2}\frac{1}{T}\left\|F^{0\prime}\hat{F}V_{NT}^{-1}\right\| \\
&= O_P\left(T^{-1}\right)O_P(1)\,O_P(1) = O_P\left(T^{-1}\right).
\end{aligned}
$$

Noting that $\max_{1\leq i\leq N}\frac{1}{NT}\sum_{j=1}^{N}\left\|\varepsilon_i'X_j\right\|^2\left\|\lambda_j^0\right\|^2 \leq \max_{1\leq i,j\leq N}\frac{1}{T}\left\|\varepsilon_i'X_j\right\|^2\frac{1}{N}\sum_{j=1}^{N}\left\|\lambda_j^0\right\|^2$ and $P(\max_{1\leq i,j\leq N} T^{-1/2}\left\|\varepsilon_i'X_j\right\| \geq C(\ln T)^{\upsilon_1}) = o\left(N^{-1}\right)$ for any $C > 0$ by Assumption A.3(ii), we have $\max_{1\leq i\leq N}P(\|A_{i2}\| \geq CT^{-1}(\ln T)^{\upsilon_1}) = o\left(N^{-1}\right)$. Similarly, we have

$$
\begin{aligned}
\|A_{i3}\| &= \frac{1}{NT^2}\left\|\sum_{j=1}^{N}\varepsilon_i'X_j\hat{b}_j\varepsilon_j'\hat{F}V_{NT}^{-1}\right\| = \frac{1}{NT^2}\left\|\sum_{j=1}^{N}V_{NT}^{-1}\hat{F}'\varepsilon_j\varepsilon_i'X_j\hat{b}_j\right\| \\
&\leq \frac{1}{T}\left\|\hat{F}V_{NT}^{-1}\right\|\eta_{nT}\left\{\frac{1}{NT^2}\sum_{j=1}^{N}\left\|\varepsilon_j\varepsilon_i'X_j\right\|^2\right\}^{1/2} \\
&= O_P(T^{-1/2})O_P(T^{-1/2})O_P(1) = O_P(T^{-1}) \text{ by Assumption A.1(viii),}
\end{aligned}
$$

$$
\begin{aligned}
\|A_{i4}\| &= \frac{1}{NT^2}\left\|\varepsilon_i' F^0 \sum_{j=1}^N \lambda_j^0 \hat{b}_j' X_j' \hat{F} V_{NT}^{-1}\right\| \\
&\leq \frac{\eta_{nT}}{T^{1/2}} \frac{1}{T^{1/2}} \left\|\varepsilon_i' F^0\right\| \left\{\frac{1}{NT}\sum_{j=1}^N \|X_j\|^2 \left\|\lambda_j^0\right\|^2\right\}^{1/2} \frac{1}{T^{1/2}} \left\|\hat{F} V_{NT}^{-1}\right\| \\
&= O_P\left(T^{-1}\right) O_P(1) O_P(1) O_P(1) = O_P\left(T^{-1}\right) \text{ by Assumptions A.1(ii), (iii) and (vii), and} \\
\|A_{i5}\| &= \frac{1}{NT^2}\left\|\varepsilon_i' \sum_{j=1}^N \varepsilon_j \hat{b}_j' X_j' \hat{F} V_{NT}^{-1}\right\| \\
&\leq \frac{1}{T^{1/2}} \left\{\max_{1\leq j\leq N} \frac{1}{T^{1/2}} \|X_j\|\right\} \eta_{nT} \left\{\frac{1}{NT}\sum_{j=1}^N \left\|\varepsilon_i'\varepsilon_j\right\|^2\right\}^{1/2} \frac{1}{T^{1/2}} \left\|\hat{F} V_{NT}^{-1}\right\| \\
&= T^{-1/2} O_P(1) O_P(T^{-1/2}) O_P(1) O_P(1) = O_P\left(T^{-1}\right) \text{ by Assumptions A.1(iii) and (viii).}
\end{aligned}
$$

Noting that $\max_{1\leq i\leq N} \frac{1}{NT^2}\sum_{j=1}^N \left\|\varepsilon_j \varepsilon_i' X_j\right\|^2 \leq \max_{1\leq i,j\leq N} \frac{1}{T} \left\|\varepsilon_i' X_j\right\|^2 \frac{1}{NT}\sum_{j=1}^N \left\|\varepsilon_j\right\|^2$ and $P(\max_{1\leq i,j\leq N} T^{-1}\left\|\varepsilon_i' X_j\right\|^2 \geq C(\ln T)^{2\upsilon_1}) = o\left(N^{-1}\right)$ for any $C > 0$ by Assumption A.3(ii), we can readily show that $P\left(\max_{1\leq i\leq N} \|A_{i3}\| \geq CT^{-1}(\ln T)^{\upsilon_1}\right) = o\left(N^{-1}\right)$. Similarly, using $P(\max_{1\leq i\leq N} T^{-1/2}\left\|\varepsilon_i' F^0\right\| \geq C(\ln T)^{\upsilon_1}) = o\left(N^{-1}\right)$ and $P\left(\max_{1\leq i\leq N} T^{-1/2}\|X_i\| \geq C(\ln T)^{\upsilon_2}\right) = o\left(N^{-1}\right)$ for any $C > 0$ by Assumption A.3(i) and A.1(iii), we can show that $P\left(\max_{1\leq i\leq N} \|A_{il}\| \geq CT^{-1}(\ln T)^{\upsilon_1}\right) = o\left(N^{-1}\right)$ for $l = 4, 5$.

By the triangle inequality, the fact that $\frac{1}{T^{1/2}}\left\|\varepsilon_i' F^0\right\| = O_P(1)$, $\frac{1}{NT^{1/2}}\sum_{j=1}^N \left\|\lambda_j^0 \varepsilon_j' F^0\right\| = O_P(1)$ and $\left\|\frac{1}{NT^{1/2}}\sum_{j=1}^N \lambda_i^0 \varepsilon_i'\right\| = O_P(1)$, Lemma A.2(i), and Theorem 3.2,

$$
\begin{aligned}
\|A_{i6}\| &\leq \frac{1}{NT^2}\left\|\varepsilon_i' F^0 \sum_{j=1}^N \lambda_j^0 \varepsilon_j' F^0 H V_{NT}^{-1}\right\| + \frac{1}{NT^2}\left\|\varepsilon_i' F^0 \sum_{j=1}^N \lambda_j^0 \varepsilon_j' \left(\hat{F} - F^0 H\right) V_{NT}^{-1}\right\| \\
&\leq \frac{1}{T}\left\{\frac{1}{T^{1/2}}\left\|\varepsilon_i' F^0\right\|\right\}\left\{\frac{1}{NT^{1/2}}\sum_{j=1}^N \left\|\lambda_j^0 \varepsilon_j' F^0\right\|\right\}\left\|H V_{NT}^{-1}\right\| \\
&\quad + \frac{1}{T^{1/2}}\left\{\frac{1}{T^{1/2}}\left\|\varepsilon_i' F^0\right\|\right\}\left\|\frac{1}{NT^{1/2}}\sum_{j=1}^N \lambda_j^0 \varepsilon_j'\right\| \frac{1}{T^{1/2}}\left\|\hat{F} - F^0 H\right\| \left\|V_{NT}^{-1}\right\| \\
&= T^{-1} O_P(1) O_P(1) O_P(1) + T^{-1/2} O_P(1) O_P(1) O_P\left(\delta_{NT}^{-1}\right) O_P(1) = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right).
\end{aligned}
$$

Using $P\left(\max_{1\leq i\leq N} T^{-1/2}\left\|\varepsilon_i' F^0\right\| \geq c(\ln T)^{\upsilon_1}\right) = o\left(N^{-1}\right)$, we can also show that $P(\max_{1\leq i\leq N} \|A_{i6}\| \geq CT^{-1/2}\delta_{NT}^{-1}(\ln T)^{\upsilon_1}) = o\left(N^{-1}\right)$.

Noting that $\left\|\frac{1}{\sqrt{NT}}\sum_{j=1}^N \varepsilon_i' \varepsilon_j \lambda_j^{0\prime}\right\|^2 = O_P\left(1 + \frac{T}{N}\right)$ by Assumption A.1(viii), we have

$$
\begin{aligned}
\|A_{i7}\| &= \frac{1}{NT^2}\left\|\varepsilon_i' \sum_{j=1}^N \varepsilon_j \lambda_j^{0\prime} F^{0\prime} \hat{F} V_{NT}^{-1}\right\| \\
&\leq \frac{1}{\sqrt{NT}}\left\|\frac{1}{\sqrt{NT}}\sum_{j=1}^N \varepsilon_i' \varepsilon_j \lambda_j^{0\prime}\right\| \frac{1}{T}\left\|F^{0\prime} \hat{F} V_{NT}^{-1}\right\| \\
&= (NT)^{-1/2} O_P(1 + \sqrt{T/N}) = O_P((NT)^{-1/2} + N^{-1}).
\end{aligned}
$$

Noting that $P\left(\max_{1\leq i\leq N}\left\|(NT)^{-1/2}\sum_{j=1}^{N}\varepsilon_i'\varepsilon_j\lambda_j^{0\prime}\right\|\geq C(1+\sqrt{T/N})\left(\ln T\right)^{\upsilon_1}\right)=o\left(N^{-1}\right)$ by Assumption A.3(iii), we have $P(\max_{1\leq i\leq N}\|A_{i7}\|\geq C((NT)^{-1/2}+N^{-1})\left(\ln T\right)^{\upsilon_1})=o\left(N^{-1}\right)$. Write

$$A_{i8}=\frac{1}{NT^2}\sum_{j=1}^{N}\varepsilon_i'\varepsilon_j\varepsilon_j'F^0HV_{NT}^{-1}+\frac{1}{NT^2}\sum_{j=1}^{N}\varepsilon_i'\varepsilon_j\varepsilon_j'(\hat{F}-F^0H)V_{NT}^{-1}\equiv A_{i8,1}+A_{i8,2}.$$

By Assumption A.1(viii), $\|A_{i8,1}\|\leq\frac{1}{T}\left\|\frac{1}{NT}\sum_{j=1}^{N}\varepsilon_i'\varepsilon_j\varepsilon_j'F^0\right\|\left\|HV_{NT}^{-1}\right\|=T^{-1}O_P\left(1\right)=O_P\left(T^{-1}\right)$. By the Cauchy-Schwarz inequality, Lemma A.7(iii)

$$\begin{aligned}
\left|A_{i8,2}'\omega_p\right| &= \left|\frac{1}{NT^2}\sum_{j=1}^{N}\varepsilon_i'\varepsilon_j\varepsilon_j'\left(\hat{F}-F^0H\right)V_{NT}^{-1}\omega_p\right| \\
&\leq \left\{\frac{1}{NT^2}\sum_{j=1}^{N}\|\varepsilon_i'\varepsilon_j\|^2\right\}^{1/2}\left\{\frac{1}{NT^2}\sum_{j=1}^{N}\omega_p'V_{NT}^{-1}\left(\hat{F}-F^0H\right)'\varepsilon_j\varepsilon_j'\left(\hat{F}-F^0H\right)V_{NT}^{-1}\omega_p\right\}^{1/2} \\
&\leq \frac{1}{T^{1/2}}\left\|V_{NT}^{-1}\right\|\left\{\frac{1}{NT}\sum_{j=1}^{N}\|\varepsilon_i'\varepsilon_j\|^2\right\}^{1/2}\left\{\frac{1}{NT^2}\sum_{j=1}^{N}\varepsilon_j'(\hat{F}-F^0H)(\hat{F}-F^0H)'\varepsilon_j\right\}^{1/2} \\
&= T^{-1/2}O_P\left(1\right)O_P\left(1\right)O_P\left(\delta_{NT}^{-2}\right)=O_P(T^{-1/2}\delta_{NT}^{-2}),
\end{aligned}$$

where we use the fact that $\omega_p'V_{NT}^{-1}(\hat{F}-F^0H)'\varepsilon_j\varepsilon_j'(\hat{F}-F^0H)V_{NT}^{-1}\omega_p=\mathrm{tr}[V_{NT}^{-1}\omega_p\omega_p'V_{NT}^{-1}(\hat{F}-F^0H)'\varepsilon_j\varepsilon_j'(\hat{F}-F^0H)]\leq\mu_{\max}\left(V_{NT}^{-1}\omega_p\omega_p'V_{NT}^{-1}\right)\mathrm{tr}[(\hat{F}-F^0H)'\varepsilon_j\varepsilon_j'(\hat{F}-F^0H)]\leq\left\|V_{NT}^{-1}\right\|^2\varepsilon_j'(\hat{F}-F^0H)(\hat{F}-F^0H)'\varepsilon_j$. It follows that $A_{i8}=O_P(T^{-1})$. Using $\left\|\frac{1}{NT}\sum_{j=1}^{N}\varepsilon_i'\varepsilon_j\varepsilon_j'F^0\right\|\leq\left\{\frac{1}{NT}\sum_{j=1}^{N}\left(\varepsilon_i'\varepsilon_j\right)^2\right\}^{1/2}\times\left\{\frac{1}{NT}\sum_{j=1}^{N}\left\|\varepsilon_j'F^0\right\|^2\right\}^{1/2}$ and $P(\max_{1\leq i,j\leq N}T^{-1}\left(\varepsilon_i'\varepsilon_j\right)^2\geq C\left(\ln T\right)^{2\upsilon_1})=o(N^{-1})$, we can show that $P\left(\max_{1\leq i\leq N}\|A_{i8,l}\|\geq C\left(\ln T\right)^{\upsilon_1}\right)$ $=o(N^{-1})$ for $l=1,2$.

In sum, we have shown that $T^{-1}\varepsilon_i'(\hat{F}-F^0H)=O_P\left(\delta_{NT}^{-2}\right)$ and $P(\max_{1\leq i\leq N}||T^{-1}\varepsilon_i'(\hat{F}-F^0H)||\geq C\delta_{NT}^{-2}\left(\ln T\right)^{\upsilon_1})=o(N^{-1})$.

(ii) The proof is analogous to that of (i) and thus omitted.

(iii) By (A.1) and the fact that $P_{F^0}=P_{F^0H}$, $-\frac{1}{\sqrt{T}}X_i'\left(M_{\hat{F}}-M_{F^0}\right)\varepsilon_i=\frac{1}{\sqrt{T}}\sum_{l=1}^{4}X_i'd_l\varepsilon_i\equiv\sum_{l=1}^{4}D_{il}$, say. By Lemma A.2 and Theorem 3.2,

$$\begin{aligned}
\|D_{i1}\| &\leq \frac{1}{T^{3/2}}\left\|X_i'(\hat{F}-F^0H)(\hat{F}-F^0H)'\varepsilon_i\right\| \\
&\leq T^{1/2}\frac{1}{T}\left\|\hat{F}-F^0H\right\|^2\frac{1}{T}\|X_i\|\|\varepsilon_i\|=T^{1/2}O_P\left(\delta_{NT}^{-2}\right),\ \text{and} \\
\|D_{i2}\| &\leq \frac{1}{T^{3/2}}\left\|X_i'(\hat{F}-F^0H)H'F^{0\prime}\varepsilon_i\right\| \\
&\leq T^{-1/2}\left\|\hat{F}-F^0H\right\|\|H\|\frac{1}{T}\|X_i\|\left\|F^{0\prime}\varepsilon_i\right\|=O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}$$

By (i),

$$\begin{aligned}
\|D_{i3}\| &= \left\|\frac{1}{T^{3/2}}X_i'F^0H(\hat{F}-F^0H)'\varepsilon_i\right\| \\
&\leq T^{1/2}\frac{1}{T}\left\|X_i'F^0H\right\|\frac{1}{T}\left\|(\hat{F}-F^0H)'\varepsilon_i\right\|=T^{1/2}O_P\left(\delta_{NT}^{-2}\right).
\end{aligned}$$

By Lemma A.2 and Theorem 3.2,

$$\begin{aligned}
\|D_{i4}\| &\leq \frac{1}{T^{3/2}} \left\| X_i' F^0 H \left(I_r - \left(T^{-1} H' F^{0\prime} F^0 H \right)^{-1} \right)' H' F^{0\prime} \varepsilon_i \right\| \\
&\leq \frac{1}{T} \left\| X_i' F^0 \right\| \|H\|^2 \left\| I_r - \left(T^{-1} H' F^{0\prime} F^0 H \right)^{-1} \right\| \frac{1}{T^{1/2}} \left\| F^{0\prime} \varepsilon_i \right\| \\
&= O_P\left(T^{-1/2} + \delta_{NT}^{-2} \right) = O_P\left(T^{-1/2} \right).
\end{aligned}$$

Consequently, $\frac{1}{\sqrt{T}} X_i' M_{\hat{F}} \varepsilon_i = \frac{1}{\sqrt{T}} X_i' M_{F^0} \varepsilon_i + O_P\left(T^{1/2} \delta_{NT}^{-2} \right)$. In addition, noting that $P(\max_{1 \leq i \leq N} T^{-1/2} \|\xi_i\| \geq C (\ln T)^{\upsilon_1}) = o\left(N^{-1} \right)$ for $\xi_i = X_i, \varepsilon_i$,and $F^{0\prime} \varepsilon_i$, we can show that

$$P\left(\max_{1 \leq i \leq N} \frac{1}{T^{1/2}} \left\| X_i' \left(M_{\hat{F}} - M_{F^0} \right) \varepsilon_i \right\| \geq C T^{1/2} \delta_{NT}^{-2} (\ln T)^{\upsilon_1} \right) = o\left(N^{-1} \right).$$

(iv) By (A.1), $-\frac{1}{\sqrt{T}} X_i' M_{\hat{F}} F^0 \lambda_i^0 = \frac{1}{\sqrt{T}} X_i' \left(M_{F^0} - M_{\hat{F}} \right) F^0 \lambda_i^0 = \frac{1}{\sqrt{T}} X_i' \left(P_{\hat{F}} - P_{F^0} \right) F^0 \lambda_i^0 = \frac{1}{T^{3/2}} \sum_{l=1}^4 X_i' d_l F^0 \lambda_i^0 \equiv \sum_{l=1}^4 \tilde{D}_{il}$, say. By Lemma A.2 and Theorem 3.2,

$$\begin{aligned}
\left\| \tilde{D}_{i1} \right\| &\leq \frac{1}{T^{3/2}} \left\| X_i' (\hat{F} - F^0 H)(\hat{F} - F^0 H)' F^0 \lambda_i^0 \right\| \\
&\leq T^{1/2} \frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 \frac{1}{T} \|X_i\| \left\| F^0 \lambda_i^0 \right\| = T^{1/2} O_P\left(\delta_{NT}^{-2} \right) = o_P(1), \\
\left\| \tilde{D}_{i3} \right\| &\leq T^{1/2} \frac{1}{T^{1/2}} \|X_i\| \frac{1}{T^{1/2}} \left\| F^0 H \right\| \frac{1}{T} \left\| (\hat{F} - F^0 H)' F^0 \right\| \left\| \lambda_i^0 \right\| \\
&= T^{1/2} O_P(T^{-1/2} + \delta_{NT}^{-2}) = O_P(1), \text{ and} \\
\left\| \tilde{D}_{i4} \right\| &\leq \frac{1}{T^{3/2}} \left\| X_i' F^0 H \left(I_r - \left(T^{-1} H' F^{0\prime} F^0 H \right)^{-1} \right)' H' F^{0\prime} F^0 \lambda_i^0 \right\| \\
&\leq T^{1/2} \frac{1}{T} \left\| X_i' F^0 \right\| \|H\|^2 \left\| I_r - \left(T^{-1} H' F^{0\prime} F^0 H \right)^{-1} \right\| \frac{1}{T} \left\| F^{0\prime} F^0 \right\| \left\| \lambda_i^0 \right\| \\
&= T^{1/2} O_P(T^{-1/2} + \delta_{NT}^{-2}) = O_P(1).
\end{aligned}$$

In addition, by (ii),

$$\begin{aligned}
\left\| \tilde{D}_{i2} \right\| &= \frac{1}{T^{3/2}} \left\| X_i' (\hat{F} - F^0 H) H' F^{0\prime} F^0 \lambda_i^0 \right\| \\
&\leq T^{1/2} \frac{1}{T} \left\| X_i' (\hat{F} - F^0 H) \right\| \frac{1}{T} \left\| H' F^{0\prime} F^0 \right\| \left\| \lambda_i^0 \right\| = T^{1/2} O_P(T^{-1/2}) = O_P(1).
\end{aligned}$$

Consequently, $\frac{1}{\sqrt{T}} X_i' M_{\hat{F}} F^0 \lambda_i^0 = O_P(1)$. In addition, noting that $P(\max_{1 \leq i \leq N} T^{-1/2} \|X_i\| \geq C (\ln T)^{\upsilon_1}) = o\left(N^{-1} \right)$ and $P\left(\max_{1 \leq i \leq N} \left\| \lambda_i^0 \right\| \geq c_{\lambda,N} \right) = o\left(N^{-1} \right)$, we can show that $P(\max_{1 \leq i \leq N} \left\| T^{-1/2} X_i' M_{\hat{F}} F^0 \lambda_i^0 \right\| \geq C (\ln T)^{\upsilon_1} c_{\lambda,N}) = o\left(N^{-1} \right)$.

(v) Noting that $\left\| T^{-1} X_i' \left(M_{\hat{F}} - M_{F^0} \right) X_i \right\| \leq T^{-1} \|X_i\|^2 \left\| P_{\hat{F}} - P_{F^0} \right\|$, $\left\| P_{\hat{F}} - P_{F^0} \right\| = O_P\left(\delta_{NT}^{-1} \right)$ by Lemma A.2(v) and Theorem 3.2, and $P\left(\max_{1 \leq i \leq N} T^{-1} \|X_i\|^2 \geq C (\ln T)^{\upsilon_2} \right) = o\left(N^{-1} \right)$ for any $C > 0$ and $\upsilon_2 > 0$, we have

$$P\left(\max_{1 \leq i \leq N} \left\| T^{-1} X_i' \left(M_{\hat{F}} - M_{F^0} \right) X_i \right\| \geq C \delta_{NT}^{-1} (\ln T)^{\upsilon_2} \right) = o\left(N^{-1} \right).$$

(vi) The proof follows from analogous arguments as used in the proof of Lemmas A.3(ii) and Lemmas A.4(iii). ■

Proof of Lemma A.5. Observe that $\eta_{NT}^2 = \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{b}_i\right\|^2 = \frac{1}{N}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k}\left\|\hat{\alpha}_k - \beta_i^0\right\|^2 + \frac{1}{N}\sum_{i\in\hat{G}_0}\left\|\hat{b}_i\right\|^2$. Noting that for any $\epsilon > 0$

$$\begin{aligned} P\left(\frac{1}{N}\sum_{i\in\hat{G}_0}\left\|\hat{b}_i\right\|^2 \geq \epsilon/(NT)\right) &\leq \sum_{k=1}^{K_0}\sum_{i\in G_k^0} P\left(i\in\hat{G}_0 | i\in G_k^0\right) \leq \sum_{k=1}^{K_0}\sum_{i\in G_k^0} P\left(i\notin\hat{G}_k | i\in G_k^0\right) \\ &= \sum_{k=1}^{K_0}\sum_{i\in G_k^0} P\left(\hat{E}_{kNT,i}\right) = o(1) \end{aligned}$$

by the proof of Theorem 3.4(i), we have $\frac{1}{N}\sum_{i\in\hat{G}_0}\left\|\hat{b}_i\right\|^2 = o_P((NT)^{-1})$. Using the fact that $\mathbf{1}\{i\in\hat{G}_k\} = \mathbf{1}\{i\in G_k^0\} + \mathbf{1}\{i\in\hat{G}_k\backslash G_k^0\} - \mathbf{1}\{i\in G_k^0\backslash\hat{G}_k\}$, we have

$$\begin{aligned} \frac{1}{N}\sum_{i\in\hat{G}_k}\left\|\hat{\alpha}_k - \beta_i^0\right\|^2 - \frac{N_k}{N}\left\|\hat{\alpha}_k - \alpha_k^0\right\|^2 &= \frac{1}{N}\sum_{i\in\hat{G}_k}\left\|\hat{\alpha}_k - \beta_i^0\right\|^2 - \frac{1}{N}\sum_{i\in G_k^0}\left\|\hat{\alpha}_k - \beta_i^0\right\|^2 \\ &= \frac{1}{N}\sum_{i\in\hat{G}_k\backslash G_k^0}\left\|\hat{\alpha}_k - \beta_i^0\right\|^2 - \sum_{i\in G_k^0\backslash\hat{G}_k}\left\|\hat{\alpha}_k - \beta_i^0\right\|^2 \\ &\equiv R_{1k} - R_{2k}, \text{ say.} \end{aligned}$$

By Theorem 3.4, $P(\|R_{1k}\| \geq \epsilon/(NT)) \leq P(\hat{F}_{kNT}) \to 0$ and $P(\|R_{2k}\| \geq \epsilon/(NT)) \leq P(\hat{E}_{kNT}) \to 0$. It follows that $\frac{1}{N}\sum_{i\in\hat{G}_k}\left\|\hat{\alpha}_k - \beta_i^0\right\|^2 - \frac{N_k}{N}\left\|\hat{\alpha}_k - \alpha_k^0\right\|^2 = o_P((NT)^{-1})$. Consequently,

$$\eta_{NT}^2 = \sum_{k=1}^{K_0}\frac{N_k}{N}\left\|\hat{\alpha}_k - \alpha_k^0\right\|^2 + o_P((NT)^{-1}) = O_P\left(\sum_{k=1}^{K_0}\left\|\hat{\alpha}_k - \alpha_k^0\right\|^2\right) + o_P((NT)^{-1}).$$

Then (i) follows by Lemma A.2(i). Similarly, we can prove (ii)-(v). ■

Proof of Lemma A.6. (i) Using Theorem 3.2(iii) and Assumption A.2(iii), we can show that $\hat{A}_{1k,l} = O_P\left(\left\|\hat{\alpha}_l - \alpha_l^0\right\|\right) = O_P\left(T^{-1/2}\right)$.

(ii) Let $A_{2k,l}^{(1)} = \frac{1}{NN_kT}\sum_{j\in\hat{G}_l}\sum_{i\in G_k^0}\nu_{ji}X_i' M_{\hat{F}} X_j$ and $A_{2k,l}^{(2)} = \frac{1}{NN_kT}\sum_{j\in G_l^0}\sum_{i\in G_k^0}\nu_{ji}X_i' M_{\hat{F}} X_j$. Then $\hat{A}_{2k,l} - A_{2k,l} = (\hat{A}_{2k,l} - A_{2k,l}^{(1)}) + (A_{2k,l}^{(1)} - A_{2k,l}^{(2)}) + (A_{2k,l}^{(2)} - A_{2k,l})$. Using the fact that $\mathbf{1}\{i\in\hat{G}_k\} = \mathbf{1}\{i\in G_k^0\} + \mathbf{1}\{i\in\hat{G}_k\backslash G_k^0\} - \mathbf{1}\{i\in G_k^0\backslash\hat{G}_k\}$, we have

$$\begin{aligned} \hat{A}_{2k,l} - A_{2k,l}^{(1)} &= \frac{1}{NN_kT}\sum_{j\in\hat{G}_l}\left[\sum_{i\in\hat{G}_k}\nu_{ji}X_i' M_{\hat{F}} X_j - \sum_{i\in G_k^0}\nu_{ji}X_i' M_{\hat{F}} X_{jj}\right] \\ &= \frac{1}{NN_kT}\sum_{j\in\hat{G}_l}\left[\sum_{i\in\hat{G}_k\backslash G_k^0}\nu_{ji}X_i' M_{\hat{F}} X_j - \sum_{i\in G_k^0\backslash\hat{G}_k}\nu_{ji}X_i' M_{\hat{F}} X_{jj}\right] \\ &\equiv R_{k,l,1} - R_{k,l,2}, \text{ say.} \end{aligned}$$

Let $\epsilon > 0$. By Theorem 3.4, $P\left(\|R_{k,l,1}\| \geq \epsilon/(NT)^{1/2}\right) \leq P(\hat{F}_{kNT}) \to 0$ and $P\left(\|R_{k,l,2}\| \geq \epsilon/(NT)^{1/2}\right)$ $\leq P(\hat{E}_{kNT}) \to 0$. It follows that $\hat{A}_{2k,l} - A_{2k,l}^{(1)} = o_P((NT)^{-1/2})$. By the same token, we can show that

$A^{(1)}_{2k,l} - A^{(2)}_{2k,l} = o_P((NT)^{-1/2})$. Now, using $|\nu_{ji}| \leq \left\|\lambda_i^0\right\| \left\|\lambda_j^0\right\| \left\|\left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\right\|$ and by Lemma A.2(v),

$$
\begin{aligned}
\left\|A^{(2)}_{2k,l} - A_{2k,l}\right\| &\leq \left\|P_{\hat{F}} - P_{F^0}\right\| \frac{1}{NN_kT} \sum_{j\in G_l^0}\sum_{i\in G_k^0} |\nu_{ji}| \left\|X_i\right\| \left\|X_j\right\| \\
&\leq \left\|P_{\hat{F}} - P_{F^0}\right\| \left\|\left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\right\| \left\{\frac{1}{N_kT^{1/2}} \sum_{i\in G_k^0} \left\|\lambda_i^0\right\| \left\|X_i\right\|\right\}^2 = O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}
$$

It follows that $\hat{A}_{2k,l} = A_{2k,l} + O_P\left(\delta_{NT}^{-1}\right)$.

(iii) By the triangle inequality,

$$
\frac{1}{NT^{3/2}} \left\|\sum_{j\in\hat{G}_l} X_j\varepsilon_j'\hat{F}G\right\| \leq \frac{1}{NT^{3/2}} \left\|\sum_{j\in\hat{G}_l} X_j\varepsilon_j'F^0HG\right\| + \frac{1}{NT^{3/2}} \left\|\sum_{j\in\hat{G}_l} X_j\varepsilon_j'(\hat{F}-F^0H)G\right\| \equiv a_{l1} + a_{l2}, \text{ say.}
$$

As in the proof of (ii), using the fact that $\mathbf{1}\{j\in\hat{G}_k\} = \mathbf{1}\{j\in G_k^0\} + \mathbf{1}\{j\in\hat{G}_k\backslash G_k^0\} - \mathbf{1}\{j\in G_k^0\backslash\hat{G}_k\}$ and Theorem 3.4, we can show that $a_{l1} = \frac{1}{NT^{3/2}}\sum_{j\in G_l^0} X_j\varepsilon_j'F^0HG + o_P((NT)^{-1/2})$. It follows that $\left\|a_{l1}\right\| \leq \frac{1}{NT^{3/2}}\left\|\sum_{j\in G_l^0} X_j\varepsilon_j'F^0\right\| \left\|HG\right\| + o_P((NT)^{-1/2}) = O_P\left(T^{-1/2}\right)$ by Assumption A.1(ix). Similarly, by Cauchy-Schwarz inequality and Lemma A.7(iii),

$$
\begin{aligned}
\left|\omega_T'a_{l2}\bar{\omega}_{R_0}\right| &\leq \frac{1}{NT^{3/2}} \left\|\sum_{j\in G_l^0} \omega_T'X_j\varepsilon_j'\left(\hat{F}-F^0H\right)G\bar{\omega}_{R_0}\right\| + o_P((NT)^{-1/2}) \\
&\leq \frac{1}{NT^{3/2}} \sum_{j\in G_l^0} \left\{\omega_T'X_jX_j'\omega_T\right\}^{1/2} \left\{\bar{\omega}_{R_0}'G'(\hat{F}-F^0H)\varepsilon_j\varepsilon_j'(\hat{F}-F^0H)G\bar{\omega}_{R_0}\right\}^{1/2} + o_P((NT)^{-1/2}) \\
&\leq \left\{\frac{1}{NT^{3/2}} \sum_{j\in G_l^0} \omega_T'X_jX_j'\omega_T\right\}^{1/2} \left\{\frac{1}{NT^{3/2}} \sum_{j\in G_l^0} \bar{\omega}_{R_0}'G'(\hat{F}-F^0H)\varepsilon_j\varepsilon_j'(\hat{F}-F^0H)G\bar{\omega}_{R_0}\right\}^{1/2} \\
&\quad + o_P((NT)^{-1/2}) \\
&\leq \left\|G\right\| \left\{\frac{1}{NT} \sum_{j\in G_l^0} \left\|X_j\right\|^2\right\}^{1/2} \left\{\frac{1}{NT^2} \sum_{j\in G_l^0} \varepsilon_j'(\hat{F}-F^0H)(\hat{F}-F^0H)'\varepsilon_j\right\}^{1/2} + o_P((NT)^{-1/2}) \\
&= O_P(1)\, O_P\left(\delta_{NT}^{-1}\right) + o_P((NT)^{-1/2}) = O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}
$$

It follows that $\frac{1}{NT^{3/2}}\left\|\sum_{j\in\hat{G}_l} X_j\varepsilon_j'\hat{F}G\right\| = O_P\left(\delta_{NT}^{-1}\right)$. Then

$$
\begin{aligned}
\left|\omega_p'A_{3k,l}\right| &= \left|\frac{1}{NN_kT^2} \sum_{i\in\hat{G}_k} \omega_p'X_i'M_{\hat{F}} \sum_{j\in\hat{G}_l} X_j\varepsilon_j'\hat{F}G\lambda_i^0\right| = \left|\frac{1}{NN_kT^2} \operatorname{tr}\left(\sum_{j\in\hat{G}_l} X_j\varepsilon_j'\hat{F}G \sum_{i\in\hat{G}_k} \lambda_i^0\omega_p'X_i'M_{\hat{F}}\right)\right| \\
&\leq \left\{\frac{1}{NT^{3/2}} \left\|\sum_{j\in\hat{G}_l} X_j\varepsilon_j'\hat{F}G\right\|\right\} \left\{\frac{1}{N_kT^{1/2}} \left\|\sum_{i\in\hat{G}_k} \lambda_i^0\omega_p'X_i'\right\|\right\} \\
&\leq \left\{\frac{1}{NT^{3/2}} \left\|\sum_{j\in\hat{G}_l} X_j\varepsilon_j'\hat{F}G\right\|\right\} \left\{\frac{1}{N_kT^{1/2}} \sum_{i=1}^{N} \left\|\lambda_i^0\right\| \left\|X_i\right\|\right\} \\
&= O_P\left(\delta_{NT}^{-1}\right) O_P(1) = O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}
$$

It follows that $\hat{A}_{3k,l} = O_P\left(\delta_{NT}^{-1}\right)$.

(iv) By Lemma A.2(i) and Assumption A.2(iii)

$$\begin{aligned}
\left\|\hat{A}_{4k,l}\right\| &= \left\|\frac{1}{NN_kT^2}\sum_{j\in\hat{G}_l}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}}(F^0H-\hat{F})H^{-1}\lambda_j^0\lambda_i^{0\prime}G'\hat{F}'X_j\right\| \\
&\leq \left\{\frac{1}{NT^{1/2}}\sum_{i\in\hat{G}_k}\|X_i\|_{\text{sp}}\left\|\lambda_i^0\right\|\right\}\left\{\frac{1}{N_kT^{1/2}}\sum_{j\in\hat{G}_l}\left\|\lambda_j^0\right\|\|X_j\|_{\text{sp}}\right\} \\
&\quad\times\frac{1}{T^{1/2}}\left\|F^0H-\hat{F}\right\|\left\|H^{-1}\right\|_{\text{sp}}\frac{1}{T^{1/2}}\left\|G'\hat{F}'\right\|_{\text{sp}} \\
&= O_P(1)\,O_P(1)\,O_P\left(\delta_{NT}^{-1}\right)O_P(1) = O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}$$

(v) By arguments as used in the proof of Lemma A.5(i),

$$\begin{aligned}
\left|\omega_p'\hat{A}_{5k,l}\bar{\omega}_p\right| &= \frac{1}{NN_kT^2}\left|\sum_{j\in\hat{G}_l}\sum_{i\in\hat{G}_k}\omega_p'X_i'M_{\hat{F}}\varepsilon_j\lambda_i^{0\prime}G'\hat{F}'X_j\bar{\omega}_p\right| \\
&= \frac{1}{NN_kT^2}\left|\text{tr}\left(\sum_{j\in G_l^0}\sum_{i\in G_k^0}\lambda_i^{0\prime}\omega_p'X_i'M_{F^0}\varepsilon_j\bar{\omega}_p'X_j'\hat{F}G\right)\right| + o_P((NT)^{-1/2}) \\
&\leq \frac{1}{NN_kT^{3/2}}\left\|\sum_{i\in G_k^0}\lambda_i^{0\prime}\omega_p'X_i'M_{F^0}\sum_{j\in G_l^0}\varepsilon_j\bar{\omega}_p'X_j'\right\|\frac{1}{T^{1/2}}\left\|\hat{F}G\right\| + o_P((NT)^{-1/2}) \\
&\leq \left\{\frac{1}{N_kT^{1/2}}\sum_{i\in G_k^0}\left\|\lambda_i^0\right\|\|X_i\|\right\}\left\{\frac{1}{NT}\left\|\sum_{j\in G_l^0}\varepsilon_j\bar{\omega}_p'X_j'\right\|\right\}\frac{1}{T^{1/2}}\left\|\hat{F}G\right\| + o_P((NT)^{-1/2}) \\
&= O_P(1)\,O_P\left(\delta_{NT}^{-1}\right)O_P(1) + o_P((NT)^{-1/2}) = O_P\left(\delta_{NT}^{-1}\right)
\end{aligned}$$

where we use the fact that $\frac{1}{NT}\left\|\sum_{j\in G_l^0}\varepsilon_j\bar{\omega}_p'X_j'\right\| = O_P\left(\delta_{NT}^{-1}\right)$ by Assumption A.1(viii). It follows that $\hat{A}_{5k,l} = O_P\left(\delta_{NT}^{-1}\right)$.

(vi) This follows from (i)-(v). ■

Proof of Lemma A.7. (i) The proof parallels that of Lemma A.4(i). By (B.1) we have $\frac{1}{NT}\sum_{i=1}^N\varepsilon_i'(\hat{F}-F^0H) = \sum_{l=1}^8\frac{1}{NT}\sum_{i=1}^N\varepsilon_i'a_lV_{NT}^{-1} \equiv \sum_{l=1}^8 S_l$, say. Using $\|a_1\| = O_P\left(T^{-1/2}\right)$, we have $\|S_1\| \leq \|a_1\|\left\|V_{NT}^{-1}\right\| \times\frac{1}{NT}\sum_{i=1}^N\|\varepsilon_i\| = O_P\left(T^{-1/2}\right)O_P\left(T^{-1/2}\right) = O_P\left(T^{-1}\right)$. Using arguments as used in the proofs of Lemmas A.5(i) and A.6(ii), we can show that $a_2 = -\frac{1}{NT}\sum_{i=1}^N X_i\hat{b}_i\lambda_i^{0\prime}F^{0\prime}\hat{F} = -\frac{1}{NT}\sum_{l=1}^{K_0}\sum_{i\in\hat{G}_l}X_i\hat{b}_i\lambda_i^{0\prime}F^{0\prime}\hat{F} + o_P((NT)^{-1}) = -\frac{1}{NT}\sum_{l=1}^{K_0}\sum_{i\in G_l^0}X_i\left(\hat{\alpha}_l-\alpha_l^0\right)\lambda_i^{0\prime}F^{0\prime}\hat{F} + o_P((NT)^{-1})$. It follows that

$$\begin{aligned}
\|S_2\| &= \left\|\frac{1}{NT}\sum_{i=1}^N\varepsilon_i'\left[\frac{1}{NT}\sum_{l=1}^{K_0}\sum_{j\in G_l^0}X_j\left(\hat{\alpha}_l-\alpha_l^0\right)\lambda_j^{0\prime}F^{0\prime}\hat{F} + o_P\left((NT)^{-1}\right)\right]V_{NT}^{-1}\right\| \\
&\leq \left\|\frac{1}{N^2T^2}\sum_{l=1}^{K_0}\sum_{j\in G_l^0}\sum_{i=1}^N\lambda_i^0\varepsilon_i'X_j\left(\hat{\alpha}_l-\alpha_l^0\right)\lambda_j^{0\prime}F^{0\prime}\hat{F}V_{NT}^{-1}\right\| + o_P((NT)^{-1}) \\
&\leq \frac{1}{T^{1/2}}\sum_{l=1}^{K_0}\left\|\hat{\alpha}_l-\alpha_l^0\right\|\left\{\frac{1}{N^2T^{1/2}}\sum_{j\in G_l^0}\sum_{i=1}^N\left\|\lambda_i^0\right\|\left\|\varepsilon_i'X_j\right\|\left\|\lambda_j^0\right\|\right\}\frac{1}{T}\left\|F^{0\prime}\hat{F}V_{NT}^{-1}\right\| + o_P((NT)^{-1}) \\
&= T^{-1/2}O_P(T^{-1/2})O_P(1)\,O_P(1) + o_P((NT)^{-1}) = O_P\left(T^{-1}\right).
\end{aligned}$$

Similarly, we can show that $\|S_s\| = O_P\left(T^{-1}\right)$ for $s = 3, 4, 5$. By the triangle inequality,

$$
\begin{aligned}
\|S_6\| &= \left\| \frac{1}{N^2T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i' F^0 \lambda_j^0 \varepsilon_j' \hat{F}\ V_{NT}^{-1} \right\| \\
&\leq \frac{1}{N^2T^2} \left\| \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i' F^0 \lambda_j^0 \varepsilon_j' F^0 H V_{NT}^{-1} \right\| + \frac{1}{N^2T^2} \left\| \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i' F^0 \lambda_j^0 \varepsilon_j' (\hat{F} - F^0 H) V_{NT}^{-1} \right\| \\
&\leq \frac{1}{T} \left\{ \frac{1}{NT^{1/2}} \sum_{i=1}^{N} \left\| \varepsilon_i' F^0 \right\| \right\} \left\{ \frac{1}{NT^{1/2}} \sum_{j=1}^{N} \left\| \lambda_j^0 \varepsilon_j' F^0 \right\| \right\} \left\| H V_{NT}^{-1} \right\| \\
&\quad + \frac{1}{T^{1/2}} \frac{1}{NT^{1/2}} \sum_{i=1}^{N} \left\| \varepsilon_i' F^0 \right\| \frac{1}{NT^{1/2}} \sum_{j=1}^{N} \left\| \lambda_j^0 \varepsilon_j' \right\| \left\{ \frac{1}{T^{1/2}} \left\| \hat{F} - F^0 H \right\| \right\} \left\| V_{NT}^{-1} \right\| \\
&= T^{-1} O_P(1) O_P(1) O_P(1) + T^{-1/2} O_P(1) O_P(1) O_P\left(\delta_{NT}^{-1}\right) O_P(1) = T^{-1/2} O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}
$$

Similarly, we can show that $S_s = O_P\left(\delta_{NT}^{-2}\right)$ for $s = 7$, 8. Thus $\frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i'(\hat{F} - F^0 H) = O_P\left(\delta_{NT}^{-2}\right)$.

(ii) The proof is analogous to that of (i) and thus omitted.

(iii) By (B.1), $T^{-1}\varepsilon_i'(\hat{F} - F^0 H) = \sum_{l=1}^{8} \frac{1}{T} \varepsilon_i' a_l V_{NT}^{-1} \equiv \sum_{l=1}^{8} A_{il}'$, say. By Cauchy-Schwarz inequality

$$
\frac{1}{NT^2} \sum_{i=1}^{N} \varepsilon_i'(\hat{F} - F^0 H)(\hat{F} - F^0 H)' \varepsilon_i = \sum_{l=1}^{8} \sum_{k=1}^{8} \frac{1}{N} \sum_{i=1}^{N} A_{il}' A_{ik} \leq 8 \sum_{l=1}^{8} \frac{1}{N} \sum_{i=1}^{N} A_{il}' A_{il}.
$$

Using $\|a_1\| = O_P\left(T^{-1/2}\right)$, we have

$$
\begin{aligned}
\frac{1}{N} \sum_{i=1}^{N} A_{i1}' A_{i1} &= \frac{1}{NT^2} \sum_{i=1}^{N} \varepsilon_i' a_1 V_{NT}^{-1} V_{NT}^{-1} a_1' \varepsilon_i \leq \frac{1}{T} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \|\varepsilon_i\|^2 \right\} \|a_1\|^2 \left\| V_{NT}^{-1} \right\|^2 \\
&= \frac{1}{T} O_P(1) O_P\left(T^{-1}\right) = O_P\left(T^{-2}\right).
\end{aligned}
$$

Using arguments as used in the proofs of Lemmas A.5(i) and A.6(ii), we can show that $a_2 = -\frac{1}{NT} \sum_{i=1}^{N} X_i \hat{b}_i \lambda_i^{0\prime} F^{0\prime} \hat{F} = -\frac{1}{NT} \sum_{l=1}^{K_0} \sum_{i \in \hat{G}_l} X_i \hat{b}_i \lambda_i^{0\prime} F^{0\prime} \hat{F} + o_P((NT)^{-1}) = -\frac{1}{NT} \sum_{l=1}^{K_0} \sum_{i \in G_l^0} X_i \left(\hat{\alpha}_l - \alpha_l^0\right) \lambda_i^{0\prime} F^{0\prime} \hat{F} + o_P((NT)^{-1})$. It follows that

$$
\begin{aligned}
\frac{1}{N} \sum_{i=1}^{N} A_{i2}' A_{i2} &= \sum_{l=1}^{K_0} \sum_{k=1}^{K_0} \frac{1}{N^3T^4} \sum_{i=1}^{N} \sum_{j \in G_l^0} \sum_{m \in G_l^0} \varepsilon_i' X_j \left(\hat{\alpha}_l - \alpha_l^0\right) \lambda_j^{0\prime} F^{0\prime} \hat{F} V_{NT}^{-2} \hat{F}' F^0 \lambda_m^0 \left(\hat{\alpha}_k - \alpha_k^0\right)' X_m' \varepsilon_i \\
&\quad + o_P((NT)^{-1}) \\
&\leq \frac{c_{NT}}{T} \sum_{l=1}^{K_0} \sum_{k=1}^{K_0} \left\| \hat{\alpha}_l - \alpha_l^0 \right\| \left\| \hat{\alpha}_m - \alpha_m^0 \right\| \left\{ \frac{1}{N^3T} \sum_{i=1}^{N} \sum_{j \in G_l^0} \sum_{m \in G_l^0} \left\| \varepsilon_i' X_j \right\| \left\| \lambda_j^0 \right\| \left\| \lambda_m^0 \right\| \left\| X_m' \varepsilon_i \right\| \right\} \\
&\quad + o_P((NT)^{-1}) \\
&\leq \frac{c_{NT}}{T} \sum_{l=1}^{K_0} \sum_{k=1}^{K_0} \left\| \hat{\alpha}_l - \alpha_l^0 \right\| \left\| \hat{\alpha}_m - \alpha_m^0 \right\| \left\{ \frac{1}{N^3T} \sum_{i=1}^{N} \sum_{j \in G_l^0} \left\| \varepsilon_i' X_j \right\|^2 \left\| \lambda_j^0 \right\|^2 \right\} + o_P((NT)^{-1}) \\
&= \frac{1}{T} O_P\left(T^{-1/2}\right) O_P\left(T^{-1/2}\right) O_P(1) + o_P((NT)^{-1}) = O_P\left(T^{-2}\right) + o_P((NT)^{-1}),
\end{aligned}
$$

where $c_{NT} = T^{-2} \left\| F^{0\prime} \hat{F} V_{NT}^{-2} \hat{F}' F^0 \right\| = O_P(1)$. Analogously, we can show that $\frac{1}{N} \sum_{i=1}^{N} A_{is}' A_{is} = O_P\left(T^{-2}\right)$ $+ o_P((NT)^{-1})$ for $s = 3, 4, 5$.

Noting that $\hat{F}\hat{F}' = (F^0H + \hat{F} - F^0H)(F^0H + \hat{F} - F^0H)' \leq 2F^0HH'F^{0\prime} + 2(\hat{F} - F^0H)(\hat{F} - F^0H)'$, we have

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} A_{i6}'A_{i6} &= \frac{1}{NT^2}\sum_{i=1}^{N}\varepsilon_i' a_6 V_{NT}^{-1}V_{NT}^{-1}a_6'\varepsilon_i \\
&\leq \left\|V_{NT}^{-1}\right\|^2 \frac{1}{N^3T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\varepsilon_i' F^0\lambda_j^0\varepsilon_j'\hat{F}\hat{F}'\varepsilon_k\lambda_k^{0\prime}F^0\varepsilon_i \\
&\leq 2\left\|V_{NT}^{-1}\right\|^2 \frac{1}{N^3T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\varepsilon_i' F^0\lambda_j^0\varepsilon_j' F^0HH'F^{0\prime}\varepsilon_k\lambda_k^{0\prime}F^0\varepsilon_i \\
&\quad +2\left\|V_{NT}^{-1}\right\|^2 \frac{1}{N^3T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\varepsilon_i' F^0\lambda_j^0\varepsilon_j'(\hat{F} - F^0H)(\hat{F} - F^0H)'\varepsilon_k\lambda_k^{0\prime}F^0\varepsilon_i.
\end{aligned}$$

The first term is bounded from above by

$$\begin{aligned}
&\frac{2}{T^2}\left\|V_{NT}^{-1}\right\|^2\|H\|^2\frac{1}{N^3T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\varepsilon_i' F^0\lambda_j^0\varepsilon_j' F^0F^{0\prime}\varepsilon_k\lambda_k^{0\prime}F^0\varepsilon_i \\
\leq\ & O_P\left(T^{-2}\right)\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{NT}\sum_{j=1}^{N}\varepsilon_i' F^0\lambda_j^0\varepsilon_j' F^0\right\|^2 = O_P\left(T^{-2}\right).
\end{aligned}$$

By (ii), the second term is bounded from above by $\frac{2}{T}\left\|V_{NT}^{-1}\right\|^2\left\|\frac{1}{NT}\sum_{j=1}^{N}\lambda_j^0\varepsilon_j'(\hat{F} - F^0H)\right\|^2\frac{1}{NT}\sum_{i=1}^{N}\left\|\varepsilon_i' F^0\right\|^2$ $= \frac{1}{T}O_P\left(\delta_{NT}^{-4}\right)$. It follows that $\frac{1}{N}\sum_{i=1}^{N}A_{i6}'A_{i6} = O_P\left(T^{-2}\right)$. Analogously, we can show that $\frac{1}{N}\sum_{i=1}^{N}A_{i7}'A_{i7}$ $= O_P\left(T^{-2}\right)$. Now,

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} A_{i8}'A_{i8} &= \frac{1}{NT^2}\sum_{i=1}^{N}\varepsilon_i' a_8 V_{NT}^{-1}V_{NT}^{-1}a_8'\varepsilon_i \\
&\leq \left\|V_{NT}^{-1}\right\|^2 \frac{1}{N^3T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\varepsilon_i'\varepsilon_j\varepsilon_j'\hat{F}\ \hat{F}'\varepsilon_k\varepsilon_k'\varepsilon_i \\
&\leq 2\left\|V_{NT}^{-1}\right\|^2 \frac{1}{N^3T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\varepsilon_i'\varepsilon_j\varepsilon_j' F^0HH'F^{0\prime}\varepsilon_k\varepsilon_k'\varepsilon_i \\
&\quad +2\left\|V_{NT}^{-1}\right\|^2 \frac{1}{N^3T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\varepsilon_i'\varepsilon_j\varepsilon_j'(\hat{F} - F^0H)(\hat{F} - F^0H)'\varepsilon_k\varepsilon_k'\varepsilon_i.
\end{aligned}$$

The first term is bounded from above by

$$\begin{aligned}
&2\left\|V_{NT}^{-1}\right\|^2\|H\|^2\frac{1}{N^3T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\varepsilon_i'\varepsilon_j\varepsilon_j' F^0F^{0\prime}\varepsilon_k\varepsilon_k'\varepsilon_i \\
=\ & \frac{2}{NT^2}\left\|V_{NT}^{-1}\right\|^2\|H\|^2\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{NT}\sum_{j=1}^{N}\varepsilon_i'\varepsilon_j\varepsilon_j' F^0\right\|^2 = O_P\left(N^{-1}T^{-2}\right).
\end{aligned}$$

For the second term, we apply (i) to obtain

$$\begin{aligned}
& 2\left\|V_{NT}^{-1}\right\|^{2}\frac{1}{N^{3}T^{4}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\varepsilon_{i}^{\prime}\varepsilon_{j}\varepsilon_{j}^{\prime}(\hat{F}-F^{0}H)(\hat{F}-F^{0}H)^{\prime}\varepsilon_{k}\varepsilon_{k}^{\prime}\varepsilon_{i} \\
\leq\ & \frac{2\left\|V_{NT}^{-1}\right\|^{2}}{T}\frac{1}{T}\left\|\hat{F}-F^{0}H\right\|^{2}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{NT}\sum_{j=1}^{N}\varepsilon_{i}^{\prime}\varepsilon_{j}\varepsilon_{j}^{\prime}\right\|^{2}=T^{-1}O_{P}\left(\delta_{NT}^{-2}\right)O_{P}\left(1\right)=O_{P}\left(T^{-1}\delta_{NT}^{-2}\right).
\end{aligned}$$

Thus $\frac{1}{N}\sum_{i=1}^{N}A_{i8}^{\prime}A_{i8}=O_{P}\left(T^{-1}\delta_{NT}^{-2}\right)$. In sum, $\frac{1}{NT^{2}}\sum_{i=1}^{N}\varepsilon_{i}^{\prime}(\hat{F}-F^{0}H)GG^{\prime}(\hat{F}-F^{0}H)^{\prime}\varepsilon_{i}=O_{P}\left(\delta_{NT}^{-4}\right)$.

(iv) The proof is analogous to that of (iii) and thus omitted.

(v) For any nonrandom $\omega_{p}\in\mathbb{R}^{p}$ with $\|\omega_{p}\|=1$, we have

$$\begin{aligned}
& \left|\frac{1}{NN_{k}T^{2}}\sum_{i\in G_{k}^{0}}\sum_{j=1}^{N}\omega_{p}^{\prime}X_{i}^{\prime}\varepsilon_{j}\varepsilon_{j}^{\prime}(\hat{F}-F^{0}H)G\lambda_{i}^{0}\right| \\
=\ & \left|\frac{1}{NN_{k}T^{2}}\sum_{i\in G_{k}^{0}}\sum_{j=1}^{N}\varepsilon_{j}^{\prime}(\hat{F}-F^{0}H)G\lambda_{i}^{0}\omega_{p}^{\prime}X_{i}^{\prime}\varepsilon_{j}\right| \\
\leq\ & \frac{1}{NN_{k}T^{2}}\sum_{i\in G_{k}^{0}}\sum_{j=1}^{N}\left\{\varepsilon_{j}^{\prime}(\hat{F}-F^{0}H)GG^{\prime}(\hat{F}-F^{0}H)^{\prime}\varepsilon_{j}\right\}^{1/2}\left\{\varepsilon_{j}^{\prime}X_{i}\omega_{p}\lambda_{i}^{0\prime}\lambda_{i}^{0}\omega_{p}^{\prime}X_{i}^{\prime}\varepsilon_{j}\right\}^{1/2} \\
\leq\ & \frac{\|G\|^{2}}{T}\left\{\frac{1}{NT}\sum_{j=1}^{N}\varepsilon_{j}^{\prime}(\hat{F}-F^{0}H)(\hat{F}-F^{0}H)^{\prime}\varepsilon_{j}\right\}^{1/2}\left\{\frac{1}{NN_{k}T}\sum_{i\in G_{k}^{0}}\sum_{j=1}^{N}\left\|\varepsilon_{j}^{\prime}X_{i}\right\|^{2}\left\|\lambda_{i}^{0}\right\|^{2}\right\}^{1/2} \\
=\ & T^{-1}O_{P}\left(\delta_{NT}^{-2}\right)O_{P}\left(1\right)=O_{P}\left(T^{-1}\delta_{NT}^{-2}\right).
\end{aligned}$$

(vi) By (B.1) we have $\frac{1}{N_{k}T}\sum_{i\in\hat{G}_{k}}X_{i}^{\prime}M_{\hat{F}}F^{0}\lambda_{i}^{0}=-\frac{1}{N_{k}T}\sum_{i\in\hat{G}_{k}}X_{i}^{\prime}M_{\hat{F}}\left(\hat{F}-F^{0}H\right)H^{-1}\lambda_{i}^{0}=-\sum_{l=1}^{8}\frac{1}{N_{k}T}$ $\sum_{i\in\hat{G}_{k}}X_{i}^{\prime}M_{\hat{F}}a_{l}G\lambda_{i}^{0}\equiv\sum_{l=1}^{8}J_{k,l}$. Using arguments like those used in the derivation of (A.23), we can readily show that

$$\begin{aligned}
J_{k,1} &= -\frac{1}{N_{k}T}\sum_{i\in\hat{G}_{k}}X_{i}^{\prime}M_{\hat{F}}\left(\frac{1}{NT}\sum_{j=1}^{N}X_{j}\hat{b}_{j}\hat{b}_{j}^{\prime}X_{j}^{\prime}\hat{F}\right)G\lambda_{i}^{0} \\
&= -\frac{1}{N_{k}T}\sum_{i\in\hat{G}_{k}}X_{i}^{\prime}M_{\hat{F}}\frac{1}{NT}\sum_{l=1}^{K_{0}}\sum_{j\in\hat{G}_{l}}X_{j}\left(\hat{\alpha}_{l}-\alpha_{l}^{0}\right)\left(\hat{\alpha}_{l}-\alpha_{l}^{0}\right)^{\prime}X_{j}^{\prime}\hat{F}G\lambda_{i}^{0}+o_{P}((NT)^{-1/2}) \\
&= \sum_{l=1}^{K_{0}}\bar{A}_{1k,l}\left(\hat{\alpha}_{l}-\alpha_{l}^{0}\right)+o_{P}((NT)^{-1/2}), \\
J_{k,2} &= \frac{1}{N_{k}T}\sum_{i\in\hat{G}_{k}}X_{i}^{\prime}M_{\hat{F}}\left(\frac{1}{NT}\sum_{j=1}^{N}X_{j}\hat{b}_{j}\lambda_{j}^{0\prime}F^{0\prime}\hat{F}\right)G\lambda_{i}^{0} \\
&= \frac{1}{N_{k}T}\sum_{i\in\hat{G}_{k}}X_{i}^{\prime}M_{\hat{F}}\frac{1}{N}\sum_{l=1}^{K_{0}}\sum_{j\in\hat{G}_{l}}X_{j}\left(\hat{\alpha}_{l}-\alpha_{l}^{0}\right)\nu_{ji}+o_{P}((NT)^{-1/2}) \\
&= \sum_{l=1}^{K_{0}}\bar{A}_{2k,l}\left(\hat{\alpha}_{l}-\alpha_{l}^{0}\right)+o_{P}((NT)^{-1/2}),
\end{aligned}$$

$$
\begin{aligned}
J_{k,3} &= \frac{1}{N_kT}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}} \left(\frac{1}{NT}\sum_{j=1}^{N} X_j \hat{b}_j \varepsilon_j' \hat{F}\right) G\lambda_i^0 \\
&= \frac{1}{N_kT}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}} \frac{1}{NT}\sum_{l=1}^{K_0}\sum_{j\in\hat{G}_l} X_j \left(\hat{\alpha}_l-\alpha_l^0\right) \varepsilon_j' \hat{F} G\lambda_i^0 + o_P((NT)^{-1/2}) \\
&= \sum_{l=1}^{K_0} \bar{A}_{3k,l}\left(\hat{\alpha}_l-\alpha_l^0\right) + o_P((NT)^{-1/2}), \\
J_{k,4} &= \frac{1}{N_kT}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}} \left(\frac{1}{NT}\sum_{j=1}^{N} F^0\lambda_j^0 \hat{b}_j' X_j' \hat{F}\right) G\lambda_i^0 \\
&= \frac{1}{N_kT}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}} \frac{1}{NT}\sum_{l=1}^{K_0}\sum_{j\in\hat{G}_l} F^0\lambda_j^0 \left(\hat{\alpha}_l-\alpha_l^0\right)' X_j' \hat{F} G\lambda_i^0 + o_P((NT)^{-1/2}) \\
&= \sum_{l=1}^{K_0} \bar{A}_{4k,l}\left(\hat{\alpha}_l-\alpha_l^0\right) + o_P((NT)^{-1/2}), \text{ and} \\
J_{k,5} &= \frac{1}{N_kT}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}} \left(\frac{1}{NT}\sum_{j=1}^{N} \varepsilon_j \hat{b}_j' X_j' \hat{F}\right) G\lambda_i^0 \\
&= \frac{1}{N_kT}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}} \frac{1}{NT}\sum_{l=1}^{K_0}\sum_{j\in\hat{G}_l} \varepsilon_j \left(\hat{\alpha}_l-\alpha_l^0\right)' X_j' \hat{F} G\lambda_i^0 + o_P((NT)^{-1/2}) \\
&= \sum_{l=1}^{K_0} \bar{A}_{5k,l}\left(\hat{\alpha}_l-\alpha_l^0\right) + o_P((NT)^{-1/2}).
\end{aligned}
$$

Note that $J_{k6} = -\frac{1}{N_kT}\sum_{i\in\hat{G}_k} X_i' M_{\hat{F}}(\frac{1}{NT}\sum_{j=1}^{N} F^0\lambda_j^0\varepsilon_j'\hat{F})G\lambda_i^0$. Using $M_{\hat{F}}F^0H = M_{\hat{F}}(F^0H-\hat{F})$, Lemma A.2(i), and Theorem 3.2, we have

$$
\begin{aligned}
\left|\omega_p' J_{k,6}\right| &= \frac{1}{N_kT}\left|\operatorname{tr}\left[\sum_{i\in\hat{G}_k}\lambda_i^0\omega_p' X_i' M_{\hat{F}}\left(F^0H-\hat{F}\right)H^{-1}\left(\frac{1}{NT}\sum_{j=1}^{N}\lambda_j^0\varepsilon_j'\hat{F}\right)G\right]\right| \\
&\leq \frac{1}{N_kT}\left\|\sum_{i\in\hat{G}_k}\lambda_i^0\omega_p' X_i' M_{\hat{F}}\left(F^0H-\hat{F}\right)\right\|\left\|\frac{1}{NT}\sum_{j=1}^{N}\lambda_j^0\varepsilon_j'\hat{F}\right\|\left\|H^{-1}\right\|\left\|G\right\| \\
&\leq \left\{\frac{1}{N_kT^{1/2}}\sum_{i\in\hat{G}_k}\left\|\lambda_i^0\right\|\left\|X_i\right\|_{\text{sp}}\right\}\left\{\frac{1}{T^{1/2}}\left\|F^0H-\hat{F}\right\|_{\text{sp}}\right\}\left\|\frac{1}{NT}\sum_{j=1}^{N}\lambda_j^0\varepsilon_j'\hat{F}\right\|\left\|H^{-1}\right\|\left\|G\right\| \\
&= O_P(1)\,O_P(\delta_{NT}^{-1})O_P(N^{-1/2}\delta_{NT}^{-1}) = o_P((NT)^{-1/2})
\end{aligned}
$$

where we use the fact that

$$
\begin{aligned}
\left\|\frac{1}{NT}\sum_{j=1}^{N}\lambda_j^0\varepsilon_j'\hat{F}\right\| &\leq \left\|\frac{1}{NT}\sum_{j=1}^{N}\lambda_j^0\varepsilon_j'F^0H\right\| + \left\|\frac{1}{NT}\sum_{j=1}^{N}\lambda_j^0\varepsilon_j'\left(\hat{F}-F^0H\right)\right\| \\
&\leq \left\|\frac{1}{NT}\sum_{j=1}^{N}\lambda_j^0\varepsilon_j'F^0\right\|\left\|H\right\| + \left\|\frac{1}{NT^{1/2}}\sum_{j=1}^{N}\lambda_j^0\varepsilon_j'\right\|\frac{1}{\sqrt{T}}\left\|\hat{F}-F^0H\right\| \\
&= O_P((NT)^{-1/2})O_P(1) + O_P(N^{-1/2})O_P(\delta_{NT}^{-1}) = O_P(N^{-1/2}\delta_{NT}^{-1}).
\end{aligned}
$$

In addition, $J_{k,7}=-\frac{1}{N_kT}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}(\frac{1}{NT}\sum_{j=1}^N\varepsilon_j\lambda_j^{0\prime}F^{0\prime}\hat{F})G\lambda_i^0=-\frac{1}{NN_kT}\sum_{i\in\hat{G}_k}\sum_{j=1}^N\nu_{ji}X_i'M_{\hat{F}}\varepsilon_j$ and $J_{k,8}=-\frac{1}{NN_kT^2}\sum_{i\in\hat{G}_k}\sum_{j=1}^N X_i'M_{\hat{F}}\varepsilon_j\varepsilon_j'\hat{F}(T^{-1}F^{0\prime}\hat{F})^{-1}(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_i^0=\mathcal{B}_{1,kNT}$. In sum, we have $\frac{1}{N_kT}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}F^0\lambda_i^0\equiv\sum_{l=1}^{K_0}\bar{A}_{k,l}\left(\hat{\alpha}_l-\alpha_l^0\right)-\frac{1}{NN_kT}\sum_{i\in\hat{G}_k}\sum_{j=1}^N\nu_{ji}X_i'M_{\hat{F}}\varepsilon_j+\mathcal{B}_{1,kNT}+o_P((NT)^{-1/2})$. ■

Proof of Lemma A.8. (i) As in the proof of Lemma A.6(ii), we can show that $\frac{1}{N_kT}\sum_{i\in\hat{G}_k}X_i'M_{\hat{F}}X_i=\frac{1}{N_kT}\sum_{i\in G_k^0}X_i'M_{\hat{F}}X_i+o_P((NT)^{-1/2})$. Then by Lemma A.2(v) and Theorem 3.2,

$$\begin{aligned}\left\|\hat{Q}_{kNT}-Q_{kNT}\right\| &= \left\|\frac{1}{N_kT}\sum_{i\in G_k^0}X_i'\left(M_{\hat{F}}-M_{F^0}\right)X_i+o_P((NT)^{-1/2})\right\| \\ &\leq \frac{1}{N_kT}\sum_{i\in G_k^0}\|X_i\|^2\left\|P_{\hat{F}}-P_{F^0}\right\|+o_P((NT)^{-1/2})=O_P\left(\delta_{NT}^{-1}\right).\end{aligned}$$

(ii) Using $M_{\hat{F}}=I_T-P_{\hat{F}}=I_T-\hat{F}\hat{F}'/T$, we have $\hat{J}_{k,8}=-\frac{1}{NN_kT^2}\sum_{i\in\hat{G}_k}\sum_{j=1}^N X_i'\varepsilon_j\varepsilon_j'\hat{F}G\lambda_i^0+\frac{1}{NN_kT^3}\sum_{i\in\hat{G}_k}\sum_{j=1}^N X_i'\hat{F}\hat{F}'\varepsilon_j\varepsilon_j'\hat{F}G\lambda_i^0\equiv-\hat{J}_{k,81}+\hat{J}_{k,82}$, say. We further decompose $\hat{J}_{k,81}$ as follows $\hat{J}_{k,81}=\frac{1}{NN_kT^2}\sum_{i\in\hat{G}_k}\sum_{j=1}^N X_i'\varepsilon_j\varepsilon_j'F^0HG\lambda_i^0+\frac{1}{NN_kT^2}\sum_{i\in\hat{G}_k}\sum_{j=1}^N X_i'\varepsilon_j\varepsilon_j'(\hat{F}-F^0H)G\lambda_i^0\equiv\hat{J}_{k,81}(1)+\hat{J}_{k,82}(2)$, say. As in the proof of Lemma A.6(ii), we can show that $\frac{1}{NN_kT^2}\sum_{i\in\hat{G}_k}\sum_{j=1}^N X_i'\varepsilon_j\varepsilon_j'F^0HG\lambda_i^0=\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j=1}^N X_i'\varepsilon_j\varepsilon_j'F^0HG\lambda_i^0+o_P((NT)^{-1})$. It follows that by Assumption A.1(viii) and (ix)

$$\begin{aligned}\left|\omega_p'\hat{J}_{k,81}(1)\right| &= \left|\frac{1}{NN_kT^2}\mathrm{tr}\left(\sum_{i\in G_k^0}\sum_{j=1}^N\lambda_i^0\omega_p'X_i'\varepsilon_j\varepsilon_j'F^0HG\right)+o_P((NT)^{-1})\right| \\ &\leq \frac{1}{T}\left\|\frac{1}{NN_kT}\sum_{i\in G_k^0}\sum_{j=1}^N\lambda_i^0\omega_p'X_i'\varepsilon_j\varepsilon_j'F^0\right\|\|HG\|+o_P((NT)^{-1}) \\ &\leq \frac{1}{T}\frac{1}{NN_kT}\sum_{i\in G_k^0}\sum_{j=1}^N\left\|\lambda_i^0\right\|\|X_i'\varepsilon_j\|\left\|\varepsilon_j'F^0\right\|\|HG\|+o_P((NT)^{-1}) \\ &\leq \frac{1}{T}\left\{\frac{1}{NN_kT}\sum_{i\in G_k^0}\sum_{j=1}^N\left\|\lambda_i^0\right\|^2\|X_i'\varepsilon_j\|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{j=1}^N\left\|\varepsilon_j'F^0\right\|^2\right\}^{1/2}\|HG\|+o_P((NT)^{-1}) \\ &= O_P(T^{-1})O_P(1)O_P(1)O_P(1)+o_P((NT)^{-1})=O_P(T^{-1}),\end{aligned}$$

By arguments as used in the proof of Lemma A.6(ii) and Lemma A.7(v), $\hat{J}_{k,82}(2)=\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j=1}^N X_i'\varepsilon_j\varepsilon_j'(\hat{F}-F^0H)G\lambda_i^0+o_P((NT)^{-1})=O_P\left(T^{-1}\delta_{NT}^{-2}\right)$. It follows that $\hat{J}_{k,8}=O_P\left(T^{-1}\right)$.

(iii) Let $\tilde{V}_{kNT}=\frac{1}{N_kT}\sum_{i\in G_k^0}X_i'M_{\hat{F}}(\varepsilon_i-\frac{1}{N}\sum_{j=1}^N\nu_{ji}\varepsilon_j)$. Following the proof of Lemma A.6(ii), we can show that $\hat{V}_{kNT}-\tilde{V}_{kNT}=o_P((NT)^{-1/2})$. Next, $\tilde{V}_{kNT}-V_{kNT}=\frac{1}{N_kT}\sum_{i\in G_k^0}X_i'\left(M_{\hat{F}}-M_{F^0}\right)\varepsilon_i-\frac{1}{NN_kT}\sum_{i\in G_k^0}\sum_{j=1}^N X_i'\left(M_{\hat{F}}-M_{F^0}\right)\nu_{ji}\varepsilon_j\equiv v_{1kNT}-v_{2kNT}$, say. By (A.1), $v_{1kNT}=\sum_{l=1}^4\frac{-1}{N_kT}\sum_{i\in G_k^0}X_i'd_l\varepsilon_i\equiv\sum_{l=1}^4 v_{1kNT,l}$, say. By Lemma A.7(iv), $\|v_{1kNT,1}\|=\left\|\frac{1}{N_kT^2}\sum_{i\in G_k^0}X_i'(\hat{F}-F^0H)(\hat{F}-F^0H)'\varepsilon_i\right\|=O_P\left(\delta_{NT}^{-3}\right)$. By Lemma A.2(ii) and Assumption A.1(ix)

$$\begin{aligned}\left|\omega_p'v_{1kNT,2}\right| &= \left|\frac{1}{N_kT^2}\mathrm{tr}\left[(\hat{F}-F^0H)H'F^{0\prime}\sum_{i\in G_k^0}\varepsilon_i\omega_p'X_i'\right]\right|\leq\frac{1}{T}\left\|(\hat{F}-F^0H)H'F^{0\prime}\right\|\left\|\frac{1}{N_kT}\sum_{i\in G_k^0}\varepsilon_i\omega_p'X_i'\right\| \\ &= O_P(T^{-1/2}+\delta_{NT}^{-2})O_P(\delta_{NT}^{-1})=O_P(T^{-1/2}\delta_{NT}^{-1}).\end{aligned}$$

It follows that $v_{1kNT,2}=O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$. Analogously, we can show that $v_{1kNT,3}=O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$. For $v_{1kNT,4}$, we have by Lemma A.2(iv) and Theorem 3.2

$$
\begin{aligned}
\|v_{1kNT,4}\| &= \left\|\frac{1}{N_kT^2}\sum_{i\in G_k^0}X_i'F^0H\left[I_{R^0}-\left(T^{-1}H'F^{0\prime}F^0H\right)^{-1}\right]H'F^{0\prime}\varepsilon_i\right\| \\
&\leq \left\{\frac{1}{T}\left\|F^0\right\|\|H\|^2\right\}\left\|I_{R^0}-\left(T^{-1}H'F^{0\prime}F^0H\right)^{-1}\right\|\left\{\frac{1}{N_kT}\sum_{i\in G_k^0}\|X_i\|\left\|F^{0\prime}\varepsilon_i\right\|\right\} \\
&= O_P(T^{-1/2})O_P(T^{-1/2}+\delta_{NT}^{-2})O_P(1)=O_P\left(T^{-1}\right).
\end{aligned}
$$

Consequently, we have $\|v_{1kNT}\|=O_P(T^{-1/2}\delta_{NT}^{-1})$.

Recall that $X_i^*=\frac{1}{N}\sum_{j=1}^N\nu_{ij}X_j$. Using (A.1), we have $v_{2kNT}=\sum_{l=1}^4\frac{-1}{N_kT}\sum_{i\in G_k^0}X_i^{*\prime}d_l\varepsilon_i\equiv\sum_{l=1}^4 v_{2kNT,l}$, say. By Cauchy-Schwarz inequality, the fact that $\frac{1}{N_kT}\sum_{i\in G_k^0}\|X_i^*\|^2=O_P(1)$, Lemmas A.2(i) and A.7(iii), and Theorem 3.2,

$$
\begin{aligned}
\left|\omega_p'v_{2kNT,1}\right| &= \left|\frac{1}{N_kT^2}\sum_{i\in G_k^0}\omega_p'X_i^{*\prime}(\hat{F}-F^0H)(\hat{F}-F^0H)'\varepsilon_i\right| \\
&\leq \left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\omega_p'X_i^{*\prime}(\hat{F}-F^0H)(\hat{F}-F^0H)'X_i^*\omega_p\right\}^{1/2} \\
&\quad\times\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\varepsilon_i'(\hat{F}-F^0H)(\hat{F}-F^0H)'\varepsilon_i\right\}^{1/2} \\
&\leq \frac{1}{T^{1/2}}\left\|\hat{F}-F^0H\right\|\left\{\frac{1}{N_kT}\sum_{i\in G_k^0}\|X_i^*\|^2\right\}^{1/2}\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\varepsilon_i'(\hat{F}-F^0H)(\hat{F}-F^0H)'\varepsilon_i\right\}^{1/2} \\
&= O_P\left(\delta_{NT}^{-1}\right)O_P(1)\,O_P\left(\delta_{NT}^{-2}\right)=O_P\left(\delta_{NT}^{-3}\right).
\end{aligned}
$$

By Lemma A.2(i) and Theorem 3.2

$$
\begin{aligned}
\left|\omega_p'v_{2kNT,2}\right| &= \left|\frac{1}{N_kT^2}\sum_{i\in G_k^0}\operatorname{tr}\left[\omega_p'X_i^{*\prime}(\hat{F}-F^0H)H'F^{0\prime}\varepsilon_i\right]\right| \\
&= \left|\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\operatorname{tr}\left[\omega_p'\sum_{j=1}^N\nu_{ji}X_j'(\hat{F}-F^0H)H'F^{0\prime}\varepsilon_i\right]\right| \\
&= \left|\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\operatorname{tr}\left[\omega_p'\lambda_i^{0\prime}\left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\sum_{j=1}^N\lambda_j^0X_j'(\hat{F}-F^0H)H'F^{0\prime}\varepsilon_i\right]\right| \\
&= \left|\frac{1}{NN_kT^2}\operatorname{tr}\left[\left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\sum_{j=1}^N\lambda_j^0X_j'(\hat{F}-F^0H)H'\sum_{i\in G_k^0}F^{0\prime}\varepsilon_i\omega_p'\lambda_i^{0\prime}\right]\right| \\
&\leq \left\|\left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\right\|\frac{1}{NT^{1/2}}\sum_{j=1}^N\left\|\lambda_j^0X_j\right\|\left\{\frac{1}{T^{1/2}}\left\|\hat{F}-F^0H\right\|\right\}\|H\|\frac{1}{N_kT}\sum_{i\in G_k^0}\left\|F^{0\prime}\varepsilon_i\right\|\left\|\lambda_i^0\right\| \\
&= O_P(1)\,O_P\left(\delta_{NT}^{-1}\right)O_P(1)\,O_P(1)\,O_P(T^{-1/2})=O_P(T^{-1/2}\delta_{NT}^{-1}).
\end{aligned}
$$

It follows that $v_{2kNT,2} = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$. Analogously, we can show that $v_{2kNT,3} = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$. For $V_{1kNT,4}$, we have by Lemma A.2(v) and Theorem 3.2

$$\begin{aligned}
\|v_{2kNT,4}\| &= \left\| \frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i^{*\prime} F^0 H \left[I_{R^0} - \left(T^{-1} H' F^{0\prime} F^0 H\right)^{-1} \right] H' F^{0\prime} \varepsilon_i \right\| \\
&\leq \left\{ \frac{1}{T} \left\|F^0\right\| \|H\|^2 \right\} \left\| I_{R^0} - \left(T^{-1} H' F^{0\prime} F^0 H\right)^{-1} \right\| \left\{ \frac{1}{N_k T} \sum_{i \in G_k^0} \|X_i^*\|^2 \right\}^{1/2} \left\{ \frac{1}{N_k T} \sum_{i \in G_k^0} \left\|F^{0\prime} \varepsilon_i\right\|^2 \right\}^{1/2} \\
&= O_P(T^{-1/2}) O_P(T^{-1/2}) O_P(1)\, O_P(1) = O_P(T^{-1}).
\end{aligned}$$

Consequently, we have $\|v_{2kNT}\| = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$. Then (iv) follows.

(iv) Note that $V_{kNT} = \frac{1}{N_k T} \sum_{i \in G_k^0} X_i' M_{F^0} \varepsilon_i - \frac{1}{N N_k T} \sum_{i \in G_k^0} X_i' M_{F^0} \sum_{j=1}^N \nu_{ji} \varepsilon_j$. The conclusion follows from Assumption A.4(ii). ■

Proof of Lemma A.9. (i) By the triangle inequality,

$$\begin{aligned}
&\frac{1}{N N_k T^2} \left\| \sum_{j=1}^N \sum_{i \in G_k^0} \hat{F}' \varepsilon_j \varepsilon_j' \varepsilon_i \omega_p' X_i' \right\| \\
\leq\ & \frac{1}{N N_k T^2} \left\| H' \sum_{j=1}^N \sum_{i \in G_k^0} F^{0\prime} \varepsilon_j \varepsilon_j' \varepsilon_i \omega_p' X_i' \right\| + \frac{1}{N N_k T^2} \left\| \sum_{j=1}^N \sum_{i \in G_k^0} (\hat{F} - F^0 H)' \varepsilon_j \varepsilon_j' \varepsilon_i \omega_p' X_i' \right\|.
\end{aligned}$$

The first term is bounded from above by $\left\| \frac{1}{NT} \sum_{j=1}^N F^{0\prime} \varepsilon_j \varepsilon_j' \right\| \left\| \frac{1}{N_k T} \sum_{i \in G_k^0} \varepsilon_i \omega_p' X_i' \right\| = O_P\left(\delta_{NT}^{-1}\right) O_P\left(\delta_{NT}^{-1}\right) = O_P\left(\delta_{NT}^{-2}\right)$, and the second term is bounded from above by

$$\left\| \hat{F} - F^0 H \right\| \left\| \frac{1}{NT} \sum_{j=1}^N \varepsilon_j \varepsilon_j' \right\| \left\| \frac{1}{N_k T} \sum_{i \in G_k^0} \varepsilon_i \omega_p' X_i' \right\| = O_P\left(T^{1/2}\delta_{NT}^{-1}\right) O_P\left(\delta_{NT}^{-1}\right) O_P\left(\delta_{NT}^{-1}\right) = O_P\left(T^{1/2}\delta_{NT}^{-3}\right).$$

So (i) follow.

(ii) By Lemma A.8,

$$\begin{aligned}
&\frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i' F^0 \left(T^{-1} F^{0\prime} F^0\right)^{-1} (\hat{F} H^{-1} - F^0)' \varepsilon_i \\
=\ & \sum_{l=1}^8 \frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i' F^0 \left(T^{-1} F^{0\prime} F^0\right)^{-1} \left(V_{NT}^{-1} H^{-1}\right)' a_l' \varepsilon_i \\
=\ & \sum_{l=1}^8 \frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i' F^0 \left(T^{-1} F^{0\prime} F^0\right)^{-1} G' a_l' \varepsilon_i \equiv \sum_{l=1}^8 L_l, \text{ say.}
\end{aligned}$$

Noting that $\|a_1\| = O_P\left(T^{1/2}\eta_{NT}^2\right) = O_P\left(T^{-1/2}\delta_{NT}^{-2}\right)$, we can show that $L_1 = O_P\left(T^{-1}\delta_{NT}^{-2}\right)$. Next, following the arguments used in the proof of Lemmas A.5(i) and A.6(ii), we can show that $a_2 = -\frac{1}{NT} \sum_{i=1}^N X_i \hat{b}_i$

$\times\lambda_i^{0\prime}F^{0\prime}\hat{F}=-\frac{1}{NT}\sum_{l=1}^{K_0}\sum_{i\in G_l^0}X_i\left(\hat{\alpha}_l-\alpha_l^0\right)\lambda_i^{0\prime}F^{0\prime}\hat{F}+o_P((NT)^{-1/2})$. It follows that

$$\begin{aligned}
\left\|\omega_p' L_2\right\| &= \frac{1}{NN_kT^3}\left\|\sum_{l=1}^{K_0}\sum_{j\in G_l^0}\sum_{i\in G_k^0}\omega_p' X_i' F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}G'(F^{0\prime}\hat{F})'\lambda_j^0\left(\hat{\alpha}_l-\alpha_l^0\right)' X_j'\varepsilon_i\right\|+o_P((NT)^{-1/2})\\
&= \frac{1}{NN_kT^2}\left|\sum_{l=1}^{K_0}\sum_{j\in G_l^0}\operatorname{tr}\left[F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_j^0\left(\hat{\alpha}_l-\alpha_l^0\right)' X_j'\left(\sum_{i\in G_k^0}\varepsilon_i\omega_p' X_i'\right)\right]\right|\\
&\quad +o_P((NT)^{-1/2})\\
&\leq c_{1NT}\sum_{l=1}^{K_0}\left\|\hat{\alpha}_l-\alpha_l^0\right\|\frac{1}{NT^{1/2}}\sum_{j\in G_l^0}\|X_j\|\left\|\lambda_j^0\right\|\left\|\frac{1}{N_kT}\sum_{i\in G_k^0}\varepsilon_i\omega_p' X_i'\right\|+o_P((NT)^{-1/2})\\
&= O_P\left(\delta_{NT}^{-2}\right)O_P\left(\delta_{NT}^{-1}\right)+o_P((NT)^{-1/2})=o_P((NT)^{-1/2}),
\end{aligned}$$

where $c_{1NT}=T^{-1/2}\left\|F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}\right\|\left\|(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\right\|$. Similarly, using $a_3=\frac{1}{NT}\sum_{i=1}^N X_i\hat{b}_i\varepsilon_i'\hat{F}=-\frac{1}{NT}\sum_{l=1}^{K_0}\sum_{i\in G_l^0}X_i\left(\hat{\alpha}_l-\alpha_l^0\right)\varepsilon_i'\hat{F}+o_P((NT)^{-1/2})$, we have

$$\begin{aligned}
\left\|\omega_p' L_3\right\| &= \frac{1}{NN_kT^3}\left\|\sum_{l=1}^{K_0}\sum_{j\in G_l^0}\sum_{i\in G_k^0}\omega_p' X_i' F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}G'\hat{F}'\varepsilon_j\left(\hat{\alpha}_l-\alpha_l^0\right)' X_j'\varepsilon_i\right\|\\
&= \frac{1}{NN_kT^3}\left\|\sum_{l=1}^{K_0}\sum_{j\in G_l^0}\operatorname{tr}\left(F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}G'\hat{F}'\varepsilon_j\left(\hat{\alpha}_l-\alpha_l^0\right)' X_j'\sum_{i\in G_k^0}\varepsilon_i\omega_p' X_i'\right)\right\|\\
&\leq c_{2NT}\sum_{l=1}^{K_0}\left\|\hat{\alpha}_l-\alpha_l^0\right\|\frac{1}{NT^{3/2}}\sum_{j\in G_l^0}\|X_j\|\left\|\hat{F}'\varepsilon\right\|\left\|\frac{1}{N_kT}\sum_{i\in G_k^0}\varepsilon_i\omega_p' X_i'\right\|+o_P((NT)^{-1/2})\\
&= O_P\left(\delta_{NT}^{-2}\right)O_P\left(\delta_{NT}^{-1}\right)O_P\left(\delta_{NT}^{-1}\right)+o_P((NT)^{-1/2})=o_P((NT)^{-1/2})
\end{aligned}$$

where $c_{2NT}=T^{-1/2}\left\|F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}G'\right\|$ and we use the fact that

$$\begin{aligned}
\frac{1}{NT^{3/2}}\sum_{j\in G_l^0}\|X_j\|\left\|\varepsilon_j'\hat{F}\right\| &\leq \frac{1}{NT^{3/2}}\sum_{j\in G_l^0}\|X_j\|\left\|\varepsilon_j' F^0 H\right\|+\frac{1}{NT^{3/2}}\sum_{j\in G_l^0}\|X_j\|\left\|\varepsilon_j'(\hat{F}-F^0H)\right\|\\
&= O_P(T^{-1/2})+\frac{1}{NT}\sum_{j\in G_l^0}\|X_j\|\|\varepsilon_j\|\frac{1}{T^{1/2}}\left\|\hat{F}-F^0H\right\|=O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}$$

Analogously, we can show that $\|L_s\|=o_P((NT)^{-1/2})$ for $s=4$, 5.

$$\begin{aligned}
\left|\omega_p' L_6\right| &= \frac{1}{NN_kT^3}\left|\sum_{i\in G_k^0}\sum_{j=1}^N\omega_p' X_i' F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}G'\hat{F}\varepsilon_j\lambda_j^{0\prime}F^{0\prime}\varepsilon_i\right|\\
&= \frac{1}{NN_kT^3}\left|\operatorname{tr}\left(F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}G'\sum_{j=1}^N\hat{F}\varepsilon_j\lambda_j^{0\prime}\sum_{i\in G_k^0}F^{0\prime}\varepsilon_i\omega_p' X_i'\right)\right|\\
&\leq c_{2NT}T^{-1/2}\left\|\frac{1}{NT}\sum_{j=1}^N\hat{F}\varepsilon_j\lambda_j^{0\prime}\right\|\left\|\frac{1}{N_kT}\sum_{i\in G_k^0}F^{0\prime}\varepsilon_i\omega_p' X_i'\right\|\\
&= T^{-1/2}O_P\left(\delta_{NT}^{-1}\right)O_P\left(\delta_{NT}^{-1}\right)=o_P((NT)^{-1/2}).
\end{aligned}$$

In addition, $L_7 = \frac{1}{N_k T^2}\sum_{i\in G_k^0}\sum_{j=1}^N X_i' F^0 \left(T^{-1}F^{0\prime}F^0\right)^{-1} G' a_7' \varepsilon_i = \frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j=1}^N X_i' F^0 \left(T^{-1}F^{0\prime}F^0\right)^{-1}$ $\times (N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_j^0 \varepsilon_j'\varepsilon_i$, and

$$
\begin{aligned}
\left\|\omega_p' L_8\right\| &= \frac{1}{N_kT^2}\left\|\sum_{i\in G_k^0}\omega_p' X_i' F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1} G' a_8' \varepsilon_i\right\| \\
&= \frac{1}{NN_kT^3}\left\|\sum_{j=1}^N\sum_{i\in G_k^0}\omega_p' X_i' F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1} G'\hat{F}'\varepsilon_j'\varepsilon_j\varepsilon_i\right\| \\
&= \frac{1}{NN_kT^3}\left\|\mathrm{tr}\left(\left(T^{-1}F^{0\prime}F^0\right)^{-1} G'\sum_{j=1}^N\sum_{i\in G_k^0}\hat{F}'\varepsilon_j\varepsilon_j'\varepsilon_i\omega_p' X_i' F^0\right)\right\| \\
&\leq c_{NT}\frac{1}{T^{1/2}}\left\|\frac{1}{NN_kT^2}\sum_{j=1}^N\sum_{i\in G_k^0}\hat{F}'\varepsilon_j\varepsilon_j'\varepsilon_i\omega_p' X_i'\right\| = T^{-1/2}O_P(T^{1/2}\delta_{NT}^{-3}) = o_P((NT)^{-1/2}),
\end{aligned}
$$

In sum, we have shown that $\frac{1}{N_kT^2}\sum_{i\in G_k^0} X_i'F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}(\hat{F}H^{-1}-F^0)'\varepsilon_i = L_7 + o_P((NT)^{-1/2})$ $= \frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j=1}^N X_i'F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}(N^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\lambda_j^0\varepsilon_j'\varepsilon_i + o_P((NT)^{-1/2})$.

(iii) The proof is analogous to that of (ii) and thus omitted. ■

Proof of Lemma A.10. (i) Noting that $Q_{1NT}\left(\boldsymbol{\beta},F_{(R)}\right) = \frac{1}{NT}\sum_{i=1}^N (Y_i - X_i\beta_i)' M_{F_{(R)}} (Y_i - X_i\beta_i)$ and $Y_i - X_i\dot{\beta}_{i,R} = X_i(\beta_i^0 - \dot{\beta}_{i,R}) + F^0\lambda_i^0 + \varepsilon_i$, we have

$$
\begin{aligned}
0 &\geq Q_{1NT}\left(\dot{\boldsymbol{\beta}}_{(R)},\dot{F}_{(R)}\right) - Q_{1NT}\left(\boldsymbol{\beta}^0,\dot{F}_{(R)}\right) \\
&= \frac{1}{NT}\sum_{i=1}^N\left[\left(Y_i - X_i\dot{\beta}_{i,R}\right)' M_{\dot{F}_{(R)}}\left(Y_i - X_i\dot{\beta}_{i,R}\right) - \left(Y_i - X_i\beta_i^0\right)' M_{\dot{F}_{(R)}}\left(Y_i - X_i\beta_i^0\right)\right] \\
&= \frac{1}{NT}\sum_{i=1}^N\left[\left(\dot{\beta}_{i,R}-\beta_i^0\right)' X_i' M_{\dot{F}_{(R)}} X_i\left(\dot{\beta}_{i,R}-\beta_i^0\right) - 2\left(\dot{\beta}_{i,R}-\beta_i^0\right)' X_i' M_{\dot{F}_{(R)}} F^0\lambda_i^0\right] \\
&\quad -\frac{2}{NT}\sum_{i=1}^N\left(\dot{\beta}_{i,R}-\beta_i^0\right)' X_i' M_{\dot{F}_{(R)}}\varepsilon_i.
\end{aligned}
$$

By Lemma A.1 (with R_0 and F replaced by R and $F_{(R)}$), we can show that $\frac{1}{NT}\sum_{i=1}^N(\dot{\beta}_{i,R}-\beta_i^0)'X_i'M_{\dot{F}_{(R)}}\varepsilon_i$ $= o_P(1)$ provided that $\frac{1}{N}\sum_{i=1}^N\left\|\dot{\beta}_{i,R}-\beta_i^0\right\|^2 = O_P(1)$, which can be shown by using similar arguments as used in the proof of Theorem 3.1. Let $\dot{d}_{\beta,R} = \mathrm{vec}(\dot{\boldsymbol{\beta}}_{(R)} - \boldsymbol{\beta}^0)$ and $\dot{d}_{F,R} = T^{-1/2}\mathrm{vec}(M_{\dot{F}_{(R)}}F^0)$. Define

$$
\dot{A}_R = \frac{1}{T}\mathrm{diag}\left(X_1' M_{\dot{F}_{(R)}} X_1, ..., X_N' M_{\dot{F}_{(R)}} X_N\right) \text{ and } \dot{C}_R = \frac{1}{T^{1/2}}\mathrm{diag}\left(\lambda_1^0 \otimes M_{\dot{F}_{(R)}} X_1, ..., \lambda_N^0 \otimes M_{\dot{F}_{(R)}} X_N\right).
$$

Then $\frac{1}{NT}\sum_{i=1}^N[(\dot{\beta}_{i,R}-\beta_i^0)'X_i'M_{\dot{F}_{(R)}}X_i(\dot{\beta}_{i,R}-\beta_i^0) - 2(\dot{\beta}_{i,R}-\beta_i^0)'X_i'M_{\dot{F}_{(R)}}F^0\lambda_i^0] = \frac{1}{N}\dot{d}_{\beta,R}'\dot{A}_R\dot{d}_{\beta,R} - \frac{2}{N}\dot{d}_{F,R}'$ $\times\dot{C}_R\dot{d}_{\beta,R}$. It follows that

$$
\begin{aligned}
0 &\geq \frac{1}{N}\dot{d}_{\beta,R}'\dot{A}_R\dot{d}_{\beta,R} - \frac{2}{N}\dot{d}_{F,R}'\dot{C}_R\dot{d}_{\beta,R} + o_P(1) \\
&\geq \frac{1}{N}\dot{d}_{\beta,R}'\dot{A}_R\dot{d}_{\beta,R} - 2\left\{\dot{d}_{F,R}'\dot{d}_{F,R}\right\}^{1/2}\left\{\frac{1}{N^2}\dot{d}_{\beta,R}'\dot{C}_R'\dot{C}_R\dot{d}_{\beta,R}\right\}^{1/2} + o_P(1).
\end{aligned}
$$

This, in conjunction with the fact that $\left\|\dot{d}_{F,R}\right\| = O_P(1)$, $N^{-1}\left\|\dot{d}_{\beta,R}\right\|^2 = O_P(1)$, and $\mu_{\max}\left(N^{-1}\dot{C}_R'\dot{C}_R\right) = o_P(1)$ (by following similar arguments as used in the proof of Theorem 3.1), implies that $0 \geq \frac{1}{N}\dot{d}_{\beta,R}'\dot{A}_R\dot{d}_{\beta,R} + o_P(1)$. It follows that

$$\frac{1}{N}\dot{d}_{\beta,R}'\dot{A}_R\dot{d}_{\beta,R} = \frac{1}{NT}\sum_{i=1}^{N}\left(\dot{\beta}_{i,R}-\beta_i^0\right)' X_i' M_{\dot{F}_{(R)}} X_i \left(\dot{\beta}_{i,R}-\beta_i^0\right) = o_P(1).$$

Note that $V(R,\dot{\boldsymbol{\beta}}_{(R)}) = \min_{\boldsymbol{\beta},F_{(R)}} Q_{1NT}\left(\boldsymbol{\beta},F_{(R)}\right)$ subject to $T^{-1}F_{(R)}'F_{(R)} = I_R$. By the results in Theorem 3.1 and Lemma A.1 and Assumption 1, we can readily show that

$$\begin{aligned} V\left(R_0,\dot{\boldsymbol{\beta}}_{(R_0)}\right) &= \frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_i\dot{\beta}_{i,R_0}\right)' M_{\dot{F}_{(R_0)}}\left(Y_i - X_i\dot{\beta}_{i,R_0}\right) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i' M_{\dot{F}_{(R_0)}}\varepsilon_i + o_P(1) = \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i'\varepsilon_i + o_P(1) \end{aligned}$$

and

$$\begin{aligned} & V\left(R,\dot{\boldsymbol{\beta}}_{(R)}\right) - V\left(R_0,\dot{\boldsymbol{\beta}}_{(R_0)}\right) \\ =\ & \frac{1}{NT}\sum_{i=1}^{N}\left[\left(Y_i - X_i\dot{\beta}_{i,R}\right)' M_{\dot{F}_{(R)}}\left(Y_i - X_i\dot{\beta}_{i,R}\right) - \left(Y_i - X_i\dot{\beta}_{i,R_0}\right)' M_{\dot{F}_{(R_0)}}\left(Y_i - X_i\dot{\beta}_{i,R_0}\right)\right] \\ =\ & \frac{1}{NT}\sum_{i=1}^{N}\left(X_i(\beta_i^0 - \dot{\beta}_{i,R}) + F^0\lambda_i^0\right)' M_{\dot{F}_{(R)}}\left(X_i(\beta_i^0 - \dot{\beta}_{i,R}) + F^0\lambda_i^0\right) + o_P(1) \\ =\ & \frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}M_{\dot{F}_{(R)}}F^0\lambda_i^0 + \frac{1}{NT}\sum_{i=1}^{N}(\beta_i^0 - \dot{\beta}_{i,R})'X_i'M_{\dot{F}_{(R)}}X_i(\beta_i^0 - \dot{\beta}_{i,R}) \\ & + \frac{2}{NT}\sum_{i=1}^{N}(\beta_i^0 - \dot{\beta}_{i,R})'X_i'M_{\dot{F}_{(R)}}F^0\lambda_i^0 + o_P(1) \\ \geq\ & \frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}M_{\dot{F}_{(R)}}F^0\lambda_i^0 + \frac{1}{NT}\sum_{i=1}^{N}(\beta_i^0 - \dot{\beta}_{i,R})'X_i'M_{\dot{F}_{(R)}}X_i(\beta_i^0 - \dot{\beta}_{i,R}) \\ & -2\left\{\frac{1}{NT}\sum_{i=1}^{N}(\beta_i^0 - \dot{\beta}_{i,R})'X_i'M_{\dot{F}_{(R)}}X_i(\beta_i^0 - \dot{\beta}_{i,R})\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0\prime}F^0M_{\dot{F}_{(R)}}F^0\lambda_i^0\right\}^{1/2} + o_P(1) \\ =\ & \frac{1}{NT}\sum_{i=1}^{N}\left(F^0\lambda_i^0\right)' M_{\dot{F}_{(R)}}F^0\lambda_i^0 + o_P(1). \end{aligned}$$

Next,

$$\begin{aligned} & \frac{1}{NT}\sum_{i=1}^{N}\left(F^0\lambda_i^0\right)' M_{\dot{F}_{(R)}}F^0\lambda_i^0 \\ =\ & \frac{1}{T}\sum_{t=R+1}^{T}\mu_r\left(\frac{1}{N}\sum_{i=1}^{N}\left(F^0\lambda_i^0\right)\left(F^0\lambda_i^0\right)'\right) = \frac{1}{T}\sum_{t=R+1}^{T}\mu_r\left(F^0\frac{\Lambda^{0\prime}\Lambda^0}{N}F^{0\prime}\right) \\ =\ & \frac{1}{T}\sum_{t=R+1}^{R_0}\mu_r\left(F^0\frac{\Lambda^{0\prime}\Lambda^0}{N}F^{0\prime}\right) \geq \mu_{\min}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\sum_{t=R+1}^{R_0}\mu_r\left(\frac{F^0F^{0\prime}}{T}\right) \\ \geq\ & (R_0 - R)\,\mu_{\min}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\mu_{\min}\left(\frac{F^{0\prime}F^0}{T}\right) = (R_0 - R)\,\mu_{\min}(\Sigma_\Lambda)\,\mu_{\min}(\Sigma_F) + o_P(1). \end{aligned}$$

It follows that w.p.a.1, $V(R, \dot{\boldsymbol{\beta}}_{(R)}) - V(R_0, \dot{\boldsymbol{\beta}}_{(R_0)}) \geq c_R$ with $c_R = (R_0 - R)\, \mu_{\min}(\Sigma_\Lambda)\, \mu_{\min}(\Sigma_F)/2 > 0$.

(ii) The proof follows from the exact arguments as used in the proof of Lemma B.4(ii) in Li et al. (2016) by reversing the role of factors and factor loadings. Another difference is that we now allow the regression coefficients to change over individuals instead of time. For brevity, we omit the details. ■

Proof of Lemma A.11. When $K \geq K_0$, following the proof of Theorem 3.2, we can show that

$$\left\|\hat{\beta}_i - \beta_i^0\right\| = O_P\left(T^{-1/2} + \kappa\right) \text{ for each } i \text{ and } \frac{1}{N}\sum_{i=1}^{N} \Pi_{k=1}^{K} \left\|\beta_i^0 - \hat{\alpha}_k\right\| = O_P\left(T^{-1/2}\right).$$

Noting that β_i^0, $i = 1, ..., N$, only take K_0 distinct values, the latter implies that the collection $\{\hat{\alpha}_k, k = 1, ..., K\}$ contains at least K_0 distinct vectors, say, $\hat{\alpha}_{(1)}, ..., \hat{\alpha}_{(K_0)}$, such that $\hat{\alpha}_{(k)} - \alpha_k^0 = O_P(T^{-1/2})$ for $k = 1, ..., K_0$. For notational simplicity, we rename the other vectors in the above collection as $\hat{\alpha}_{(K_0+1)}, ..., \hat{\alpha}_{(K)}$. As before, we classify $i \in \hat{G}_k(K, \kappa)$ if $\left\|\hat{\beta}_i - \hat{\alpha}_{(k)}\right\| = 0$ for $k = 1, ..., K$, and $i \in \hat{G}_0(K, \kappa)$ otherwise. Using arguments like those used in the proof of Theorem 3.4, we can show that

$$\sum_{i \in G_k^0} P\left(\hat{E}_{kNT,i}\right) = o(1) \text{ for } k = 1, ..., K_0 \text{ and } \sum_{i \in \hat{G}_k(K,\kappa)} P\left(\hat{F}_{kNT,i}\right) = o(1) \text{ for } k = 1, ..., K_0.$$

The first part implies that $\sum_{i=1}^{N} P\left(i \in \hat{G}_0(K, \kappa) \cup \hat{G}_{K_0+1}(K, \kappa) \cup \cdots \cup \hat{G}_K(K, \kappa)\right) = o(1)$.

Using the fact that $\mathbf{1}\{i \in \hat{G}_k\} = \mathbf{1}\{i \in G_k^0\} + \mathbf{1}\{i \in \hat{G}_k \backslash G_k^0\} - \mathbf{1}\{i \in G_k^0 \backslash \hat{G}_k\}$, we have

$$\begin{aligned}
\hat{\sigma}^2_{\hat{G}(K,\kappa)} &= \frac{1}{NT}\sum_{k=1}^{K}\sum_{i \in \hat{G}_k(K,\kappa)}\sum_{t=1}^{T}\left[\bar{\varepsilon}_{it}^{(K,\kappa)}\right]^2 \\
&= \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i \in G_k^0}\sum_{t=1}^{T}\left[\bar{\varepsilon}_{it}^{(K,\kappa)}\right]^2 + \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i \in \hat{G}_k(K,\kappa)\backslash G_k^0}\sum_{t=1}^{T}\left[\bar{\varepsilon}_{it}^{(K,\kappa)}\right]^2 \\
&\quad - \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i \in G_k^0 \backslash \hat{G}_k(K,\kappa)}\sum_{t=1}^{T}\left[\bar{\varepsilon}_{it}^{(K,\kappa)}\right]^2 + \frac{1}{NT}\sum_{k=K_0+1}^{K}\sum_{i \in \hat{G}_k(K,\kappa)}\sum_{t=1}^{T}\left[\bar{\varepsilon}_{it}^{(K,\kappa)}\right]^2 \\
&\equiv D_{1NT} + D_{2NT} - D_{3NT} + D_{4NT},
\end{aligned}$$

where $\bar{\varepsilon}_{it}^{(K,\kappa)} = Y_{it} - \tilde{\alpha}'_{\hat{G}_k(K,\kappa)} X_{it} - \tilde{\lambda}_i^{(K,\kappa)\prime} \tilde{F}_t^{(K,\kappa)}$. Following the proof of Theorem 3.7, for $k = 1, ..., K_0$, we have $\tilde{\alpha}_{\hat{G}_k(K,\kappa)} - \alpha_k^0 = O_P\left(\delta_{NT}^{-2}\right)$. In addition, we can show that $\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t^{(K,\kappa)} - H' F_t^0\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$ and $\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i^{(K,\kappa)} - H^{-1}\lambda_i^0\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$. With these results, we can readily show that $D_{1NT} = \bar{\sigma}^2_{G^0} + O_P\left(\delta_{NT}^{-2}\right)$. For D_{2NT}, D_{3NT}, and D_{4NT}, we have that for any $\epsilon > 0$,

$$\begin{aligned}
P\left(D_{2NT} \geq \delta_{NT}^{-2}\epsilon\right) &\leq \sum_{i=1}^{K_0} P\left(\hat{F}_{kNT}\right) \to 0, \\
P\left(D_{3NT} \geq \delta_{NT}^{-2}\epsilon\right) &\leq \sum_{i=1}^{K_0} P\left(\hat{E}_{kNT}\right) \to 0, \text{ and} \\
P\left(D_{4NT} \geq \delta_{NT}^{-2}\epsilon\right) &\leq \sum_{i=1}^{N} \Pr\left(i \in \cup_{K_0+1 \leq k \leq K} \hat{G}_k(K, \kappa)\right) \to 0.
\end{aligned}$$

It follows that $\hat{\sigma}^2_{\hat{G}(K,\kappa)} = \bar{\sigma}^2_{G^0} + O_P\left(\delta_{NT}^{-2}\right)$ for all $K_0 \leq K \leq K_{\max}$. ■

C Verification of Assumption A.4(ii) under Assumptions B.1-B.2

In this appendix, we verify Assumption A.4(ii) under Assumptions B.1-B.2.

(i) We verify that $V_{1kNT} \equiv \frac{1}{NT}\sum_{i\in G_k^0} X_i' M_{F^0}\varepsilon_i = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$. Write $V_{kNT} = \frac{1}{N_k T}\sum_{i\in G_k^0} X_i' M_{F^0}\varepsilon_i - \frac{1}{NN_kT}\sum_{i\in G_k^0} X_i' M_{F^0}\sum_{j=1}^{N}\nu_{ji}\varepsilon_j \equiv V_{1kNT} - V_{2kNT}$, say. We further decompose V_{1kNT} as follows $V_{1kNT} = \frac{1}{N_kT}\sum_{i\in G_k^0} X_i'\varepsilon_i - \frac{1}{N_kT^2}\sum_{i\in G_k^0} X_i' F^0\left(T^{-1}F^{0\prime}F^0\right)^{-1}F^{0\prime}\varepsilon_i \equiv V_{1kNT,1} - V_{1kNT,2}$. Apparently, $V_{1kNT,1} = O_P\left((N_kT)^{-1/2}\right)$ under Assumption A.1(vii). Let $\mu_{ts} = F_t^{0\prime}\left(T^{-1}F^{0\prime}F^0\right)^{-1}F_s^0$. Note that μ_{ts} is measurable with respect to $\mathcal{D}$ and $V_{1kNT,2} = \frac{1}{N_kT^2}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T} X_{it}\varepsilon_{is}\mu_{ts}$. For notational simplicity, we assume $p=1$ in this proof (otherwise we can study each element in the $p\times 1$ vector $V_{1kNT,2}$). Observe that

$$
\begin{aligned}
\left|E_{\mathcal{D}}\left(V_{1kNT,2}\right)\right| &= \left|\frac{1}{N_kT^2}\sum_{i\in G_k^0}\sum_{t=2}^{T}\sum_{s=1}^{t} E_{\mathcal{D}}\left(X_{it}\varepsilon_{is}\right)\mu_{ts}\right| \\
&\leq \frac{c_{NT}}{N_kT^2}\sum_{i\in G_k^0}\sum_{t=2}^{T}\sum_{s=1}^{t}\left\|X_{it}\right\|_{2+\sigma,\mathcal{D}}\left\|\varepsilon_{is}\right\|_{2+\sigma,\mathcal{D}}\left\|F_t^0\right\|\left\|F_s^0\right\|\alpha_{NT}^{\mathcal{D}}\left(t-s\right)^{(2+\sigma)/\sigma} \\
&\leq \frac{c_{NT}}{2N_kT^2}\sum_{i\in G_k^0}\sum_{t=2}^{T}\sum_{s=1}^{t}\left\|X_{it}\right\|_{2+\sigma,\mathcal{D}}^2\left\|F_t^0\right\|^2\alpha_{NT}^{\mathcal{D}}\left(t-s\right)^{(2+\sigma)/\sigma} \\
&\quad + \frac{c_{NT}}{2N_kT^2}\sum_{i\in G_k^0}\sum_{t=2}^{T}\sum_{s=1}^{t}\left\|\varepsilon_{is}\right\|_{2+\sigma,\mathcal{D}}^2\left\|F_s^0\right\|^2\alpha_{NT}^{\mathcal{D}}\left(t-s\right)^{(2+\sigma)/\sigma} \\
&\equiv v_{1,1} + v_{1,2}, \text{ say,}
\end{aligned}
$$

where $c_{NT} = \left\|\left(T^{-1}F^{0\prime}F^0\right)^{-1}\right\|$, the first inequality follows from the Davydov inequality for conditional strong mixing process, and the second inequality follows from the Cauchy-Schwarz inequality. By Assumption B.2(i), $\sum_{\tau=1}^{\infty}\alpha_{NT}^{\mathcal{D}}\left(\tau\right)^{(2+\sigma)/(1+\sigma)} = O_P\left(1\right)$.

$$
v_{1,1} \leq \frac{c_{NT}}{2T}\left\{\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\left\|X_{it}\right\|_{2+\sigma,\mathcal{D}}^2\left\|F_t^0\right\|^2\right\}\sum_{\tau=1}^{\infty}\alpha_{NT}^{\mathcal{D}}\left(\tau\right)^{(2+\sigma)/(1+\sigma)} = T^{-1}O_P\left(1\right)O_P\left(1\right) = O_P\left(T^{-1}\right).
$$

By the same token, $v_{1,2} = O_P\left(T^{-1}\right)$. It follows that $E_{\mathcal{D}}\left(V_{1kNT,2}\right) = O_P\left(T^{-1}\right)$. Next,

$$
\begin{aligned}
E_{\mathcal{D}}\left(V_{1kNT,2}^2\right) &= \frac{1}{N_k^2T^4}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T} E_{\mathcal{D}}\left(X_{it}\varepsilon_{is}X_{ir}\varepsilon_{iq}\right)\mu_{ts}\mu_{rq} \\
&\quad + \frac{1}{N_k^2T^4}\sum_{i\in G_k^0}\sum_{j\in G_k^0, j\neq i}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T} E_{\mathcal{D}}\left(X_{it}\varepsilon_{is}\right)E_{\mathcal{D}}\left(X_{jr}\varepsilon_{jq}\right)\mu_{ts}\mu_{rq} \\
&= \frac{1}{N_k^2T^4}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T} E_{\mathcal{D}}\left(X_{it}\varepsilon_{is}X_{ir}\varepsilon_{iq}\right)\mu_{ts}\mu_{rq} + \left[E_{\mathcal{D}}\left(V_{1kNT,2}\right)\right]^2 \\
&\quad - \frac{1}{N_k^2T^4}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T} E_{\mathcal{D}}\left(X_{it}\varepsilon_{is}\right)E_{\mathcal{D}}\left(X_{ir}\varepsilon_{iq}\right)\mu_{ts}\mu_{rq} \\
&\equiv v_{2,1} + v_{2,2} - v_{2,3}, \text{ say.}
\end{aligned}
$$

To analyze $v_{2,1}$, we distinguish two cases: (a) $\#\{t,s,r,q\}\leq 3$, and (b) $\#\{t,s,r,q\}=4$. In case (a), we can apply Assumption B.2(i) and Davydov and Hölder inequalities and readily show that $v_{2,1}=O_P\left(N^{-1}T^{-1}\right)$ as there are only $O\left(NT^3\right)$ terms in the summation. In case (b), noting that $E_{\mathcal{D}}\left(X_{it}\varepsilon_{is}X_{ir}\varepsilon_{iq}\right)=0$ if either s or q is largest among $\{t,s,r,q\}$, without loss of generality assume t is largest among the four time indices and $s>q$. Then it suffices to consider three subcases: (a1) $t>s>r>q$, (a2) $t>r>s>q$, and (a3) $t>s>q>r$. We define $v_{2,1}\left(s\right)$ analogously to $v_{2,1}$ but with the time indices restricted to subcase s for $s=a1$, $a2$, and $a3$.

$$
\begin{aligned}
\left|v_{2,1}\left(a1\right)\right| &\leq \frac{1}{N_k^2T^4}\sum_{i\in G_k^0}\sum_{T\geq t>s>r>q\geq 1}\left|E_{\mathcal{D}}\left(X_{it}\varepsilon_{is}X_{ir}\varepsilon_{iq}\right)\right|\left|\mu_{ts}\mu_{rq}\right| \\
&\leq \frac{c_{NT}^2}{N_k^2T^4}\sum_{i\in G_k^0}\sum_{T\geq t>s>r>q\geq 1}\left\|X_{it}\varepsilon_{is}X_{ir}\right\|_{\frac{8+4\sigma}{3},\mathcal{D}}\left\|\varepsilon_{iq}\right\|_{8+4\sigma,\mathcal{D}}\alpha_{NT}^{\mathcal{D}}\left(r-q\right)^{(2+\sigma)/(1+\sigma)} \\
&\quad\times\left\|F_t^0\right\|\left\|F_s^0\right\|\left\|F_r^0\right\|\left\|F_q^0\right\| \\
&\leq \frac{c_{NT}^2}{2N_k^2T^4}\sum_{i\in G_k^0}\sum_{T\geq t>s>r>q\geq 1}\left\|X_{it}\varepsilon_{is}X_{ir}\right\|_{\frac{8+4\sigma}{3},\mathcal{D}}^2\left(\left\|F_t^0\right\|\left\|F_s^0\right\|\left\|F_r^0\right\|\right)^2\alpha_{NT}^{\mathcal{D}}\left(r-q\right)^{(2+\sigma)/(1+\sigma)} \\
&\quad+\frac{c_{NT}^2}{2N_k^2T^4}\sum_{i\in G_k^0}\sum_{T\geq t>s>r>q\geq 1}\left\|\varepsilon_{iq}\right\|_{8+4\sigma,\mathcal{D}}^2\left\|F_q^0\right\|^2\alpha_{NT}^{\mathcal{D}}\left(r-q\right)^{(2+\sigma)/(1+\sigma)} \\
&\equiv v_{2,11}\left(a1\right)+v_{2,12}\left(a1\right),\text{ say.}
\end{aligned}
$$

By Hölder inequality

$$
\begin{aligned}
\left\|X_{it}\varepsilon_{is}X_{ir}\right\|_{\frac{8+4\sigma}{3},\mathcal{D}}^2 &= \left[E_{\mathcal{D}}\left\|X_{it}\varepsilon_{is}X_{ir}\right\|^{\frac{8+4\sigma}{3}}\right]^{\frac{6}{8+4\sigma}} \\
&\leq \left\{\left[E_{\mathcal{D}}\left\|X_{it}\right\|^{8+4\sigma}\right]^{1/3}\left[E_{\mathcal{D}}\left\|\varepsilon_{is}X_{ir}\right\|^{4+2\sigma}\right]^{2/3}\right\}^{\frac{6}{8+4\sigma}} \\
&\leq \left\{\left[E_{\mathcal{D}}\left\|X_{it}\right\|^{8+4\sigma}\right]^{1/3}\left[E_{\mathcal{D}}\left\|\varepsilon_{is}\right\|^{8+4\sigma}\right]^{1/3}\left[E_{\mathcal{D}}\left\|X_{ir}\right\|^{8+3\sigma}\right]^{1/3}\right\}^{\frac{6}{8+4\sigma}} \\
&= \left\|X_{it}\right\|_{8+4\sigma,\mathcal{D}}^2\left\|\varepsilon_{is}\right\|_{8+4\sigma,\mathcal{D}}^2\left\|X_{ir}\right\|_{8+4\sigma,\mathcal{D}}^2.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
v_{2,11}\left(a1\right) &\leq \frac{c_{NT}^2}{2N_k^2T^4}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\left\|X_{it}\varepsilon_{is}X_{ir}\right\|_{\frac{8+4\sigma}{3},\mathcal{D}}^2\left(\left\|F_t^0\right\|\left\|F_s^0\right\|\left\|F_r^0\right\|\right)^2\sum_{\tau=1}^{\infty}\alpha_{NT}^{\mathcal{D}}\left(\tau\right)^{(2+\sigma)/(1+\sigma)} \\
&\leq \frac{O_P\left(1\right)}{N_kT}\frac{c_{NT}^2}{2N_k}\sum_{i\in G_k^0}\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|X_{it}\right\|_{8+4\sigma,\mathcal{D}}^2\left\|F_t^0\right\|^2\right\}^2\frac{1}{T}\sum_{s=1}^{T}\left\|\varepsilon_{is}\right\|_{8+4\sigma,\mathcal{D}}^2\left\|F_s^0\right\|^2 \\
&= O_P\left(N^{-1}T^{-1}\right).
\end{aligned}
$$

In addition

$$
\begin{aligned}
v_{2,11}\left(a1\right) &\leq \frac{c_{NT}^2}{2N_k^2T^2}\sum_{i\in G_k^0}\sum_{q=1}^{T}\left\|\varepsilon_{iq}\right\|_{8+4\sigma,\mathcal{D}}^2\left\|F_q^0\right\|^2\sum_{\tau=1}^{\infty}\alpha_{NT}^{\mathcal{D}}\left(\tau\right)^{(2+\sigma)/(1+\sigma)} \\
&\leq \frac{O_P\left(1\right)}{N_kT}\frac{c_{NT}^2}{2N_kT}\sum_{i\in G_k^0}\sum_{q=1}^{T}\left\|\varepsilon_{iq}\right\|_{8+4\sigma,\mathcal{D}}^2\left\|F_q^0\right\|^2=O_P\left(N^{-1}T^{-1}\right).
\end{aligned}
$$

So $v_{2,1}(a1) = O_P\left(N^{-1}T^{-1}\right)$. Similarly we can show that $v_{2,1}(a2) = O_P\left(N^{-1}T^{-1}\right)$ and $v_{2,1}(a3) = O_P\left(N^{-1}T^{-1}\right)$. So $v_{2,1} = O_P\left(N^{-1}T^{-1}\right)$. By the same token, we can show that $v_{2,3} = O_P\left(N^{-1}T^{-1}\right)$. By the analysis of $E_{\mathcal{D}}\left(V_{1kNT,2}\right)$, $v_{2,2} = O_P\left(T^{-2}\right)$. It follows that $E_{\mathcal{D}}(V_{1kNT,2}^2) = O_P(N^{-1}T^{-1} + T^{-2})$ and $V_{1kNT,2} = O_P(N^{-1/2}T^{-1/2} + T^{-1}) = O_P(T^{-1/2}\delta_{NT}^{-1})$ by Markov inequality. In sum, $V_{1kNT} = O_P(T^{-1/2}\delta_{NT}^{-1})$.

(ii) We verify that $V_{2kNT} \equiv \frac{1}{NT}\sum_{i\in G_k^0} X_i' M_{F^0}\varepsilon_i^* = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$. Write $V_{2kNT} = \frac{1}{NN_kT}\sum_{i\in G_k^0} X_i' \sum_{j=1}^N \nu_{ji}\varepsilon_j - \frac{1}{NN_kT}\sum_{i\in G_k^0} X_i' P_{F^0}\sum_{j=1}^N \nu_{ji}\varepsilon_j \equiv V_{2kNT,1} - V_{2kNT,2}$, say. Noting that

$$E_{\mathcal{D}}\left(V_{2kNT,1}\right) = \frac{1}{NN_kT}\sum_{i\in G_k^0}\sum_{j=1}^N\sum_{t=1}^T E_{\mathcal{D}}\left(X_{it}\varepsilon_{jt}\right)\nu_{ji} = \frac{1}{NN_kT}\sum_{i\in G_k^0}\sum_{t=1}^T E_{\mathcal{D}}\left(X_{it}\varepsilon_{it}\right)\nu_{ji} = 0,$$

and

$$\begin{aligned}
E_{\mathcal{D}}\left(V_{2kNT,1}^2\right) &= \frac{1}{N^2N_k^2T^2}\sum_{i\in G_k^0}\sum_{j=1}^N\sum_{k\in G_k^0}\sum_{l=1}^N\sum_{t=1}^T\sum_{s=1}^T E_{\mathcal{D}}\left(X_{it}\varepsilon_{jt}X_{ks}\varepsilon_{ls}\right)\nu_{ji}\nu_{kl} \\
&= \frac{1}{N^2N_k^2T^2}\sum_{i\in G_k^0}\sum_{j=1}^N\sum_{k\in G_k^0}\sum_{t=1}^T\sum_{s=1}^T E_{\mathcal{D}}\left(X_{it}\varepsilon_{jt}X_{ks}\varepsilon_{js}\right)\nu_{ji}\nu_{kl} \\
&= \frac{1}{N^2N_k^2T^2}\sum_{i\in G_k^0}\sum_{j=1,j\neq i}^N\sum_{k\in G_k^0,k\neq j,i}\sum_{t=1}^T\sum_{s=1}^T E_{\mathcal{D}}\left(X_{it}\varepsilon_{jt}X_{ks}\varepsilon_{js}\right)\nu_{ji}\nu_{kl} + O_P\left(N^{-1}T^{-1}\right) \\
&= \frac{1}{N^2N_k^2T^2}\sum_{i\in G_k^0}\sum_{j=1,j\neq i}^N\sum_{k\in G_k^0,k\neq j,i}\sum_{t=1}^T E_{\mathcal{D}}\left(X_{it}\right)E_{\mathcal{D}}\left(X_{ks}\right)E_{\mathcal{D}}\left(\varepsilon_{jt}^2\right)\nu_{ji}\nu_{kl} + O_P\left(N^{-1}T^{-1}\right) \\
&= O_P\left(N^{-1}T^{-1}\right),
\end{aligned}$$

we have $V_{2kNT,1} = O_P((NT)^{-1/2})$. For $V_{2kNT,2}$, we have

$$\begin{aligned}
\left|E_{\mathcal{D}}\left(V_{2kNT,2}\right)\right| &= \left|\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T E_{\mathcal{D}}\left(X_{it}\varepsilon_{js}\right)\mu_{ts}\nu_{ji}\right| \\
&= \left|\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{t=2}^T\sum_{s=1}^t E_{\mathcal{D}}\left(X_{it}\varepsilon_{is}\right)\mu_{ts}\nu_{ii}\right| \\
&\leq \frac{c_{NT}}{NN_kT^2}\sum_{i\in G_k^0}\sum_{t=2}^T\sum_{s=1}^t \left\|X_{it}\right\|_{4+2\sigma,\mathcal{D}}\left\|\varepsilon_{is}\right\|_{4+2\sigma,\mathcal{D}}\left\|F_t^0\right\|\left\|F_s^0\right\|\left\|\lambda_i^0\right\|^2\alpha_{NT}^{\mathcal{D}}\left(t-s\right)^{(2+\sigma)/(1+\sigma)} \\
&\leq \frac{c_{NT}}{2NN_kT^2}\sum_{i\in G_k^0}\sum_{t=2}^T\sum_{s=1}^t \left\|X_{it}\right\|_{4+2\sigma,\mathcal{D}}^2\left\|F_t^0\right\|^2\left\|\lambda_i^0\right\|^2\alpha_{NT}^{\mathcal{D}}\left(t-s\right)^{(2+\sigma)/(1+\sigma)} \\
&\quad + \frac{c_{NT}}{2NN_kT^2}\sum_{i\in G_k^0}\sum_{t=2}^T\sum_{s=1}^t \left\|\varepsilon_{is}\right\|_{4+2\sigma,\mathcal{D}}^2\left\|F_s^0\right\|^2\left\|\lambda_i^0\right\|^2\alpha_{NT}^{\mathcal{D}}\left(t-s\right)^{(2+\sigma)/(1+\sigma)} \\
&= O_P\left(N^{-1}\right) + O_P\left(N^{-1}\right) = O_P\left(N^{-1}\right),
\end{aligned}$$

where $c_{NT}=\left\|\left(T^{-1}F^{0\prime}F^{0}\right)^{-1}\right\|\left\|\left(T^{-1}\Lambda^{0\prime}\Lambda^{0}\right)^{-1}\right\|$. In addition,

$$
\begin{aligned}
E_{\mathcal{D}}\left(V_{2kNT,2}^{2}\right) &= \frac{1}{N^{2}N_{k}^{2}T^{4}}\sum_{i\in G_{k}^{0}}\sum_{j=1}^{N}\sum_{k\in G_{k}^{0}}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}E_{\mathcal{D}}\left(X_{it}\varepsilon_{js}X_{kr}\varepsilon_{lq}\right)\mu_{ts}\nu_{ji}\mu_{rq}\nu_{kl}\\
&= \frac{1}{N^{2}N_{k}^{2}T^{4}}\sum_{i\in G_{k}^{0}}\sum_{j=1}^{N}\sum_{k\in G_{k}^{0}}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}E_{\mathcal{D}}\left(X_{it}\varepsilon_{js}X_{kr}\varepsilon_{jq}\right)\mu_{ts}\nu_{ji}\mu_{rq}\nu_{kj}\\
&= \frac{1}{N^{2}N_{k}^{2}T^{4}}\sum_{i\in G_{k}^{0}}\sum_{j=1,j\neq i}^{N}\sum_{k\in G_{k}^{0},k\neq i,j}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}E_{\mathcal{D}}\left(X_{it}\right)E_{\mathcal{D}}\left(X_{kr}\right)E_{\mathcal{D}}\left(\varepsilon_{js}^{2}\right)\mu_{ts}\nu_{ji}\mu_{rq}\nu_{kj}\\
&\quad+\frac{1}{N^{2}N_{k}^{2}T^{4}}\sum_{i\in G_{k}^{0}}\sum_{k\in G_{k}^{0},k\neq i}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}E_{\mathcal{D}}\left(X_{it}\varepsilon_{is}X_{kr}\varepsilon_{iq}\right)\mu_{ts}\nu_{ii}\mu_{rq}\nu_{ki}\\
&\quad+\frac{1}{N^{2}N_{k}^{2}T^{4}}\sum_{i\in G_{k}^{0}}\sum_{j=1,j\neq i}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}E_{\mathcal{D}}\left(X_{it}X_{ir}\varepsilon_{jq}\right)E_{\mathcal{D}}\left(\varepsilon_{js}^{2}\right)\mu_{ts}\nu_{ji}\mu_{rq}\nu_{ij}\\
&\quad+\frac{1}{N^{2}N_{k}^{2}T^{4}}\sum_{i\in G_{k}^{0}}\sum_{k\in G_{k}^{0},k\neq i}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}E_{\mathcal{D}}\left(X_{it}\right)E_{\mathcal{D}}\left(\varepsilon_{ks}X_{kr}\varepsilon_{kq}\right)\mu_{ts}\nu_{ki}\mu_{rq}\nu_{kk}\\
&\quad+O_{P}\left(N^{-3}\right)\\
&= O_{P}(N^{-1}T^{-1})+O_{P}(N^{-2}T^{-1})+O_{P}(N^{-2}T^{-1})+O_{P}(N^{-2}T^{-1})+O_{P}(N^{-2}T^{-1})+O_{P}(N^{-3})\\
&= O_{P}(N^{-1}T^{-1}).
\end{aligned}
$$

It follows that $V_{2kNT}=O_{P}(N^{-1/2}T^{-1/2})$. In sum,we have $V_{2kNT}=O_{P}(N^{-1/2}T^{-1/2})$. This completes the proof. ■

D Numerical algorithm

In this appendix, we present the numerical algorithm to obtain the PPC estimates $\hat{\boldsymbol{\alpha}}=(\hat{\alpha}_{1},...,\hat{\alpha}_{K_{0}})$ and $\hat{\boldsymbol{\beta}}=(\hat{\beta}_{1},...,\hat{\beta}_{N})$. The algorithm is iterative and goes as follows:

1. Start with arbitrary initial values $\hat{\boldsymbol{\alpha}}^{(0)}=(\hat{\alpha}_{1}^{(0)},...,\hat{\alpha}_{K_{0}}^{(0)})$ and $\hat{\boldsymbol{\beta}}^{(0)}=(\hat{\beta}_{1}^{(0)},...,\hat{\beta}_{N}^{(0)})$ such that $\sum_{i=1}^{N}||\hat{\beta}_{i}^{(0)}-\hat{\alpha}_{k}^{(0)}||\neq 0$ for each $k=2,...,K_{0}$.[3]
2. Suppose that we have obtained $\hat{\boldsymbol{\alpha}}^{(r-1)}\equiv(\hat{\alpha}_{1}^{(r-1)},...,\hat{\alpha}_{K_{0}}^{(r-1)})$ and $\hat{\boldsymbol{\beta}}^{(r-1)}\equiv(\hat{\beta}_{1}^{(r-1)},\,...,\hat{\beta}_{N}^{(r-1)})$. In Step $r\geq 1$, we first choose $(\boldsymbol{\beta},\alpha_{1})$ to minimize

$$Q_{K_{0}NT}^{(r,1)}\left(\boldsymbol{\beta},\alpha_{1}\right)=Q_{1,NT}\left(\boldsymbol{\beta}\right)+\frac{\kappa}{N}\sum_{i=1}^{N}\left\|\beta_{i}-\alpha_{1}\right\|\Pi_{k\neq 1}^{K_{0}}\left\|\hat{\beta}_{i}^{(r-1)}-\hat{\alpha}_{k}^{(r-1)}\right\|$$

[3]For static and dynamic panel data models with IFEs, we propose to use the CCE estimators of Pesaran (2006) and Chudik and Pesaran (2015), respectively, as the initial estimates $\{\hat{\beta}_{i}^{(0)}\}$. Under the regularity conditions stated in these papers, these estimates are $\sqrt{T}$-consistent. In addition, one can simply set $\hat{\alpha}_{k}^{(0)}$'s as zero or the average of $\hat{\beta}_{i}^{(0)}$'s.

and obtain the updated estimate $(\hat{\boldsymbol{\beta}}^{(r,1)}, \hat{\alpha}_1^{(r)})$ of $(\boldsymbol{\beta}, \alpha_1)$. Then we choose $(\boldsymbol{\beta}, \alpha_2)$ to minimize

$$Q_{K_0NT}^{(r,2)}(\boldsymbol{\beta}, \alpha_2) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\kappa}{N} \sum_{i=1}^{N} \|\beta_i - \alpha_2\| \left\|\hat{\beta}_i^{(r,1)} - \hat{\alpha}_1^{(r)}\right\| \Pi_{k \neq 1,2}^{K_0} \left\|\hat{\beta}_i^{(r-1)} - \hat{\alpha}_k^{(r-1)}\right\|$$

to obtain the updated estimate $(\hat{\boldsymbol{\beta}}^{(r,2)}, \hat{\alpha}_2^{(r)})$ of $(\boldsymbol{\beta}, \alpha_2)$. Repeat this procedure until we choose $(\boldsymbol{\beta}, \alpha_{K_0})$ to minimize

$$Q_{K_0NT}^{(r,K_0)}(\boldsymbol{\beta}, \alpha_{K_0}) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\kappa}{N} \sum_{i=1}^{N} \|\beta_i - \alpha_K\| \Pi_{k=1}^{K_0-1} \left\|\hat{\beta}_i^{(r,K_0-1)} - \hat{\alpha}_k^{(r)}\right\|$$

to obtain the updated estimate $(\hat{\boldsymbol{\beta}}^{(r,K_0)}, \hat{\alpha}_{K_0}^{(r)})$ of $(\boldsymbol{\beta}, \alpha_{K_0})$. Let $\hat{\boldsymbol{\beta}}^{(r)} = \hat{\boldsymbol{\beta}}^{(r,K_0)}$ and $\hat{\boldsymbol{\alpha}}^{(r)} = (\hat{\alpha}_1^{(r)}, ..., \hat{\alpha}_{K_0}^{(r)})$.

3. Repeat the above step until certain convergence criterion is met, say, when

$$\frac{\sum_{i=1}^{N} \left\|\hat{\beta}_i^{(r)} - \hat{\beta}_i^{(r-1)}\right\|^2}{\sum_{i=1}^{N} \left\|\hat{\beta}_i^{(r-1)}\right\|^2 + 0.0001} < \epsilon_{tol} \text{ and } \frac{\sum_{k=1}^{K_0} \left\|\hat{\alpha}_k^{(r)} - \hat{\alpha}_k^{(r-1)}\right\|^2}{\sum_{k=1}^{K_0} \left\|\hat{\alpha}_k^{(r-1)}\right\|^2 + 0.0001} < \epsilon_{tol},$$

where ϵ_{tol} is some prescribed tolerance level (e.g., 0.0001). Define the final iterative estimate of $\boldsymbol{\alpha}$ as $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1^{(R)}, ..., \hat{\alpha}_{K_0}^{(R)})$ for sufficiently large R such that the convergence criterion is met. The final iterative estimate of $\boldsymbol{\beta}$ is defined as $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, ..., \hat{\beta}_N)$ where

$$\begin{aligned} \hat{\beta}_i &= \sum_{k=1}^{K_0} \hat{\alpha}_k^{(R)} \mathbf{1}\left\{\hat{\beta}_i^{(R,l)} = \hat{\alpha}_k^{(R)} \text{ for some } l = 1, ..., K_0\right\} \\ &\quad + \hat{\beta}_i^{(R,K_0)} \left[1 - \sum_{k=1}^{K_0} \mathbf{1}\left\{\hat{\beta}_i^{(R,l)} = \hat{\alpha}_k^{(R)} \text{ for some } l = 1, ..., K_0\right\}\right] \end{aligned} \tag{D.1}$$

where $\hat{\beta}_i^{(R,l)}$ denotes the ith column of $\hat{\boldsymbol{\beta}}^{(R,l)}$ for $l = 1, 2, ..., K$. Intuitively speaking, we classify individual i to group $\hat{G}_k$ if $\hat{\beta}_i^{(R,l)} = \hat{\alpha}_k^{(R)}$ for some $l = 1, ..., K_0$; otherwise it is left unclassified so that $\hat{\beta}_i$ is defined as $\hat{\beta}_i^{(R,K_0)}$.

References

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

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

Bernstein, D. S., 2005. Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory. Princeton University Press, Princeton.

Bertsekas, D.P., 1995. Nonlinear Programming. Athena Scientific, Belmont, MA.

Chudik, A., Pesaran M. H., 2015. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics 188, 393-420.

Li, D., Qian, J., Su, L., 2016. Panel data models with interactive fixed effects and multiple structural breaks. Journal of the American Statistical Association, forthcoming.

Lu, X., Su, L., 2016. Shrinkage estimation of dynamic panel data models with interactive fixed effects. Journal of Econometrics 190, 148-175.

Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967-1012.

Pollard, D., 1984. Convergence of Stochastic Processes. Springer-Verlag, New York.

Su, L., Shi, Z., Phillips, P.C.B., 2016. Identifying latent structures in panel data. Econometrica, forthcoming.