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

Liangjun Su\textsuperscript{a} and Gaosheng Ju\textsuperscript{b}

\textsuperscript{a} School of Economics, Singapore Management University
\textsuperscript{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 (IFE$s$) 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 (AFE$s$) and IFE$s$, 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 AFE$s$. 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 IFE$s$:

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

where $X_{it}$ is a $p \times 1$ vector of predetermined variables that may include lagged dependent variables,
\( \beta_i^0 \) is a \( p \times 1 \) vector of unknown slope coefficients, \( \lambda_i^0 \) and \( F_i^0 \) 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 \( \varepsilon_{it} \) is the idiosyncratic error term. Here, parameters with the superscript 0 denote the true values. We assume that the slope coefficients \( \beta^0 \)'s exhibit a certain grouped pattern so that they only take \( K_0 \) distinct values, say, \( \{\alpha_k^0, k = 1, ..., K_0\} \) 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 \( \beta_i \)'s to the unknown group-specific parameter vectors \( \alpha_k \)'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 \( \beta_i \)'s as the asymptotics for the estimates of \( \lambda_i \) and \( F_t \) follow directly from Bai (2003). When we assume that \( \beta_i \)'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 \( \lambda_0 \) 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 $\|A\|_{sp}$, 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 $r$th largest eigenvalue by counting multiple eigenvalues multiple times; we also use $\mu_{\text{max}} ( \cdot )$ and $\mu_{\text{min}} (A)$ to denote $A$’s largest and smallest eigenvalues, respectively. Let $0_{p \times 1}$ denote a $p \times 1$ vector of zeros and $1 \{ \cdot \}$ the usual indicator function. We use “p.d.” and “p.s.d.” to abbreviate “positive definite” and “positive semidefinite”, respectively. The operator $\overset{P}{\rightarrow}$ denotes convergence in probability, $\overset{D}{\rightarrow}$ convergence in distribution, and plim probability limit. We use $(N,T) \rightarrow \infty$ 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 $i = 1, \ldots, N$, and $t = 1, \ldots, T$. We consider the following panel structure model

$$Y_{it} = \beta_i^0 X_{it} + \lambda^0 F^0 + \varepsilon_{it}, \quad (2.1)$$

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

$$\beta_i^0 = \begin{cases} 
\alpha_i^0 & \text{if } i \in G^0_1 \\
\vdots & \vdots \\
\alpha_i^0 & \text{if } i \in G^0_{K_0}
\end{cases} \quad (2.2)$$

Here $\alpha_i^0 \neq \alpha_j^0$ for any $j \neq k$, $\cup_{k=1}^{K_0} G^0_k = \{1, 2, \ldots, N\}$, and $G^0_k \cap G^0_j = \emptyset$ for any $j \neq k$. Let $N_k = \# G^0_k$, the cardinality of the set $G^0_k$. 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

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

The true values of $\alpha, \beta, \Lambda$, and $F$ are denoted as $\alpha^0, \beta^0, \Lambda^0$, and $F^0$ respectively. We are interested in inferring each individual’s group identity and estimating $(\alpha^0, \Lambda^0, F^0)$ consistently.
2.2 Penalized principal component estimation

In this subsection we consider the PPC estimation of \((\beta^0, \alpha^0, \Lambda^0, F^0)\) 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

\[
Q^{(K_0)}_{0\text{NT}, \kappa} (\beta, \alpha, \Lambda, F) = Q_{0\text{NT}} (\beta, \Lambda, F) + \frac{\kappa}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \| \beta_i - \alpha_k \|,
\]

(2.4)

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

\[
Q^{(K_0)}_{1\text{NT}, \kappa} (\beta, \alpha, F) = Q_{1\text{NT}} (\beta, F) + \frac{\kappa}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \| \beta_i - \alpha_k \|,
\]

(2.5)

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

\[
Q^{(K_0)}_{NT, \kappa} (\beta, \alpha) = Q_{NT} (\beta) + \frac{\kappa}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \| \beta_i - \alpha_k \|,
\]

(2.6)

where \(Q_{NT} (\beta) = \frac{1}{N} \sum_{r=R_0+1}^{T} \mu_r \left[ \frac{1}{N} \sum_{i=1}^{N} (Y_i - X_i \beta_i)' (Y_i - X_i \beta_i) \right] \). Minimizing the profile PPC criterion function in (2.6) produces the Classifier Lasso (C-Lasso) estimators \(\hat{\beta}\) and \(\hat{\alpha}\) of \(\beta\) and \(\alpha\), respectively. Let \(\hat{\beta}_i\) and \(\hat{\alpha}_k\) denote the \(i\)th and \(k\)th columns of \(\hat{\beta}\) and \(\hat{\alpha}\), respectively, i.e., \(\hat{\alpha} \equiv (\hat{\alpha}_1, \ldots, \hat{\alpha}_{K_0})\) and \(\hat{\beta} \equiv (\hat{\beta}_1, \ldots, \hat{\beta}_N)\). We then obtain the estimate \((\hat{\Lambda}, \hat{F})\) of \((\Lambda, F)\) 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

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

(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 \(\beta_i\) to a particular
unknown group-level parameter vector $\alpha_k$, 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 $(\beta, \alpha)$ even though it is convex in $\alpha_k$ when one fixes $\alpha_j$ for $j \neq k$. In the supplemental Appendix D, we propose an iterative algorithm to obtain the estimates $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha}_{K_0})$ and $\hat{\beta} = (\hat{\beta}_1, ..., \hat{\beta}_N)$.

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 $\hat{\beta}$ and $P_\ell$. Let $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ and $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)'$. 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 $\zeta > 0$, we define $\mathcal{N} = \{1, 2, ..., N\}$ and

$$ S_{N,F,\zeta} = \{i \in \mathcal{N} : P(\mu_{\min}(T^{-1}X_i'X_i) \geq \zeta) = 1 - o(N^{-1})\}. $$

Let $S_{N,F,\zeta}^* = \mathcal{N} \setminus S_{N,F,\zeta}$. 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) $\max_{1 \leq t \leq T} E \left\| F_{it} \right\|^2 \leq C$ and $T^{-1} F_{it}F_{it} \rightarrow \Sigma_{F_{it}} > 0$ as $T \rightarrow \infty$.

(ii) $\max_{1 \leq i \leq N} E(\lambda^0_{it}) \leq C$ and $N^{-1} \Lambda^0 \Lambda^0 \rightarrow \Sigma_{\Lambda^0} > 0$ as $N \rightarrow \infty$.

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

(iv) $E(\varepsilon_{it}) = 0$, $\|\varepsilon\|_p = O_P(\sqrt{N} + \sqrt{T})$, and $\max_{1 \leq i \leq N} \max_{1 \leq t \leq T} E(\varepsilon_{it}^2) \leq C$.

(v) Let $\sigma_{ij,ts} = E(\varepsilon_{it} \varepsilon_{js})$, $\max_{1 \leq i \leq N} T^{-1} \sum_{t=1}^T \sum_{s=1}^T \sigma_{ij,ts} \leq C$, 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) $\max_{1 \leq i \leq N} T^{-1} \sum_{t=1}^T \sum_{s=1}^T \max_{1 \leq \varepsilon_{it} \leq \varepsilon_{is}} \left| \sum_{t=1}^T \varepsilon_{it} \varepsilon_{is} - E(\varepsilon_{it} \varepsilon_{is}) \right| \leq C$.

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

(viii) $\max_{1 \leq i \leq N} \left| \sum_{t=1}^T \varepsilon_{it} \varepsilon_{is} - E(\varepsilon_{it} \varepsilon_{is}) \right|^2 \leq C$.

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

(x) $\max_{1 \leq i \leq N} T^{-1} \sum_{t=1}^T \sum_{s=1}^T \max_{1 \leq \varepsilon_{it} \leq \varepsilon_{is}} \left| \sum_{t=1}^T \varepsilon_{it} \varepsilon_{is} - E(\varepsilon_{it} \varepsilon_{is}) \right|^2 \leq C$.
(x) There exists a constant $\xi_{XX} > 0$ such that $P\left(\min_{1 \leq i \leq N} \mu_{\min}\left(T^{-1}X_i'M_{F_0}X_i\right) \geq \xi_{XX}\right) = 1 - o\left(N^{-1}\right)$.

(xi) There exists a constant $\gamma > 0$ such that $\sup_{F \in \mathcal{F}} \#S_{N, F}^* / N = o(1)$, $N^{-1} \sum_{i \in N_s} \lambda_i 0^y P \to \Sigma_s > 0$ for any subset of $N_s$ of $N$ with $\#N_s / N \to 1$, and $\max_{1 \leq i \leq N} T^{-1} \epsilon_i M_{F_0} \epsilon_i = O_P(1)$.

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

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

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

A.1(i)-(iii) impose standard moment conditions on $F_{F_0}$, $\lambda_1$, 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 $F_{F_0}$ but require that $\lambda_1$ be uniformly bounded. A.1(iv)-(vi) impose conditions on the error processes $\{\epsilon_t\}$. 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 $F_0$. 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 $\|\lambda^0 F_0\| = O_P(N^{1/2}T^{1/2})$ by Chebyshev inequality, which further implies that $\|\lambda^0 \epsilon_0\|^2 = O_P(NT)$ and $\|\epsilon F_0\|^2 = \sum_{i=1}^{N} \|F_0 \epsilon_i\|^2 = O_P(NT)$ 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 $N^{-1} \sum_{i=1}^{N} \|\beta_i - \beta_1^0\|^2$ 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 $\alpha_k^0, k = 1, \ldots, K_0$, 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 $\kappa$ 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) $N^{-1} \sum_{i=1}^{N} \left(\hat{\beta}_i - \beta_i^0\right)' \left(T^{-1}X_i'M_{F_0}X_i\right) \left(\hat{\beta}_i - \beta_i^0\right) = o_P(1)$,

(ii) $T^{-1}F_0^{-1} \hat{F}$ is invertible and $\|P_{\hat{F}} - P_{F_0}\| = o_P(1)$,

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

8
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 \( Q_{1\rightarrow N}^{(K_0)} (\beta, \alpha, F) \) 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 \( \beta \) 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 \( \beta^0_i \), 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 \( \beta^0_i \).

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

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

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

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, \( \alpha^0_1, \ldots, \alpha^0_{K_0} \), can be estimated consistently by \( \hat{\alpha}_1, \ldots, \hat{\alpha}_{K_0} \) subject to suitable permutation. Our findings are similar to those in SSP who show that the pointwise convergence of \( \hat{\beta}_i \) depends on \( \kappa \) 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 \( \alpha^0_k \)'s, and define

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

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

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) \( P \left( \max_{1 \leq i \leq N} \left\| T^{-1/2} \sum_{t=1}^{T} [\xi_{it} - E (\xi_{it})] \right\| \geq C \left( \ln T \right)^{v_1} \right) = o(N^{-1}) \forall C > 0 \) for some \( v_1 \geq 1 \) and for \( \xi_{it} = X_{it}X_{it}' + \varepsilon_{it}, X_{it} \varepsilon_{it}, X_{it}F^0_{it}, \) and \( F^0_{it} \varepsilon_{it} \).

(ii) \( P \left( \max_{1 \leq i, j \leq N} \left\| T^{-1/2} \sum_{t=1}^{T} [\xi_{it} - \xi_{jt}] \right\| \geq C \left( \ln T \right)^{v_1} \right) = o(N^{-1}) \forall C > 0 \) for \( \xi_{it} = X_{it} \) and \( \varepsilon_{it} \).

(iii) \( P \left( \max_{1 \leq i \leq N} \left\| (NT)^{-1/2} \sum_{t=1}^{T} \varepsilon_{it} \right\| \geq C(1 + \sqrt{T/N}) \left( \ln T \right)^{v_1} \right) = o(N^{-1}) \forall C > 0 \).
(iv) \( P\left(\max_{1 \leq i \leq N} \|\lambda_i^0\| \geq c_{\lambda,N}\right) = o(N^{-1}) \) 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 \( \{(X_{it}, \varepsilon_{it}, F_i^0, \lambda_i^0)\} \). 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 \( P\left(\max_{1 \leq i \leq N} T^{-1/2}X_i^\top M_{F_i \varepsilon_i} \geq C (\ln T)^{v_1}\right) = o\left(N^{-1}\right) \) 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 \( \lambda_i^0 \)'s are uniformly bounded, then A.3(iv) is satisfied with some \( c_{\lambda,N} \) that does not depend on \( N \). If \( \max_{1 \leq i \leq N} E[\exp\left(c_\lambda \|\lambda_i^0\|\right)] \leq C < \infty \) 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 \( v_2 > 0 \),

\[
P\left(\max_{1 \leq i \leq N} \|\hat{\beta}_i - \beta_i^0\| \geq C[T^{-1/2} (\ln T)^{v_1} + \kappa (\ln T)^{v_2}]\right) = o\left(N^{-1}\right) .
\]

### 3.2 Classification consistency

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

\[
\hat{E}_{kNT,i} = \left\{ i \notin \hat{G}_k \mid i \in G^0_k \right\} \quad \text{and} \quad \hat{F}_{kNT,i} = \left\{ i \notin \hat{G}_k \mid i \in \hat{G}_k \right\} ,
\]

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

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

**Assumption A.2 (iv)** \( T\kappa^2/([\ln T]^{2v_1} c_{\lambda,N}^2) \rightarrow \infty \) as \((N,T) \rightarrow \infty\).
Assumption A.2(iv) strengthens the conditions on the tuning parameter $\kappa$. If $\lambda_i^0$’s are sub-gaussian, Assumptions A.2(iii)-(iv) indicate that we can specify $\kappa$ such that

$$\kappa \propto T^{-a}$$

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) $P\left(\bigcup_{k=1}^{K_0} \tilde{E}_{k,NT} \right) \leq \sum_{k=1}^{K_0} P\left(\tilde{E}_{k,NT}\right) \to 0$ as $(N, T) \to \infty$,

(ii) $P\left(\bigcup_{k=1}^{K_0} \tilde{F}_{k,NT} \right) \leq \sum_{k=1}^{K_0} P\left(\tilde{F}_{k,NT}\right) \to 0$ as $(N, T) \to \infty$.

Theorem 3.4 implies that for all individuals within a group, say, $G^0_k$, 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 ($G^0_k$) w.p.a.1. Define the events $\hat{H}_{i,NT} = \{i \in \hat{G}_0\}$. Theorem 3.4(i) implies that $P(\cup_{1 \leq i \leq N} \hat{H}_{i,NT}) \leq \sum_{k=1}^{K_0} \hat{P}(\hat{E}_{k,NT}) \to 0$. 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

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

Since the event $\sum_{k=1}^{K_0} I \{\hat{\beta}_i = \hat{\alpha}_k\} = 0$ 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 $\tilde{N}_k = \#\hat{G}_k$ for $k = 0, 1, ..., K_0$. Based on Theorem 3.4, SSP also show that $\tilde{N}_k = N_k + o_P(1)$ for $k = 1, ..., K_0$.

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

Let $\nu_{ij} = \lambda_i^{0r} (N^{-1}\Lambda^r\Lambda^0)^{-1} \lambda_j^0$ and $\mu_{ts} = F_t^0 (T^{-1}F_0^rF_0)^{-1} F_s^0$. Let $Q_{k,NT} = \frac{1}{N_k T} \sum_{i \in G_k} X_i^rM_{F_0}X_i$, $A_{k,l} = \frac{1}{N_{N_k T}} \sum_{i \in G_k} \sum_{j \in G_l} \nu_{ij}X_i^rM_{F_0}X_j$, and $\hat{V}_{k,NT} = \frac{1}{N_{k T}} \sum_{i \in \hat{G}_k} X_i^rM_{\hat{F}}(\varepsilon_i - \frac{1}{N} \sum_{j=1}^{N} \nu_{ji} \varepsilon_j)$ where $k, l = 1, ..., K_0$. Define

$$Q_{NT} = \begin{pmatrix}
Q_{1NT} - A_{1,1} & -A_{1,2} & \cdots & -A_{1,K_0} \\
-A_{2,1} & Q_{2NT} - A_{2,2} & \cdots & -A_{2,K_0} \\
\vdots & \vdots & \ddots & \vdots \\
-A_{K_0,1} & -A_{K_0,2} & \cdots & Q_{K_0,NT} - A_{K_0,K_0}
\end{pmatrix} \text{ and } \hat{V}_{NT} = \begin{pmatrix}
\hat{V}_{1NT} \\
\hat{V}_{2NT} \\
\vdots \\
\hat{V}_{K_0,NT}
\end{pmatrix} \tag{3.3}$$
Let $D \equiv \sigma \left(F^0, \lambda^0\right)$, the sigma-field generated by $(F^0, \lambda^0)$, and $E_D(A) \equiv E(A|D)$. Let$^2$

$$B_{1,kNT} = - \frac{1}{NN_k T^2} \sum_{i \in G_k} X_i' \beta i \varepsilon_i \hat{F}(T^{-1} F^{0 \prime} \hat{F})^{-1} (N^{-1} \Lambda^{0 \prime} \Lambda^0)^{-1} \lambda_i^0,$$

$$B_{2,kNT} = - \frac{1}{NN_k T^2} \sum_{i \in G_k} \sum_{j=1}^N (X_i - X_i^*)' E_i F^0 \left( \hat{F}(T^{-1} F^{0 \prime} F^0)^{-1} \left(N^{-1} \Lambda^{0 \prime} \Lambda^0\right)^{-1} \lambda_{ij}^0 \varepsilon_j \varepsilon_i \right),$$

$$B_{3,kNT} = \frac{1}{N_k T} \sum_{i \in G_k^0} \hat{v}_i,$$

$$\mathcal{V}_{kNT} = \frac{1}{NT} \sum_{i=1}^N Z_{k,i}^\prime \varepsilon_i,$$

where $X_i^* = \frac{1}{N} \sum_{j=1}^N \nu_{ij} X_j$ with the $t$th row given by $X_{it}^*$ and $Z_{k,i} = \frac{N}{\nu_i} [X_i - F_0 E_D(X_i)] 1 \{ i \in G_k \}$.

$$- \frac{1}{N_k} \sum_{j \in G_k} \nu_{ji} E_D(X_j).$$

Let $Z_i^\prime$ denote the $t$th row of $Z_i = (Z_{1,i}, ..., Z_{K_0,i})$. Let $B_{s,NT} = (B_{s,1NT}, ..., B_{s,K_0NT})^\prime$ for $s = 1, 2, 3$ and $\mathcal{V}_{NT} = (\mathcal{V}_{1NT}, ..., \mathcal{V}_{K_0NT})^\prime$. Note that $\mathcal{V}_{NT} = \frac{1}{NT} \sum_{i=1}^N Z_i^\prime \varepsilon_i = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z_{it} \varepsilon_{it}$.

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

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

(ii) For each $k = 1, ..., K_0$, $\frac{1}{NT} \sum_{i \in G_k^0} X_i^0 \beta i = O_P \left(T^{-1/2} \delta_{NT}^{-1}\right)$ and $\frac{1}{NT} \sum_{i \in G_k^0} X_i^0 \beta i = O_P \left(T^{-1/2} \delta_{NT}^{-1}\right)$ 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 that 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 $D$. 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 $\text{vec}(\hat{\alpha} - \alpha^0) = Q_{NT}^{-1} [\hat{V}_{NT} + \mathcal{B}_{1,NT}] + o_P((NT)^{-1/2}) = O_P \left(T^{-1/2} \delta_{NT}^{-1}\right)$.
Theorem 3.5 gives the key Bahadur-type representation of \( \vec{\alpha} - \alpha^0 \). As we demonstrate in the proof of Theorem 3.5, \( B_{1,NT} \) 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 \( V_{kNT} = \frac{1}{N} \sum_{i \in G_k} X_i^0 M_{F0}(\varepsilon_i - \frac{1}{N} \sum_{j=1}^N \nu_{ji}\varepsilon_j) \) and \( V_{NT} = (V_{1NT}, \ldots, V_{K_0 NT})' \). To study the asymptotic distribution of the PPC estimator \( \{\hat{\alpha}_k\} \), we add the following assumption.

**Assumption A.5.** (i) For each \( k = 1, \ldots, K_0 \), \( V_{kNT} = V_{kNT} + B_{3,kNT} + o_P((NT)^{-1/2}) \).

(ii) For each \( k = 1, \ldots, K_0 \), \( B_{3,kNT} \) is independent of \( X_{js}, \lambda_{js}^0, \) and \( F_{s}^0 \) for all \( i, j, s, t, \) and \( s \) (i.e., Assumption D in Bai (2009) is maintained), it is easy to verify that \( A.5 \) holds with \( V_{kNT} = V_{kNT} \) and \( B_{3,kNT} = 0 \). If in addition \( K_0 = 1 \), noticing that \( V_{NT} = \frac{1}{N} \sum_{i=1}^N (X_i - X_i^*)' M_{F0}\varepsilon_i \) in this case and \( M_{F0}(X_i - X_i^*) \) is defined as \( Z_i \) in Bai (2009), \( A.5(1)-(3) \) become identical to Assumption E in Bai (2009). Nevertheless, we allow \( X_{it} \) to contain lagged dependent variables and \( \varepsilon_{it} \) to be dependent on \( F_t \). In this case, \( B_{3,kNT} \) 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 \( \sqrt{NT}[\vec{\alpha} - \alpha^0 - Q_{NT}^{-1}(B_{1,NT} + B_{2,NT} + B_{3,NT})] \frac{D}{N}(0, Q_0^{-1} \Theta_0 Q_0^{-1}) \).

Clearly, Theorem 3.6 indicates that \( \vec{\alpha} \) has three bias terms: \( Q_{NT}^{-1}B_{1,NT}, Q_{NT}^{-1}B_{2,NT}, \) and \( Q_{NT}^{-1}B_{3,NT} \). If Assumption D in Bai (2009) holds true, then \( B_{3,NT} = 0 \). Maintaining such an assumption, we can follow the analysis of Bai (2009) and show that \( B_{2,NT} \) is absent from the above expression in the absence of both *cross-sectional* correlation and heteroskedasticity in the error terms and that \( B_{1,NT} \) 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 \( \hat{\alpha}_k \) of \( \alpha_k^0 \) by utilizing such group information. Let \( \tilde{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{K_0}) \). Under Assumptions A.1, A.2(i) and A.4-A.5, we can show that \( \sqrt{NT}[\vec{\alpha} - \alpha^0 - Q_{NT}^{-1}(B_{1,NT} + B_{2,NT} + B_{3,NT})] \frac{D}{N}(0, Q_0^{-1} \Theta_0 Q_0^{-1}) \).

13
$N(0, Q_0^{-1}\Theta_0 Q_0^{-1})$. 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 $\{\tilde{\alpha}_{G_k}\}$ denote the post-Lasso (non-penalized) PC estimators of $\{\alpha_k^0\}$ based on the estimated group structure in $\{\hat{G}_k\}$. Let $\tilde{\alpha} = (\tilde{\alpha}_{G_1}, \ldots, \tilde{\alpha}_{G_{K_0}})$. The following theorem reports the asymptotic distribution of the post-Lasso PC estimators $\{\tilde{\alpha}_{G_k}\}$.

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

That is, the post-Lasso PC estimators $\{\tilde{\alpha}_{G_k}\}$ 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 $\{\lambda^0_i\}$ to exhibit the same grouped pattern as $\{\beta^0_i\}$, we can let $\gamma^0_k$, $k = 1, \ldots, K_0$, denote the group-specific factor loadings. In this case, one can consider the following penalized PC objective function

$$
\hat{Q}_{0NT,0}(\beta, \alpha, \Lambda, F) = Q_{0NT}(\beta, \Lambda, F) + \frac{K}{N} \sum_{i=1}^N \Pi_{k=1}^{K_0} ||(\beta^i, \lambda^i) - (\alpha^i_k, \gamma^i_k)||,
$$

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

$$
\hat{Q}_{1NT,0}(\beta, \alpha, \Lambda) = \hat{Q}_{1NT}(\beta, \Lambda) + \frac{K}{N} \sum_{i=1}^N \Pi_{k=1}^{K_0} ||(\beta^i, \lambda^i) - (\alpha^i_k, \gamma^i_k)||,
$$

where $\hat{Q}_{1NT}(\beta, \Lambda)$ is analogously defined as $Q_{1NT}(\beta, F)$ by inverting the roles of $F$ and $\Lambda$. 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 $\|A\|_{q,D} \equiv [E_D(||A||^q)]^{1/q}$.

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

(ii) $\max_{1 \leq t \leq N} E(||\lambda^0_t||) \leq C$ and $N^{-1} \lambda^0 \lambda^0 \xrightarrow{P} \Sigma_{\lambda^0}$ as $N \to \infty$.

(iii) $\max_{1 \leq t \leq N, 1 \leq \ell \leq T} E(||\zeta_{it}||) \leq C$ for $\zeta_{it} = X_{it}$ and $\epsilon_{it}$.
(iv) \( \max_{1 \leq i \leq N} T^{-1} \sum_{t=1}^{T} E_{D} \| \zeta_{it} \|^2 = O_{P}(1) \) and \( \max_{1 \leq t \leq T} N^{-1} \sum_{i=1}^{N} E_{D} \| \zeta_{it} \|^2 = O_{P}(1) \) for \( \zeta_{it} = X_{it} \) and \( \varepsilon_{it} \).

(v) \( \max_{1 \leq t \leq T} N^{-1} \sum_{i=1}^{N} \| \zeta_{it} \|_{8+4\sigma,D}^4 = O_{P}(1) \) for \( \zeta_{it} = X_{it} \) and \( \varepsilon_{it} \).

Note that Assumptions B.1(i)-(iii) strengthen the moment conditions in Assumptions A.1(i)-(iii) and require finite eighth plus moments for \( F_{1}^{0} \), \( \lambda_{1}^{0} \), \( X_{it} \), and \( \varepsilon_{it} \) 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 \( F_{1}^{0} \), \( \lambda_{1}^{0} \) and \( X_{it} \) and finite eighth moments for \( \varepsilon_{it} \); but he assumes independence between \( \varepsilon_{it} \) 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-\( \sigma \)-algebra of \( \mathcal{A} \). Let \( P_{\mathcal{B}}(\cdot) \equiv P(\cdot|\mathcal{B}) \). 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 \( \alpha^{B}(t) \) converging to 0 a.s. as \( t \to \infty \) such that

\[
|P_{\mathcal{B}}(A \cap B) - P_{\mathcal{B}}(A) P_{\mathcal{B}}(B)| \leq \alpha^{B}(t) \quad \text{a.s.} \quad (3.6)
\]

for all \( A \in \sigma(\xi_{1}, \ldots, \xi_{k}) \), \( B \in \sigma(\xi_{k+1}, \xi_{k+2}, \ldots) \) and \( k \geq 1 \), \( t \geq 1 \), we refer to it as the \( \mathcal{B} \)-mixing strong coefficient.

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

(ii) \( (\varepsilon_{i}, X_{i}) \), \( i = 1, \ldots, N \), are mutually independent of each other conditional on \( \mathcal{D} \).

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

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

B.2(i) requires that each individual time series \( \{(X_{it}, \varepsilon_{it}) : t = 1, 2, \ldots\} \) 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 $\sigma$ 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 $\varepsilon_{it}$ exhibits conditional heteroskedasticity: $\varepsilon_{it} = \sigma_0(Y_{i,t-1}) \varepsilon_{it}$ where $\sigma_0(\cdot)$ 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 $D$ such that B.2(ii) is still satisfied. B.2(iii) requires that the error term $\varepsilon_{it}$ be a martingale difference sequence (m.d.s.) with respect to the filter $F_{NT,t}$ which allows for lagged dependent variables in $X_{it}$, and conditional heteroskedasticity, skewness, or kurtosis of an unknown form in $\varepsilon_{it}$. In contrast, both Bai (2009) and Pesaran (2006) assume that $\varepsilon_{it}$ is independent of $X_{js}$, $\lambda_j$, and $F_s$ for all $i$, $t$, $j$ and $s$; Moon and Weidner (2015b) allow dynamics but assume that $\varepsilon_{it}$’s are independent conditional on $D$ 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 $\Theta_0 = \text{plim}_{(N,T) \to \infty} \Theta_{NT}$ and $\Theta_{NT} = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} Z_{it} Z_{it}^T$. To conduct inference, we need to estimate $Q_{NT}$, $B_{1,NT}$, $B_{2,NT}$, $B_{3,NT}$, and $\Theta_{NT}$. Let $\hat{\psi}_{ij} = \hat{\lambda}_i (N^{-1} \hat{\Lambda})^{-1} \hat{\lambda}_j$, $\hat{Q}_{kNT} = \frac{1}{NNT} \sum_{i \in \hat{G}_k} \hat{X}_i \hat{M}_F \hat{X}_i$, and $\hat{A}_{k,l} = \frac{1}{NNT} \sum_{i \in \hat{G}_k} \sum_{j \in \hat{G}_i} \hat{\nu}_{ij} \hat{X}_i \hat{M}_F \hat{X}_j$ where $k$, $l = 1, \ldots, K_0$. Define

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

(3.7)

Let $\hat{\varepsilon}_{it} = Y_{it} - \hat{X}_i \hat{\alpha}_k - \hat{\Lambda} \hat{F}_t$ for $i \in \hat{G}_k$. Let $\hat{\Psi}_{NT} \equiv \text{diag}(\hat{\psi}_{1T}, \ldots, \hat{\psi}_{NT})$ and $\hat{\Phi}_{NT} \equiv \text{diag}(\hat{\varphi}_{1N}, \ldots, \hat{\varphi}_{TN})$ where $\hat{\psi}_{iT} = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2$ and $\hat{\varphi}_{iN} = N^{-1} \sum_{i=1}^{N} \hat{\varepsilon}_{it}^2$. For $k = 1, \ldots, K_0$, 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 $\hat{X}_i^* = \frac{1}{\hat{N}_k} \sum_{j=1}^{N} \hat{\nu}_{ij} \hat{X}_j$, and $\hat{Z}_{k,i} = \frac{N}{\hat{N}_k} [X_i - \hat{P}_F \hat{X}_i] 1 \{i \in \hat{G}_k\} - \frac{1}{\hat{N}_k} \hat{M}_F \sum_{j \in \hat{G}_k} \hat{\nu}_{ij} \hat{X}_j$. Note that we can write the $l$th element
of $B_{2,kNT}$ and $B_{3,NT}$ respectively as

$$
\begin{align*}
B_{2,kNT,l} &= -\frac{1}{N_k T^2} \text{tr} \left\{ F^0 \left( T^{-1} F^0 F^0 \right)^{-1} \left( N^{-1} \Lambda_0 \Lambda_0^0 \right)^{-1} \Lambda_0^0 \varepsilon \varepsilon^T (k) \right\}, \\
B_{3,kNT,l} &= \frac{1}{N_k T} \text{tr} \left\{ P_{F_0} E \left[ \varepsilon (k)^T X_l (k) \right] \right\},
\end{align*}
$$

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

$$
\begin{align*}
\hat{B}_{1,kNT} &= -\frac{1}{N_k T^2} \sum_{i \in G_k} X_i^T M_F \hat{\phi}_{NT} \hat{F} \left( N^{-1} \hat{\Lambda} \hat{\Lambda}' \right)^{-1} \hat{\lambda}_i, \\
\hat{B}_{2,kNT,l} &= -\frac{1}{N_k T} \text{tr} \left\{ \hat{F} \left( N^{-1} \hat{\Lambda} \hat{\Lambda}' \right)^{-1} \hat{\Lambda} \hat{\Lambda}' \hat{\phi}_{kNT} \hat{X}_l (k) - \hat{X}_l^* (k) \right\}, \\
\hat{B}_{3,kNT,l} &= \frac{1}{N_k T} \text{tr} \left\{ P_{\hat{F}} \left[ \hat{\varepsilon} (k)^T \hat{X}_l (k) \right] \right\}_\text{trunc},
\end{align*}
$$

where $A^{\text{trunc}} \equiv \sum_{t=1}^{T-M} \sum_{s=t+1}^{t+M} A_{ts}$ 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 $T \to \infty$, $\hat{\varepsilon} (k)$ denotes the $\hat{N}_k \times T$ matrix with typical elements $\hat{\varepsilon}_{it,l}$ for $i \in \hat{G}_k$ and $t = 1, ..., T$, and $\hat{Z}_{it,l}$ denotes the $l$th row of $\hat{Z}_l$. Let $\hat{B}_{s,kNT} = (\hat{B}_{s,kNT,1}, ..., \hat{B}_{s,kNT,p})'$ for $s = 2, 3$ and $k = 1, ..., K_0$. Let $\hat{B}_{s,NT} = (\hat{B}_{s,1NT}, ..., \hat{B}_{s,K_0NT})'$ for $s = 1, 2, 3$. Define the bias-corrected PPC estimator as

$$
\text{vec} \left( \hat{\alpha}^{bc} \right) = \text{vec} \left( \hat{\alpha} \right) - Q_{NT}^{-1} \left( \hat{B}_{1,NT} + \hat{B}_{2,NT} + \hat{B}_{3,NT} \right),
$$

where $\hat{\alpha}^{bc} = (\hat{\alpha}_1^{bc}, ..., \hat{\alpha}_K^{bc})$ 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) $\sqrt{NT} \text{vec} (\hat{\alpha}^{bc} - \alpha^0) \overset{D}{\to} N (0, Q_0^{-1} \Theta_0 Q_0^{-1})$, (ii) $Q_{NT} = Q_{NT} + O_p \left( \delta_{NT}^{-1} \right)$, and (iii) $\hat{\Theta}_{NT} = \Theta_{NT} + o_p (1)$.

Similarly, one can obtain a bias-corrected version for the post Lasso PC estimator $\hat{\alpha}$. The only difference is that now one relies on the post-Lasso estimates ($\hat{\alpha}_k, \hat{\lambda}_l, \hat{F}_l$) and residuals ($\hat{\varepsilon}_{it} = Y_{it} - X_{it,k} \hat{\alpha}_k - \hat{\lambda}_l \hat{F}_l$ for $i \in \hat{G}_k$) to construct estimates of $Q_{NT}, B_{1,NT}, B_{2,NT}, B_{3,NT},$ and $\Theta_{NT}$. 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 $R_{\text{max}}$ and propose a BIC-type information criterion to determine $R_0$ before determining $K_0$.

Let $\hat{\beta}_{i,R}$, $\hat{F}_{t,R}$ and $\hat{\lambda}_{i,R}$ denote the PCA estimators (without the penalization device) of $\beta_i$, $F_{t,R}$ 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 $\hat{\beta}_{(R)} = \left(\hat{\beta}_{1,R}', ..., \hat{\beta}_{N,R}'\right)'$. Define

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

(4.1)

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

$$IC_{1NT}(R) = \ln V(R, \hat{\beta}_{(R)}) + \rho_{1NT} R,$$

(4.2)

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

$$P \left( \hat{R} = R_0 \right) \to 1 \text{ as } (N, T) \to \infty.$$
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 $\rho_{1NT}$. Following Bai and Ng (2002), we can set

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

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 $K_{\text{max}}$ 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

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

where $1 \leq K \leq K_{\text{max}}$. By minimizing the above objective function, we can obtain the C-Lasso estimates \( \{ \hat{\beta}_i(K, \kappa), \hat{\alpha}_k(K, \kappa), \hat{\lambda}_i^{(K,\kappa)}, \hat{F}_i^{(K,\kappa)} \} \) of \( \{ \beta_0, \alpha_k^0, \lambda_0^0, F_0^0 \} \), where we make the dependence of \( \hat{\beta}_i, \hat{\alpha}_k, \hat{\lambda}_i, \) and \( \hat{F}_i \) on \( (K, \kappa) \) 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.,

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

Let $\hat{G}(K, \kappa) = \{ \hat{G}_1(K, \kappa), \ldots, \hat{G}_K(K, \kappa) \}$. Based on (4.4), we denote the post-Lasso estimate of \( \{ \alpha_k^0, \lambda_0^0, F_0^0 \} \) as \( \{ \hat{\alpha}_{\hat{G}_k(K, \kappa)}, \hat{\lambda}_i^{(K,\kappa)}, \hat{F}_i^{(K,\kappa)} \} \). Let \( \hat{\sigma}_{\hat{G}(K, \kappa)}^2 = \frac{1}{NT} \sum_{k=1}^{K} \sum_{i \in \hat{G}_k(K, \kappa)} \sum_{t=1}^{T} [Y_{it} - \hat{\beta}_i^{(K,\kappa)} X_{it}]^2 / \left( \hat{\lambda}_i^{(K,\kappa)} \right)^2 \). We propose to select the number of groups by choosing $K$ to minimize the following information criterion:

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

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

$$\mathcal{K}_0 = \{ K \in \mathcal{K} : K = K_0 \}, \quad \mathcal{K}_- = \{ K \in \mathcal{K} : K < K_0 \}, \quad \text{and} \quad \mathcal{K}_+ = \{ K \in \mathcal{K} : K > K_0 \}.$$
Clearly, \( K_0, K_- \) and \( K_+ \) denote the three subsets of \( K \) in which the true, under-, and over-fitted models are produced, respectively. Let \( G^{(K)} = (G_{K,1}, \ldots, G_{K,K}) \) be any \( K \)-partition of the set of individual indices \( \{1, 2, \ldots, N\} \). Let \( \mathcal{G}_K \) denote the collection of such partitions. Let

\[
\hat{\sigma}^2_{G^{(K)}} = \frac{1}{NT} \sum_{k=1}^{K} \sum_{i \in G_{k,k}} \sum_{t=1}^{T} [Y_{it} - \bar{\alpha}'_{G_{k,k}} X_{it} - \bar{\lambda}^{(K)}_{i} P_{t}^{(K)}]^2, \]

where \( \{\bar{\alpha}_{G_{k,k}}, \bar{\lambda}^{(K)}_{i}, P_{t}^{(K)}\} \) denote the estimates of \( \{\alpha^0_{k}, \lambda^0_{i}, P^0_{t}\} \) given the group structure specified in \( G^{(K)} \).

To proceed, we add the following two assumptions.

**Assumption A.7.** As \( (N, T) \to \infty \), \( \min_{1 \leq K < K_0} \inf_{G^{(K)} \in \mathcal{G}_K} \hat{\sigma}^2_{G^{(K)}} \xrightarrow{P} \sigma^2 > \sigma^2_0 \), where \( \sigma^2_0 = \text{plim}_{(N, T) \to \infty} \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in G^0_k} \sum_{t=1}^{T} (Y_{it} - \alpha^0_{k} X_{it} - \lambda^0_{i} P^0_{t})^2 \).

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

A.7 is intuitively clear. It requires that all under-fitted models yield asymptotic mean square errors that are larger than \( \sigma^2_0 \), which is delivered by the true model. A.8 reflects the usual conditions for the consistency of model selection. The penalty coefficient \( \rho_{2NT} \) 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

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

Let \( K_0(\kappa) = \arg \min_{1 \leq K \leq K_{\max}} IC_{2NT} (K, \kappa) \). As Theorem 4.2 indicates, as long as \( \kappa \) 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 \( \kappa \). In practice, it is desirable to have a data-driven method to choose \( \kappa \). For this purpose, define

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

Then we can choose the tuning parameter to be \( \hat{\kappa} = \arg \min_{\kappa \in K} IC^*_{2NT} (\kappa) \), where \( K = \{\kappa : \kappa \propto T^{-a} \text{ for any } a \in (1/4, 1/2)\} \) provided some conditions on the moments of \( \|X_{it} \| \) 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

\[
Y_{it} = \beta_0^{10} Y_{i,t-1} + \beta_1^{12} X_{it,1} + \beta_2^{12} X_{it,2} + \lambda_i^0 F_t^0 + \varepsilon_{it}.
\]

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

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

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

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 \( \lambda_i^0 \) 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

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

where \( \sigma \) is set to 2.74 so that the signal-noise ratio is 4.

**DGP2** (Dynamic panel model). The model is

\[
Y_{it} = \beta_0^{10} Y_{i,t-1} + \beta_1^{11} X_{it,1} + \beta_2^{12} X_{it,2} + \lambda_i^0 F_t^0 + \varepsilon_{it}.
\]

The common factor vector \( F_t^0 \) has 2 dimensions and follows an AR(1) process: \( F_t^0 = 0.2 + 0.5 F_{t-1}^0 + \zeta_t \), where \( \zeta_t, t = 1, \ldots, T, \) 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
\[ \varepsilon_{it} = \sigma_{it} \varepsilon_{it}, \sigma_{it} = \sigma \left[ 0.25 + 0.05(X_{it,1}^2 + X_{it,2}^2) \right]^{1/2}, \] \[ \varepsilon_{it} \sim IID N(0, 1), \]
where \( \sigma \) 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 \( \rho_{1NT} \) and \( \rho_{2NT} \) in the information criteria. Following Bai and Ng (2002) and Li et al. (2016), we set \( \rho_{1NT} = \frac{\bar{N}_T + T}{N_T} \ln (\delta_{NT}^2) \). Following SSP, we set \( \rho_{2NT} = \frac{1}{25} \ln(N(T)/\min(N, T)) \).

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 \( \kappa \) is set to be \( 0.2S_{\varepsilon}T^{-0.45} \), where \( S_{\varepsilon} \) 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%. 

22
Table 1: Frequency of selecting $R$ ($R_0 = 2$; $R = 1, 2, \ldots, 6$)

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$R = 1$</th>
<th>$R = 2$</th>
<th>$R = 3$</th>
<th>$R = 4$</th>
<th>$R = 5$</th>
<th>$R = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP1</td>
<td>100</td>
<td>50</td>
<td>0.004</td>
<td>0.996</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>150</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$K = 1$</th>
<th>$K = 2$</th>
<th>$K = 3$</th>
<th>$K = 4$</th>
<th>$K = 5$</th>
<th>$K = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP1</td>
<td>100</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0.990</td>
<td>0.010</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0</td>
<td>0.164</td>
<td>0.828</td>
<td>0.006</td>
<td>0.002</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0.990</td>
<td>0.010</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>150</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$K = 1$</th>
<th>$K = 2$</th>
<th>$K = 3$</th>
<th>$K = 4$</th>
<th>$K = 5$</th>
<th>$K = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP1</td>
<td>100</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0.990</td>
<td>0.010</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0</td>
<td>0.164</td>
<td>0.828</td>
<td>0.006</td>
<td>0.002</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0.990</td>
<td>0.010</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>150</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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:

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

(5.1)

where $\hat{\sigma}^2_{G(K, \kappa, R)}$ is defined similar to the counterpart in (4.5) except that it is now dependent on $R$. Specifying $\kappa$, $\rho_{1,NT}$, and $\rho_{NT}$ 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 $T \geq 100$. 

23
Table 3: Frequency of selecting $K$ and $R$ ($R_0 = 2$, $K_0 = 3$; $K = 1, 2, \ldots, 6$; $R = 1, 2, 3, 4$)

<table>
<thead>
<tr>
<th>DGP</th>
<th>$N$</th>
<th>$T$</th>
<th>$R$</th>
<th>$K = 1$</th>
<th>$K = 2$</th>
<th>$K = 3$</th>
<th>$K = 4$</th>
<th>$K = 5$</th>
<th>$K = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>50</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.004</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.986</td>
<td>0.010</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>50</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.052</td>
<td>0.224</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.096</td>
<td>0.620</td>
<td>0.006</td>
<td>0.002</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>150</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.034</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.956</td>
<td>0.010</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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 $\kappa$ is set to be $C_\kappa S_\kappa T^{-0.45}$, where $S_\kappa$ denotes the sample variance of $\varepsilon_{it}$ and $C_\kappa$ is a sequence of geometrically increasing constants. We try 5 values for $C_\kappa$, namely, 0.05, 0.1, 0.2, 0.4, 0.8. Regarding the initial values of $\{\hat{\beta}^{(0)}_{i} \}_{i=1}^{N}$, 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 $\{\hat{\alpha}^{(0)}_{k} \}_{k=1}^{K_0}$ are all set to the average of $\{\hat{\delta}^{(0)}_{i} \}_{i=1}^{N}$. 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 \( R = 250 \) 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

\[
\bar{P}(E) = \frac{1}{NR} \sum_{k=1}^{K_0} \sum_{i \in G_k^0} \sum_{r=1}^{\mathcal{R}} 1 \{ i \notin \hat{G}_k^{(r)} \},
\]

where \( \hat{G}_k^{(r)} \) denotes the estimated \( k \)th group in the \( r \)th replication. The behavior of the Type II classification error \( \bar{P}(\mathcal{F}) = \frac{1}{NR} \sum_{k=1}^{K_0} \sum_{i \in G_k^{(r)}} \sum_{r=1}^{\mathcal{R}} 1 \{ i \notin G_k^0 \} \) 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 \( C_\kappa \). 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 \( \text{RMSE}(\hat{\beta}_1) \) and \( \text{Bias}(\hat{\beta}_1) \) respectively as

\[
\text{RMSE}(\hat{\beta}_1) = \frac{1}{N} \sum_{k=1}^{K_0} N_k \sqrt{\frac{1}{\mathcal{R}} \sum_{r=1}^{\mathcal{R}} (\hat{\alpha}_{k1,r} - \alpha_{k1})^2},
\]

\[
\text{Bias}(\hat{\beta}_1) = \frac{1}{N} \sum_{k=1}^{K_0} N_k \left| \frac{1}{\mathcal{R}} \sum_{r=1}^{\mathcal{R}} (\hat{\alpha}_{k1,r} - \alpha_{k1}) \right|,
\]

where \( \mathcal{R} = 250 \) is the number of replications and \( \alpha_{k1} \) 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)

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

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. To save computing time, we focus on the samples with $N = 100$ and fix $C_\kappa$ 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 $\beta_{1,1}$ in DGP 1 by PPC ($K_0$ and $R_0$ are Known)

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$C_{K}$</th>
<th>0.05</th>
<th></th>
<th></th>
<th>0.1</th>
<th></th>
<th></th>
<th>0.2</th>
<th></th>
<th></th>
<th>0.4</th>
<th></th>
<th></th>
<th>0.8</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>C-Lasso</td>
<td>0.0708</td>
<td>0.0282</td>
<td>0.0739</td>
<td>0.0303</td>
<td>0.0776</td>
<td>0.0346</td>
<td>0.0909</td>
<td>0.0431</td>
<td>0.1037</td>
<td>0.0551</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.0646</td>
<td>0.0279</td>
<td>0.0661</td>
<td>0.0310</td>
<td>0.0689</td>
<td>0.0352</td>
<td>0.0788</td>
<td>0.0452</td>
<td>0.0966</td>
<td>0.0595</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0709</td>
<td>0.0283</td>
<td>0.0740</td>
<td>0.0304</td>
<td>0.0776</td>
<td>0.0347</td>
<td>0.0910</td>
<td>0.0432</td>
<td>0.1037</td>
<td>0.0552</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0647</td>
<td>0.0286</td>
<td>0.0662</td>
<td>0.0311</td>
<td>0.0690</td>
<td>0.0353</td>
<td>0.0789</td>
<td>0.0453</td>
<td>0.0967</td>
<td>0.0596</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0502</td>
<td>0.0020</td>
<td>0.0502</td>
<td>0.0020</td>
<td>0.0502</td>
<td>0.0020</td>
<td>0.0502</td>
<td>0.0020</td>
<td>0.0502</td>
<td>0.0020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>C-Lasso</td>
<td>0.0423</td>
<td>0.0165</td>
<td>0.0431</td>
<td>0.0176</td>
<td>0.0454</td>
<td>0.0199</td>
<td>0.0516</td>
<td>0.0251</td>
<td>0.0657</td>
<td>0.0403</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.0360</td>
<td>0.0073</td>
<td>0.0363</td>
<td>0.0078</td>
<td>0.0368</td>
<td>0.0093</td>
<td>0.0391</td>
<td>0.0131</td>
<td>0.0476</td>
<td>0.0246</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0424</td>
<td>0.0166</td>
<td>0.0432</td>
<td>0.0177</td>
<td>0.0455</td>
<td>0.0200</td>
<td>0.0517</td>
<td>0.0252</td>
<td>0.0658</td>
<td>0.0403</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0361</td>
<td>0.0073</td>
<td>0.0364</td>
<td>0.0079</td>
<td>0.0369</td>
<td>0.0093</td>
<td>0.0393</td>
<td>0.0131</td>
<td>0.0477</td>
<td>0.0247</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0334</td>
<td>0.0066</td>
<td>0.0334</td>
<td>0.0066</td>
<td>0.0334</td>
<td>0.0066</td>
<td>0.0334</td>
<td>0.0066</td>
<td>0.0334</td>
<td>0.0066</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>C-Lasso</td>
<td>0.0556</td>
<td>0.0295</td>
<td>0.0569</td>
<td>0.0313</td>
<td>0.0610</td>
<td>0.0352</td>
<td>0.0716</td>
<td>0.0420</td>
<td>0.0871</td>
<td>0.0497</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.0471</td>
<td>0.0263</td>
<td>0.0492</td>
<td>0.0292</td>
<td>0.0542</td>
<td>0.0356</td>
<td>0.0651</td>
<td>0.0457</td>
<td>0.0850</td>
<td>0.0603</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0556</td>
<td>0.0296</td>
<td>0.0569</td>
<td>0.0314</td>
<td>0.0611</td>
<td>0.0352</td>
<td>0.0717</td>
<td>0.0421</td>
<td>0.0872</td>
<td>0.0498</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0471</td>
<td>0.0264</td>
<td>0.0493</td>
<td>0.0293</td>
<td>0.0543</td>
<td>0.0357</td>
<td>0.0652</td>
<td>0.0458</td>
<td>0.0850</td>
<td>0.0603</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0340</td>
<td>0.0018</td>
<td>0.0340</td>
<td>0.0018</td>
<td>0.0340</td>
<td>0.0018</td>
<td>0.0340</td>
<td>0.0018</td>
<td>0.0340</td>
<td>0.0018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>C-Lasso</td>
<td>0.0337</td>
<td>0.0161</td>
<td>0.0346</td>
<td>0.0173</td>
<td>0.0369</td>
<td>0.0198</td>
<td>0.0429</td>
<td>0.0257</td>
<td>0.0590</td>
<td>0.0432</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.0249</td>
<td>0.0064</td>
<td>0.0252</td>
<td>0.0073</td>
<td>0.0261</td>
<td>0.0095</td>
<td>0.0285</td>
<td>0.0136</td>
<td>0.0382</td>
<td>0.0257</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0337</td>
<td>0.0161</td>
<td>0.0346</td>
<td>0.0173</td>
<td>0.0369</td>
<td>0.0198</td>
<td>0.0429</td>
<td>0.0258</td>
<td>0.0590</td>
<td>0.0432</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0249</td>
<td>0.0064</td>
<td>0.0252</td>
<td>0.0074</td>
<td>0.0261</td>
<td>0.0095</td>
<td>0.0285</td>
<td>0.0136</td>
<td>0.0382</td>
<td>0.0257</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0225</td>
<td>-0.0002</td>
<td>0.0225</td>
<td>-0.0002</td>
<td>0.0225</td>
<td>-0.0002</td>
<td>0.0225</td>
<td>-0.0002</td>
<td>0.0225</td>
<td>-0.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>150</td>
<td>C-Lasso</td>
<td>0.0275</td>
<td>0.0118</td>
<td>0.0280</td>
<td>0.0126</td>
<td>0.0292</td>
<td>0.0142</td>
<td>0.0344</td>
<td>0.0200</td>
<td>0.0441</td>
<td>0.0317</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.0196</td>
<td>0.0013</td>
<td>0.0197</td>
<td>0.0015</td>
<td>0.0198</td>
<td>0.0021</td>
<td>0.0204</td>
<td>0.0037</td>
<td>0.0226</td>
<td>0.0079</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0275</td>
<td>0.0118</td>
<td>0.0280</td>
<td>0.0126</td>
<td>0.0292</td>
<td>0.0142</td>
<td>0.0344</td>
<td>0.0200</td>
<td>0.0441</td>
<td>0.0317</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0196</td>
<td>0.0013</td>
<td>0.0197</td>
<td>0.0016</td>
<td>0.0198</td>
<td>0.0021</td>
<td>0.0204</td>
<td>0.0037</td>
<td>0.0226</td>
<td>0.0079</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0192</td>
<td>-0.0003</td>
<td>0.0192</td>
<td>-0.0003</td>
<td>0.0192</td>
<td>-0.0003</td>
<td>0.0192</td>
<td>-0.0003</td>
<td>0.0192</td>
<td>-0.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6: Estimation of $\beta_{1,1}$ in DGP 2 by PPC ($K_0$ and $R_0$ are Known)

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$C_\kappa$</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
<th>0.4</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>C-Lasso</td>
<td>0.0924</td>
<td>0.0390</td>
<td>0.0934</td>
<td>0.0406</td>
<td>0.0956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.0914</td>
<td>0.0517</td>
<td>0.0950</td>
<td>0.0553</td>
<td>0.1002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0857</td>
<td>-0.0016</td>
<td>0.0871</td>
<td>0.0005</td>
<td>0.0895</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0805</td>
<td>0.0138</td>
<td>0.0829</td>
<td>0.0176</td>
<td>0.0872</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0650</td>
<td>0.0116</td>
<td>0.0658</td>
<td>0.0116</td>
<td>0.0650</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>C-Lasso</td>
<td>0.0620</td>
<td>0.0249</td>
<td>0.0635</td>
<td>0.0269</td>
<td>0.0675</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.0523</td>
<td>0.0198</td>
<td>0.0529</td>
<td>0.0218</td>
<td>0.0563</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0580</td>
<td>0.0014</td>
<td>0.0596</td>
<td>0.0034</td>
<td>0.0636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0496</td>
<td>-0.0028</td>
<td>0.0494</td>
<td>-0.0008</td>
<td>0.0510</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0447</td>
<td>0.0041</td>
<td>0.0447</td>
<td>0.0041</td>
<td>0.0447</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>C-Lasso</td>
<td>0.0376</td>
<td>0.0048</td>
<td>0.0379</td>
<td>0.0357</td>
<td>0.0745</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.0687</td>
<td>0.0450</td>
<td>0.0709</td>
<td>0.0469</td>
<td>0.0759</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0674</td>
<td>-0.0041</td>
<td>0.0671</td>
<td>-0.0031</td>
<td>0.0695</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0558</td>
<td>0.0082</td>
<td>0.0573</td>
<td>0.0101</td>
<td>0.0614</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0433</td>
<td>0.0042</td>
<td>0.0433</td>
<td>0.0042</td>
<td>0.0433</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>C-Lasso</td>
<td>0.0375</td>
<td>0.0191</td>
<td>0.0391</td>
<td>0.0210</td>
<td>0.0428</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.0276</td>
<td>0.0077</td>
<td>0.0281</td>
<td>0.0092</td>
<td>0.0293</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0342</td>
<td>0.0036</td>
<td>0.0357</td>
<td>0.0055</td>
<td>0.0394</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0281</td>
<td>0.0075</td>
<td>0.0279</td>
<td>-0.0059</td>
<td>0.0278</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0249</td>
<td>0.0012</td>
<td>0.0249</td>
<td>0.0012</td>
<td>0.0249</td>
</tr>
</tbody>
</table>

28
Table 7: Estimation of $\beta_{1,1}$ in DGP 1
(Misspecification of $K$ and/or $R$; $C_{a} = 0.2$; $(R_{0}, K_{0}) = (2, 3)$)

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

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$K$</th>
<th>$R$</th>
<th>RMSE</th>
<th>Bias</th>
<th>RMSE</th>
<th>Bias</th>
<th>RMSE</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>50</td>
<td>3</td>
<td>2</td>
<td>0.1843</td>
<td>0.1071</td>
<td>0.1378</td>
<td>-0.0195</td>
<td>0.1418</td>
<td>-0.0031</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1853</td>
<td>0.1093</td>
<td>0.1370</td>
<td>-0.0047</td>
<td>0.1393</td>
<td>0.0048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1843</td>
<td>0.1072</td>
<td>0.1377</td>
<td>-0.0194</td>
<td>0.1407</td>
<td>-0.0030</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1854</td>
<td>0.1093</td>
<td>0.1370</td>
<td>-0.0046</td>
<td>0.1383</td>
<td>0.0048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1649</td>
<td>0.1480</td>
<td>0.0502</td>
<td>0.0020</td>
<td>0.0538</td>
<td>0.0025</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>4</td>
<td>2</td>
<td>0.1368</td>
<td>0.0808</td>
<td>0.0938</td>
<td>-0.0464</td>
<td>0.0981</td>
<td>-0.0443</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1408</td>
<td>0.1040</td>
<td>0.0738</td>
<td>-0.0104</td>
<td>0.0782</td>
<td>-0.0106</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1368</td>
<td>0.0808</td>
<td>0.0938</td>
<td>-0.0463</td>
<td>0.0972</td>
<td>-0.0441</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1409</td>
<td>0.1040</td>
<td>0.0738</td>
<td>-0.0104</td>
<td>0.0777</td>
<td>-0.0107</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1588</td>
<td>0.1491</td>
<td>0.0334</td>
<td>0.0006</td>
<td>0.0358</td>
<td>0.0008</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>4</td>
<td>2</td>
<td>0.1257</td>
<td>0.0900</td>
<td>0.0714</td>
<td>-0.0303</td>
<td>0.0740</td>
<td>-0.0337</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1351</td>
<td>0.1198</td>
<td>0.0486</td>
<td>0.0019</td>
<td>0.0512</td>
<td>-0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1257</td>
<td>0.0900</td>
<td>0.0714</td>
<td>-0.0303</td>
<td>0.0731</td>
<td>-0.0335</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1351</td>
<td>0.1198</td>
<td>0.0486</td>
<td>0.0019</td>
<td>0.0508</td>
<td>-0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1556</td>
<td>0.1479</td>
<td>0.0266</td>
<td>0.0014</td>
<td>0.0290</td>
<td>0.0019</td>
</tr>
</tbody>
</table>

29
Table 8: Estimation of $\beta_{1,1}$ in DGP 2
(Misspecification of $K$ and/or $R$: $C_n = 0.2$; $(R_0, K_0) = (2, 3)$)

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$K = 3$</th>
<th>$R = 1$</th>
<th>$R = 2$</th>
<th>$R = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>C-Lasso</td>
<td>0.1856</td>
<td>0.1634</td>
<td>0.0956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.2007</td>
<td>0.1803</td>
<td>0.1002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.1720</td>
<td>0.1470</td>
<td>0.0895</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.1883</td>
<td>0.1655</td>
<td>0.0872</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.1526</td>
<td>0.1354</td>
<td>0.0650</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>C-Lasso</td>
<td>0.1617</td>
<td>0.1487</td>
<td>0.0675</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.1670</td>
<td>0.1555</td>
<td>0.0563</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.1530</td>
<td>0.1388</td>
<td>0.0636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.1588</td>
<td>0.1463</td>
<td>0.0510</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.1418</td>
<td>0.1309</td>
<td>0.0447</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>C-Lasso</td>
<td>0.1532</td>
<td>0.1446</td>
<td>0.0510</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.1544</td>
<td>0.1465</td>
<td>0.0393</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.1474</td>
<td>0.1382</td>
<td>0.0477</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.1489</td>
<td>0.1405</td>
<td>0.0379</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.1399</td>
<td>0.1324</td>
<td>0.0353</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$K = 4$</th>
<th>$R = 1$</th>
<th>$R = 2$</th>
<th>$R = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>C-Lasso</td>
<td>0.1680</td>
<td>0.1038</td>
<td>0.1306</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.1801</td>
<td>0.1043</td>
<td>0.1529</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.1683</td>
<td>0.1035</td>
<td>0.1311</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.1805</td>
<td>0.1040</td>
<td>0.1531</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.1526</td>
<td>0.1354</td>
<td>0.0650</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>C-Lasso</td>
<td>0.1574</td>
<td>0.0903</td>
<td>0.1123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.1580</td>
<td>0.1099</td>
<td>0.1030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.1576</td>
<td>0.0904</td>
<td>0.1123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.1583</td>
<td>0.1099</td>
<td>0.1031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.1418</td>
<td>0.1309</td>
<td>0.0447</td>
</tr>
<tr>
<td>100</td>
<td>150</td>
<td>C-Lasso</td>
<td>0.1288</td>
<td>0.0713</td>
<td>0.0956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso</td>
<td>0.1332</td>
<td>0.1045</td>
<td>0.0747</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.1288</td>
<td>0.0714</td>
<td>0.0956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.1333</td>
<td>0.1044</td>
<td>0.0749</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.1399</td>
<td>0.1324</td>
<td>0.0353</td>
</tr>
</tbody>
</table>

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

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

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 \( \beta_i \) 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 \( \varepsilon_{it} \) 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 \( \kappa \), 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 \( C_\kappa = 0.2 \), 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, Qinhuangdao, Dandong, Jinhzhou, Wuxi, Xuzhou, Wenzhou, Ganzhou, Yantai, Jining, Luoyang, Yueyang, Shaoguan, Beihai, Sanya, Luzhou, Nanchong, Dali;

\(^4\)The classification is not sensitive to the choice of \( C_\kappa \).
- 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 $\beta_1$. The housing prices in the first group are most persistent, with $\beta_1$ 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>
<thead>
<tr>
<th>$\mathcal{IC}_{1NT}(R)$</th>
<th>R=2</th>
<th>R=3</th>
<th>R=4</th>
<th>R=5</th>
<th>R=6</th>
<th>R=7</th>
<th>R=8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.8295</td>
<td>4.8012</td>
<td>4.7895</td>
<td><strong>4.7809</strong></td>
<td>4.7859</td>
<td>4.7949</td>
<td>4.8002</td>
</tr>
</tbody>
</table>

Table 10: Information criteria for $K_0$ (Method I)

<table>
<thead>
<tr>
<th>$C_\kappa$</th>
<th>K=1</th>
<th>K=2</th>
<th>K=3</th>
<th>K=4</th>
<th>K=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>4.3342</td>
<td>4.3098</td>
<td><strong>4.2977</strong></td>
<td>4.3100</td>
<td>4.3082</td>
</tr>
<tr>
<td>0.1</td>
<td>4.4146</td>
<td>4.3098</td>
<td><strong>4.2977</strong></td>
<td>4.3100</td>
<td>4.3082</td>
</tr>
<tr>
<td>0.2</td>
<td>4.4146</td>
<td>4.3098</td>
<td><strong>4.2977</strong></td>
<td>4.3100</td>
<td>4.3082</td>
</tr>
<tr>
<td>0.4</td>
<td>4.4146</td>
<td>4.3098</td>
<td><strong>4.2966</strong></td>
<td>4.3100</td>
<td>4.3082</td>
</tr>
<tr>
<td>0.8</td>
<td>4.4146</td>
<td>4.3098</td>
<td><strong>4.2966</strong></td>
<td>4.3100</td>
<td>4.3082</td>
</tr>
</tbody>
</table>
### Table 11: Information criteria for \((R_0, K_0)\) (Method II)

<table>
<thead>
<tr>
<th>K=</th>
<th>R=2</th>
<th>R=3</th>
<th>R=4</th>
<th>R=5</th>
<th>R=6</th>
<th>K=</th>
<th>R=2</th>
<th>R=3</th>
<th>R=4</th>
<th>R=5</th>
<th>R=6</th>
</tr>
</thead>
</table>

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

<table>
<thead>
<tr>
<th>Group#</th>
<th>( \beta_1 )</th>
<th>t-value</th>
<th>( \beta_1 )</th>
<th>t-value</th>
<th>( \beta_1 )</th>
<th>t-value</th>
<th>( \beta_1 )</th>
<th>t-value</th>
<th>( \beta_1 )</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-Lasso bias-corrected estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.5115</td>
<td>13.9621</td>
<td>0.5116</td>
<td>13.9658</td>
<td>0.5118</td>
<td>13.9711</td>
<td>0.5108</td>
<td>14.8863</td>
<td>0.5114</td>
<td>15.3179</td>
</tr>
<tr>
<td>2</td>
<td>0.3134</td>
<td>8.6061</td>
<td>0.3132</td>
<td>8.5996</td>
<td>0.3128</td>
<td>8.5866</td>
<td>0.3052</td>
<td>8.0482</td>
<td>0.3039</td>
<td>8.0280</td>
</tr>
<tr>
<td>3</td>
<td>-0.0065</td>
<td>-0.1558</td>
<td>-0.0061</td>
<td>-0.1478</td>
<td>-0.0065</td>
<td>-0.1569</td>
<td>-0.0064</td>
<td>-0.1538</td>
<td>-0.0064</td>
<td>-0.1455</td>
</tr>
<tr>
<td>Post-Lasso bias-corrected estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.5563</td>
<td>15.0382</td>
<td>0.5563</td>
<td>15.0382</td>
<td>0.5563</td>
<td>15.0382</td>
<td>0.5464</td>
<td>15.7838</td>
<td>0.5314</td>
<td>15.8928</td>
</tr>
<tr>
<td>2</td>
<td>0.2697</td>
<td>7.2744</td>
<td>0.2697</td>
<td>7.2744</td>
<td>0.2697</td>
<td>7.2744</td>
<td>0.2625</td>
<td>6.8102</td>
<td>0.2778</td>
<td>7.2771</td>
</tr>
<tr>
<td>3</td>
<td>0.0074</td>
<td>0.1793</td>
<td>0.0074</td>
<td>0.1793</td>
<td>0.0074</td>
<td>0.1793</td>
<td>0.0075</td>
<td>0.1804</td>
<td>-0.0371</td>
<td>-0.8577</td>
</tr>
</tbody>
</table>

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


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


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


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


Supplemental Material for

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

(Not for Publication)

Liangjun Su\textsuperscript{a} and Gaosheng Ju\textsuperscript{b}

\textsuperscript{a} School of Economics, Singapore Management University
\textsuperscript{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

\[
P_\hat{\beta} - P_{F_0H} = T^{-1} \hat{\beta} \hat{\beta}' - F_0^0 H (H' F_0^0 F_0^0 H)^{-1} H' F_0^0
\]

\[
= T^{-1} (\hat{\beta} - F_0^0 H) (\hat{\beta} - F_0^0 H)' - T^{-1} (\hat{\beta} - F_0^0 H) H' F_0^0
\]

\[
+ T^{-1} F_0^0 H (\hat{\beta} - F_0^0 H)' + T^{-1} F_0^0 H (I_{R_0} - (T^{-1} H' F_0^0 F_0^0 H)^{-1})' H' F_0^0
\]

\[
= d_1 + d_2 + d_3 + d_4, \text{ say, (A.1)}
\]

where \( H = (N^{-1} \Lambda_0^0 \Lambda_0^0) (T^{-1} F_0^0 \hat{\beta}) V_{N,T}^{-1} \). Let \( G = V_{N,T}^{-1} H^{-1} = (T^{-1} F_0^0 \hat{\beta})^{-1} (N^{-1} \Lambda_0^0 \Lambda_0^0)^{-1} \), \( \mu_{ts} = F_i^0 \times (T^{-1} F_0^0 F_0^0)^{-1} F_s^0 \) and \( \nu_{ij} = \lambda_j^0 (N^{-1} \Lambda_0^0 \Lambda_0^0)^{-1} \lambda_i^0 \). Let \( \eta_{N,T} = \frac{1}{N} \sum_{i=1}^{N} ||b_i||^2 \). Let \( \omega_p, \bar{\omega}_p \in \mathbb{R}^p \) 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) \( \sup_{N^{-1/2} \leq b \leq L} \sup_{F \in \mathcal{F}} \left| \frac{1}{N^2 T} \sum_{i=1}^{N} b_i X_i' M F \varepsilon_i \right| = o_P (1) \),

(ii) \( \sup_{F \in \mathcal{F}} \left| \frac{1}{N^2 T} \sum_{i=1}^{N} \lambda_i^0 F_0^0 M F \varepsilon_i \right| = o_P (1) \),

(iii) \( \sup_{F \in \mathcal{F}} \left| \frac{1}{N^2 T} \sum_{i=1}^{N} \varepsilon_i' F \varepsilon_i \right| = o_P (1) \),

(iv) \( \frac{1}{N^2 T} \sum_{i=1}^{N} \varepsilon_i' P F \varepsilon_i = o_P (T^{-1}) \),

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

\footnote{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 \( o(\text{min}(N^{1/4}, T^{1/4})) \).}
Proof of Theorem 3.1. Let \( Q_{1NT,i} (\beta_i, F) = \frac{1}{T} (Y_i - X_i \beta_i)' M_F (Y_i - X_i \beta_i) \) and \( Q_{1NT,k}^{(K_k)} (\beta_i, \alpha, F) = Q_{1NT,i} (\beta_i, F) + \kappa \Pi_{k=1}^{K_k} \| \beta_i - \alpha_k \| \). Note that \( Q_{1NT,k}^{(K_k)} (\beta, \alpha, F) = \frac{1}{N} \sum_{i=1}^{N} Q_{1NT,k}^{(K_k)} (\beta_i, \alpha, F) \). We prove the theorem under Assumptions A.1(i)-(x) by assuming that the minimizer \( \hat{\beta} \) for \( \beta \) can only occur in the L-neighborhood of \( \beta^0 = (\beta^0_1, ..., \beta^0_N) : N_L (\beta^0) = \{ \beta \in \mathbb{R}^{p \times N} : \| \beta - \beta^0 \| \leq L \} \), 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 \( N^{-1/2} \| \hat{\beta} - \beta^0 \| \) is divergent.

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

\[
Q_{1NT,i} (\beta_i, F) = Q_{1NT,i} (\beta^0_i, F^0)
\]

where

\[
S_{NT,i} (\beta_i, F) = \frac{1}{T} b_i' X_i' M_F \varepsilon_i + \frac{2}{T} \lambda_0' F^{0'} M_F \varepsilon_i - \frac{1}{T} \varepsilon_i' (P_F - P_F^0) \varepsilon_i,
\]

and

\[
\left[ Q_{1NT,i} (\beta_i, F) - Q_{1NT,i} (\beta^0_i, F^0) \right] = \frac{1}{N} \sum_{i=1}^{N} [Q_{1NT,i} (\beta_i, F) - Q_{1NT,i} (\beta^0_i, F^0)]
\]

Let \( A = T^{-1} \text{diag}(X'_1 M_F X_1, ..., X'_N M_F X_N) \), \( B = (\Lambda_0' \Lambda_0) \otimes I_T \), and \( C = T^{-1/2} ((\lambda_0^0 \otimes M_F X_1), ..., (\lambda_N^0 \otimes M_F X_N)) \). Let \( \eta = \frac{1}{1+\tau} \text{vec}(M_F F^0) \). Then

\[
\frac{1}{N} \sum_{i=1}^{N} S_{NT,i} (\beta_i, F) = \frac{1}{N} \hat{b}' A \hat{b} + \text{tr} \left[ \left( \frac{1}{T^{1/2}} M_F F^0 \right) \left( \frac{1}{N} \Lambda_0' \Lambda_0 \right) \left( \frac{1}{T^{1/2}} F^{0'} M_F \right) \right] - \frac{2}{NT} \sum_{i=1}^{N} \text{tr} (M_F F^0 \lambda_0^0 b_i' X_i' M_F)
\]

Let \( \vartheta = \eta - B^{-1} C b \), the second equality follows from the fact that

\[
\text{tr} (B_1 B_2 B_3) = \text{vec} (B_1)' (B_2 \otimes I) \text{vec} (B_3)' \quad \text{and} \quad \text{tr} (B_1 B_2 B_3 B_4) = \text{vec} (B_1)' (B_2 \otimes B_3) \text{vec} (B_4) \]

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 \( \beta \) such that \( \beta \in N_L (\beta^0) \) and \( N^{-1} b' b = O (1) \). Observe that \( N^{-1} b' C' B^{-1} C b \leq \mu_{\max} (C' B^{-1} C) N^{-1} b' b = o_p (1) \) for any \( N^{-1} b' b = O (1) \) provided
$\mu_{\text{max}}(C'B^{-1}C) = o_P(1)$. Note that $\mu_{\text{max}}(C'B^{-1}C) \leq [\mu_{\text{min}}(N^{-1}B)]^{-1} \mu_{\text{max}}(N^{-1}C'C) = \zeta_A^{-1} \mu_{\text{max}}(N^{-1} \times C'C)$, where $\zeta_A \equiv \mu_{\text{min}}(A^0A^0/N)$. Define the upper block-triangular matrix

$$C_1 = T^{-1} \begin{pmatrix} \lambda_1^0 \lambda_1^0 X'_1 M_F X_1 & \lambda_1^0 \lambda_2^0 X'_1 M_F X_2 & \cdots & \lambda_1^0 \lambda_N^0 X'_1 M_F X_N \\ 0 & \lambda_2^0 \lambda_2^0 X'_2 M_F X_2 & \cdots & \lambda_2^0 \lambda_N^0 X'_2 M_F X_N \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N^0 \lambda_N^0 X'_N M_F X_N \end{pmatrix}.$$ 

Noting that the $Np \times Np$ matrix $C'C$ has a typical $p \times p$ block submatrix $T^{-1} \lambda_i^0 \lambda_i^0 X'_i M_F X_j$, we have $C'C = C_1 + C_1' - C_d$ where $C_d = T^{-1} \text{diag}(\lambda_1^0 \lambda_1^0 X'_1 M_F X_1, ..., \lambda_N^0 \lambda_N^0 X'_N M_F X_N)$. 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

$$N^{-1} \mu_{\text{max}}(C'C) \leq N^{-1} \{2\mu_{\text{max}}(C_1) - \mu_{\text{min}}(C_d)\} \leq 2N^{-1} \max_{1 \leq i \leq N} \|\lambda^0_i\|^2 \mu_{\text{max}}(T^{-1}X'_i M_F X_i) \leq 2N^{-1} \max_{1 \leq i \leq N} \|\lambda^0_i\|^2 \max_{1 \leq i \leq N} T^{-1} \|X_i\|^2 = O_P(N^{-1} o_P(N^{1/4}) = o_P(N^{-3/4}), \quad (A.5)$$

where the third inequality follows from the fact that $\mu_{\text{max}}(T^{-1}X'_i M_F X_i) \leq \mu_{\text{max}}(T^{-1}X'_i X_i) = T^{-1} \|X_i\|^2_{\text{sp}} \leq T^{-1} \|X_i\|^2$ as $\mu_{\text{max}}(M_F) = 1$ and the first equality follows from the fact that $\max_{1 \leq i \leq N} \|\lambda^0_i\|^2 = O_P(N^{1/4})$ under Assumption A.1(ii) by the Markov inequality. It follows that $\mu_{\text{max}}(C'B^{-1}C) = o_P(N^{-3/4})$ and $N^{-1}b'C'B^{-1}Cb = o_P(1)$ uniformly in $b$ such that $N^{-1}b'b = O(1)$. This, in conjunction with (A.2)-(A.4) and the fact that $Q^{(K_0)}_{1^{\text{NT}},h}(\hat{\theta}, \hat{\alpha}, \hat{F}) - Q^{(K_0)}_{1^{\text{NT}},h}(\beta^0, \alpha^0, F^0) \leq 0$ implies that $N^{-1} \hat{b}' \hat{A} \hat{b} = o_P(1)$, 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

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

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

$$0 \geq Q^{(K_0)}_{1^{\text{NT}},h}(\hat{\theta}, \hat{\alpha}, \hat{F}) - Q^{(K_0)}_{1^{\text{NT}},h}(\beta^0, \alpha^0, F^0) = \frac{1}{N} \sum_{i=1}^N S_{NT,i}(\hat{\beta}_i, \hat{F}) + o_P(1)$$

$$= \frac{1}{N} \hat{b}' \hat{A} \hat{b} + \frac{1}{N \text{Tr}} \left[ (F^0 M_F F^0) (\Lambda^0 \Lambda^0) \right] - \frac{2}{N \text{Tr}} \sum_{i=1}^N \hat{b}'_i X'_i M_F F^0 \Lambda^0_i + o_P(1)$$

$$\geq \frac{1}{N} \hat{b}' \hat{A} \hat{b} + \frac{1}{N \text{Tr}} \left[ (F^0 M_F F^0) (\Lambda^0 \Lambda^0) \right] - 2 \left\{ \frac{1}{N} \hat{b}' \hat{A} \hat{b} \right\}^{1/2} \left\{ \frac{1}{N \text{Tr}} \left[ (F^0 M_F F^0) (\Lambda^0 \Lambda^0) \right] \right\}^{1/2} + o_P(1)$$

$$= o_P(1) + \frac{1}{N \text{Tr}} \left[ (F^0 M_F F^0) (\Lambda^0 \Lambda^0) \right] - 2o_P(1) \left\{ \frac{1}{N \text{Tr}} \left[ (F^0 M_F F^0) (\Lambda^0 \Lambda^0) \right] \right\}^{1/2}.$$

It follows that $\frac{1}{N \text{Tr}} \left[ (F^0 M_F F^0) (\Lambda^0 \Lambda^0) \right] = o_P(1)$. As in Bai (2009, p.1265), this further implies that $\frac{1}{N \text{Tr}} F^0 F^0 = o_P(1)$ under Assumption A.1(ii), $\frac{1}{N \text{Tr}} F^0 = \text{invertible}$, and $\|P_F - P_{F_0}\| = o_P(1)$. 

3
(iii) By Assumption A.1(iii) and the result in part (ii),

$$\left| N^{-1} \sum_{i=1}^{N} \left( \hat{\beta}_i - \beta_0 \right)' \left( T^{-1} X_i' (M_{F^0} - M_{\hat{F}}) X_i \right) \left( \hat{\beta}_i - \beta_0 \right) \right|$$

$$\leq \max_{1 \leq i \leq N} T^{-1} \| X_i \| \| P_{\hat{F}} - P_{F^0} \| \sum_{i=1}^{N} \left( \hat{\beta}_i - \beta_0 \right)' \left( \hat{\beta}_i - \beta_0 \right)$$

$$\leq \max_{1 \leq i \leq N} T^{-1} \| X_i \| \| P_{\hat{F}} - P_{F^0} \| \sum_{i=1}^{N} \left\| \hat{\beta}_i - \beta_0 \right\|^2$$

$$= o_P (1) \sum_{i=1}^{N} \left\| \hat{\beta}_i - \beta_0 \right\|^2.$$

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

$$o_P (1) = N^{-1} \sum_{i=1}^{N} \left( \hat{\beta}_i - \beta_0 \right)' \left( T^{-1} X_i' (M_{F^0} X_i) \right) \left( \hat{\beta}_i - \beta_0 \right)$$

$$= N^{-1} \sum_{i=1}^{N} \left( \hat{\beta}_i - \beta_0 \right)' \left( T^{-1} X_i' (M_{F^0} X_i) \right) \left( \hat{\beta}_i - \beta_0 \right)$$

$$- N^{-1} \sum_{i=1}^{N} \left( \hat{\beta}_i - \beta_0 \right)' \left( T^{-1} X_i' (M_{F^0} - M_{\hat{F}}) X_i \right) \left( \hat{\beta}_i - \beta_0 \right)$$

$$\geq \left[ \min_{1 \leq i \leq N} \mu_{\min} \left( T^{-1} X_i' (M_{F^0} X_i) \right) - o_P (1) \right] N^{-1} \sum_{i=1}^{N} \left\| \hat{\beta}_i - \beta_0 \right\|^2.$$

Then by Assumption A.1(x), $N^{-1} \sum_{i=1}^{N} \left\| \hat{\beta}_i - \beta_0 \right\|^2 = o_P (1).$

Now, we argue that Assumption A.1(xi) rules out the possibility that $N^{-1/2} \left\| \hat{\beta} - \beta^0 \right\|$ is divergent so that it is sufficient to consider minimization of $Q_{1NT,\kappa}^{(K_0)}$ over $\beta \in \mathcal{N}_L (\beta^0)$. In the following analysis, we will show that $N^{-1} \left\| \hat{\beta} - \beta^0 \right\|^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{b}_i \hat{b}_i = o_P (1)$ without restricting $\beta \in \mathcal{N}_L (\beta^0)$ 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 $\hat{\alpha}$ are the estimates of $F^0$ and $\alpha^0$, respectively. By (A.2)-(A.3) and Lemmas A.1(ii)-(iv)

$$Q_{1NT,\kappa}^{(K_0)} \left( \beta, \hat{\alpha}, \hat{F} \right) - Q_{1NT,\kappa}^{(K_0)} \left( \beta^0, \alpha^0, F^0 \right)$$

$$\geq \frac{1}{N} \sum_{i \in \mathcal{S}_{N,F,\kappa}} \left[ Q_{1NT,i} \left( \beta_i, \hat{F} \right) - Q_{1NT,i} \left( \beta^0_i, F^0 \right) \right] + \frac{1}{N} \sum_{i \in \mathcal{S}_{N,F,\kappa}} \left[ Q_{1NT,i} \left( \beta_i, \hat{F} \right) - Q_{1NT,i} \left( \beta^0_i, F^0 \right) \right]$$

$$\geq \frac{1}{N} \sum_{i \in \mathcal{S}_{N,F,\kappa}} \left[ Q_{1NT,i} \left( \beta_i, \hat{F} \right) - Q_{1NT,i} \left( \beta^0_i, F^0 \right) \right] - \frac{1}{N} \sum_{i \in \mathcal{S}_{N,F,\kappa}} \left[ Q_{1NT,i} \left( \beta^0_i, F^0 \right) \right]$$

$$\geq \frac{1}{N} \sum_{i \in \mathcal{S}_{N,F,\kappa}} \left( S_{NT,i}(\beta_i, \hat{F}) - \frac{2}{T} b_i X_i' M_{\hat{F}} \hat{e}_i \right).$$
where the last inequality follows because by Assumption A.1(xi)

\[
\frac{1}{N} \sum_{i \in \mathcal{S}_{N,F,\xi}} Q_{1NT,i}^{(K_0)} \left( \beta_{i,1}^0, F^0 \right) = \frac{1}{N} \sum_{i \in \mathcal{S}_{N,F,\xi}} \frac{\varepsilon_i^j M_{F^0 \xi}^i}{T} \leq \frac{\# \mathcal{S}_{N,F,\xi}^*}{N} \max_{1 \leq i \leq N} \frac{\varepsilon_i^j M_{F^0 \xi}^i}{T} = o_P(1).
\]

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

\[
0 \geq \frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \tilde{b}_i^j X_i' M_{\hat{F}} X_i \hat{b}_i - \frac{2}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \hat{b}_i^j X_i' M_{\hat{F}} \left( F^0 \lambda_i^0 + \varepsilon_i \right) + o_P(1) \]

\[
+ \frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \lambda_i^0 F^{0 \prime} M_{\hat{F}} F^0 \lambda_i^0 \geq \frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \tilde{b}_i^j X_i' M_{\hat{F}} X_i \hat{b}_i - \frac{2}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \hat{b}_i^j X_i' M_{\hat{F}} \left( F^0 \lambda_i^0 + \varepsilon_i \right) + o_P(1) \]

\[
\geq \frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \tilde{b}_i^j X_i' M_{\hat{F}} X_i \hat{b}_i - \frac{2}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \hat{b}_i^j X_i' M_{\hat{F}} \left( F^0 \lambda_i^0 + \varepsilon_i \right) + o_P(1) \]

\[
\geq \left\{ \frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \tilde{b}_i^j X_i' M_{\hat{F}} X_i \hat{b}_i \right\}^{1/2} \left\{ \frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \hat{b}_i^j X_i' M_{\hat{F}} \left( F^0 \lambda_i^0 + \varepsilon_i \right) \right\}^{1/2} + o_P(1) \]

where \(\Gamma_{NT} = \frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \left( F^0 \lambda_i^0 + \varepsilon_i \right) \left( F^0 \lambda_i^0 + \varepsilon_i \right) = o_P(1)\). It follows that \(\frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \tilde{b}_i^j X_i' M_{\hat{F}} X_i \hat{b}_i = O_P(1)\), which further implies that \(\frac{1}{N} \sum_{i \in \mathcal{S}_{N,F,\xi}} \tilde{b}_i \tilde{b}_i = O_P(1)\) under Assumption A.1(xi). But when \(\frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \tilde{b}_i \tilde{b}_i = O_P(1)\), we have \(\frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \tilde{b}_i \tilde{b}_i = O_P(1)\) 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

\[
0 \geq \frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \tilde{b}_i^j X_i' M_{\hat{F}} X_i \hat{b}_i - \frac{2}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \hat{b}_i^j X_i' M_{\hat{F}} \left( F^0 \lambda_i^0 \right) + \frac{1}{NT} \sum_{i \in \mathcal{S}_{N,F,\xi}} \lambda_i^0 F^{0 \prime} M_{\hat{F}} F^0 \lambda_i^0 + o_P(1) \]

\[
= N^{-1} \tilde{b}^\prime \tilde{A} \hat{b} + N^{-1} \tilde{b}^\prime \hat{B} \hat{b} - 2N^{-1} \tilde{b}^\prime \hat{C} \hat{b} \tag{A.7}
\]

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

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

\[
0 \geq Q_{1NT,\kappa}^{(K_0)} \left( \hat{\beta}, \hat{\alpha}, \hat{F} \right) - Q_{1NT,\kappa}^{(K_0)} \left( \beta^0, \alpha^0, F^0 \right) \]

\[
= \frac{1}{N} \sum_{i = 1}^{N} \left[ Q_{1NT,i} \left( \hat{\beta}_i, \hat{F} \right) - Q_{1NT,i} \left( \beta_i^0, F_i^0 \right) \right] \]

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

This, in conjunction with the fact that \(\frac{1}{NT} \sum_{i=1}^{N} \lambda_i^0 F^0 M_{F^0} F^0 \lambda_i^0 = o_P(1)\) and \(\frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i' \varepsilon_i = O_P(1)\), implies that \(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \hat{b}_i'X_i'M_{F^0}X_i \hat{b}_i = O_P(1)\). Noting that
\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \hat{b}_i'X_i'M_{F^0}X_i \hat{b}_i \geq \min_{1 \leq i \leq N} \mu_{\min} \left( \frac{1}{T} X_i'M_{F^0}X_i \right) \frac{1}{N} \sum_{i=1}^{N} \hat{b}_i' \hat{b}_i
\]
and
\[
\mu_{\min} \left( \frac{1}{T} X_i'M_{F^0}X_i \right) = \mu_{\min} \left( \frac{1}{T} X_i'(M_{F^0} - M_{F^0}) X_i \right)
\geq \mu_{\min} \left( \frac{1}{T} X_i'X_i \right) - \frac{1}{T} \| X_i \| \| P_{F^0} - P_{F^0} \|
\geq \xi_{XX} - o_P(1) \text{ w.p.a.1},
\]
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 \(\hat{\beta}\) of \(\beta\) satisfies \(\hat{\beta} \in N_{L}(\beta^0)\) with probability at least \(1 - \epsilon\). By 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) \(T^{-1} \| \hat{F} - F^0 \|^2 = O_P(\eta_{NT}^2) + O_P(\delta_{NT}^2)\),
(ii) \(T^{-1} (\hat{F} - F^0)' F^0 H = O_P(\eta_{NT}) + O_P(\delta_{NT}^2)\),
(iii) \(T^{-1} (\hat{F} - F^0)' \hat{F} = O_P(\eta_{NT}) + O_P(\delta_{NT}^2)\),
(iv) \(T^{-1} (\hat{F} - F^0)' H (\hat{F} - F^0) H = I_{R_0} - T^{-1} H (\hat{F} - F^0) H = O_P(\eta_{NT}) + O_P(\delta_{NT}^2)\),
(v) \(\| P_{F^0} - H_{F^0} \| = O_P(\eta_{NT}) + O_P(\delta_{NT}^2)\).

**Lemma A.3** Let \(R_{1i} = \frac{1}{T} X_i'(M_{F^0} - M_{F^0}) \varepsilon_i\) and \(R_{2i} = \frac{1}{T} X_i'M_{F^0} (\hat{F} - F^0) H^{-1} \chi_i' + \frac{1}{NT} \sum_{j=1}^{N} X_j'M_{F^0} X_j \nu_{ji} \hat{b}_j\). Suppose Assumptions A.1-A.2 hold. Then
(i) \(R_{1i} = O_P(\zeta_{NT})\) for each \(i = 1, \ldots, N\), and \(N^{-1} \sum_{i=1}^{N} R_{1i}^2 = O_P(\zeta_{NT}^2)\),
(ii) \(R_{2i} = O_P(\zeta_{NT})\) for each \(i = 1, \ldots, N\), and \(N^{-1} \sum_{i=1}^{N} R_{2i}^2 = O_P(\zeta_{NT}^2)\),
where \(\zeta_{NT} = O_P(\eta_{NT}^2) + O_P(\delta_{NT}^2)\).

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

\begin{equation}
\alpha_i = \beta_i - \frac{\hat{\alpha}_i}{||\beta_i - \hat{\alpha}_i||} \neq 0 \text{ and } ||\alpha_i|| \leq 1 \text{ if } ||\beta_i - \hat{\alpha}_i|| = 0.
\end{equation}

Let \( \hat{Q}_{X_i^jX_i} = \frac{1}{T} X_i^j M_{i^j} X_i \). Noting that \( Y_i = X_i \beta_i^0 + \hat{F} H^{-1} \lambda_0^0 \in \varepsilon^i + (F^0 - \hat{F} H^{-1}) \lambda_0^0 \), (A.8) implies that

\begin{equation}
\hat{Q}_{X_i^jX_i} (\hat{\beta}_i - \beta_i^0) = \frac{1}{T} X_i^j M_{i^j} \varepsilon_i + \frac{1}{T} X_i^j M_{i^j} (F^0 - \hat{F} H^{-1}) \lambda_0^0 - \frac{\kappa_0}{T} \sum_{j=1}^{K_0} \hat{\epsilon}_{ij} \Pi_{i=1,j \neq j} K_i \hat{\beta}_i - \hat{\alpha}_i \|
\end{equation}

where \( \hat{\epsilon}_{ij} = \frac{\beta_i - \hat{\alpha}_i}{||\beta_i - \hat{\alpha}_i||} \). Using the definition of \( \hat{b}_i(= \hat{\beta}_i - \beta_i^0) \), (A.9) can be rewritten as

\begin{equation}
\hat{Q}_{X_i^jX_i} \hat{b}_i = \frac{1}{N^T} \sum_{j=1}^{N} X_i^j M_{i^j} X_j \nu_{ij} \hat{b}_i + R_i,
\end{equation}

where \( R_i = \frac{1}{T} X_i^j M_{i^j} P_{i^j} = 1 - R_{\hat{f}i} - R_{\hat{g}i} + R_{\hat{g}i} \), where \( R_{\hat{f}i} \) and \( R_{\hat{g}i} \) are defined in Lemma A.3, and \( R_{\hat{g}i} = -\frac{\kappa_0}{4T} \sum_{j=1}^{K_0} \hat{\epsilon}_{ij} \Pi_{i=1,j \neq j} K_i \hat{\beta}_i - \hat{\alpha}_i \). Noting that \( ||X_i^j M_{i^j} \varepsilon_i||^2 \leq 2 ||X_i^j \varepsilon_i||^2 + 2 ||X_i^j P_{i^j} \varepsilon_i||^2 \) and \( ||X_i^j P_{i^j} \varepsilon_i||^2 = \text{tr}(X_i^j P_{i^j} \varepsilon_i P_{i^j} \varepsilon_i') \leq \text{tr}(X_i^j X_i^j) \text{tr}(P_{i^j} \varepsilon_i P_{i^j} \varepsilon_i) \leq \text{tr}(X_i^j X_i^j) \text{tr}(\varepsilon_i P_{i^j} \varepsilon_i) \leq \mu_{\min}(T^{-1} F^0 F^0)^{-1} T^{-1} \mu_{\max}(T^{-1} X_i^j X_i^j) \),

\begin{align*}
\frac{1}{N} \sum_{i=1}^{N} ||R_i||^2 &= O_P \left( \frac{\eta_{NT}}{N T} \right) + O_P \left( \frac{\delta_{NT}}{N T} + \kappa^2 \right)
\end{align*}

by Lemma A.3. Let \( \hat{Q}_i = \text{diag}(\hat{Q}_{X_i^jX_i}, ..., \hat{Q}_{X_i^jX_i}) \). Define \( \hat{Q}_2 \) as an \( N \times N \) matrix with typical blocks \( \frac{1}{N^T} X_i^j M_{i^j} P_{i^j} \). Then (A.10) implies that

\begin{equation}
(\hat{Q}_1 - \hat{Q}_2) \hat{b} = R.
\end{equation}
by Theorem 3.1(ii) and Assumptions A.1(iii) and (x). It follows that 
\( \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{b}_i \right\|^2 = \frac{1}{N} \left\| b \right\|^2 \leq 4 \xi_{XX}^2 \frac{1}{N} \sum_{i=1}^{N} \left\| R_i \right\|^2 = O_P \left( \eta_{NT}^2 \right) + O_P \left( T^{-1} + \delta^2_{NT} + \kappa^2 \right) = O_P \left( T^{-1} + \kappa^2 \right) \) 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 \( \beta = \beta^0 + T^{-1/2} v \), where \( v = (v_1, ..., v_N) \) is a \( p \times N \) matrix. We want to show that for any given \( \epsilon^* > 0 \), there exists a large constant \( L = L (\epsilon^*) \) such that for sufficiently large \( N \) and \( T \) we have

\[
P \left\{ \inf_{\hat{\alpha}, \hat{F}} \frac{Q^{(K_0)}_{1NT, \kappa} (\beta^0 + T^{-1/2} v, \hat{\alpha}, \hat{F}) - Q^{(K_0)}_{1NT, \kappa} (\beta^0, \alpha^0, \hat{F})}{\left\| \hat{b} - b \right\|^2} \geq 1 - \epsilon^* \right\}
\]

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

\[
T \left[ Q^{(K_0)}_{1NT, \kappa} (\beta^0 + T^{-1/2} v, \hat{\alpha}, \hat{F}) - Q^{(K_0)}_{1NT, \kappa} (\beta^0, \alpha^0, \hat{F}) \right] 
\geq T \left[ Q^{(K_0)}_{1NT, \kappa} (\beta^0 + T^{-1/2} v, \hat{\alpha}, \hat{F}) - Q^{(K_0)}_{1NT, \kappa} (\beta^0, \hat{\alpha}, \hat{F}) \right] 
\]

where \( v = \text{vec}(v) \), \( A^0 \) 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 \( 2 (NT)^{-1} X_i' M_{F^0} X_i \nu_{ji} \). As in the proof of Theorem 3.1, we can show that w.p.a.1

\[
\mu_{\min} (A^0 - J^0) \geq \mu_{\min} (A^0) - \mu_{\max} (J^0)
\]
\[
\geq \min_{1 \leq i \leq N} \mu_{\min} (T^{-1} X_i' M_{F^0} X_i) - 4 \max_{1 \leq i \leq N} \mu_{\max} \left( (NT)^{-1} X_i' M_{F^0} X_i \nu_{ji} \right)
\]
\[
\geq \min_{1 \leq i \leq N} \mu_{\min} (T^{-1} X_i' M_{F^0} X_i) - 4 N^{-1} \max_{1 \leq i \leq N} \left\| \lambda_{i0} \right\|^2 \left[ \mu_{\min} (N^{-1} A^0 A^0) \right]^{-1} \max_{1 \leq i \leq N} \mu_{\max} (T^{-1} X_i' M_{F^0} X_i)
\]
\[
\geq \xi_{XX}/2,
\]
where we use the fact that \( \nu_{ii} = \lambda_i^0 (N^{-1} \Lambda_i^0 \Lambda_0)^{-1} \lambda_i^0 \leq [\min_i (N^{-1} \Lambda_i^0 \Lambda_0)]^{-1} \| \lambda_i^0 \|^2 \) and that \( \max_{1 \leq i \leq N} \| \lambda_i^0 \|^2 = o(N^{1/2}) \) under Assumption A.1(ii). Then \( \frac{1}{N} v' (A^0 - J^0) v \geq \frac{\lambda_0}{2N} \| v \|^2 \) 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

\[
\frac{1}{N \sqrt{T}} \sum_{i=1}^{N} v_i' X_i' M \hat{\varepsilon}_i \leq \left( \frac{1}{N} \sum_{i=1}^{N} \| v_i \|^2 \right)^{1/2} \left( \frac{1}{NT} \sum_{i=1}^{N} \| \varepsilon_i' M_{X_i X_i'} \| \right)^{1/2} = N^{-1/2} \| v \| C_1^{1/2} \{ 1 + o_p(1) \},
\]

where \( C_1 = \min_{(N,T)} \frac{1}{N} \sum_{i=1}^{N} \| \varepsilon_i'^2 M_{X_i X_i'} \varepsilon_i \), and we use the fact that \( \frac{1}{N} \sum_{i=1}^{N} \| R_{2i} \|^2 = o_p(1) \) by Lemma A.3(ii), the above analysis, and Assumptions A.2(iii)-(iii). It follows that

\[
T \left[ Q_{1NT,n}^{(K)} \left( \beta^0 + T^{-1/2} v, \hat{\alpha}, \hat{F} \right) - Q_{1NT,n}^{(K)} \left( \beta^0, \alpha^0, \hat{F} \right) \right] \geq \mu_{\min} \left( \frac{A^0 - J^0}{N} \right) \| \varepsilon \|^2 \{ 1 + o_p(1) \} - 2N^{-1/2} \| v \| C_1^{1/2} \{ 1 + o_p(1) \} + o_p(1),
\]

where \( \mu_{\min} (A^0 - J^0) \geq \epsilon \log \log \frac{N}{T} > 0 \) w.p.a.1. The first term dominates the second term in the last display for sufficiently large \( L \). That is \( T Q_{1NT,n}^{(K)} \left( \beta^0 + T^{-1/2} v, \hat{\alpha}, \hat{F} \right) - Q_{1NT,n}^{(K)} \left( \beta^0, \alpha^0, \hat{F} \right) > 0 \) for sufficiently large \( L \). Consequently, the minimizer \( \hat{\beta} \) must satisfy \( \frac{1}{N} \sum_{i=1}^{N} \| b_i \|^2 = o_p \left( T^{-1} \right) \).

By Lemma A.3(ii), the fact that \( \frac{1}{N} \sum_{i=1}^{N} X_i' M \varepsilon_i = O_p \left( T^{-1/2} \right) \) under Assumptions A.1(i) and (vi), and the result in (i), we have

\[
\frac{1}{T} X_i' M \hat{\varepsilon}_i = O_p \left( \eta_{NT} \right) + O_p \left( T^{-1/2} \eta_{NT} \right) + O_p \left( \delta_{NT}^{-2} \right) + O_p \left( T^{-1/2} \right) = O_p \left( T^{-1/2} \right).
\]

By Lemma A.3(ii), the fact that \( \frac{1}{N} \sum_{i=1}^{N} X_i' M X_i' \beta_j = O_p (\eta_{NT}) \), and (i),

\[
\frac{1}{T} \sum_{i=1}^{N} X_i' M (F^0 - \hat{F} H^{-1}) = O_p (\eta_{NT}) + O_p \left( T^{-1/2} \eta_{NT} \right) + O_p \left( \delta_{NT}^{-2} \right) = O_p \left( T^{-1/2} \right).
\]

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

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

\[
\| \beta_i - \alpha_k \| - \Pi_{k=1}^{K_0} \| \beta_{i-1} - \alpha_k \| \leq \tilde{c}_{iNT} (\alpha) \| \beta_i - \beta_i^0 \| \quad \text{and} \quad \tilde{c}_{iNT} (\alpha) \leq C_{k,0NT} (\alpha) \left( 1 + \| \beta_i - \beta_i^0 \| \right),
\]

where \( C_{k,0NT} (\alpha) = \max_{1 \leq k \leq K_0} \max_{1 \leq k \leq K_0} \Pi_{k=1}^{K_0} \| \beta_i - \alpha_k \| \) and \( \max_{1 \leq k \leq K_0} \max_{1 \leq k \leq K_0} \Pi_{k=1}^{K_0} \| \beta_i - \alpha_k \| = O(1) \) and \( \alpha_k \) is finite integers. It follows that as \( (N,T) \to \infty \),

\[
\left| P_{NT} (\hat{\beta}, \alpha) - P_{NT} (\beta^0, \alpha) \right| \leq C_{k,0NT} (\alpha) \left( \frac{1}{N} \sum_{i=1}^{N} \| b_i \| \right)^2 + \frac{1}{N} \sum_{i=1}^{N} \| b_i \| \leq C_{k,0NT} (\alpha) \left( \frac{1}{N} \sum_{i=1}^{N} \{ 1 + o_p (T^{-1/2}) \} \right)^{1/2} + o_p (T^{-1/2}) = O_p (T^{-1/2}). \quad \text{(A.13)}
\]
By (A.13) and the fact that \( P_{NT}(\beta^0, \alpha^0) = 0 \), we have

\[
0 \geq P_{NT}(\hat{\beta}, \hat{\alpha}) - P_{NT}(\beta^0, \alpha^0) = P_{NT}(\beta^0, \hat{\alpha}) - P_{NT}(\beta^0, \alpha^0) + O_P(T^{-1/2})
\]

\[
\begin{align*}
&= \frac{1}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \|\beta_i^0 - \alpha_i^0\| + O_P(T^{-1/2}) \\
&= \frac{N_1}{N} \Pi_{k=1}^{K_0} \|\hat{\alpha}_k - \alpha_0^0\| + \ldots + \frac{N_K}{N} \Pi_{k=1}^{K_0} \|\hat{\alpha}_k - \alpha_0^0\| + O_P(T^{-1/2}).
\end{align*}
\]

(A.14)

Then by Assumption A.2(i), we have \( \Pi_{k=1}^{K_0} \|\hat{\alpha}_k - \alpha_0^0\| = O_P(T^{-1/2}) \) for \( l = 1, \ldots, K_0 \). It follows that \((\hat{\alpha}(1), \ldots, \hat{\alpha}(K_0)) - (\alpha_0^1, \ldots, \alpha_0^0) = O_P(T^{-1/2})\) for some suitable permutation \((\hat{\alpha}(1), \ldots, \hat{\alpha}(K_0))\) of \((\hat{\alpha}_1, \ldots, \hat{\alpha}_{K_0})\).

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

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

1. \( T^{-1} \varepsilon_i'(\hat{F} - F^0)H = O_P(\delta_{NT}^2) \) and \( P(\max_{1 \leq i \leq N} \left\| T^{-1} \varepsilon_i'(\hat{F} - F^0)H \right\| \geq C \delta_{NT}^2 (\ln T)^{\nu_1}) = o(N^{-1}) \),

2. \( T^{-1/2} X_i'(\hat{F} - F^0)H = O_P(1) \) and \( P(\max_{1 \leq i \leq N} \left\| T^{-1/2} X_i'(\hat{F} - F^0)H \right\| \geq C(\ln T)^{\nu_1}) = o(N^{-1}) \),

3. \( T^{-1} X_i'M_{i\varepsilon_i} = T^{-1} X_i'M_{i\varepsilon_i} + O_P(\delta_{NT}^2) \) and \( P(\max_{1 \leq i \leq N} \|R_{1i}\| \geq C \delta_{NT}^2 (\ln T)^{\nu_1}) = o(N^{-1}) \),

4. \( T^{-1/2} X_i'M_{i\varepsilon_i} F^{-1} = O_P(1) \) and \( P(\max_{1 \leq i \leq N} \left\| T^{-1/2} X_i'M_{i\varepsilon_i} F^{-1} \right\| \geq C(\ln T)^{\nu_1}) = o(N^{-1}) \),

5. \( P(\max_{1 \leq i \leq N} \left\| T^{-1} X_i'M_{i\varepsilon_i} X_i \right\| \geq C \delta_{NT}^2 (\ln T)^{\nu_2}) = o(N^{-1}) \),

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 \( \hat{Q}_1, \hat{Q}_2, R \), 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 \( S_i R = R_i \). Then \( \hat{b}_i = S_i^T (\hat{Q}_1 - \hat{Q}_2)^{-1} R \) by (A.11). It follows that

\[
\begin{align*}
\| \hat{b}_i \|^2 &= \text{tr} \left( S_i^T S_i \left( \hat{Q}_1 - \hat{Q}_2 \right)^{-1} RR^T \left( \hat{Q}_1 - \hat{Q}_2 \right)^{-1} \right) \\
&= \mu_{\max} \left( \left( \hat{Q}_1 - \hat{Q}_2 \right)^{-1} \otimes \left( \hat{Q}_1 - \hat{Q}_2 \right)^{-1} \right) \text{vec}(S_i^T S_i) \text{vec}(RR^T) \\
&\leq \mu_{\max} \left( \left( \hat{Q}_1 - \hat{Q}_2 \right)^{-1} \otimes \left( \hat{Q}_1 - \hat{Q}_2 \right)^{-1} \right) \left( S_i S_i^T \right) \text{vec}(RR^T) \\
&= \left[ \mu_{\min} \left( \hat{Q}_1 - \hat{Q}_2 \right) \right]^{-2} \text{tr} \left( S_i^T S_i^T RR^T \right) \\
&= \left[ \mu_{\min} \left( \hat{Q}_1 - \hat{Q}_2 \right) \right]^{-2} \|R_i\|^2 ,
\end{align*}
\]

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

\[
\begin{align*}
\mu_{\min}(\hat{Q}_1 - \hat{Q}_2) &\geq \mu_{\min}(\hat{Q}_1) - \mu_{\max}(\hat{Q}_2) \\
&\geq \min_{1 \leq i \leq N} \mu_{\min}(\hat{Q}_1 X_i X_i) - 2 \gamma_1^{-1} \max_{1 \leq i \leq N} \lambda_i^0 \|X_i\|^2 \max_{1 \leq i \leq N} T^{-1} \|X_i\|^2.
\end{align*}
\]
Under Assumptions A.3(i) and (iv), we have $P(\max_{1 \leq i \leq N} N^{-1} \| \hat{\lambda}_i \|^2 \max_{1 \leq i \leq N} T^{-1} \| X_i \|^2 \geq C) = o(N^{-1})$ for any small $C > 0$. This, in conjunction with Lemma A.4(v), and Assumption A.1(x), implies that $P(\mu_{\min}(\hat{Q}_1 - \hat{Q}_2) \geq \xi_X/2) = 1 - o(N^{-1})$. On the other hand,

$$\| R_i \|^2 = \left\| \frac{1}{T} X_i' M_{P_0 \xi_i} + R_{1i} - R_{2i} \right\|^2 \leq 2 \left\| \frac{1}{T} X_i' M_{P_0 \xi_i} + R_{1i} - R_{2i} \right\|^2 + 2 \| R_{3i} \|^2.$$

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

$$P \left( \max_{1 \leq i \leq N} \left\| \frac{1}{T} X_i' M_{P_0 \xi_i} + R_{1i} - R_{2i} \right\| \geq CT^{-1/2} (\ln T)^{\nu_1} \right) = o(N^{-1}) \text{ for any } C > 0.$$

By the triangle inequality and the fact that $\Pi_{i=1, l \neq j}^{K_0} \| \hat{\beta}_i - \hat{\alpha}_l \| - \Pi_{l=1, l \neq j}^{K_0} \| \hat{\beta}_0^0 - \hat{\alpha}_l \| \leq C_{K_0} (\hat{\alpha}) (1 + 2 \| \hat{\beta}_i - \hat{\beta}_i^0 \|)$ for some $C_{K_0} (\hat{\alpha}) = O_p(1)$ (see, e.g., the argument used in the proof of Theorem 3.2), we have

$$\| R_{3i} \| \leq \frac{\kappa}{2} \sum_{j=1}^{K_0} \Pi_{l=1, l \neq j}^{K_0} \| \hat{\beta}_i - \hat{\alpha}_l \| \leq \frac{\kappa}{2} \sum_{j=1}^{K_0} \Pi_{l=1, l \neq j}^{K_0} \| \hat{\beta}_0^0 - \hat{\alpha}_l \| + \frac{\kappa}{2} \sum_{j=1}^{K_0} \Pi_{l=1, l \neq j}^{K_0} \| \hat{\beta}_i - \hat{\alpha}_l \| \leq \kappa K_0 C (\hat{\alpha}) \| \hat{b}_i \| + r_{NT}. $$

where $r_{NT} = \frac{\kappa}{2} [K_0 C_{K_0} (\hat{\alpha}) + \max_{1 \leq l \leq N} \sum_{j=1}^{K_0} \Pi_{l=1, l \neq j}^{K_0} \| \beta_0^0 - \alpha_l \|] = O(\kappa)$ as $\beta_i^0$'s can only take finite $K_0$ values. It follows that

$$\| \hat{b}_i \|^2 \leq \left[ \mu_{\min}(\hat{Q}_1 - \hat{Q}_2) \right]^{-2} \| R_i \|^2 \leq \left[ \mu_{\min}(\hat{Q}_1 - \hat{Q}_2) \right]^{-2} \left\{ 2 \left\| \frac{1}{T} X_i' M_{P_0 \xi_i} + R_{1i} - R_{2i} \right\|^2 + 2 \| R_{3i} \|^2 \right\} \leq \left[ \mu_{\min}(\hat{Q}_1 - \hat{Q}_2) \right]^{-2} \left\{ 2 \left\| \frac{1}{T} X_i' M_{P_0 \xi_i} + R_{1i} - R_{2i} \right\|^2 + 4 [\kappa K_0 C (\hat{\alpha})]^{2} \| \hat{b}_i \|^2 + 4 r_{NT}^2 \right\}.$$ 

That is, $\| \hat{b}_i \|^2 \leq \left[ \mu_{\min}(\hat{Q}_1 - \hat{Q}_2) \right]^{-2} \left\{ 2 \left\| \frac{1}{T} X_i' M_{P_0 \xi_i} + R_{1i} - R_{2i} \right\|^2 + 4 r_{NT}^2 \right\}$. It follows that for sufficiently small $\kappa$ (as $\kappa \to 0$), we have

$$P \left( \max_{1 \leq i \leq N} \left\| \hat{b}_i \right\|^2 \geq C \left[ T^{-1} (\ln T)^{2\nu_1} + \kappa^2 (\ln T)^{2\nu_2} \right] \right) \leq P \left( \max_{1 \leq i \leq N} \left\| \hat{b}_i \right\|^2 \geq C \left[ T^{-1} (\ln T)^{2\nu_1} + \kappa^2 (\ln T)^{2\nu_2} \right] \right) \leq P \left( \max_{1 \leq i \leq N} \left\| \hat{b}_i \right\|^2 \geq C T^{-1} (\ln T)^{2\nu_1} \right) + o(N^{-1})$$

$$= \left\{ P \left( 44 \frac{\kappa^2}{\kappa^2} (\ln T)^{2\nu_2} \xi_X^2 / 8 \right) + o(N^{-1}) \right\} = \left\{ P \left( 44 \frac{\kappa^2}{\kappa^2} (\ln T)^{2\nu_2} \xi_X^2 / 8 \right) + o(N^{-1}) \right\} = o(N^{-1}) + o(N^{-1}) = o(N^{-1}).$$
where we use the fact that \( P \left( \max_{1 \leq i \leq N} \| T^{-1/2} X'_i M_{p0} \varepsilon_i \| \geq C \left( \ln T \right)^{v_1} \right) = o \left( N^{-1} \right) \) \( \forall C > 0 \) by Assumptions A.1(i) and A.3(i). That is, \( P \left( \max_{1 \leq i \leq N} \| \hat{b}_i \| \geq C \left( T^{-1/2} \left( \ln T \right)^{v_1} + \kappa \left( \ln T \right)^{v_2} \right) \right) = o \left( N^{-1} \right) \) for any \( C > 0 \).

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

\[
0_{p \times 1} = - \frac{2}{\sqrt{T}} X'_i M_{p0} (Y_i - X_i \beta_i) + \sqrt{T} K \sum_{j=1}^{K_0} \hat{\epsilon}_{ij} \Pi_{l=1,l \neq j}^{K_0} \beta_i - \hat{\alpha}_l \| I_p (\hat{\beta}_i - \hat{\alpha}_k) 
\]

\[
= - \frac{2}{\sqrt{T}} X'_i (M_{p0} - M_{p}) \varepsilon_i - \frac{2}{\sqrt{T}} X'_i (M_{p0} - M_{p}) F^{0} \lambda_i^0 + \sqrt{T} K \sum_{j=1}^{K_0} \hat{\epsilon}_{ij} \Pi_{l=1,l \neq j}^{K_0} \beta_i - \hat{\alpha}_l \|
\]

\[
\equiv \hat{B}_{i1} + \hat{B}_{i2} + \hat{B}_{i3} + \hat{B}_{i4} + \hat{B}_{i5} + \hat{B}_{i6}, \text{ say,} \quad (A.15)
\]

where \( \hat{c}_{ij} = \frac{\hat{\beta}_i - \hat{\alpha}_j}{\| \beta_i - \hat{\alpha}_j \|} \) if \( \| \hat{\beta}_i - \hat{\alpha}_j \| \neq 0 \) and \( \| \hat{c}_{ij} \| \leq 1 \) otherwise.

Let \( \chi_{NT} = T^{-1/2} \left( \ln T \right)^{v_1} + \kappa \left( \ln T \right)^{v_2} \) where \( 0 < v_2 \leq v_1 \). By Theorem 3.3,

\[
P \left( \max_{1 \leq i \leq N} \| \hat{b}_i \| \geq C \chi_{NT} \right) = o \left( N^{-1} \right) \text{ for any } C > 0. \quad (A.16)
\]

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

\[
P \left( \| \hat{\alpha}_k - \alpha_k^0 \| \geq CT^{-1/2} \left( \ln T \right)^{v_1} \right) = o \left( N^{-1} \right) \quad \text{and} \quad P \left( \max_{i \in G_k^0} \| \hat{c}_i - c_k^0 \| \geq c_k^0/2 \right) = o \left( N^{-1} \right). \quad (A.17)
\]

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

\[
P \left( \max_{i \in G_k^0} \| \hat{B}_{i3} \| > C \chi_{NT} \right) = o \left( N^{-1} \right) \quad \text{and} \quad P \left( \max_{i \in G_k^0} \| \hat{B}_{i6} \| > C \chi_{NT} \right) = o \left( N^{-1} \right).
\]

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

\[
\Xi_{k,NT} = \left\{ \max_{i \in G_k^0} \| \hat{c}_i - c_k^0 \| \leq c_k^0/2 \right\} \cap \left\{ \max_{i \in G_k^0} \| \hat{B}_{i1} \| \leq CT^{1/2} \delta_{NT} \right\}
\]

\[
\cap \left\{ \max_{i \in G_k^0} \| \hat{B}_{i4} \| \leq C \left( \ln T \right)^{v_1} \right\} \cap \left\{ \max_{i \in G_k^0} \| \hat{B}_{i5} \| \leq C \left( \ln T \right)^{v_1} \right\}
\]

\[
\cap \left\{ \max_{i \in G_k^0} \| \hat{B}_{i6} \| \leq C \sqrt{T} \kappa \chi_{NT} \right\}.
\]
Then conditional on $\Xi_{kNT}$, we have that uniformly in $i \in G_k^0$,
\[
\left\| \left( \hat{\beta}_i - \alpha_k \right)' \left( \hat{B}_{i2} + \hat{B}_{i3} + \hat{B}_{i4} + \hat{B}_{i5} + \hat{B}_{i6} \right) \right\| \\
\geq \left\| \left( \hat{\beta}_i - \alpha_k \right)' \hat{B}_{i1} \right\| - \left\| \left( \hat{\beta}_i - \alpha_k \right)' \left( \hat{B}_{i3} + \hat{B}_{i4} + \hat{B}_{i5} + \hat{B}_{i6} \right) \right\| \\
\geq \sqrt{T} \kappa \hat{c}_k \left\| \hat{\beta}_i - \alpha_k \right\| - C \left\| \hat{\beta}_i - \alpha_k \right\| \left\{ 3 (\ln T)^{\nu_1} c_{\lambda,N} + \sqrt{T} \kappa \chi_{NT} \right\} \\
\geq \sqrt{T} \kappa \hat{c}_k \left\| \hat{\beta}_i - \alpha_k \right\| / 4 \text{ for sufficiently large } (N,T),
\]
where the last inequality follows because $\hat{c}_k \geq c_k^0 / 2$ on $\Xi_{kNT}$, $\sqrt{T} \kappa \gg 3 (\ln T)^{\nu_1} c_{\lambda,N} + \sqrt{T} \kappa \chi_{NT}$ by Assumptions A.2(iii)-(iv). It follows that for all $i \in G_k^0$
\[
P(\hat{\epsilon}_{kNT,i} \mid i \in G_k^0) = P \left( \left\| \hat{\beta}_i - \alpha_k \right\| \leq 1 / \sqrt{T} \kappa \hat{c}_k \right) + P(\Xi_{kNT}^c) \\
\to 0 \text{ as } (N,T) \to \infty,
\]
where $\Xi_{kNT}^c$ denotes the complement of $\Xi_{kNT}$ 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 $P(\Xi_{kNT}^c) = o \left( N^{-1} \right)$. Consequently, we can conclude that with probability $1 - o \left( N^{-1} \right)$, $\hat{\beta}_i - \alpha_k$ must be in a position where $\left\| \hat{\beta}_i - \alpha_k \right\|$ is not differentiable with respect to $\beta_i$ for any $i \in G_k^0$. That is, $P \left( \left\| \hat{\beta}_i - \alpha_k \right\| = 0 \mid i \in G_k^0 \right) = 1 - o \left( N^{-1} \right)$ 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) $T^{-1} \left\| \hat{F} - F^0 H \right\|^2 = O_P \left( \sum_{k=1}^{K_0} \left\| \hat{\alpha}_k - \alpha_k^0 \right\|^2 \right) + O_P \left( \delta_{NT}^{-2} \right)$,
(ii) $T^{-1} (\hat{F} - F^0 H)'F^0 H = O_P \left( \sum_{k=1}^{K_0} \left\| \hat{\alpha}_k - \alpha_k^0 \right\| \right) + O_P \left( \delta_{NT}^{-2} \right)$,
(iii) $T^{-1} (\hat{F} - F^0 H)' \hat{F} = O_P \left( \sum_{k=1}^{K_0} \left\| \hat{\alpha}_k - \alpha_k^0 \right\| \right) + O_P \left( \delta_{NT}^{-2} \right)$,
(iv) $T^{-1} (\hat{F}' \hat{F} - H' F^0 H) = I_{R_0} - T^{-1} H' F^0 H = O_P \left( \sum_{k=1}^{K_0} \left\| \hat{\alpha}_k - \alpha_k^0 \right\| \right) + O_P \left( \delta_{NT}^{-2} \right)$,
(v) $\left\| P_{\hat{F}} - P_{F^0 H} \right\| = O_P \left( \sum_{k=1}^{K_0} \left\| \hat{\alpha}_k - \alpha_k^0 \right\| \right) + O_P \left( \delta_{NT}^{-1} \right)$.

To state the next lemma, for $l = 1, \ldots, K_0$, define
\[
\tilde{A}_{1k,l} = -\frac{1}{NN_k T^2} \sum_{j \in G_l} \sum_{i \in G_k} X_i' M_{ik} X_j \left( \hat{\alpha}_l - \alpha_l^0 \right) \lambda_l^0 G' \hat{F}' X_j, \\
\tilde{A}_{2k,l} = \frac{1}{NN_k T} \sum_{j \in G_l} \sum_{i \in G_k} \nu_{ji} X_i' M_{ik} X_j, \\
\tilde{A}_{3k,l} = \frac{1}{NN_k T^2} \sum_{j \in G_l} \sum_{i \in G_k} X_i' M_{ik} X_j \varepsilon_j' \hat{F} G \lambda_l^0,
\]

13
$$\tilde A_{ik,l} = \frac{1}{NN_{k}T^2} \sum_{j \in \mathcal{G}_i} \sum_{i \in \mathcal{G}_k} X'_i M_{ik} F_0 \lambda_i^0 \lambda_i^0 G' \hat F' X_j, \quad \text{and}$$

$$\tilde A_{sk,l} = \frac{1}{NN_{s}T^2} \sum_{j \in \mathcal{G}_i} \sum_{i \in \mathcal{G}_s} X'_i M_{ik} \varepsilon_j \lambda_i^0 G' \hat F' X_j.$$

Let $\tilde A_{k,l} = \sum_{s=1}^5 \tilde A_{sk,l}$ and $A_{k,l} = \frac{1}{NN_{k}T} \sum_{j \in \mathcal{G}_{s}} \sum_{i \in \mathcal{G}_k} \nu_{ij} X'_i M_{F_0} X_j.$

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

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

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

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

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

(v) $\tilde A_{5k,l} = O_P \left( \delta_{NT}^{-1} \right)$.

Let $Q_{kNT} = \frac{1}{NN_{k}T} \sum_{i \in \mathcal{G}_k} X'_i M_{F_0} X_i$, and $V_{kNT} = \frac{1}{NN_{k}T} \sum_{i \in \mathcal{G}_k} X'_i M_{F_0} (\varepsilon_i - \frac{1}{T} \sum_{j=1}^T \nu_{ij} \varepsilon_j)$.

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

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

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

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

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

(v) $\frac{1}{NT} \sum_{i=1}^N \varepsilon_i' (\hat F - F^0 H) \lambda_i^0 = O_P \left( T^{-1} \delta_{NT}^{-2} \right)$.

**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 $i = 1, \ldots, N$ (resp. $k = 1, \ldots, K_0$), $0_{p \times 1}$ belongs to the subdifferential of $Q_{iNT}(\beta, \alpha, \hat F')$ with respect to $\beta_i$ (resp. $\alpha_k$) evaluated at $\{\hat \beta_i\}$ and $\{\hat \alpha_k\}$. That is, for each $i = 1, \ldots, N$ and $k = 1, \ldots, K_0$, we have

$$0_{p \times 1} = - \frac{2}{NT} X'_i M_{F_0} (Y_i - X_i \hat \beta_i) + \frac{\mu}{N} \sum_{j=1}^{K_0} \hat e_{ij} \Pi_{i=1, l \neq j}^{K_0} \left\| \hat \beta_i - \hat \alpha_l \right\|,$$  

and

$$0_{p \times 1} = \frac{K}{N} \sum_{i=1}^N \hat e_{ik} \Pi_{i=1, l \neq k}^{K_0} \left\| \hat \beta_i - \hat \alpha_l \right\|,$$

The above expressions hold for $0_{p \times 1}$.
where \( \hat{e}_{ij} = \frac{\hat{\beta}_i - \hat{\alpha}_j}{\| \hat{\beta}_i - \hat{\alpha}_j \|} \) if \( \| \hat{\beta}_i - \hat{\alpha}_j \| \neq 0 \) and \( \| \hat{e}_{ij} \| \leq 1 \) if \( \| \hat{\beta}_i - \hat{\alpha}_j \| = 0 \). Fix \( k \in \{1, ..., K_0 \} \). Observe that (a) \( \| \hat{\beta}_i - \hat{\alpha}_k \| = 0 \) 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 \( \hat{e}_{ij} = \frac{\hat{\beta}_i - \hat{\alpha}_j}{\| \hat{\beta}_i - \hat{\alpha}_j \|} = \frac{\hat{\alpha}_k - \hat{\alpha}_j}{\| \alpha_k^0 - \alpha_j \|} \) w.p.a.1 for any \( i \in \hat{G}_k \) and \( j \neq k \). This further implies w.p.a.1

\[
\sum_{i \in \hat{G}_k} \sum_{j=1, j \neq k}^{K_0} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \left\| \hat{\beta}_i - \hat{\alpha}_l \right\| = \sum_{i \in \hat{G}_k} \sum_{j=1, j \neq k}^{K_0} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \left\| \hat{\beta}_i - \hat{\alpha}_l \right\| = \sum_{i \in \hat{G}_k} \sum_{j=1, j \neq k}^{K_0} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \left\| \hat{\beta}_j - \hat{\alpha}_l \right\|
\]

Then by (A.18) we have

\[
\frac{2}{NT} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} (Y_i - X_i \hat{\alpha}_k) + \frac{K}{N} \sum_{i \in \hat{G}_0} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \left\| \hat{\beta}_i - \hat{\alpha}_l \right\| = 0_{p \times 1}.
\]

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

\[
\frac{1}{NT} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} Y_i = \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} X_i \alpha_k^0 + \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} F^0 \lambda_i^0 + \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} \varepsilon_i
\]

\[
= \frac{1}{NT} \sum_{i \in G_k^0} X_i^\prime M_{ki} X_i \alpha_k^0 + \frac{1}{NT} \sum_{i \in \hat{G}_k \backslash G_k^0} X_i^\prime M_{ki} X_i \alpha_k^0 - \frac{1}{NT} \sum_{i \in G_k^0 \backslash \hat{G}_k} X_i^\prime M_{ki} X_i \alpha_k^0 + \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} F^0 \lambda_i^0 + \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} \varepsilon_i.
\]

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

\[
\frac{1}{N_k T} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} X_i (\hat{\alpha}_k - \alpha_i^0) = \frac{1}{N_k T} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} F^0 \lambda_i^0 + \frac{1}{N_k T} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} \varepsilon_i + \hat{R}_{1k} - \hat{R}_{2k} + \hat{R}_{3k},
\]

where \( \hat{R}_{1k} = \frac{1}{NT} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} X_i \alpha_k^0, \hat{R}_{2k} = \frac{1}{NT} \sum_{i \in \hat{G}_k \backslash \hat{G}_k^0} X_i^\prime M_{ki} X_i \alpha_k^0, \) and \( \hat{R}_{3k} = \frac{\max}{\min} \sum_{i \in \hat{G}_0} \hat{e}_{ik} \Pi_{l=1, l \neq k}^{K_0} \| \hat{\beta}_i - \hat{\alpha}_l \| \). Noting that for any \( \epsilon > 0 \), by Theorem 3.4 and the proof of Theorem 2.2 in SSP, we have \( P( \| \hat{R}_{1k} \| \geq \epsilon / (NT)^{1/2} ) \leq P( \hat{f}_{k, NT} ) \rightarrow 0, P( \| \hat{R}_{2k} \| \geq \epsilon / (NT)^{1/2} ) \leq P( \hat{f}_{k, NT} ) \rightarrow 0, \) and \( P( \| \hat{R}_{3k} \| \geq \epsilon / (NT)^{1/2} ) \leq \sum_{k=1}^{K_0} P( i \in \hat{G}_k | i \in G_k^0 ) \leq \sum_{k=1}^{K_0} \sum_{i \in G_k} P( \hat{e}_{i,k} ) = o(1) \). It follows that

\[
\| \hat{R}_{1k} - \hat{R}_{2k} + \hat{R}_{3k} \| = o_P((NT)^{-1/2}).
\]

Next, by Lemma A.7(vi),

\[
\frac{1}{N_k T} \sum_{i \in \hat{G}_k} X_i^\prime M_{ki} F^0 \lambda_i^0 = \sum_{i=1}^{K_0} \hat{A}_{k, i} (\hat{\alpha}_i - \alpha_i^0) - \frac{1}{N N_k T} \sum_{i \in \hat{G}_k} \sum_{j=1}^{N} \nu_{ij} X_i^\prime M_{kj} \varepsilon_j + B_{1, k, NT} + o_P((NT)^{-1/2}).
\]
Combining (A.22)-(A.24) yields
\[
\hat{Q}_{kNT} (\hat{\alpha}_k - \alpha_k^0) = \sum_{i=1}^{K_0} \hat{A}_{k,i} (\hat{\alpha}_i - \alpha_i^0) + \hat{V}_{kNT} + B_{1,kNT} + o_P((NT)^{-1/2}),
\]
where \(\hat{Q}_{kNT} = \frac{1}{N_{kT}} \sum_{i \in G_k} X_i' M_{\rho} X_i\) and \(\hat{V}_{kNT} = \frac{1}{N_{kT}} \sum_{i \in G_k} X_i' M_{\rho} (\varepsilon_i - \frac{1}{N} \sum_{j=1}^N \nu_{ji} \varepsilon_j)\). The above result holds for \(k = 1, \ldots, K_0\). It follows that
\[
\text{vec} (\hat{\alpha} - \alpha^0) = \hat{Q}_{NT}^{-1} \begin{pmatrix} \hat{V}_{NT} + B_{1,NT} + o_P((NT)^{-1/2}) \end{pmatrix}, \tag{A.25}
\]
where
\[
\hat{Q}_{NT} = \begin{pmatrix} \hat{Q}_{1NT} - \hat{A}_{1,1} & \cdots & \cdots & \cdots & -\hat{A}_{1,K_0} \\ -\hat{A}_{2,1} & \hat{Q}_{2NT} - \hat{A}_{2,2} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\hat{A}_{K_0,1} & \cdots & \cdots & \cdots & \hat{Q}_{K_0,NT} - \hat{A}_{K_0,K_0} \end{pmatrix}, \hat{V}_{NT} = \begin{pmatrix} \hat{V}_{1NT} \\ \vdots \\ \hat{V}_{K_0,NT} \end{pmatrix}, \tag{A.26}
\]
and \(B_{1,NT} = \left( B_{1,1,NT}, \ldots, B_{1,K_0,NT} \right)'\). By Lemmas A.7(vi) and A.8(i), \(\hat{Q}_{NT} = Q_{NT} + o_P (\delta_{NT}^{-1})\) where \(Q_{NT}\) is defined in (3.3). By Lemmas A.8(ii)-(iv), \(B_{1,NT} = O_P(T^{-1})\) and \(\hat{V}_{NT} = o_P(T^{-1/2}\delta_{NT}^{-1}).\) Then by (A.25), \(\text{vec} (\hat{\alpha} - \alpha^0) = \left[Q_{NT}^{-1} + O_P \left( \delta_{NT}^{-1} \right) \right] [\hat{V}_{NT} + B_{1,NT}] + o_P((NT)^{-1/2}) = Q_{NT}^{-1} [O_P(T^{-1/2}\delta_{NT}^{-1}) + O_P(T^{-1})] + o_P((NT)^{-1/2}) = O_P \left( T^{-1/2}\delta_{NT}^{-1} \right).\)

**Lemma A.9** Suppose Assumptions A.1-A.5 hold. Let \(\omega_p = (\omega_{1p}, \ldots, \omega_{pp})'\) be an arbitrary \(p \times 1\) nonrandom vector with \(\|\omega_p\| = 1\). Then
\[
(i) \quad \left\| \frac{1}{N_{kT}} \sum_{i \in G_k} X_i' F_{\epsilon}_i \right\| = O_P \left( T^{1/2} \delta_{NT}^{-3/2} \right),
\]
\[
(ii) \quad \frac{1}{N_{kT}} \sum_{i \in G_k} X_i' F_{\epsilon_i} \left( T^{-1} \bar{F}_0 F_0 \right)^{-1} \left( T^{-1} \bar{F}_0 - F_0 \right)^{\prime} = \frac{1}{N_{KNT}} \sum_{i \in G_k} \sum_{j=1}^N X_i' F_0 \left( T^{-1} \bar{F}_0 F_0 \right)^{-1} \times (N^{-1} \Lambda^0 \Lambda^0)^{-1} \lambda_0 \varepsilon_j \varepsilon_i + o_P((NT)^{-1/2}),
\]
\[
(iii) \quad \frac{1}{N_{KNT}} \sum_{i \in G_k} X_i' F_{\epsilon} \left( T^{-1} \bar{F}_0 F_0 \right)^{-1} \left( T^{-1} \bar{F}_0 - F_0 \right)^{\prime} = \frac{1}{N_{KNT}} \sum_{i \in G_k} \sum_{j=1}^N X_i' F_0 \left( T^{-1} \bar{F}_0 F_0 \right)^{-1} \times (N^{-1} \Lambda^0 \Lambda^0)^{-1} \lambda_0 \varepsilon_j \varepsilon_i + o_P((NT)^{-1/2}).
\]

**Proof of Theorem 3.6.** By (A.25), the fact that \(\hat{Q}_{NT} = Q_{NT} + o_P (\delta_{NT}^{-1})\), and Assumptions A.4-A.5, it suffices to prove the theorem by showing that (i) \(\hat{V}_{NT} = V_{NT} + B_{2,NT} + o_P((NT)^{-1/2})\), and (ii) \(B_{NT} = B_{1,NT} + B_{2,NT} + B_{3,NT} = O_P(T^{-1/2}\delta_{NT}^{-1}).\) This follows because (i)-(ii), in conjunction with (A.25) and Assumptions A.4(i) and A.4(ii)-(iii), imply that
\[
\sqrt{NT} \left[ \text{vec} (\hat{\alpha} - \alpha^0) - Q_{NT}^{-1} B_{NT} \right] = \hat{Q}_{NT}^{-1} \sqrt{NT} \left[ \hat{V}_{NT} + B_{1,NT} + o_P((NT)^{-1/2}) \right] - \sqrt{NT} Q_{NT}^{-1} B_{NT}
\]
\[
= \hat{Q}_{NT}^{-1} \sqrt{NT} \left[ V_{NT} + B_{2,NT} + o_P((NT)^{-1/2}) \right] - \sqrt{NT} Q_{NT}^{-1} B_{NT}
\]
\[
= \hat{Q}_{NT}^{-1} \sqrt{NT} V_{NT} + o_P \left( \delta_{NT}^{-1} \right) \sqrt{NT} O_P \left( T^{-1/2} \delta_{NT}^{-1} \right) + o_P(1)
\]
\[
= \hat{Q}_{NT}^{-1} \sqrt{NT} V_{NT} + o_P(1) \frac{d}{N} N \left( 0, Q_0^{-1} \Theta_0 Q_0^{-1} \right).
\]

We first show (i). Let \(\hat{V}_{kNT} = \frac{1}{N_{kT}} \sum_{i \in G_k} X_i' M_{\rho} (\varepsilon_i - \frac{1}{N} \sum_{j=1}^N \nu_{ji} \varepsilon_j)\). Following the proof of Lemma A.6(ii), we can show that \(\hat{V}_{kNT} - \hat{V}_{kNT} = o_P((NT)^{-1/2}).\) Next,
\[
\hat{V}_{kNT} - \hat{V}_{kNT} = \frac{1}{N_{kT}} \sum_{i \in G_k} X_i' (M_{\rho} - M_{F0}) \varepsilon_i - \frac{1}{N_{kT}} \sum_{i \in G_k} X_i' (M_{\rho} - M_{F0}) \nu_{ji} \varepsilon_j \equiv v_{kNT} - v_{kNT}, \text{ say.}
\]
Using (A.1), we have
\[ v_{1kNT} = \sum_{i=1}^{4} \sum_{g_k} X_i' d_3 \varepsilon_i = \sum_{i=1}^{4} v_{1kNT,l} \] and
\[ v_{2kNT} = \sum_{i=1}^{4} \sum_{g_k} X_i' d_3 \varepsilon_i \]
where \( X_i' \) is the column vector of elements in \( G_k \). We have shown in the proof of Lemma A.8(iv) that \( v_{1kNT,1} = O_P(\delta_N^{-3}) \) and \( v_{2kNT,1} = O_P(\delta_N^{-2}) \) under the preliminary convergence rate \( \text{vec}(\hat{\alpha} - \alpha^0) = O_P(T^{-1/2}) \). With the faster convergence rate, \( \text{vec}(\hat{\alpha} - \alpha^0) = O_P(T^{-1/2}\delta_N^{-1}) \), obtained in Theorem 3.5, we have
\[ \frac{1}{T} \left\| \hat{F} - F^0 H' F^0 \right\| = O_P(T^{-1/2}\delta_N^{-1} + \delta_N^{-2}) = O_P(\delta_N^{-2}) \] and \( \left\| I_{k} - (T^{-1} H' F^0 F^0 H) \right\| = O_P(T^{-1/2}\delta_N^{-1} + \delta_N^{-2}) \) 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 \( v_{2kNT,4} : v_{1kNT,2} = O_P(\delta_N^{-3}), v_{1kNT,4} = O_P(T^{-1/2}\delta_N^{-1}), \) and \( v_{2kNT,4} = O_P(T^{-1/2}\delta_N^{-2}) \). We now study \( v_{1kNT,3} \) and \( v_{2kNT,3} \). For \( v_{1kNT,3} \), we make the following decomposition
\[ v_{1kNT,3} = -\frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i' F^0 H H' (\hat{F} H - F^0)' \varepsilon_i \]
\[ = -\frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i' F^0 (T^{-1} F^0 F^0)^{-1} (\hat{F} H - F^0)' \varepsilon_i \]
\[ + \frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i' F^0 (T^{-1} F^0 F^0)^{-1} (1 - H H') (\hat{F} H - F^0)' \varepsilon_i \]
\[ = v_{1kNT,3}(1) + v_{1kNT,3}(2) \]
\( v_{1kNT,3}(1) \) 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 \( (T^{-1} F^0 F^0)^{-1} - H H' = O_P(T^{-1/2}\delta_N^{-1} + \delta_N^{-2}) = O_P(\delta_N^{-2}) \). Then by Lemma A.7(iii)
\[ |\omega_p' v_{1kNT,3}(2)| \leq \frac{1}{N_k T^2} \sum_{i \in G_k^0} \omega_p' X_i' F^0 \left[ (T^{-1} F^0 F^0)^{-1} - H H' \right] (\hat{F} H - F^0)' \varepsilon_i \]
\[ \leq \left\{ \frac{1}{N_k T^2} \sum_{i \in G_k^0} \omega_p' X_i' F^0 \left[ (T^{-1} F^0 F^0)^{-1} - H H' \right] (T^{-1} F^0 F^0)^{-1} - H H' \right\}^{1/2} \]
\[ \times \left\{ \frac{1}{N_k T^2} \sum_{i \in G_k^0} \varepsilon_i (\hat{F} H - F^0)' \varepsilon_i \right\}^{1/2} \]
\[ \leq \left\| (T^{-1} F^0 F^0)^{-1} - H H' \right\| \left\{ \frac{1}{N_k T^2} \sum_{i \in G_k^0} \left\| X_i' F^0 \right\|^2 \right\}^{1/2} O_P(\delta_N^{-2}) \]
\[ = O_P(\delta_N^{-2}) O_P(\delta_N^{-2}) = O_P(\delta_N^{-2}) \cdot \]

It follows that \( v_{1kNT,3} = -\frac{1}{N_k T^2} \sum_{i \in G_k^0} \sum_{j=1}^{N} X_i' F^0 \left( T^{-1} F^0 F^0 \right)^{-1} \left( N^{-1} \Lambda^0 \Lambda^0 \right)^{-1} \delta_j^0 \varepsilon_i + o_P((NT)^{-1/2}) \).

Analogously, by Lemma A.9 and using \( (T^{-1} F^0 F^0)^{-1} - H H' = O_P(\delta_N^{-2}) \), we can show that
\[ v_{2kNT,3} = -\frac{1}{N N_k T^2} \sum_{i \in G_k^0} \sum_{j=1}^{N} X_i' F^0 \left( T^{-1} F^0 F^0 \right)^{-1} \left( N^{-1} \Lambda^0 \Lambda^0 \right)^{-1} \delta_j^0 \varepsilon_i + o_P((NT)^{-1/2}) \].
It follows that \( \hat{V}_{k,NT} - V_{k,NT} = B_{2,k,NT} + o_P((NT)^{-1/2}) \) and (i) follows. Now, by Lemma A.8(ii), \( B_{1,NT} = O_P(T^{-1}) \). Lemmas A.8(iii)-(iv), in conjunction with the result in (i), imply that \( B_{2,NT} = O_P(T^{-1/2}\delta_{NT}^{-1}) \). By Assumption A.5(ii), \( B_{3,NT} = O_P(T^{-1/2}\delta_{NT}^{-1}) \). It follows that \( B_{NT} = O_P(T^{-1/2}\delta_{NT}^{-1}) \). That is, (ii) follows.

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

\[
\frac{1}{N_k T} \sum_{i \in G_k} X_i' M_{\hat{F}} X_i \hat{\alpha}_k^0 = \frac{1}{N_k T} \sum_{i \in G_k} X_i' M_{\hat{F}} Y_i, \tag{A.27}
\]

where \( \hat{V}_{NT} \) is a diagonal matrix consisting of the \( R \) largest eigenvalues of the above matrix in the square bracket, arranged in descending order, and \( \hat{\lambda}_i = T^{-1} \hat{F}' (Y_i - X_i \hat{\alpha}_k) \) 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

\[
\frac{1}{N_k T} \sum_{i \in G_k} X_i' M_{\hat{F}} X_i (\hat{\alpha}_k - \alpha_k^0) = \frac{1}{N_k T} \sum_{i \in G_k} X_i' M_{\hat{F}} \nu_{ij}^0 \nu_{ij} + \frac{1}{N_k T} \sum_{i \in G_k} X_i' M_{\hat{F}} \varepsilon_i + \hat{R}_{1k} - \hat{R}_{2k}. \tag{A.29}
\]

where \( \hat{R}_{1k} = \frac{1}{NT} \sum_{i \in G_k \setminus G_k^0} X_i' M_{\hat{F}} X_i \alpha_k^0 \) and \( \hat{R}_{2k} = \frac{1}{NT} \sum_{i \in G_k^0 \setminus G_k} X_i' M_{\hat{F}} X_i \beta_i^0 \). By arguments like those used in the derivation of (A.23), we can show that \( \hat{R}_{sk} = o_P((NT)^{-1/2}) \) for \( s = 1, 2 \), and \( k = 1, ..., K_0 \). By the uniform selection consistency obtained in Theorem 3.4, we can show that \( \hat{F} \) shares the same first order asymptotic properties as \( \tilde{F} \) 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

\[
\text{vec}(\hat{\alpha} - \alpha_0^0) = \hat{Q}_N^{-1} \left[ \hat{V}_{NT} + \hat{B}_{1,NT} + o_P((NT)^{-1/2}) \right],
\]

where \( \hat{Q}_{NT} \), \( \hat{V}_{NT} \), and \( \hat{B}_{1,NT} \) are analogously defined as \( \hat{Q}_{NT} \), \( \hat{V}_{NT} \), and \( \hat{B}_{1,NT} \) 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 \( V_{k,NT} = V_{k,NT} + B_{3,NT} + o_P((NT)^{-1/2}) \). Observe that

\[
V_{k,NT} = \frac{1}{N_k T} \sum_{i \in G_k^0} X_i' M_{\hat{F}^0} \varepsilon_i - \frac{1}{N N_k T} \sum_{i \in G_k^0} X_i' M_{\hat{F}^0} \sum_{j=1}^N \nu_{ji} \varepsilon_j
\]

\[
= \frac{1}{N_k T} \sum_{i \in G_k^0} X_i' \varepsilon_i - \frac{1}{N_k T} \sum_{i \in G_k^0} X_i' P_{\hat{F}^0} \varepsilon_i - \frac{1}{NT} \sum_{i=1}^N \left[ \frac{1}{N_k} \sum_{j \in G_k^0} \nu_{ji} X_j' M_{\hat{F}^0} \right] \varepsilon_i
\]

\[
= \left\{ \frac{1}{N_k T} \sum_{i \in G_k^0} E_D(X_i') P_{\hat{F}^0} \varepsilon_i + \frac{1}{N_k T} \sum_{i \in G_k^0} \{[X_i - E_D(X_i)] P_{\hat{F}^0} \varepsilon_i - E_D(X_i' P_{\hat{F}^0} \varepsilon_i) \} \right\}
\]

\[
+ \frac{1}{NT} \sum_{i=1}^N \left\{ \frac{1}{N_k} \sum_{j \in G_k^0} \nu_{ji} [X_j' - E_D(X_j')] M_{\hat{F}^0} \right\} \varepsilon_i
\]

\[
= \mathcal{V}_{k,NT} + \mathcal{B}_{3,NT} + \mathcal{R}_{1,NT} + \mathcal{R}_{2,NT},
\]

18
where \( \mathcal{V}_{k,NT} \) and \( \mathcal{B}_{3,kNT} \) are defined in Section 3.3, \( Z_{ki} = \frac{k}{N_k} [X_i - P_{F_0} E_D(X_i) 1 \{ i \in G_k^0 \}] - \frac{1}{N_k} M_{P_{F_0}} \sum_{j \in G_k^0} \nu_{ji} E_D(X_j) \), \( R_{1,kNT} = \frac{1}{N_k} \sum_{i \in G_k^0} \{ [X_i - E_D(X_i)] P_{F_0} \nu_{i} - E_D(X_i | P_{F_0} \nu_{i}) \} \), and \( R_{2,kNT} = \frac{1}{N_k} \sum_{i=1}^{N} \{ \frac{1}{N_k} \sum_{j \in G_k^0} \nu_{ji} [X_j' - E_D(X_j')] M_{P_{F_0}} \nu_{ij} \} \). It suffices to prove that (i) \( R_{1,kNT} = o_p((NT)^{-1/2}) \) and (ii) \( R_{2,kNT} = o_p((NT)^{-1/2}) \). For the proof of (i), see the proof of Corollary 3.4 in Lu and Su (2016). To prove (ii), 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 \( \bar{X}_j' \) denote the \( t \)th row of \( X_j - E_D(X_j) \). Then

\[
ER = E_D \left[ \omega'_p \mathcal{R}_{2,kNT,1} \right]^2 = \frac{1}{N_k^2 N^2 T^2} \sum_{i=1}^{N} \sum_{j \in G_k^0} \sum_{l=1}^{T} \sum_{t=1}^{T} \nu_{ji} \nu_{lj} E_D \left[ \omega'_p \bar{X}_j' \omega'_p \bar{X}_l \right] E_D \left[ \varepsilon_i \varepsilon_l \right]
\]

\[
= \frac{1}{N_k^2 N^2 T^2} \sum_{i=1}^{N} \sum_{j \in G_k^0} \sum_{l=1}^{T} \nu_{ji} \nu_{lj} E_D \left[ \varepsilon_i \varepsilon_l \omega'_p \bar{X}_j' \omega'_p \bar{X}_l \right] E_D \left[ \varepsilon_i^2 \right] + o_P \left( N^{-2} T^{-1} \right),
\]

where the second inequality holds because \( E_D \left[ \varepsilon_i \varepsilon_l \omega'_p \bar{X}_j' \omega'_p \bar{X}_l \right] = 0 \) if \( t \neq s, \) \( E_D \left[ \varepsilon_i \varepsilon_l \omega'_p \bar{X}_j' \omega'_p \bar{X}_l \right] = 0 \) if \( t \geq s, \) and \( E_D \left[ \omega'_p \bar{X}_j' \varepsilon_i \varepsilon_l \right] = 0 \) if \( s \geq t. \) Next,

\[
ER \left( c \right) = \frac{1}{N_k^2 N^2 T^2} \sum_{i=1}^{N} \sum_{j \in G_k^0} \sum_{l=1}^{T} \nu_{ji}^2 \nu_{lj}^2 E_D \left[ \varepsilon_i^2 \varepsilon_l^2 \omega'_p \bar{X}_j' \omega'_p \bar{X}_l \right]
\]

\[
= \frac{1}{N_k^2 N^2 T^2} \sum_{i=1}^{N} \sum_{j \in G_k^0} \sum_{l=1}^{T} \nu_{ji}^2 \nu_{lj}^2 E_D \left[ \varepsilon_i^2 \varepsilon_l^2 \right] = o_P \left( N^{-3} T^{-1} \right).
\]

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

(ii) We verify that \( \mathcal{B}_{3,kNT} = O_P \left( T^{-1} \right). \) Observe that \( \mathcal{B}_{3,kNT} = \frac{1}{N_k} \sum_{i \in G_k^0} E_D \left( X_i | P_{F_0} \nu_{i} \right) = \frac{1}{N_k} \sum_{i=1}^{N} \sum_{s=1}^{T} \mu_{ts} E_D \left( X_i \varepsilon_i \right) \). By Davydov inequality for conditional strong mixing processes and Cauchy-
Schwarz inequality,

$$|\omega_p'\mathcal{B}_{3,k,NT}| \leq \frac{1}{N_k T^2} \sum_{i \in G_k^0} \left[ \sum_{t=2}^{T} \sum_{s=1}^{i} \mu_{ts} E_D[\omega_p^i X_{it} \varepsilon_{it}] \right]$$

$$\leq \frac{8}{N_k T^2} \sum_{i \in G_k^0} \sum_{t=2}^{T} \sum_{s=1}^{i} \left\| F_0^t \right\| \left\| F_0^s \right\| \left\| \bar{X}_{it} \right\|_{8+4\sigma, D} \| \varepsilon_{is} \|_{8+4\sigma, D} \alpha_{NT}^D (t-s)^{(4+2\sigma)/(3+2\sigma)}$$

$$\leq \frac{4}{N_k T^2} \sum_{i \in G_k^0} \sum_{t=2}^{T} \sum_{s=1}^{i} \left\| F_0^t \right\|^2 \left\| \bar{X}_{it} \right\|^2_{8+4\sigma, D} \sum_{t=1}^{\infty} \alpha_{NT}^D (t) (4+2\sigma)/(3+2\sigma)$$

$$+ \frac{4}{N_k T^2} \sum_{i \in G_k^0} \sum_{t=1}^{T-1} \sum_{s=1}^{i} \left\| F_0^t \right\|^2 \left\| \varepsilon_{is} \right\|^2_{8+4\sigma, D} \sum_{t=1}^{\infty} \alpha_{NT}^D (t) (4+2\sigma)/(3+2\sigma)$$

$$= O_P(T^{-1}) + O_P(T^{-1}) = O_P(T^{-1}).$$

Thus $\mathcal{B}_{3,k,NT} = O_P(T^{-1}).$

(iii) We show that $\sqrt{\lambda T} V_{NT} \xrightarrow{d} N(0, \Omega_N).$ Let $Z_{k, it}$ denote the $t$th row of $Z_{k, it} \equiv \frac{X_i}{\sqrt{N_k}} - P_F (E_D(X_i))$ \times 1 \{i \in G_k^0\} - $\frac{1}{N_k} M_{F_0} \sum_{j \in G_k^0} \nu_{ji} E_D(X_{ji})$. That is $Z_{k, it} \equiv \frac{N_k}{N_k} [X_{it} - \frac{1}{T} \sum_{s=1}^{T} \mu_{st} E_D(X_{is})] 1 \{i \in G_k^0\} -$ $\frac{1}{N_k} \sum_{i=1}^{N} \left[1 \{t = s\} - \mu_{st}\right] \sum_{j \in G_k^0} \nu_{ji} E_D(X_{ji})$. Let $Z_{it} = (Z_{i,1,t}, \ldots, Z_{i, k_i, t})'$ and $Z_i = (Z_{1,1,t}, \ldots, Z_{k_i,1,t})$. Then $\sqrt{\lambda T} V_{NT} = \frac{1}{\sqrt{\lambda T}} \sum_{t=1}^{T} \sum_{i=1}^{N} \varepsilon_{it} = \frac{1}{\sqrt{\lambda T}} \sum_{t=1}^{T} \sum_{i=1}^{N} Z_{it} = \sum_{t=1}^{T} \sum_{i=1}^{N} \varepsilon_{it}$, where $\xi_{NT, t} = \frac{1}{\sqrt{\lambda T}} \sum_{t=1}^{T} \sum_{i=1}^{N} Z_{it}$.

By Assumption B.2(iii), $E(\xi_{NT, t} | F_{NT, t-1}) = \frac{1}{\sqrt{\lambda T}} \sum_{t=1}^{T} \sum_{i=1}^{N} Z_{it} E(\varepsilon_{it} | F_{NT, t-1}) = 0$. That is, $\{\xi_{NT, t}, F_{NT, t}\}$ is an m.d.s. By the martingale CLT (e.g., Pollard, 1984, p. 171) and Cramér-Wold device, it suffices to show that

$$Z_1 \equiv \sum_{t=1}^{T} E\left[|\omega' \xi_{NT, t}|^4 | F_{NT, t-1}\right] = o_P(1)$$

and $Z_2 \equiv \sum_{t=1}^{T} |\omega' \xi_{NT, t}|^2 - \omega' \Omega_{NT} \omega = o_P(1)$

(A.30) for any nonrandom $K_0 P \times 1$ real vector $\omega$ with $\|\omega\| = 1$, where $E_{F_{NT, t-1}}$ denote expectation conditional on $F_{NT, t-1}$. Observing that $Z \geq 0$, it suffices to show the first part of (A.30) by showing that $E_{D}(Z_2) = o_P(1)$ by conditional Markov inequality. Noting that $\{Z_{it, \varepsilon_{it}}\}$ are independent across $i$ given $D$ by Assumption B.2(iii) and $E_D(Z_{it, \varepsilon_{it}}) = 0$, we have

$$E_D(Z_1) = \frac{1}{N^2 T^2} \sum_{t=1}^{T} \sum_{i=1}^{N} E_D \left[ \sum_{t=1}^{N} \omega' Z_{it, \varepsilon_{it}} \right]^4$$

$$= \frac{3}{N^2 T^2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} E_D[(\omega' Z_{it})^2 e_{ij}^2 E_D[(\omega' Z_{ij})^2 e_{ij}^2] + \frac{1}{N^2 T^2} \sum_{t=1}^{T} \sum_{i=1}^{N} E_D[(\omega' Z_{it})^4 e_{it}^4]$$

$$= O_P(T^{-1}) + O_P(N^{-1} T^{-1}) = o_P(1).$$

Note that

$$Z_2 = \omega' \sum_{t=1}^{T} \sum_{1 \leq i < j \leq N} Z_{it} Z_{jt} \epsilon_{it} \epsilon_{jt} + \frac{1}{N T} \omega' \sum_{t=1}^{T} \sum_{1 \leq j < i \leq N} Z_{it} Z_{jt} \epsilon_{it} \epsilon_{jt} \omega = Z_{21} + Z_{22},$$

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

\[
E_D (Z_{21}^2) = \frac{1}{N^2 T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{1 \leq i < j \leq N} \sum_{1 \leq k < l \leq N} E_D (\omega' Z_{it} Z'_j \omega' Z'_{ls} Z_{ks} \omega' Z_{kls}) \\
= \frac{1}{N^2 T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{1 \leq i < j \leq N} \sum_{s=1}^{N} \sum_{1 \leq k < l \leq N} E_D (\omega' Z_{it} \omega' Z_{is} \omega' Z_{js} \omega' Z_{ls}) \\
= \frac{1}{N^2 T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{1 \leq i < j \leq N} E_D ([\omega' Z_{it}]^2 \epsilon_{ij}^2) E_D ([\omega' Z_{jt}]^2 \epsilon_{ij}^2) = O_P (T^{-1}).
\]

Similarly, $E_D (Z_{22}^2) = O_P (T^{-1})$. Thus $Z_2 = \omega' \sum_{t=1}^{T} \xi_{NT, t} \hat{\epsilon}'_{NT, t} \omega' \hat{\epsilon}_{NT} \omega = o_P (1)$ and (iii) follows. 

(iv) We verify that \[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} F^{0*} \epsilon_{ij} \epsilon_{ij}' = O_P (\delta_{NT}^{-1})\]. Note that 

\[
E_D \left( \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} F^{0*} \epsilon_{ij} \epsilon_{ij}' \right)^2 = \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} E_D [\text{tr}(F^{0*} \epsilon_{ij} \epsilon_{ij}' F^0)] \\
= \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E_D (\epsilon_{it} \epsilon_{is} \epsilon_{jt} \epsilon_{js}) F^{0*} F^0_t \\
= \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E_D (\epsilon_{it} \epsilon_{is} \epsilon_{jt} \epsilon_{js}) F^{0*} F^0_t \\
+ \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E_D (\epsilon_{is} \epsilon_{is}') E_D (\epsilon_{js} \epsilon_{js}') F^{0*} F^0_s.
\]

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

**Proof of Corollary 3.9.** In view of the fact that 

\[
\sqrt{NT} \text{vec} (\hat{\alpha}^{bc} - \alpha^0) = \sqrt{NT} \left[ \text{vec} (\hat{\alpha} - \alpha^0) - Q_{NT}^{-1} (B_{1, NT} + B_{2, NT} + B_{3, NT}) \right] \\
= -\sqrt{NT} \left[ Q_{NT}^{-1} (\hat{B}_{1, NT} + \hat{B}_{2, NT} + \hat{B}_{3, NT}) - Q_{NT}^{-1} (B_{1, NT} + B_{2, NT} + B_{3, NT}) \right],
\]

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

First, we prove (i) $\hat{B}_{1, NT} - B_{1, NT} = o_P ((NT)^{-1/2})$. Let $\hat{B}_{1, k, NT}$ and $B_{1, k, NT}$ denote the kth block of $\hat{B}_{1, NT}$ and $B_{1, NT}$, respectively. Let $\hat{B}_{1, k, NT} = -\frac{1}{N_k T} \sum_{i \in \hat{G}_k} X_i M_{k} \Phi_{NT} \hat{F} G_i^0$. We prove (i) by showing that (i) $\sqrt{N_k T} (\hat{B}_{1, k, NT} - B_{1, k, NT}) = o_P (1)$ and (ii) $\sqrt{N_k T} (\hat{B}_{1, k, NT} - B_{1, k, NT}) = o_P (1)$ for $k = 1, ..., K_0$. 

21
We make the following decomposition:

\[
\sqrt{N_k T} (\bar{B}_{1,kNT} - B_{1,kNT}) = \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in \hat{G}_k} X_i' M_{\hat{F}} (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) \tilde{F} G \lambda_i^0
\]

\[
+ \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in \hat{G}_k} X_i' (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 H G \lambda_i^0 + \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in \hat{G}_k} X_i' (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) (\hat{F} - F^0 H) G \lambda_i^0
\]

\[
- \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in \hat{G}_k} X_i' F^0 (F^0 H)^{-1} F^0 (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 H G \lambda_i^0
\]

\[
+ \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in \hat{G}_k} X_i' (M_{\hat{F}} - M_{F^0}) (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 H G \lambda_i^0
\]

\[
= b_{11} + b_{12} + b_{13} - b_{14} + b_{15}, \text{ say.}
\]

Define \( \bar{b}_{1s} \) analogously as \( b_{1s} \) for \( s = 1, \ldots, 5 \), 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 \( s = 1, \ldots, 5 \). Let \( \omega_p = (\omega_{1p}, \ldots, \omega_{pp})' \) be an arbitrary \( p \times 1 \) nonrandom vector with \( \|\omega_p\| = 1 \). Let \( X_{ik} \) denote an \( N_k \times T \) matrix with a typical element \( X_{it,l} \) for \( i \in G_k^0, t = 1, \ldots, T \), and \( l = 1, \ldots, p \), where \( X_{it,l} \) denotes the \( l \)th element in \( X_{it} \). Let \( \Lambda_i^0 \) denote an \( N_k \times R \) matrix with a typical row \( \lambda_i^{0r} \) for \( i \in G_k^0 \). Then

\[
|\omega_p' \bar{b}_{11}| = \left| \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \omega_p' X_i' (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 H G \lambda_i^0 + o_P(1) \right|
\]

\[
\leq \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \omega_p' \frac{1}{T^{3/2}} \|X_{ik} (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 H G \lambda_i^0 \| + o_P(1)
\]

\[
\leq \left\{ \sum_{i = 1}^p \left( \frac{1}{T^{3/2}} \|X_{ik} (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 H G \lambda_i^0 \| \right) \right\} O_P(1) O_P(1) + o_P(1) = o_P(1),
\]

where we use the fact that \( \frac{1}{T^{3/2}} \|X_{ik} (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 H G \lambda_i^0 \| = O_P(N_k^{-1/2} + o_P(T^{-(7+4\gamma)/(16+8\gamma)}) \) by Lemma D.4(iii) in Lu and Su (2016). It follows that \( b_{11} = o_P(1) \). Similarly, using \( \frac{1}{T^{3/2}} \|X_{ik} (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 H G \lambda_i^0 \| = O_P(N_k^{-1/2} + o_P(T^{-(7+4\gamma)/(16+8\gamma)}) \), Lemma A.2(i), and Theorem 3.2,

\[
|\omega_p' \bar{b}_{12}| = \left| \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \omega_p' X_i' (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) (\hat{F} - F^0 H) G \lambda_i^0 \right| + o_P(1)
\]

\[
\leq p T^{1/2} \left\{ \sum_{i = 1}^p \left( \frac{1}{T^{3/2}} \|X_{ik} (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) (\hat{F} - F^0 H) G \lambda_i^0 \| \right) \right\} + o_P(1)
\]

\[
= T^{1/2} O_P(1) \left\{ O_P(N_k^{-1/2}) + o_P(T^{-(7+4\gamma)/(16+8\gamma)}) \right\} + o_P(1) = o_P(1).\]
Using the fact that $T^{-1} \| F^{0'} (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 \| = O_P(\delta_{NT}^{-1})$ by Lemma D.4(i) in Lu and Su (2016), we have
\[
|\omega'_{p,13}| = \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i=1}^{P} \omega_{ip} \left[ X_{ik} F_0^{-1} (T^{-1} F^{0'} F_0)^{-1} F^{0'} (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 H G A_k^0 \right] + O_P(1)
\]
\[
\leq \frac{N_k^{1/2}}{T^{1/2}} \frac{1}{N_k^{1/2} T} \left\| X_{ik} F_0 \right\| \left\| (T^{-1} F^{0'} F_0)^{-1} \right\| \frac{1}{T} \left\| F^{0'} (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) F^0 \right\| \frac{1}{N_k^{1/2}} \left\| H G A_k^0 \right\| + o_P(1)
\]
\[
= N_k^{1/2} T^{-1/2} O_P(1) O_P(1) O_P(\delta_{NT}^{-1}) + o_P(1) = o_P(1).
\]

Similarly, using the fact that $T^{-1} \| F^{0'} (N^{-1} \varepsilon' \varepsilon - \Phi_{NT}) \| = O_P(\delta_{NT}^{-1})$, we can readily show that $b_{14} = o_P(1)$ and $b_{15} = o_P(1)$. It follows that $\sqrt{N_k T} (B_{1,kNT} - B_{1,kNT}) = o_P(1)$. This proves (i1).

To show (i2), notice that
\[
\sqrt{N_k T} (\tilde{B}_{1,kNT} - \tilde{B}_{1,kNT})
\]
\[
= \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in \mathcal{G}_k} \left[ X_i' M_\hat{F} (\hat{\Phi}_{NT} - \Phi_{NT}) \hat{F} \hat{G} \hat{\lambda}_i - X_i' M_\hat{F} (\Phi_{NT} - \Phi_{NT}) \hat{F} \hat{G} \hat{\lambda}_i \right]
+ \frac{\hat{N}_k^{-1/2} - N_k^{-1/2}}{T^{3/2}} \sum_{i \in \mathcal{G}_k} X_i' M_\hat{F} (\hat{\Phi}_{NT} \hat{F} \hat{G} \hat{\lambda}_i - G \lambda_i^0)
+ \frac{\hat{N}_k^{-1/2} - N_k^{-1/2}}{T^{3/2}} \sum_{i \in \mathcal{G}_k} X_i' M_\hat{F} (\Phi_{NT} \hat{F} \hat{G} \hat{\lambda}_i)
= b_{16} + b_{17} + b_{18}, \text{ say},
\]
where $G = (T^{-1} \hat{F} \hat{F})^{-1} (N^{-1} \hat{\Lambda}' \hat{\Lambda})^{-1} = (N^{-1} \hat{\Lambda}' \hat{\Lambda})^{-1}$. Using the fact that $\| \Phi_{NT} - \Phi_{NT} \|_{sp} = O_P(\delta_{NT}^{-1} T^{1/(8+4\sigma)})$ by Lemma D.4(ii) in Lu and Su (2016), we have
\[
\| b_{16} \|
\leq \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in \mathcal{G}_k} \left\| X_i' M_\hat{F} (\Phi_{NT} - \Phi_{NT}) \hat{F} \hat{G} \hat{\lambda}_i \right\|_{sp}
\leq \frac{N_k^{1/2}}{T^{1/2}} \left\| \Phi_{NT} - \Phi_{NT} \right\|_{sp} \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in \mathcal{G}_k} \left\| X_i \right\| \left\| \hat{\lambda}_i \right\| \frac{1}{T^{1/2}} \left\| \hat{F} \hat{G} \right\|
= N_k^{1/2} T^{-1/2} O_P(\delta_{NT}^{-1} T^{1/(8+4\sigma)}) O_P(1) = o_P(1),
\]
where we also use the fact that $\frac{1}{N_k^{1/2} T^{1/2}} \sum_{i \in \mathcal{G}_k} \left\| X_i \right\| \left\| \hat{\lambda}_i \right\| \leq \left\{ \frac{1}{N_k T} \sum_{i \in \mathcal{G}_k} \left\| X_i \right\|^2 \right\}^{1/2} \left\{ \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in \mathcal{G}_k} \left\| \hat{\lambda}_i \right\|^2 \right\}^{1/2} + o_P(1) = o_P(1)$ by the mean-square convergence of $\{ \hat{\lambda}_i \}$. Note that
\[
\hat{G} \hat{\lambda}_i - G \lambda_i^0 = (T^{-1} \hat{F} \hat{F})^{-1} (N^{-1} \hat{\Lambda}' \hat{\Lambda})^{-1} \hat{\lambda}_i - (T^{-1} \hat{F} \hat{F})^{-1} (N^{-1} \hat{\Lambda}' \hat{\Lambda})^{-1} (N^{-1} \hat{\Lambda}' \hat{\Lambda})^{-1} H^{-1} \lambda_i^0
= \varsigma_1 \hat{\lambda}_i + \varsigma_2 (\hat{\lambda}_i - H^{-1} \lambda_i^0),
\]
where $\varsigma_1 = (T^{-1} \hat{F} \hat{F})^{-1} (N^{-1} \hat{\Lambda}' \hat{\Lambda})^{-1} - (T^{-1} \hat{F} \hat{F})^{-1} (N^{-1} \hat{\Lambda}' \hat{\Lambda})^{-1} (N^{-1} \hat{\Lambda}' \hat{\Lambda})^{-1} = O_P(\delta_{NT}^{-1})$ and $\varsigma_2 = (T^{-1} \hat{F} \hat{F})^{-1} (N^{-1} \hat{\Lambda}' \hat{\Lambda})^{-1} (N^{-1} \hat{\Lambda}' \hat{\Lambda})^{-1} = O_P(1)$. By the triangle and Jensen inequalities, the fact that
Consequently, we have arguments as used in the proof of Lemma D.3(vi) in Lu and Su (2016), we can show that

\[ \frac{1}{N_k} \sum_{i \in G_k} \left\| \tilde{X}_i \right\|^2 = O_P \left( T^{-1} \right) \text{ and } \| \Phi_{NT} \|_{sp} = O_P \left( 1 \right), \]

we have

\[
\|b_{17}\| = \frac{1}{N_k^{1/2}T^{3/2}} \left\| \sum_{i \in G_k} X_i' M_k \Phi_{NT} \hat{F} \left( \hat{G} \tilde{X}_i - G \lambda_i^0 \right) \right\|_{sp} \leq \frac{1}{N_k^{1/2}T^{3/2}} \left\| \sum_{i \in G_k} X_i' M_k \Phi_{NT} \hat{F} \tilde{G} \right\|_{sp} \leq \frac{1}{N_k^{1/2}T^{3/2}} \| \Phi_{NT} \|_{sp} \left( \sum_{i \in G_k} \| X_i \|^2 \right)^{1/2} \left( \sum_{i \in G_k} \| \lambda_i - H^{-1} \lambda_i^0 \|^2 \right)^{1/2} = \frac{N_k^{1/2}}{T^{3/2}} \| \Phi_{NT} \|_{sp} \left( O_P \left( 1 \right) \right) + O_P \left( 1 \right) = O_P \left( 1 \right).
\]

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

\[
\|b_{18}\| = \frac{\| \hat{N}_k - N_k \|}{N_k^{1/2}N_k^{1/2} \left( N_k^{1/2} + N_k^{1/2} \right)^{3/2}} \left\| \sum_{i \in G_k} X_i' M_k \hat{\Phi}_{NT} \hat{F} \hat{G} \lambda_i \right\| \leq O_P \left( N_k^{-3/2}T^{-3/2} \right) \left\| \sum_{i \in G_k} \| X_i \| \| \lambda_i \| \right\| \left\| \hat{F} \hat{G} \right\| \| \hat{\Phi}_{NT} \|_{sp} = O_P \left( N_k^{-3/2}T^{-3/2} \right) O_P \left( \delta_{NT}^{-1} \right) O_P \left( 1 \right) = O_P \left( N_k^{-3/2}T^{-1/2} \right).
\]

Consequently, we have \( \sqrt{N_kT} \left( \hat{B}_{2,kNT} - B_{2,kNT} \right) = o_P \left( 1 \right) \).

Next, we prove (ii) \( \hat{B}_{2,NT} - B_{2,NT} = o_P \left( \left( NT \right)^{-1/2} \right) \). Let \( \hat{B}_{2,kNT,l} = -\frac{1}{N_k} \text{tr} \{ F^0 \left( T^{-1} F_0^0 F_0^0 \right)^{-1} \left( N^{-1} \Lambda^0 \Lambda^0 \right)^{-1} \Lambda^0 \Psi_{kNT} \left[ X_l \left( k \right) - \tilde{X}_i^* \left( k \right) \right] \} \} \) for \( l = 1, \ldots, p \). We prove (ii) by showing that (ii1) \( \sqrt{N_kT} \left( \hat{B}_{2,kNT,l} - B_{2,kNT,l} \right) = o_P \left( 1 \right) \) and (ii2) \( \sqrt{N_kT} \left( \hat{B}_{2,kNT,l} - B_{2,kNT,l} \right) = o_P \left( 1 \right) \) for \( k = 1, \ldots, K_0 \) and \( l = 1, \ldots, p \). Using arguments as used in the proof of Lemma D.3(vi) in Lu and Su (2016), we can show that

\[
N^{-3/2} \left\| \Lambda^0 \left( T^{-1} \epsilon \epsilon \left( k \right) - \Psi_{kNT} \right) \left[ X_l \left( k \right) - \tilde{X}_i^* \left( k \right) \right] \right\| = O_P \left( \delta_{NT}^{-1} \right). \]

It follows that

\[
\frac{1}{N_k} \frac{1}{T^{1/2}} \left\{ \text{tr} \{ F^0 \left( T^{-1} F_0^0 F_0^0 \right)^{-1} \left( N^{-1} \Lambda^0 \Lambda^0 \right)^{-1} \Lambda^0 \Psi_{kNT} \left[ X_l \left( k \right) - \tilde{X}_i^* \left( k \right) \right] \} \right\} \leq \frac{N_k^{1/2}}{T^{3/2}} \left\| F^0 \left( T^{-1} F_0^0 F_0^0 \right)^{-1} \left( N^{-1} \Lambda^0 \Lambda^0 \right)^{-1} \Lambda^0 \left( T^{-1} \epsilon \epsilon \left( k \right) - \Psi_{kNT} \right) \left[ X_l \left( k \right) - \tilde{X}_i^* \left( k \right) \right] \right\| \leq O_P \left( 1 \right) O_P \left( \delta_{NT}^{-1} \right) = O_P \left( 1 \right). \]
This proves (ii1). The proof of (ii2) is analogous to that of (i2) except that now \( \left\| \Psi_{kNT} - \Psi_{kNT} \right\|_{sp} = o_P(\delta_{NT}^{-1}N^{1/(8+4\sigma)}) \) 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 \( \text{plim inf}_{(N,T) \to \infty} \left[ V(R, \hat{\beta}_1(R)) - V(R_0, \hat{\beta}_1(R_0)) \right] \) \( \geq c_R \) for each \( R \) with \( 1 \leq R < R_0 \),

(ii) \( V(R, \hat{\beta}_1(R)) - V(R_0, \hat{\beta}_1(R_0)) = O_P(\delta_{NT}^{-2}) \) 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 \( V(R) = V(R, \hat{\beta}_1(R)) \) for all \( R \). Note that

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

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

For case (a), by Lemma A.10(i), \( V(R) / V(R_0) > 1 + \epsilon_0 \) and thus \( \ln \left[ \frac{V(R)}{V(R_0)} \right] \geq \epsilon_0 / 2 \) 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

\[
P \left( IC_{1NT}(R) - IC_{1NT}(R_0) > 0 \right) = P \left( \frac{\ln [V(R)/V(R_0)]}{(R - R_0) \rho_{1NT}} > 0 \right) = P \left( O_P(1) + (R - R_0) \rho_{1NT} \sigma^2_{NT} > 0 \right) \to 1
\]

as \((N,T) \to \infty \) for any \( R_0 < R \leq R_{\text{max}} \). 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 \( R \in [1, R_{\text{max}}] \) 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 \( \hat{\sigma}_{G_0}^2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \hat{\varepsilon}_{it}^2 \). Then \( \max_{K_0 \leq K \leq K_{\text{max}}} \left| \hat{\sigma}_{G(K_0,\kappa)}^2 - \sigma_{G_0}^2 \right| = O_P(\delta_{NT}^{-2}) \).

**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

\[
IC_{2NT}(K_0, \kappa) = \ln \left( \hat{\sigma}_{G(K_0,\kappa)}^2 \right) + \rho_{2NT} p K_0
\]

\[
= \ln \left\{ \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k(K_0,\kappa)} \left[ Y_{it} - \hat{\alpha}_{G_k(K_0,\kappa)} Z_{it} \right] \right\} + o(1)
\]

\[
\to \ln \left( \sigma_{G_0}^2 \right).
\]
Combining the results in Cases 1 and 2, we complete the proof. Noting that
\[
\hat{\sigma}^2_{G_k(K,\kappa)} = \frac{1}{NT} \sum_{k=1}^{K} \sum_{i \in G_k(K,\kappa)} \sum_{t=1}^{T} \left[ Y_{it} - \hat{\alpha}'_{G_k(K,\kappa)} X_{it} - \hat{\lambda}_i^{(K,\kappa)^\prime} \tilde{F}_i^{(K,\kappa)} \right]^2
\]
\[
\geq \min_{1 \leq k < K_0} \frac{1}{NT} \sum_{k=1}^{K} \sum_{i \in G_k(K,\kappa)} \sum_{t=1}^{T} \left[ Y_{it} - \hat{\alpha}'_{G_k,k} X_{it} - \lambda_i^{(K,\kappa)^\prime} \tilde{F}_i^{(K,\kappa)} \right]^2 = \min_{1 \leq k < K_0} \frac{\hat{\sigma}^2_{G(k)}}{G(k)} \min_{1 \leq k < K_0} \frac{\hat{\sigma}^2_{G(k)}}{G(k)}
\]
we have by Assumptions A.7-A.8 and the Slutsky Lemma
\[
\min_{1 \leq k < K_0} IC_{2NT}(K, \kappa) \geq \min_{1 \leq k < K_0} \frac{\hat{\sigma}^2_{G(k)}}{G(k)} \min_{1 \leq k < K_0} \frac{\hat{\sigma}^2_{G(k)}}{G(k)} \min_{1 \leq k < K_0} \frac{\hat{\sigma}^2_{G(k)}}{G(k)} = \min_{1 \leq k < K_0} \frac{\hat{\sigma}^2_{G(k)}}{G(k)} \min_{1 \leq k < K_0} \frac{\hat{\sigma}^2_{G(k)}}{G(k)} \min_{1 \leq k < K_0} \frac{\hat{\sigma}^2_{G(k)}}{G(k)} \min_{1 \leq k < K_0} \frac{\hat{\sigma}^2_{G(k)}}{G(k)}
\]
It follows that \( P(\min_{K \in \Omega} IC_{2NT}(K, \kappa) > IC_{2NT}(K_0, \kappa)) \rightarrow 1 \).

Case 2: Under-fitted model \((K > K_0)\). Let \( K \in \Omega_+ \). By Lemma A.13 and the fact that \( \hat{\sigma}^2_{NT} \rho_{2NT} \rightarrow \infty \) under Assumption A.8, we have
\[
P \left( \min_{K \in \Omega_+} IC_{2NT}(K, \kappa) > IC_{2NT}(K_0, \kappa) \right)
= P \left( \min_{K \in \Omega_+} \delta^2_{NT} \left( \hat{\sigma}^2_{G(k)} - \hat{\sigma}^2_{G(k_0)} \right) / \hat{\sigma}^2_{G(k_0)} + \delta^2_{NT} \rho_{2NT} (K - K_0) > 0 \right)
= P \left( \min_{K \in \Omega_+} \delta^2_{NT} \left( \hat{\sigma}^2_{G(k)} - \hat{\sigma}^2_{G(k_0)} \right) / \hat{\sigma}^2_{G(k_0)} + \delta^2_{NT} \rho_{2NT} (K - K_0) + o_P(1) > 0 \right)
\rightarrow 1 \text{ as } (N, T) \rightarrow \infty.
\]
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 \( \frac{1}{\sqrt{T}} \sum_{i=1}^{N} b_i^2 X_i^\prime M F \varepsilon_i = \frac{1}{\sqrt{T}} \sum_{i=1}^{N} b_i^2 X_i^\prime \varepsilon_i \) as \( T^{-1/2} F F = I_{R_0} \), we have by Assumption A.1(2(iii)),
\[
\left| A_1 (b) \right| = \left| \frac{1}{\sqrt{T}} \sum_{i=1}^{N} b_i^2 X_i^\prime M F \varepsilon_i \right| \leq \left( \frac{1}{N} \sum_{i=1}^{N} b_i^2 \right) \left( \frac{1}{N} \sum_{i=1}^{N} \| X_i \|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \| Y_i \|^2 \right)^{1/2}
\]
\[
= \sqrt{R_0} \max_{1 \leq i \leq N} T^{-1/2} \| X_i \| \frac{1}{\sqrt{T}} \sum_{i=1}^{N} \| b_i \| \| F \varepsilon_i \|
\]
\[
\leq O_P(1) \left( \frac{1}{N} \sum_{i=1}^{N} \| b_i \|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \| F \varepsilon_i \|^2 \right)^{1/2}
\]
\[
= O_P(1) O_P(1) \frac{1}{N} \sum_{i=1}^{N} \| F \varepsilon_i \|^2 = o_P(1) \text{ uniformly in } (b, F)
\]

Noting that \( T^{-1/2} \| b_i^2 X_i^\prime F \| \leq \sqrt{T_0} T^{-1/2} \| b_i \| \| X_i \| \) as \( T^{-1/2} F F = I_{R_0} \), we have by Assumption A.1(2(iii)),
\[
\left| A_2 (b, F) \right| = \left| \frac{1}{\sqrt{T}} \sum_{i=1}^{N} b_i^2 X_i^\prime M F \varepsilon_i \right| \leq \left( \frac{1}{N} \sum_{i=1}^{N} b_i^2 \right) \left( \frac{1}{N} \sum_{i=1}^{N} \| X_i \|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \| Y_i \|^2 \right)^{1/2}
\]
\[
= \sqrt{R_0} \max_{1 \leq i \leq N} T^{-1/2} \| X_i \| \frac{1}{\sqrt{T}} \sum_{i=1}^{N} \| b_i \| \| F \varepsilon_i \|
\]
\[
\leq O_P(1) \left( \frac{1}{N} \sum_{i=1}^{N} \| b_i \|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \| F \varepsilon_i \|^2 \right)^{1/2}
\]
\[
= O_P(1) O_P(1) \frac{1}{N} \sum_{i=1}^{N} \| F \varepsilon_i \|^2 = o_P(1) \text{ uniformly in } (b, F)
\]
where we have used the fact that \( \frac{1}{NT} \sum_{i=1}^{N} \|F'\varepsilon_i\|^2 = O_P \left( N^{-1/2} + T^{-1/2} \right) \) uniformly in \( F \) by the proof of Lemma A.1(i) in Bai (2009). Then (i) follows.

Using \( P_{F0} = F^0(F^0'F^0)^{-1}F^0 \) and the fact that \( \frac{1}{NT} \sum_{i=1}^{N} E \|F^0\varepsilon_i\|^2 = O(1) \), we have by Markov inequality and Assumptions A.1 (i) and (vii)

\[
\frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i P_{F0} \varepsilon_i \leq \frac{1}{T} \left\| (T^{-1}F^0F^0)^{-1} \right\| \frac{1}{NT} \sum_{i=1}^{N} \|F^0\varepsilon_i\|^2 = O_P \left( T^{-1} \right).
\]

That is, (iv) follows. ■

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

\[
\hat{F}V_{NT} - F^0H\hat{V}_{NT} = \left[ \frac{1}{NT} \sum_{i=1}^{N} \left( -X_i\hat{\beta}_i + F^0\lambda_i^0 + \varepsilon_i \right) \left( -X_i\hat{\beta}_i + F^0\lambda_i^0 + \varepsilon_i \right)' \right] \hat{F} - F^0H\hat{V}_{NT}
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} X_i\hat{\beta}_i\hat{\beta}_i'X_i' \hat{F} - \frac{1}{NT} \sum_{i=1}^{N} X_i\hat{\beta}_i\lambda_i^0F^0\hat{F} - \frac{1}{NT} \sum_{i=1}^{N} X_i\hat{\beta}_i\varepsilon_i' \hat{F} + \frac{1}{NT} \sum_{i=1}^{N} X_i\hat{\beta}_i\varepsilon_i' \hat{F}
\]

\[
+ \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i\lambda_i^0F^0\hat{F} + \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i\varepsilon_i' \hat{F}
\]

\[
\equiv a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8, \text{ say. (B.1)}
\]

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

\[
T^{-1/2} \|a_2\| \leq \sqrt{R_0} \max_{1 \leq i \leq N} T^{-1/2} \|X_i\| T^{-1/2} \|\lambda_i^0\| \frac{1}{NT} \sum_{i=1}^{N} \|\hat{b}_i\| \|\lambda_i^0\| \frac{1}{NT} \sum_{i=1}^{N} \|\hat{b}_i\|^2 \frac{1}{NT} \sum_{i=1}^{N} \|\lambda_i^0\|^2 \frac{1}{NT} \sum_{i=1}^{N} \|\lambda_i^0\|^2 \frac{1}{NT} \sum_{i=1}^{N} \|\varepsilon_i\|^2 = O_P (\eta^2_{NT}).
\]

Analogously, we can show that \( T^{-1/2} \|a_3\| = O_P (\eta^2_{NT}) \) for \( s = 3, 4, 5 \). By Bai and Ng (2002), \( T^{-1/2} \|a_4\| = O_P \left( \delta^{-1}_{NT} \right) \) for \( s = 6, 7, 8 \). It follows that \( T^{-1/2} \|\hat{F}V_{NT} - F^0H\hat{V}_{NT}\| = O_P \left( \eta_{NT} + \delta^{-1}_{NT} \right) \). 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 \( T^{-1} \left\| \hat{F} - F^0H \right\|^2 = O_P \left( \eta^2_{NT} + \delta^{-2}_{NT} \right) \).

(ii) Writing \( T^{-1}(\hat{F} - F^0H)'F^0H = \sum_{l=1}^{8} T^{-1}V_{NT}^{-1}\alpha_l'F^0H \equiv \sum_{l=1}^{8} A_l \), it suffices to show that \( \|A_l\| = O_P (\eta_{NT}) + O_P (\delta^{-2}_{NT}) \) for \( l = 1, 2, \ldots, 8 \). By the definitions of \( a_i \)‘s in (B.1), we have \( \|A_1\| \leq \{T^{-1/2} \|a_1\|\} \times \|V_{NT}^{-1}\| \|F^0H\| = O_P (\eta^2_{NT}) \) and \( \|A_2\| \leq \{T^{-1/2} \|a_2\|\} \|V_{NT}^{-1}\| \|T^{-1/2} \|F^0H\| = O_P (\eta_{NT}) \). For \( A_3 \), we have

\[
\|A_3\| = N^{-1}T^{-2} \left\| V_{NT}^{-1} \hat{F}' \sum_{i=1}^{N} \varepsilon_i\hat{b}_iX_i'F^0H \right\|
\]

\[
\leq N^{-1}T^{-2} \left\| V_{NT}^{-1} \left( \hat{F} - F^0H \right)' \sum_{i=1}^{N} \varepsilon_i\hat{b}_iX_i'F^0H \right\| + N^{-1}T^{-2} \left\| V_{NT}^{-1}H' \sum_{i=1}^{N} F^0\varepsilon_i\hat{b}_iX_i'F^0H \right\|
\]

\[
\equiv A_{3,1} + A_{3,2}, \text{ say.}
\]
Noting that $N^{-1} \sum_{i=1}^{N} \| \varepsilon_i \| \left\| \hat{b}_i \right\| \leq \left\{ N^{-1} \sum_{i=1}^{N} \| \varepsilon_i \| \right\}^{1/2} \left\{ N^{-1} \sum_{i=1}^{N} \| \hat{b}_i \| \right\}^{1/2} = O_P (\eta_{NT})$, we have

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

Similarly,

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

Hence $\| A_3 \| = O_P (\eta_{NT} + \delta_{NT}^{-1}) O_P (\eta_{NT})$. As in the analysis of $A_2$ and $A_3,2$, we can readily show that $\| A_4 \| = N^{-1} T^{-2} \left\| V_{NT}^{-1} \bar{F} \sum_{i=1}^{N} X_i \hat{b}_i \Lambda^0 \right\| = O_P (\eta_{NT})$ and $\| A_5 \| = N^{-1} T^{-2} \left\| V_{NT}^{-1} \bar{F} \sum_{i=1}^{N} X_i \hat{b}_i \varepsilon_i \right\| = O_P \left( T^{-1/2} \eta_{NT} \right)$. In addition,

$$\| A_6 \| = \frac{1}{N T^2} \left\| V_{NT}^{-1} \bar{F} \sum_{i=1}^{N} \varepsilon_i \Lambda^0 \right\| = N^{-1} T^{-2} \left\| V_{NT}^{-1} \bar{F} \sum_{i=1}^{N} \varepsilon_i \Lambda^0 \right\| = O_P (\eta_{NT}^{1/2} + O_P (\delta_{NT}^{-2})).$$

$$\| A_7 \| = N^{-1} T^{-2} \left\| V_{NT}^{-1} \bar{F} \sum_{i=1}^{N} \varepsilon_i \varepsilon_i \right\| = N^{-1} T^{-2} \left\| V_{NT}^{-1} \bar{F} \sum_{i=1}^{N} \varepsilon_i \varepsilon_i \right\| = O_P (\eta_{NT}^{1/2} + O_P (\delta_{NT}^{-2})).$$

It follows that $T^{-1} (\bar{F} - F^0 H) (\bar{F} - F^0 H)^T = O_P (\eta_{NT} + \delta_{NT}^{-1}) O_P (\eta_{NT})$, the result follows from (i) and (ii).

(iii) Noting that $T^{-1} (\bar{F} - F^0 H) (\bar{F} - F^0 H)^T = T^{-1} (\bar{F} - F^0 H) (\bar{F} - F^0 H) + T^{-1} (\bar{F} - F^0 H) (\bar{F} - F^0 H) + T^{-1} (\bar{F} - F^0 H) (\bar{F} - F^0 H)$, the result follows from (i) and (ii).

(v) By (A.1), the result follows from (i) and the fact that $T^{-1/2} \left\| F^0 H \right\| = O_P (1)$. ■
\[ R_{11} \leq \frac{1}{NT^2} \sum_{i=1}^{N} \left\| X_i' (F - F^0 H)(\hat{F} - F^0 H)' \varepsilon_i \right\|^2 \]
\leq T^{-2} \left\| F - F^0 H \right\|^4 \left\| X_i \right\|^2 \leq \frac{1}{NT^2} \sum_{i=1}^{N} \left\| X_i \right\|^2 \left\| \varepsilon_i \right\|^2 = O_P (\eta_{NT}^4) + O_P \left( T^{-1} \eta_{NT}^2 + \delta_{NT}^4 \right) , \quad \text{and} \]
\[ R_{12} \leq \frac{1}{NT^2} \sum_{i=1}^{N} \left\| X_i'(\hat{F} - F^0 H)H'F^0 \varepsilon_i \right\| \]
\leq T^{-2} \left\| \hat{F} - F^0 H \right\|^2 \| H \|^2 \left\| X_i \right\|^2 \left\| F^0 \varepsilon_i \right\|^2 = T^{-1} O_P \left( \eta_{NT}^2 + \delta_{NT}^{-2} \right) .
\]

For \( R_{13} \), we apply (B.1) and the \( C_r \) inequality to obtain
\[ R_{13} = \frac{1}{N T^2} \sum_{i=1}^{N} \left\| X_i' F^0 H (\hat{F} - F^0 H)' \varepsilon_i \right\|^2 \leq 8 \frac{1}{N T^2} \sum_{i=1}^{N} \left\| X_i' F^0 H V^{-1}_{NT} a_i \varepsilon_i \right\|^2 = 8 \sum_{i=1}^{N} R_{13,i} , \quad \text{say.} \]

For \( R_{13,l} \), we use the fact that \( \| a_1 \| = O_P (\eta_{NT}^2) \) to obtain the rough bound \( R_{12,l} = O_P \left( \eta_{NT}^4 \right) \). Next by Assumption A.2 \( (i) \)
\[ R_{13,2} = \frac{1}{N^3 T^6} \sum_{i=1}^{N} \left\| X_i' F^0 H V^{-1}_{NT} \hat{F}' F^0 \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j X_j \varepsilon_i \right\|^2 \]
\leq T^{-3} \left\| F^0 H V^{-1}_{NT} \hat{F}' F^0 \right\|^2 \max_{1 \leq i \leq N} T^{-1} \left\| X_i \right\|^2 \frac{1}{N T^2} \sum_{i=1}^{N} \left\| \lambda_j^0 \varepsilon_j X_j \right\|^2 \leq \frac{8}{N^2 T^2} \sum_{i=1}^{N} \left\| \lambda_j^0 \varepsilon_j X_j \right\|^2 = 8 \sum_{i=1}^{N} R_{13,l} , \quad \text{say.} \]

Similarly, we can show that \( R_{13,l} = O_P (T^{-1} \eta_{NT}^2) \) for \( l = 3, 4, 5 \). For \( R_{12,6} \), we have
\[ R_{12,6} = \frac{1}{N^3 T^6} \sum_{i=1}^{N} \left\| X_i' F^0 H V^{-1}_{NT} \hat{F}' F^0 \sum_{j=1}^{N} \lambda_j \varepsilon_j F^0 \varepsilon_i \right\|^2 \]
\leq \frac{2}{N^3 T^6} \sum_{i=1}^{N} \left\| X_i' F^0 H V^{-1}_{NT} \hat{F}' F^0 \sum_{j=1}^{N} \lambda_j \varepsilon_j F^0 \varepsilon_i \right\|^2 + \frac{2}{N^3 T^6} \sum_{i=1}^{N} \left\| X_i' F^0 H V^{-1}_{NT} (\hat{F} - F^0 H)' \sum_{j=1}^{N} \lambda_j F^0 \varepsilon_i \right\|^2 .
\]

The first term is bounded by
\[ N^{-1} T^{-2} \left\{ T^{-1} \left\| F^0 H V^{-1}_{NT} \hat{F}' \right\|^2 \right\} \left\{ \frac{1}{N T^2} \sum_{i=1}^{N} \left\| X_i \right\|^2 \left\| F^0 \varepsilon_i \right\|^2 \right\} \left\{ \frac{1}{N T} \left\| F^0 \varepsilon' \lambda^0 \right\|^2 \right\} = O_P \left( N^{-1} T^{-2} \right) ,
\]
and the second term is bounded by

\[ N^{-1} - T^{-1} \left\| F^0 H V_{NT}^{-1} \right\| T^{-1} \left\| \hat{F} - F^0 H \right\|^2 \left\{ \frac{1}{N^2 T^2} \sum_{i=1}^N \left\| X_i \right\|^2 \left\| F^0 \right\|^2 \right\} \frac{1}{N T} \left\| \sum_{j=1}^N \varepsilon_j \lambda_j^0 \varepsilon_{i,j} \right\|^2 \]

\[ = N^{-1} T^{-1} \left[ O_P \left( \eta_{NT}^2 \right) + O_P \left( \delta_{NT}^2 \right) \right]. \]

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,

\[ \mathcal{R}_{13,7} = \frac{1}{N^3 T^2} \sum_{i=1}^N \left\| X_i F^0 H V_{NT}^{-1} \hat{F} \left( \sum_{j=1}^N \lambda_j^0 \varepsilon_{i,j} \varepsilon_{i,j} \right) \right\|^2 \]

\[ \leq N^{-1} T^{-1} \max_{1 \leq i \leq N} T^{-1} \left\| X_i \right\|^2 \left\{ T^{-3} \left\| F^0 H V_{NT}^{-1} \hat{F} \right\|^2 \right\} \frac{1}{N^2 T^2} \sum_{i=1}^N \left\| \lambda_j^0 \varepsilon_{i,j} \varepsilon_{i,j} \right\|^2 \]

\[ = N^{-1} T^{-1} O_P \left( 1 \right) O_P \left( 1 \right) O_P \left( 1 \right) = O_P \left( N^{-1} T^{-1} \right). \]

Similarly,

\[ \mathcal{R}_{13,8} = \frac{1}{N^3 T^2} \sum_{i=1}^N \left\| X_i F^0 H V_{NT}^{-1} \hat{F} \left( \sum_{j=1}^N \lambda_j^0 \varepsilon_{i,j} \varepsilon_{i,j} \right) \right\|^2 \]

\[ \leq \left\{ \frac{1}{T^2} \left\| F^0 H V_{NT}^{-1} \hat{F} \right\|^2 \right\} \frac{1}{N^2 T^2} \left\| \max_{1 \leq i \leq N} T^{-1} \left\| X_i \right\| \right\|^2 \frac{1}{T^2} \left\| H^0 \right\|^2 \left\{ \frac{1}{N T} \sum_{i=1}^N \left\| \lambda_j^0 \varepsilon_{i,j} \varepsilon_{i,j} \right\|^2 \right\} \]

\[ = O_P \left( 1 \right) O_P \left( 1 \right) O_P \left( N^{-2} + T^{-2} \right) O_P \left( 1 \right) = O_P \left( \delta_{NT}^2 \right). \]

In addition, using \( \hat{F}^\prime \hat{F} / T = I_R \) and Lemma A.2(i),

\[ \mathcal{R}_{14} \leq \frac{1}{N T^4} \sum_{i=1}^N \left\| X_i F^0 H \left( I_r - (T^{-1} H^r F^0 H^0)^{-1} \right)^\prime H^r F^0 \varepsilon_{i} \right\|^2 \]

\[ \leq T^{-1} \left\| I_r - (T^{-1} H^r F^0 H^0)^{-1} \right\| T^{-1} \left\| F^0 H \right\|^2 \frac{1}{N T} \sum_{i=1}^N \left\| X_i \right\|^2 \left\| F^0 \varepsilon_{i} \right\|^2 \]

\[ = T^{-1} \left[ O_P \left( \eta_{NT}^2 \right) + O_P \left( \delta_{NT}^2 \right) \right]. \]

It follows that \( \mathcal{R}_{14} = T^{-1} \left[ O_P \left( \eta_{NT}^2 \right) + O_P \left( \delta_{NT}^2 \right) \right] \). 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), \( \frac{1}{N} X_i^\prime M_{\hat{F}} (\hat{F} - F^0 H) H^{-1} \lambda_i^0 = \sum_{i=1}^8 \frac{1}{N} X_i^\prime M_{\hat{F}} g_i V_{NT}^{-1} H^{-1} \lambda_i^0 = \sum_{i=1}^8 R_{2i}, \) say. Noting that \( R_{2i} = -\frac{1}{N T^2} \sum_{j=1}^N \sum_{i=1}^N X_i^\prime M_{\hat{F}} X_i \hat{b}_j X_j^\prime \hat{b}_j X_j^\prime \hat{F} V_{NT}^{-1} H^{-1} \lambda_i^0 = -\frac{1}{N T^2} \sum_{j=1}^N X_i^\prime M_{\hat{F}} X_j \nu_j \hat{b}_j \), we have \( \mathcal{R}_2 = N^{-1} \sum_{i=1}^N \left\| R_{2i} \right\|^2 \leq 7 \sum_{i=1,j=1}^N \sum_{i=1}^N \left\| R_{2i} \right\|^2 \). Following the analyses of \( \mathcal{R}_{11} \) and \( \mathcal{R}_{13}, \) we can readily show that

\[ \mathcal{R}_{21} = \frac{1}{N^3 T^4} \sum_{i=1}^N \left\| X_i^\prime M_{\hat{F}} \sum_{j=1}^N X_j \hat{b}_j \hat{b}_j X_j^\prime \hat{F} V_{NT}^{-1} H^{-1} \lambda_i^0 \right\|^2 \]

\[ \leq T^{-1} \left\| \hat{F} V_{NT}^{-1} H^{-1} \right\|^2 \left\{ \max_{1 \leq i \leq N} T^{-3} \left\| X_i \right\| \right\} \left\{ \frac{1}{N} \sum_{i=1}^N \left\| \lambda_i^0 \right\|^2 \right\} \left\{ \frac{1}{N} \sum_{j=1}^N \left\| \hat{b}_j \right\|^2 \right\} \]

\[ = O_P \left( \eta_{NT}^4 \right), \]
and

\[
\mathcal{R}_{23} = \frac{1}{N^3 T^4} \sum_{i=1}^{N} \left\| X_i' M_{\hat{F}} \sum_{j=1}^{N} X_j \hat{b}_j \lambda_j^{0'} F^{0'} \hat{F} V^{-1}_{NT} H^{-1} \lambda_i^0 \right\|^2
\]

\[
\leq T^{-1} \left\{ \max_{1 \leq i \leq N} T^{-1} \left\| X_i \right\|^2 \right\} T^{-2} \left\| F^{0'} \hat{F} V^{-1}_{NT} H^{-1} \right\|^2 \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\| \lambda_i^0 \right\|^2 \right\} \left\{ \frac{1}{N} \sum_{j=1}^{N} \left\| \hat{b}_j \right\| \left\| \lambda_j^0 \right\| \right\}^2
\]

\[
\leq T^{-1} O_P \left( 1 \right) \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\| \lambda_i^0 \right\|^2 \right\}^2 \left\{ \frac{1}{N} \sum_{j=1}^{N} \left\| \hat{b}_j \right\| \right\}^2 = T^{-1} O_P \left( \eta_{NT}^2 \right).
\]

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

\[
\mathcal{R}_{24} \leq \frac{1}{N^3 T^4} \sum_{i=1}^{N} \left\| X_i' M_{\hat{F}} \left( F^{0} - \hat{F} H^{-1} \right) \sum_{j=1}^{N} \lambda_j^0 \hat{b}_j X_j \hat{F} V^{-1}_{NT} H^{-1} \lambda_i \right\|^2
\]

\[
\leq T^{-1} \left\| F^{0} - \hat{F} H^{-1} \right\|^2 T^{-1} \left\| \hat{F} V^{-1}_{NT} H^{-1} \right\|^2 \left\{ \frac{1}{N T} \sum_{i=1}^{N} \left\| X_i \right\|^2 \left\| \lambda_i^0 \right\|^2 \right\} \left\{ \frac{1}{N T^2} \sum_{j=1}^{N} \left\| \lambda_j^0 \right\| \left\| \hat{b}_j \right\| \left\| X_j \right\| \right\}^2
\]

\[
\leq O_P \left( \eta_{NT}^2 + \delta_{NT}^4 \right) O_P \left( 1 \right) \left\{ \frac{1}{N T} \sum_{i=1}^{N} \left\| \lambda_i^0 \right\|^2 \left\| X_i \right\|^2 \right\}^2 \left\{ \frac{1}{N} \sum_{j=1}^{N} \left\| \hat{b}_j \right\| \right\}^2 = O_P \left( \eta_{NT}^2 + \delta_{NT}^4 \right) O_P \left( \eta_{NT}^2 \right).
\]

By the fact that \( M_{\hat{F}} = (M_{\hat{F}} - M_{F_0}) + I_T - P_{F_0} \) and Cauchy-Schwarz inequality,

\[
\mathcal{R}_{25} \leq \frac{3}{N^3 T^4} \sum_{i=1}^{N} \left\| X_i' M_{\hat{F}} \left( F^{0} - \hat{F} H^{-1} \right) \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j X_j \hat{F} V^{-1}_{NT} H^{-1} \lambda_i \right\|^2
\]

\[
+ \frac{3}{N^3 T^4} \sum_{i=1}^{N} \left\| X_i' F^{0} \left( F^{0'} F^{0} \right)^{-1} \sum_{j=1}^{N} \lambda_j^{0'} \varepsilon_j^{0'} X_j \hat{F} V^{-1}_{NT} H^{-1} \lambda_i \right\|^2
\]

\[
= \eta_{NT}^2 O_P \left( \eta_{NT}^2 + \delta_{NT}^4 \right) + O_P \left( T^{-1} \eta_{NT}^2 \right) + O_P \left( T^{-1} \eta_{NT}^2 \right) = O_P \left( T^{-1} \eta_{NT}^2 \right) + O_P \left( \eta_{NT}^4 + \delta_{NT}^4 \right),
\]

and

\[
\mathcal{R}_{26} \leq \frac{1}{N^3 T^4} \sum_{i=1}^{N} \left\| X_i' M_{\hat{F}} \left( F^{0} - \hat{F} H^{-1} \right) \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j \hat{F} V^{-1}_{NT} H^{-1} \lambda_i \right\|^2
\]

\[
\leq \frac{2}{N^3 T^4} \sum_{i=1}^{N} \left\| X_i' M_{\hat{F}} \left( F^{0} - \hat{F} H^{-1} \right) \Lambda^{0'} \varepsilon F^{0'} H V^{-1}_{NT} H^{-1} \lambda_i \right\|^2
\]

\[
+ \frac{2}{N^3 T^4} \sum_{i=1}^{N} \left\| X_i' M_{\hat{F}} \left( F^{0} - \hat{F} H^{-1} \right) \Lambda^{0'} \varepsilon \left( \hat{F} - F^{0} H \right) V^{-1}_{NT} H^{-1} \lambda_i \right\|^2
\]

\[
\leq T^{-1} O_P \left( \eta_{NT}^2 + \delta_{NT}^2 \right) + O_P \left( \eta_{NT}^4 + \delta_{NT}^4 \right).
\]
Using $M_{\hat{F}} = (M_{\hat{F}} - M_{F^0}) + I_T - P_{F^0}$ and Cauchy-Schwarz inequality,
\[
\mathcal{R}_{27} \leq \frac{1}{N^3T^4} \sum_{i=1}^{N} \left\| X_i' M_{\hat{F}} \sum_{j=1}^{N} \varepsilon_j \lambda_j' \right\|^2
\]
\[
\leq \frac{3}{N^3T^4} \sum_{i=1}^{N} \left\| X_i' (M_{\hat{F}} - M_{F^0}) \varepsilon_j \lambda_j' \right\| + \frac{3}{N^3T^4} \sum_{i=1}^{N} \left\| X_i' \varepsilon_j \lambda_j' \right\|^2
\]
\[
= T^{-1} O_P \left( \sigma^2_{N,T} + \sigma^2_{N,T} \right) + O_P \left( N^{-1} T^{-1} \right) = T^{-1} O_P \left( \sigma^2_{N,T} \right) + O_P \left( \sigma^2_{N,T} \right)
\]
where, e.g., the third term is bounded by
\[
3N^{-1}T^{-1} \left\{ T \right\| F^0 \left( F^0 \varepsilon_j \right)^{-1} \left\|^2 \right\| T^{-2} \left\| F^0 \hat{F} V^{-1} \left\|^2 \right\| \max_{1 \leq i \leq N} T^{-1} \left\| X_i \right\|^2
\]
\[
\times \left\{ \frac{1}{N^3T^4} \sum_{i=1}^{N} \left\| \varepsilon_i X_j \right\|^2 \right\} \left\{ \frac{1}{N^3T^4} \sum_{i=1}^{N} \left\| \varepsilon_i X_j \right\|^2 \right\} \right)^2 = O_P \left( N^{-1} T^{-1} \right).
\]

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

**Proof of Lemma A.4.** (i) By (B.1), $T^{-1/2} (\hat{F} - F^0) = \sum_{i=1}^{N} \frac{1}{N} \varepsilon_i X_j \lambda_j', \quad \sum_{i=1}^{N} \frac{1}{N} \varepsilon_i X_j \lambda_j' = \sum_{i=1}^{N} A_i$, say. For $A_{i1}$, using $\|a_i\| = O_P \left( T^{1/2} \sigma_{N,T} \right) = O_P \left( T^{1/2} \right)$ by Theorem 3.2, we can readily show that $\|A_{i1}\| = T^{-1/2} \|\varepsilon_i\| T^{-1/2} \|A_i\| \leq T^{-1} \|\varepsilon_i\| T^{-1} \sum_{i=1}^{T} E(\varepsilon_i^2) + T^{-1} \sum_{i=1}^{T} (\varepsilon_i - E(\varepsilon_i)) \leq c_T + T^{-1} \sum_{i=1}^{T} (1 - E(\varepsilon_i))$, where $c_T = \max_{1 \leq i \leq N} T^{-1} \sum_{i=1}^{T} E(\varepsilon_i^2) \leq C < \infty$ under Assumption A.1(iv), we can readily show that $P(\max_{1 \leq i \leq N} T^{-1} \|\varepsilon_i\|^2 \geq 3c_T) = o \left( N^{-1} \right)$ by Assumption A.3(i). Then $P \left( \max_{1 \leq i \leq N} \|A_{i1}\| \geq CT^{-1} (\ln T)^{v_2} \right) = o \left( N^{-1} \right)$ for any $C > 0$ and $v_2 > 0$.

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

\[
\|A_{i2}\| = \frac{1}{N^2T^2} \left\| \sum_{j=1}^{N} \varepsilon_j X_j \hat{b}_j \lambda_j' \right\| \leq \frac{1}{N^2T^2} \left\| \sum_{j=1}^{N} \varepsilon_j X_j \lambda_j' \right\| \leq \frac{1}{N^2T^2} \left\| \sum_{j=1}^{N} \varepsilon_j X_j \lambda_j' \right\|^2
\]
\[
\leq \frac{1}{N^2T^2} \left\| \sum_{j=1}^{N} \varepsilon_j X_j \lambda_j' \right\| \leq \frac{1}{N^2T^2} \left\| \sum_{j=1}^{N} \varepsilon_j X_j \lambda_j' \right\|^2
\]
\[
= O_P \left( T^{-1} \right) O_P \left( 1 \right) O_P \left( 1 \right) = O_P \left( T^{-1} \right).
\]

Noting that $\max_{1 \leq i \leq N} \frac{1}{N^2T^2} \sum_{j=1}^{N} \varepsilon_j X_j \lambda_j' \leq \max_{1 \leq i \leq N} \frac{1}{T^2} \|\varepsilon_i X_j \lambda_j' \| \leq \frac{1}{T^2} \|\varepsilon_i X_j \lambda_j' \| \leq \frac{1}{T} \|\varepsilon_i X_j \lambda_j' \|^2$, and $P(\max_{1 \leq i \leq N} T^{-1/2} \|\varepsilon_i X_j \lambda_j' \| \geq C (\ln T)^{v_1}) = o \left( N^{-1} \right)$ for any $C > 0$ by Assumption A.3(ii), we have $\max_{1 \leq i \leq N} P(\|A_{i2}\| \geq CT^{-1} (\ln T)^{v_1}) = o \left( N^{-1} \right)$. Similarly, we have

\[
\|A_{i3}\| = \frac{1}{N^2T^2} \left\| \sum_{j=1}^{N} \varepsilon_j X_j \hat{b}_j \varepsilon_j' \right\| \leq \frac{1}{N^2T^2} \left\| \sum_{j=1}^{N} \varepsilon_j X_j \lambda_j' \right\|^2
\]
\[
= O_P \left( T^{-1/2} \right) O_P \left( T^{-1/2} \right) O_P \left( 1 \right) = O_P \left( T^{-1} \right).
\]

32
\[
\|A_{44}\| = \frac{1}{NT^2} \left\| \epsilon'_i F^0 \sum_{j=1}^{N} \lambda'_j b_j X_j \tilde{F} V_{NT}^{-1} \right\|
\leq \eta_{nT} \left\{ \frac{1}{T^{1/2}} \left( \epsilon'_i F^0 \right) \left\| \sum_{j=1}^{N} \|X_j\| \right\| \left\| \lambda'_j \right\| \right\} \left\| \epsilon'_i \right\| \frac{1}{T^{1/2}} \left\| \tilde{F} V_{NT}^{-1} \right\|
= O_P \left(T^{-1}\right) O_P(1) O_P(1) O_P\left(T^{-1}\right) = O_P \left(T^{-1}\right)
\]
by Assumptions A.1(ii), (iii) and (vii), and
\[
\|A_{55}\| = \frac{1}{NT^2} \left\| \epsilon'_i \sum_{j=1}^{N} \epsilon_j b'_j X'_j \tilde{F} V_{NT}^{-1} \right\|
\leq \frac{1}{T^{1/2}} \left\{ \max_{1 \leq j \leq N} \frac{1}{T^{1/2}} \left\| X_j \right\| \right\} \eta_{nT} \left\{ \frac{1}{NT^2} \sum_{j=1}^{N} \left\| \epsilon'_j \right\| \right\} \left\| \epsilon'_i \right\| \frac{1}{T^{1/2}} \left\| \tilde{F} V_{NT}^{-1} \right\|
= O_P \left(T^{-1/2}\right) O_P(1) O_P(T^{-1/2}) O_P(1) O_P(1) = O_P \left(T^{-1}\right)
\]
by Assumptions A.1(iii) and (viii).

Noting that max_{1 \leq i \leq N} \sum_{j=1}^{N} \left\| \epsilon_j \epsilon'_j X_j \right\|^2 \leq \max_{1 \leq i, j \leq N} \frac{1}{T^{1/2}} \left\| \epsilon'_j \right\| \left\| \epsilon_j \right\| \left\| X_j \right\| \left\| \lambda'_j \right\| \left\| \lambda_j \right\| \left\| F^0 \right\| \left\| \nu \right\|.
and P(max_{1 \leq i \leq N} T^{-1/2} \left\| \epsilon'_i \right\| \left\| \epsilon_j \right\| \left\| X_j \right\| \left\| F^0 \right\| \left\| \nu \right\| \geq c \left( \ln T \right)^{-1/2}
= o \left(N^{-1}\right)
. Similarly, using P(max_{1 \leq i \leq N} T^{-1/2} \left\| \epsilon_j \right\| \left\| X_j \right\| \geq C \left( \ln T \right)^{-1/2}
= o \left(N^{-1}\right)
. and P(max_{1 \leq i \leq N} T^{-1/2} \left\| X_i \right\| \geq C \left( \ln T \right)^{-1/2}
= o \left(N^{-1}\right)
. for any C > 0 by Assumption A.3(i) and A.1(iii), we can show that P(max_{1 \leq i \leq N} \| A_{ii} \| \geq \epsilon'_i \left\| \epsilon_j \right\| \left\| X_j \right\| \left\| F^0 \right\| \left\| \nu \right\| \left\| \lambda_j \right\| \left\| \lambda'_j \right\| \geq C \left( \ln T \right)^{-1/2}
= o \left(N^{-1}\right)
. for l = 4, 5.

By the triangle inequality, the fact that \frac{1}{T^{1/2}} \left\| \epsilon'_i \right\| = O_P(1), \frac{1}{T^{1/2}} \sum_{j=1}^{N} \left\| \lambda_j \epsilon'_j \right\| = O_P(1) and \frac{1}{T^{1/2}} \sum_{j=1}^{N} \lambda'_j \epsilon'_j = O_P(1), Lemma A.2(i), and Theorem 3.2,
\[
\|A_{66}\| \leq \frac{1}{NT^2} \left\{ \frac{1}{T^{1/2}} \left\| \epsilon'_i F^0 \right\| \right\} \left\{ \frac{1}{NT^2} \sum_{j=1}^{N} \left\| \lambda'_j \epsilon_j \right\| \right\} \left\| H V_{NT}^{-1} \right\|
\leq \frac{1}{T^{1/2}} \left\{ \frac{1}{T^{1/2}} \left\| \epsilon'_i F^0 \right\| \right\} \left\{ \frac{1}{NT^2} \sum_{j=1}^{N} \left\| \lambda'_j \epsilon_j \right\| \right\} \left\| H \right\| \left\| V_{NT}^{-1} \right\|
= O_P(1) O_P(1) + T^{-1/2} O_P(1) O_P(1) = O_P \left(\frac{1}{T^{1/2}} \right)
\right).
Noting that $P \left( \max_{1 \leq i \leq N} \left\| (NT)^{-1/2} \sum_{j=1}^{N} \varepsilon_i \varepsilon_j \lambda_j^0 \right\| \geq C(1+\sqrt{T/N}) (\ln T)^{\nu_1} \right) = o(N^{-1})$ by Assumption A.3(iii), we have $P(\max_{1 \leq i \leq N} \| A_{ij} \| \geq C((NT)^{-1/2} + N^{-1}) (\ln T)^{\nu_1}) = o(N^{-1})$. Write

$$A_{is} = \frac{1}{NT^2} \sum_{j=1}^{N} \varepsilon_i \varepsilon_j \varepsilon_j F^0 \bar{V}_{NT}^{-1} + \frac{1}{NT^2} \sum_{j=1}^{N} \varepsilon_i \varepsilon_j \varepsilon_j (\hat{F} - F^0 H) \bar{V}_{NT}^{-1}$$

$$= A_{is,1} + A_{is,2}.$$

By Assumption A.1(viii), $\| A_{is,1} \| \leq \frac{1}{T^2} \| \frac{1}{NT} \sum_{j=1}^{N} \varepsilon_i \varepsilon_j \varepsilon_j F^0 \| \| \bar{V}_{NT}^{-1} \| = T^{-1} O_P(1) = O_P(T^{-1})$. By the Cauchy-Schwarz inequality, Lemma A.7(iii)

$$|A'_{is,2}\omega_p| = \left\{ \frac{1}{NT^2} \sum_{j=1}^{N} \left\| \varepsilon_i \varepsilon_j \varepsilon_j \right\| \right\}^{1/2} \left\{ \frac{1}{NT^2} \sum_{j=1}^{N} \omega_j' \bar{V}_{NT}^{-1} (\hat{F} - F^0 H) \varepsilon_j \varepsilon_j' (\hat{F} - F^0 H) \bar{V}_{NT}^{-1} \right\}^{1/2} \leq \left\{ \frac{1}{NT^2} \sum_{j=1}^{N} \left\| \varepsilon_i \varepsilon_j \varepsilon_j \right\| \right\}^{1/2} \left\{ \frac{1}{NT^2} \sum_{j=1}^{N} \varepsilon_j (\hat{F} - F^0 H) (\hat{F} - F^0 H)' \varepsilon_j \right\}^{1/2} \leq T^{-1/2} O_P(1),$$

where we use the fact that $\omega_j' \bar{V}_{NT}^{-1} (\hat{F} - F^0 H)' \varepsilon_j \varepsilon_j' (\hat{F} - F^0 H) \bar{V}_{NT}^{-1} \omega_p = \text{tr}[\bar{V}_{NT}^{-1} \omega_j' \omega_p' \bar{V}_{NT}^{-1} (\hat{F} - F^0 H)' \varepsilon_j \varepsilon_j' (\hat{F} - F^0 H)] \leq \mu_{\text{max}} \left( \bar{V}_{NT}^{-1} \omega_j' \omega_p' \bar{V}_{NT}^{-1} \right) \text{tr}[(\hat{F} - F^0 H)' \varepsilon_j \varepsilon_j' (\hat{F} - F^0 H)] \leq \| \bar{V}_{NT}^{-1} \| \varepsilon_j (\hat{F} - F^0 H)' \varepsilon_j$. It follows that $A_{is} = O_P(T^{-1})$. Using $\| \bar{V}_{NT}^{-1} \sum_{j=1}^{N} \varepsilon_i \varepsilon_j \varepsilon_j \| \leq \left\{ \frac{1}{NT^2} \sum_{j=1}^{N} (\varepsilon_i^2) \right\}^{1/2} \times \left\{ \frac{1}{NT} \sum_{j=1}^{N} \| \varepsilon_j F^0 \| \right\}^{1/2}$ and $P(\max_{1 \leq i \leq N} T^{-1} (\varepsilon_i^2) \geq C (\ln T)^{2\nu_1}) = o(N^{-1})$, we can show that $P(\max_{1 \leq i \leq N} \| A_{is,1} \| \geq C (\ln T)^{\nu_1}) = o(N^{-1})$ for $l = 1, 2$.

(iii) By (A.1) and the fact that $P_{F^0} = P_{F^0 H, \frac{1}{T} \sum_{i=1}^{N} \| X_i \| \| \varepsilon_i \| = T^{1/2} O_P(\delta_{NT}^2)$, and

$$\| D_{i1} \| \leq \frac{1}{T^{3/2}} \| X_i (\hat{F} - F^0 H)' \varepsilon_i \| \leq T^{1/2} \frac{1}{T} \| \hat{F} - F^0 H \| \frac{1}{T} \| X_i \| \| \varepsilon_i \| = T^{1/2} O_P(\delta_{NT}^2),$$

$$\| D_{i2} \| \leq \frac{1}{T^{3/2}} \| X_i (\hat{F} - F^0 H)' H' F^0 \varepsilon_i \| \leq T^{1/2} \| \hat{F} - F^0 H \| \frac{1}{T} \| X_i \| \| F^0 \varepsilon_i \| = O_P(\delta_{NT}^{-1}).$$

By (i),

$$\| D_{i3} \| \leq \frac{1}{T^{3/2}} \| X_i F^0 \| \frac{1}{T} \| (\hat{F} - F^0 H)' \varepsilon_i \| \leq T^{1/2} \frac{1}{T} \| X_i F^0 \| \frac{1}{T} \| (\hat{F} - F^0 H)' \varepsilon_i \| = T^{1/2} O_P(\delta_{NT}^2).$$
Consequently, \( \frac{1}{\sqrt{T}} X'_i M_{\rho} \varepsilon_i = \frac{1}{\sqrt{T}} X'_i M_{\rho} 0 \varepsilon_i + O_P \left( \frac{T^{1/2} \delta_{NT}^{-2}}{\delta_{NT}} \right) \). In addition, noting that \( P(\max_{1 \leq i \leq N} T^{-1/2} \| \varepsilon_i \| \geq C (\ln T)^{\nu_1} ) = o \left( N^{-1} \right) \) for \( \varepsilon_i = X_i, \varepsilon_i \), and \( F^0 \varepsilon_i \), we can show that

\[
P \left( \max_{1 \leq i \leq N} \frac{1}{T^{1/2}} \| X'_i (M_{\rho} - M_{F^0}) \varepsilon_i \| \geq C T^{1/2} \delta_{NT}^{-2} (\ln T)^{\nu_1} \right) = o \left( N^{-1} \right) .
\]

(iv) By (A.1),

\[
\frac{1}{\sqrt{T}} X'_i M_{\rho} F^0 \lambda_i^0 = \frac{1}{\sqrt{T}} X'_i (M_{F^0} - M_{\rho}) F^0 \lambda_i^0 = \frac{1}{\sqrt{T}} X'_i (P_{\rho} - P_{F^0}) F^0 \lambda_i^0 = \frac{1}{\sqrt{T}} \sum_{l=1}^{4} X'_i d_l F^0 \lambda_i^0 = \sum_{l=1}^{4} \tilde{D}_{il}, \text{ say.}
\]

By Lemma A.2 and Theorem 3.2,

\[
\left\| \tilde{D}_{i1} \right\| = \frac{1}{T^{1/2}} \left\| X'_i (\hat{F} - F^0 H)(\hat{F} - F^0 H)' F^0 \lambda_i^0 \right\| 
\leq T^{1/2} \frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 \frac{1}{T} \| X_i \| \| F^0 \lambda_i^0 \| = T^{1/2} O_P \left( \frac{1}{\delta_{NT}^2} \right) = o (1),
\]

\[
\left\| \tilde{D}_{i3} \right\| 
\leq T^{1/2} \frac{1}{T^{1/2}} \left\| X_i \right\| \frac{1}{T^{1/2}} \left\| F^0 H \right\| \frac{1}{T} \left\| (\hat{F} - F^0 H)' F^0 \right\| \| \lambda_i^0 \|
= T^{1/2} O_P (T^{-1/2} + \delta_{NT}^{-2}) = O_P (1), \text{ and}
\]

\[
\left\| \tilde{D}_{i4} \right\| 
\leq \frac{1}{T^{1/2}} \left\| X'_i (\hat{F} - F^0 H) (I_r - (T^{-1} H' F^0 F^0 H)^{-1})' H' F^0 \lambda_i^0 \right\|
\leq T^{1/2} \frac{1}{T} \left\| X'_i F^0 \right\| \| H_i \| \left\| I_r - (T^{-1} H' F^0 F^0 H)^{-1} \right\| \frac{1}{T} \left\| F^0 \right\| \left\| \lambda_i^0 \right\|
= T^{1/2} O_P (T^{-1/2} + \delta_{NT}^{-2}) = O_P (1).
\]

In addition, by (ii),

\[
\left\| \tilde{D}_{i2} \right\| = \frac{1}{T^{1/2}} \left\| X'_i (\hat{F} - F^0 H) H' F^0 \lambda_i^0 \right\|
\leq T^{1/2} \frac{1}{T} \left\| X'_i (\hat{F} - F^0 H) \right\| \frac{1}{T} \left\| H' F^0 \lambda_i^0 \right\| = T^{1/2} O_P (T^{-1/2}) = O_P (1).
\]

Consequently, \( \frac{1}{\sqrt{T}} X'_i M_{\rho} F^0 \lambda_i^0 = O_P (1) \). In addition, noting that \( P(\max_{1 \leq i \leq N} T^{-1/2} \| X_i \| \geq C (\ln T)^{\nu_1} ) = o \left( N^{-1} \right) \) and \( P(\max_{1 \leq i \leq N} \| \lambda_i^0 \| \geq c_{\lambda,N} ) = o \left( N^{-1} \right) \), we can show that \( P(\max_{1 \leq i \leq N} T^{-1/2} X'_i M_{\rho} F^0 \lambda_i^0 \geq C (\ln T)^{\nu_1} \) by Lemma A.2(v) and Theorem 3.2, and

\[
P \left( \max_{1 \leq i \leq N} T^{-1} X'_i (M_{\rho} - M_{F^0}) X_i \right) \geq C \delta_{NT}^{-1} (\ln T)^{\nu_2} = o \left( N^{-1} \right) \text{ for any } C > 0 \text{ and } \nu_2 > 0,
\]

we have

\[
P \left( \max_{1 \leq i \leq N} T^{-1} X'_i (M_{\rho} - M_{F^0}) X_i \right) \geq C \delta_{NT}^{-1} (\ln T)^{\nu_2} = o \left( N^{-1} \right) .
\]

(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 \( \epsilon > 0 \)

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

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 \( 1 \{ i \in \hat{G}_k \} = 1 \{ i \in G_k^0 \} + 1 \{ i \in \hat{G}_k \setminus G_k^0 \} - 1 \{ i \in G_k^0 \setminus \hat{G}_k \}, \) we have

\[
\frac{1}{N} \sum_{i \in \hat{G}_k} \left\| \hat{\alpha}_k - \beta_i^0 \right\|^2 - \frac{N_k}{N} \left\| \hat{\alpha}_k - \alpha_k^0 \right\|^2 = \frac{1}{N} \sum_{i \in \hat{G}_k} \left\| \hat{\alpha}_k - \beta_i^0 \right\|^2 - \frac{1}{N} \sum_{i \in G_k^0} \left\| \hat{\alpha}_k - \alpha_k^0 \right\|^2
= \frac{1}{N} \sum_{i \in \hat{G}_k \setminus G_k^0} \left\| \hat{\alpha}_k - \beta_i^0 \right\|^2 - \sum_{i \in G_k^0 \setminus \hat{G}_k} \left\| \hat{\alpha}_k - \beta_i^0 \right\|^2
= R_{1k} - R_{2k}, \text{ say.}
\]

By Theorem 3.4, \( P(\|R_{1k}\| \geq \epsilon / (NT)) \leq P(\hat{E}_{k,NT}) \to 0 \) and \( P(\|R_{2k}\| \geq \epsilon / (NT)) \leq P(\hat{E}_{k,NT}) \to 0. \) 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). \( \square \)

Proof of Lemma A.6. (i) Using Theorem 3.2(iii) and Assumption A.2(iii), we can show that \( \hat{A}_{1k,l} = O_p \left( \left\| \hat{\alpha}_1^l - \alpha_1^0 \right\| \right) = O_p \left( T^{-1/2} \right). \)

(ii) Let \( A_{2k,l}^{(1)} = \frac{1}{N_{N_{k}T}} \sum_{j \in \hat{G}_k} \sum_{i \in G_k^0} \nu_{ji} X'_i M_F X_j \) and \( A_{2k,l}^{(2)} = \frac{1}{N_{N_{k}T}} \sum_{j \in \hat{G}_k} \sum_{i \in G_k^0} \nu_{ji} X'_i M_F X_j. \) Then \( \hat{A}_{2k,l} - A_{2k,l} = (A_{2k,l}^{(1)} - A_{2k,l}^{(2)}). \)

Let \( \epsilon > 0. \) By Theorem 3.4, \( P \left( \|R_{1k}\| \geq \epsilon / (NT)^{1/2} \right) \leq P(\hat{E}_{k,NT}) \to 0 \) and \( P \left( \|R_{1k}\| \geq \epsilon / (NT)^{1/2} \right) \leq P(\hat{E}_{k,NT}) \to 0. \) It follows that \( \hat{A}_{2k,l} - A_{2k,l} = o_p((NT)^{-1/2}). \) By the same token, we can show that
\[ A^{(1)}_{2k,t} - A^{(2)}_{2k,t} = o_P(NT)^{-1/2} \]. Now, using \(|\nu_{ji}| \leq \|\lambda_i^0\| \|\lambda_j^0\| \left\| (N^{-1}\Lambda^0\Lambda^0)^{-1} \right\| \) and by Lemma A.2(v),

\[
\left\| A^{(2)}_{2k,t} - A_{2k,t} \right\| \leq \|P_F - P_{F^0}\| \left\{ \frac{1}{Nk_T} \sum_{j \in G_k^0} \sum_{i \in G_k^0} |\nu_{ji}| \|X_i\| \|X_j\| \right\}^2 = o_P(\delta_{NT}^{-1}).
\]

It follows that \( \hat{A}_{2k,t} = A_{2k,t} + o_P(\delta_{NT}^{-1}) \).

(iii) By the triangle inequality,

\[
\frac{1}{NT^{3/2}} \left\| \sum_{j \in G_k} X_j \varepsilon_j^T \hat{F} G \right\| \leq \frac{1}{NT^{3/2}} \left\| \sum_{j \in G_k} X_j \varepsilon_j^T F^0 H G \right\| + \frac{1}{NT^{3/2}} \left\| \sum_{j \in G_k} X_j \varepsilon_j^T (\hat{F} - F^0 H) G \right\| \equiv a_{11} + a_{12}, \text{ say.}
\]

As in the proof of (ii), using the fact that \(1 \{j \in \hat{G}_k\} = 1 \{j \in G_k^0\} + 1 \{j \in \hat{G}_k \setminus \hat{G}_k\} \) and Theorem 3.4, we can show that \(a_{11} = \frac{1}{NT^{3/2}} \sum_{j \in G_k^0} X_j \varepsilon_j^T F^0 H G + o_P((NT)^{-1/2})\). It follows that \(\|a_{11}\| \leq \frac{1}{NT^{3/2}} \left\| \sum_{j \in G_k^0} X_j \varepsilon_j^T F^0 H G \right\| \leq o_P((NT)^{-1/2})\) by Assumption A.1(ix). Similarly, by Cauchy-Schwarz inequality and Lemma A.7(iii),

\[
\left| \omega_T a_{12} \hat{\omega}_{R_0} \right| \leq \frac{1}{NT^{3/2}} \left\| \sum_{j \in G_k^0} \omega_T X_j \varepsilon_j^T (\hat{F} - F^0 H) G \hat{\omega}_{R_0} \right\| + o_P((NT)^{-1/2})
\]

\[
\leq \frac{1}{NT^{3/2}} \left\| \sum_{j \in G_k^0} \omega_T X_j \varepsilon_j^T X_j^T \omega_T \right\| \left\{ \frac{1}{NT^{3/2}} \left\| \sum_{j \in G_k^0} \omega_{R_0} \hat{G}(\hat{F} - F^0 H) \varepsilon_j \varepsilon_j^T (\hat{F} - F^0 H) \hat{G} \hat{\omega}_{R_0} \right\| \right\}^{1/2} + o_P((NT)^{-1/2})
\]

\[
\leq \left\{ \frac{1}{NT^{3/2}} \sum_{j \in G_k^0} \left\| X_j \right\|^2 \right\}^{1/2} + o_P((NT)^{-1/2})
\]

\[
\leq \left\{ \frac{1}{NT^{3/2}} \sum_{j \in G_k^0} \omega_{R_0} \hat{G}(\hat{F} - F^0 H) \varepsilon_j \varepsilon_j^T (\hat{F} - F^0 H) \hat{G} \hat{\omega}_{R_0} \right\}^{1/2} + o_P((NT)^{-1/2})
\]

\[
= O_P(1) O_P(\delta_{NT}^{-1}) + o_P((NT)^{-1/2}) = O_P(\delta_{NT}^{-1}).
\]

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

\[
\left| \omega_T A^{(3)}_{k,l} \right| = \left\| \frac{1}{Nk_T^2} \sum_{j \in G_k} \omega_T X_j^T F \hat{G} \lambda_i^0 \right\| \leq \frac{1}{Nk_T^2} \left\| \sum_{j \in G_k} X_j^T \hat{F} G \sum_{i \in G_k} \lambda_i^0 \omega_T X_i^T M_F \right\|
\]

\[
\leq \left\{ \frac{1}{NT^{3/2}} \left\| \sum_{j \in G_k} X_j \varepsilon_j^T \hat{F} G \right\| \right\} \left\{ \frac{1}{NT^{3/2}} \left\| \sum_{i \in G_k} \lambda_i^0 \omega_T X_i^T \right\| \right\} \leq \left\{ \frac{1}{NT^{3/2}} \left\| \sum_{j \in G_k} X_j \varepsilon_j^T \hat{F} G \right\| \right\} \left\{ \frac{1}{NT^{3/2}} \left\| \sum_{i=1}^N \lambda_i^0 X_i \right\| \right\}
\]

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

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

$$\left\| \hat{A}_{4k,l} \right\| = \left\| \frac{1}{N N_k T^2} \sum_{j \in G_1} \sum_{i \in G_k} X_i^t M_{\hat{p}} (F^0 H - \hat{F}) H^{-1} \lambda_j^0 \lambda_i^0 G^t \hat{F} X_j \right\|$$

$$\leq \left\{ \frac{1}{N T^{1/2}} \sum_{i \in G_k} \| X_i \|_{sp} \| \lambda_i^0 \| \right\} \left\{ \frac{1}{N \bar{k} T^{1/2}} \sum_{j \in G_i} \| \lambda_j^0 \| \| X_j \|_{sp} \right\}$$

$$\times \frac{1}{T^{1/2}} \left\| F^0 H - \hat{F} \right\| H^{-1} \| X_i \|_{sp} \frac{1}{T^{1/2}} \left\| G^t \hat{F} X_j \right\|_{sp}$$

$$= O_P(1) O_P(1) O_P(\delta_{NT}^{-1}) O_P(1) = O_P(\delta_{NT}^{-1}) .$$

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

$$\left| \omega_{p} \hat{A}_{5k,l} \hat{w}_{p} \right| = \frac{1}{N N_k T^2} \left\| \sum_{j \in G_1} \sum_{i \in G_k} \omega_{p} X_i^t M_{\hat{p}}^i \lambda_i^0 \lambda_j^0 G^t \hat{F} X_j \hat{w}_{p} \right\|$$

$$\leq \frac{1}{N N_k T^{3/2}} \left\| \sum_{i \in G_k} \lambda_i^0 \omega_{p} X_i^t M_{\hat{p}} \sum_{j \in G_i} \varepsilon_j \hat{w}_{p} \varepsilon_j \right\| \left\{ \frac{1}{N T^{1/2}} \left\| \hat{F} G \right\| + o_P((NT)^{-1/2}) \right\}$$

$$\leq \left\{ \frac{1}{N_k T^{1/2}} \sum_{i \in G_k} \lambda_i^0 \omega_{p} X_i^t M_{\hat{p}} \sum_{j \in G_i} \varepsilon_j \hat{w}_{p} \varepsilon_j \right\} \left\{ \frac{1}{N T} \sum_{j \in G_i} \| \hat{w}_{p} \| X_j \right\} \frac{1}{T^{1/2}} \| \hat{F} G \| + o_P((NT)^{-1/2})$$

$$= O_P(1) O_P(\delta_{NT}^{-1}) O_P(1) + o_P((NT)^{-1/2}) = O_P(\delta_{NT}^{-1}) .$$

where we use the fact that $\frac{1}{N T} \| \sum_{j \in G_i} \varepsilon_j \hat{w}_{p} \varepsilon_j \| = O_P(\delta_{NT}^{-1})$ by Assumption A.1(viii). It follows that $\hat{A}_{5k,l} = O_P(\delta_{NT}^{-1})$.

(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 $\frac{1}{N T} \sum_{i=1}^{N} \varepsilon_i^t (\hat{F} - F^0 H) = \sum_{i=1}^{N} \varepsilon_i^t \alpha_i V_{NT}^{-1} = \sum_{i=1}^{N} S_i$, say. Using $\| a_i \| = O_P(\alpha_i T^{-1/2})$, we have $\| S_i \| \leq \| a_i \| \| V_{NT}^{-1} \| \times \frac{1}{N T} \sum_{i=1}^{N} \| \varepsilon_i \| = O_P(\alpha_i T^{-1/2}) O_P(\alpha_i T^{-1/2}) = O_P(\alpha_i T^{-1/2})$. Using arguments as used in the proofs of Lemmas A.5(i) and A.6(ii), we can show that $a_2 = -\frac{1}{N T} \sum_{i=1}^{N} X_i \hat{b}_i \lambda_i^0 F^0 \hat{F} = -\frac{1}{N T} \sum_{i=1}^{N} X_i \hat{b}_i \lambda_i^0 F^0 \hat{F} + o_P((NT)^{-1})$. It follows that

$$\| S_2 \| = \frac{1}{N T} \sum_{i=1}^{N} \varepsilon_i^t \left( \frac{1}{N T} \sum_{j=1}^{N} X_j (\hat{a}_i - \lambda_i^0) \lambda_j^0 F^0 \hat{F} + o_P((NT)^{-1}) \right) \| V_{NT}^{-1} \|$$

$$\leq \frac{1}{N T} \sum_{i=1}^{N} \varepsilon_i^t \left( \frac{1}{N T} \sum_{j=1}^{N} X_j (\hat{a}_i - \lambda_i^0) \lambda_j^0 F^0 \hat{F} \right) \| V_{NT}^{-1} \| + o_P((NT)^{-1})$$

$$\leq \frac{1}{N T} \sum_{i=1}^{N} \varepsilon_i^t \lambda_i^0 X_i (\hat{a}_i - \lambda_i^0) \lambda_j^0 F^0 \hat{F} \| V_{NT}^{-1} \| + o_P((NT)^{-1})$$

$$\leq \frac{1}{N T} \sum_{i=1}^{N} \varepsilon_i^t \lambda_i^0 X_i (\hat{a}_i - \lambda_i^0) \lambda_j^0 F^0 \hat{F} \| V_{NT}^{-1} \| + o_P((NT)^{-1})$$

$$= T^{-1/2} O_P(1) O_P(1) + o_P((NT)^{-1}) = O_P(T^{-1/2}) .$$
Similarly, we can show that \( \|S_s\| = O_P(T^{-1}) \) for \( s = 3, 4, 5 \). By the triangle inequality,
\[
\|S_6\| = \left\| \frac{1}{N^2T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i F_0 \lambda_j \varepsilon_j \hat{F} V^{-1}_{NT} \right\| \\
\leq \frac{1}{N^2T^2} \left\| \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i F_0 \lambda_j \varepsilon_j F_0 \right\| H V^{-1}_{NT} + \frac{1}{N^2T^2} \left\| \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i F_0 \lambda_j \varepsilon_j (\hat{F} - F_0 H) V^{-1}_{NT} \right\| \\
\leq \frac{1}{T} \left\{ \frac{1}{NT^{1/2}} \sum_{i=1}^{N} \| \varepsilon_i F_0 \| \right\} \left\{ \frac{1}{NT^{1/2}} \sum_{j=1}^{N} \| \lambda_j \varepsilon_j F_0 \| \right\} \| H V^{-1}_{NT} \| \\
+ \frac{1}{T^{1/2}} \frac{1}{NT^{1/2}} \sum_{i=1}^{N} \| \varepsilon_i F_0 \| \sum_{j=1}^{N} \| \lambda_j \varepsilon_j \| \left\{ \frac{1}{T^{1/2}} \| \hat{F} - F_0 H \| \right\} \| V^{-1}_{NT} \| \\
= T^{-1} O_P(1) O_P(1) + T^{-1/2} O_P(1) O_P(1) O_P(\delta_{NT}^{-1}) = T^{-1/2} O_P(\delta_{NT}^{-1}).
\]
Similarly, we can show that \( S_s = O_P(\delta_{NT}^{-1}) \) for \( s = 7, 8 \). Thus \( \frac{1}{\sqrt{T}} \sum_{i=1}^{N} \varepsilon_i (\hat{F} - F_0 H) = O_P(\delta_{NT}^{-1}) \).

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

(iii) By (B.1), \( T^{-1} \varepsilon_i (\hat{F} - F_0 H) = \sum_{i=1}^{8} \frac{1}{\sqrt{T}} \varepsilon_i V^{-1}_{NT} \varepsilon_i \). By Cauchy-Schwarz inequality
\[
\frac{1}{NT^2} \sum_{i=1}^{N} \varepsilon_i (\hat{F} - F_0 H) (\hat{F} - F_0 H)' \varepsilon_i \\
= \sum_{i=1}^{8} \sum_{k=1}^{N} \sum_{i=1}^{N} \frac{1}{N} A'_{ik} A_{ik} \\
\leq 8 \sum_{i=1}^{N} \frac{1}{N} \sum_{k=1}^{N} A'_{ik} A_{ik} \leq 8 \sum_{i=1}^{N} \frac{1}{N} \sum_{k=1}^{N} A'_{ik} A_{ik}.
\]
Using \( \|a_1\| = O_P(T^{-1/2}) \), we have
\[
\frac{1}{N} \sum_{i=1}^{N} A'_{i1} A_{i1} = \frac{1}{NT^2} \sum_{i=1}^{N} \varepsilon_i a_1 V^{-1}_{NT} V^{-1}_{NT} a_1 \varepsilon_i \\
\leq \frac{1}{NT^2} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \| \varepsilon_i \|^2 \right\} \|a_1\|^2 \| V^{-1}_{NT} \|^2 \\
= \frac{1}{T} O_P(1) O_P(T^{-1}) = O_P(T^{-2}).
\]

Using arguments as used in the proofs of Lemmas A.5(i) and A.6(ii), we can show that \( a_2 = -\frac{1}{NT} \sum_{i=1}^{N} \sum_{\ell=1}^{K_0} \sum_{j=1}^{G_{\ell}^0} \sum_{m=1}^{G_{\ell}^0} \varepsilon_i X_j (\hat{\alpha}_i - \alpha_i^0) \lambda_j \varepsilon_j F_0 \hat{F} V^{-2}_{NT} \hat{F} V^{-2}_{NT} F_0 \lambda_m (\hat{\alpha}_k - \alpha_k^0)' X_m' \varepsilon_i \\
+ o_P((NT)^{-1}) \)
\[
\leq -\frac{c_{NT}}{NT} \sum_{i=1}^{N} \sum_{k=1}^{K_0} \| \hat{\alpha}_i - \alpha_i^0 \| \| \hat{\alpha}_m - \alpha_m^0 \| \left\{ \frac{1}{N^3T^2} \sum_{i=1}^{N} \sum_{j=1}^{K_0} \sum_{m=1}^{G_{\ell}^0} \| \varepsilon_i X_j \| \| \lambda_j \| \| \lambda_m \| \| X_m' \varepsilon_i \| \right\} \\
+ o_P((NT)^{-1}) \\
\leq -\frac{c_{NT}}{NT} \sum_{i=1}^{N} \sum_{k=1}^{K_0} \| \hat{\alpha}_i - \alpha_i^0 \| \| \hat{\alpha}_m - \alpha_m^0 \| \left\{ \frac{1}{N^3T} \sum_{i=1}^{N} \sum_{j=1}^{K_0} \sum_{m=1}^{G_{\ell}^0} \| \varepsilon_i X_j \| \| \lambda_j \| \| \lambda_m \| \| X_m' \varepsilon_i \| \right\} \right \} + o_P((NT)^{-1}) \\
\leq \frac{1}{T} O_P(T^{-1/2}) O_P(T^{-1/2}) O_P(1) + o_P((NT)^{-1}) = O_P(T^{-2}) + o_P((NT)^{-1}),
\]
where \( c_{NT} = T^{-2} \| F_0 V^{-2}_{NT} \hat{F} V^{-2}_{NT} \| = O_P(1) \). Analogously, we can show that \( \frac{1}{N} \sum_{i=1}^{N} A'_{is} A_{is} = O_P(T^{-2}) + o_P((NT)^{-1}) \) for \( s = 3, 4, 5 \).
Noting that \( \hat{F}' \hat{F}' = (F^0 H + \hat{F} - F^0 H)(F^0 H + \hat{F} - F^0 H)' \leq 2F^0 H H' F^0 + 2(\hat{F} - F^0 H)(\hat{F} - F^0 H)' \),
we have

\[
\frac{1}{N} \sum_{i=1}^{N} A'_{i6} A_{i6} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \varepsilon'_i \varepsilon'_j \varepsilon_k \varepsilon_k \varepsilon_i
\]

\[
\leq \|V_{NT}^{-1}\| \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \varepsilon'_i \varepsilon'_j \varepsilon_k \varepsilon_k \varepsilon_i
\]

\[
\leq 2 \|V_{NT}^{-1}\| \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \varepsilon'_i \varepsilon'_j \varepsilon_k \varepsilon_k \varepsilon_i
\]

\[
\leq 2 \|V_{NT}^{-1}\| \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \varepsilon'_i \varepsilon'_j \varepsilon_k \varepsilon_k \varepsilon_i
\]

The first term is bounded from above by

\[
\leq O_P \left( T^{-2} \right)
\]

By (ii), the second term is bounded from above by \( \frac{2}{T} \|V_{NT}^{-1}\| \|H\|^2 \). It follows that \( \frac{2}{N} \sum_{i=1}^{N} A'_{i6} A_{i6} = O_P \left( T^{-2} \right) \). Analogously, we can show that \( \frac{2}{N} \sum_{i=1}^{N} A'_{i7} A_{i7} = O_P \left( T^{-2} \right) \). Now,

\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} A'_{i8} A_{i8} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon'_i \varepsilon'_j \varepsilon_k \varepsilon_k \varepsilon_i
\]

\[
\leq 2 \|V_{NT}^{-1}\| \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \varepsilon'_i \varepsilon'_j \varepsilon_k \varepsilon_k \varepsilon_i
\]

\[
\leq 2 \|V_{NT}^{-1}\| \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \varepsilon'_i \varepsilon'_j \varepsilon_k \varepsilon_k \varepsilon_i
\]

\[
\leq 2 \|V_{NT}^{-1}\| \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \varepsilon'_i \varepsilon'_j \varepsilon_k \varepsilon_k \varepsilon_i
\]

The first term is bounded from above by

\[
\leq O_P \left( N^{-1} T^{-2} \right)
\]
For the second term, we apply (i) to obtain

\[
2 \left\| V_{NT}^{-1} \right\|^2 \frac{1}{N^3 T^4} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \epsilon_i \epsilon_j \epsilon_k \epsilon_l (\hat{F} - F^0 H) (\hat{F} - F^0 H)^t \epsilon_k \epsilon_i \\
\leq \frac{2 \left\| V_{NT}^{-1} \right\|^2}{T} \frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{NT} \sum_{j=1}^{N} \epsilon_i \epsilon_j \epsilon_k \epsilon_l \right\|^2 = T^{-1} O_P \left( \delta_{NT}^{-2} \right) O_P \left( 1 \right) = O_P \left( T^{-1} \delta_{NT}^{-2} \right).
\]

Thus \( \frac{1}{N} \sum_{i=1}^{N} A_{1i} A_{i8} = O_P \left( T^{-1} \delta_{NT}^{-2} \right) \). In sum, \( \frac{1}{NT} \sum_{i=1}^{N} \epsilon_i (\hat{F} - F^0 H) G G' (\hat{F} - F^0 H)^t \epsilon_i = O_P \left( \delta_{NT}^{-4} \right) \).

(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

\[
\frac{1}{NN_k T^2} \sum_{i \in G_k} \sum_{j=1}^{N} \omega'_p X'_i \epsilon_j \epsilon_j (\hat{F} - F^0 H) G \lambda_i^0 \left| \epsilon'_j (\hat{F} - F^0 H) G \lambda_i^0 \omega'_p X'_i \epsilon_j \right| \\
\leq \frac{1}{NN_k T^2} \sum_{i \in G_k} \sum_{j=1}^{N} \left\| \epsilon'_j (\hat{F} - F^0 H) G G' (\hat{F} - F^0 H)^t \epsilon_j \right\|^{1/2} \left\{ \epsilon'_j X_i \omega_p \lambda_i^0 \lambda_i^0 \omega'_p X'_i \epsilon_j \right\}^{1/2} \\
\leq \frac{\|G\|^2}{T} \left\{ \frac{1}{NT} \sum_{j=1}^{N} \epsilon'_j (\hat{F} - F^0 H) (\hat{F} - F^0 H)^t \epsilon_j \right\}^{1/2} \left\{ \frac{1}{NN_k T} \sum_{i \in G_k} \sum_{j=1}^{N} \epsilon'_j X_i \epsilon_j \epsilon_j \epsilon_j \right\}^{1/2} \\
= T^{-1} O_P \left( \delta_{NT}^{-2} \right) O_P \left( 1 \right) = O_P \left( T^{-1} \delta_{NT}^{-2} \right).
\]

(vi) By (B.1) we have \( \sum_{i \in G_k} X'_i M_F^t F^0 \lambda_i^0 = - \frac{1}{NT} \sum_{j=1}^{N} X_j (\hat{a}_i - \alpha_i^0)^t (\hat{a}_i - \alpha_i^0)^t X_j \hat{F} G \lambda_i^0 + o_P((NT)^{-1/2}) \).

Using arguments like those used in the derivation of (A.23), we can readily show that

\[
J_{k,1} = - \frac{1}{NT} \sum_{i \in G_k} X'_i M_F \left( \frac{1}{NT} \sum_{j=1}^{N} X_j \hat{b}_j \hat{b}_j^t X'_j \hat{F} \right) G \lambda_i^0 \]

\[
= - \frac{1}{NT} \sum_{i \in G_k} X'_i M_F \left( \frac{1}{NT} \sum_{j=1}^{N} X_j (\hat{a}_i - \alpha_i^0)^t (\hat{a}_i - \alpha_i^0)^t X_j \hat{F} G \lambda_i^0 + o_P((NT)^{-1/2}) \right) \\
= \sum_{i=1}^{K_0} \hat{A}_{1i,k} \left( \hat{a}_i - \alpha_i^0 \right) + o_P((NT)^{-1/2}),
\]

\[
J_{k,2} = \frac{1}{NT} \sum_{i \in G_k} X'_i M_F \left( \frac{1}{NT} \sum_{j=1}^{N} X_j \hat{b}_j \lambda_j^0 \nu_j \hat{F} \right) G \lambda_i^0 \]

\[
= \frac{1}{NT} \sum_{i \in G_k} X'_i M_F \left( \frac{1}{NT} \sum_{j=1}^{N} X_j (\hat{a}_i - \alpha_i^0)^t \nu_j + o_P((NT)^{-1/2}) \right) \\
= \sum_{i=1}^{K_0} \hat{A}_{2i,k} \left( \hat{a}_i - \alpha_i^0 \right) + o_P((NT)^{-1/2}),
\]
\[ J_{k,3} = \frac{1}{NkT} \sum_{i \in G_k} X_i' M_F \left( \frac{1}{NT} \sum_{j=1}^{N} X_j \hat{b}_j \varepsilon_j' \hat{F} \right) G \lambda_i^0 \]

\[ = \frac{1}{NkT} \sum_{i \in G_k} X_i' M_F \frac{1}{NT} \sum_{l=1}^{K_0} \sum_{j \in G_l} X_j \left( \hat{a}_l - a_l^0 \right)' \varepsilon_j' \hat{F} G \lambda_l^0 + o_P((NT)^{-1/2}) \]

\[ = \sum_{l=1}^{K_0} \tilde{A}_{3k,l} \left( \hat{a}_l - a_l^0 \right) + o_P((NT)^{-1/2}), \]

\[ J_{k,4} = \frac{1}{NkT} \sum_{i \in G_k} X_i' M_F \left( \frac{1}{NT} \sum_{j=1}^{N} F^0 \lambda_j^0 \varepsilon_j' X_j' \hat{F} \right) G \lambda_i^0 \]

\[ = \frac{1}{NkT} \sum_{i \in G_k} X_i' M_F \frac{1}{NT} \sum_{l=1}^{K_0} \sum_{j \in G_l} F^0 \lambda_j^0 \left( \hat{a}_l - a_l^0 \right)' X_j' \hat{F} G \lambda_l^0 + o_P((NT)^{-1/2}) \]

\[ = \sum_{l=1}^{K_0} \tilde{A}_{4k,l} \left( \hat{a}_l - a_l^0 \right) + o_P((NT)^{-1/2}), \text{ and} \]

\[ J_{k,5} = \frac{1}{NkT} \sum_{i \in G_k} X_i' M_F \left( \frac{1}{NT} \sum_{j=1}^{N} \varepsilon_j' \hat{b}_j \varepsilon_j' \hat{F} \right) G \lambda_i^0 \]

\[ = \frac{1}{NkT} \sum_{i \in G_k} X_i' M_F \frac{1}{NT} \sum_{l=1}^{K_0} \sum_{j \in G_l} \varepsilon_j \left( \hat{a}_l - a_l^0 \right)' X_j' \hat{F} G \lambda_l^0 + o_P((NT)^{-1/2}) \]

\[ = \sum_{l=1}^{K_0} \tilde{A}_{5k,l} \left( \hat{a}_l - a_l^0 \right) + o_P((NT)^{-1/2}). \]

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

\[ |\omega_p' J_{k,6}| = \frac{1}{NkT} \left| \text{tr} \left( \sum_{i \in G_k} \lambda_i^0 \omega_p' X_i' M_F \left( F^0 H - \hat{F} \right) H^{-1} \left( \frac{1}{NT} \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j' \hat{F} \right) G \right) \right| \]

\[ \leq \frac{1}{NkT} \left\| \sum_{i \in G_k} \lambda_i^0 \omega_p' X_i' M_F \left( F^0 H - \hat{F} \right) H^{-1} \left\| \frac{1}{NT} \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j' \hat{F} \right\| G \right\| \]

\[ \leq \left\{ \frac{1}{NkT} \left( \sum_{i \in G_k} \| \lambda_i^0 \| \| X_i \|_{sp} \right) \right\} \left\{ \frac{1}{NT^{1/2}} \left\| F^0 H - \hat{F} \right\|_{sp} \right\} \left\| \frac{1}{NT} \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j' \hat{F} \right\| \left\| H^{-1} \right\| \| G \| \]

\[ = O_P(1) O_P(\delta_{NT}^{-1}) O_P(NT^{-1/2} \delta_{NT}^{-1}) = o_P((NT)^{-1/2}), \]

where we use the fact that

\[ \left\| \frac{1}{NT} \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j' \hat{F} \right\| \leq \left\| \frac{1}{NT} \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j' F^0 H \right\| + \left\| \frac{1}{NT} \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j' \left( \hat{F} - F^0 H \right) \right\| \]

\[ \leq \left\| \frac{1}{NT} \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j' F^0 H \right\| + \left\| \frac{1}{NT^{1/2}} \sum_{j=1}^{N} \lambda_j^0 \varepsilon_j' \left( \hat{F} - F^0 H \right) \right\| \frac{1}{\sqrt{T}} \left\| \hat{F} - F^0 H \right\| \]

\[ = O_P((NT)^{-1/2}) O_P(1) + O_P(NT^{-1/2}) O_P(\delta_{NT}^{-1}) = O_P(NT^{-1/2} \delta_{NT}^{-1}). \]
In addition, $J_{k,7} = -\frac{1}{N_k T} \sum_{i \in G_k} X'_i M_{\bar{F}} (\frac{N}{N_k T} \sum_{j=1}^{N} \varepsilon_j \lambda_j^{0} \tilde{F}' \tilde{F} G \lambda_i^{0}) = -\frac{1}{N_k T} \sum_{i \in G_k} \sum_{j=1}^{N} \nu_{j,i} X'_i M_{\bar{F}} \varepsilon_j$ and $J_{k,8} = -\frac{1}{N_k T} \sum_{i \in G_k} \sum_{j=1}^{N} X_j^i M_{\bar{F}} \varepsilon_j \varepsilon_j^* F (T^{-1} \tilde{F}^0 \tilde{F})^{-1} \left( (N^{-1} \bar{A}^{0} \lambda^0)^{-1} \lambda_i^0 \right) = B_{1,k,NT}$. In sum, we have $\frac{1}{N_k T} \sum_{i \in G_k} X'_i M_{\bar{F}} F^0 \lambda_i^0 = \sum_{i=1}^{K_0} \tilde{A}_{k,i} (\tilde{g}_i - \alpha_i) - \frac{1}{N_k T} \sum_{i \in G_k} \sum_{j=1}^{N} \nu_{j,i} X'_i M_{\bar{F}} \varepsilon_j + B_{1,k,NT} + o_P((NT)^{-1/2})$.

**Proof of Lemma A.8.** (i) As in the proof of Lemma A.6(ii), we can show that $\frac{1}{N_k T} \sum_{i \in G_k} X'_i M_{\bar{F}} X_i = \frac{1}{N_k T} \sum_{i \in G_k} X'_i M_{\bar{F}} X_i + o_P((NT)^{-1/2})$. Then by Lemma A.2(v) and Theorem 3.2,

$$\left\| Q_{k,NT} - Q_{k,NT} \right\| = \left\| \frac{1}{N_k T} \sum_{i \in G_k} X'_i (M_{\bar{F}} - M_{F^0}) X_i + o_P((NT)^{-1/2}) \right\| \leq \frac{1}{N_k T} \sum_{i \in G_k} \left\| X'_i \right\|^2 \left\| P_{\bar{F}} - P_{F^0} \right\| + o_P((NT)^{-1/2}) = O_p(\delta_{NT})$$

(ii) Using $M_{\bar{F}} = I_T - P_{\bar{F}} = I_T - \tilde{F} \tilde{F}' / T$, we have $\hat{J}_{k,8} = -\frac{1}{N_k T^2} \sum_{i \in G_k} \sum_{j=1}^{N} X'_i \tilde{F}' \tilde{F} \varepsilon_j \tilde{F} G \lambda_i^{0} + \frac{1}{N_k T} \sum_{i \in G_k} X'_i \tilde{F}' \tilde{F} \varepsilon_j \tilde{F} G \lambda_i^{0} = -\hat{J}_{k,81} + \hat{J}_{k,82}$, say. We further decompose $\hat{J}_{k,81}$ as follows

$$\hat{J}_{k,81} = \frac{1}{N_k T} \sum_{i \in G_k} \sum_{j=1}^{N} X'_i \tilde{F}' \tilde{F} \varepsilon_j \tilde{F} G \lambda_i^{0} + \frac{1}{N_k T} \sum_{i \in G_k} X'_i \tilde{F}' \tilde{F} \varepsilon_j \tilde{F} G \lambda_i^{0} = \hat{J}_{k,81}(1) + \hat{J}_{k,82}(2),$$

As in the proof of Lemma A.6(ii), we can show that $\frac{1}{N_k T} \sum_{i \in G_k} \sum_{j=1}^{N} X'_i \tilde{F}' \tilde{F} \varepsilon_j \tilde{F} G \lambda_i^{0} = \frac{1}{N_k T} \sum_{i \in G_k} \sum_{j=1}^{N} X'_i \tilde{F}' \tilde{F} \varepsilon_j \tilde{F} G \lambda_i^{0} = O_P((NT)^{-1})$. It follows that by Assumption A.1(viii) and (ix)

$$|\omega_{P}^T \hat{J}_{k,81}(1)| = \left\| \frac{1}{N_k T^2} \left( \sum_{i \in G_k} \sum_{j=1}^{N} \lambda_i^0 \omega_{P}^T X'_i \varepsilon_j \varepsilon_j^* F^0 H \right) \right\| + o_P((NT)^{-1}) \leq \frac{1}{T} \left\| \frac{1}{N_k T} \sum_{i \in G_k} \sum_{j=1}^{N} \lambda_i^0 \sum_{j=1}^{N} \left\| X'_i \varepsilon_j \right\| \left\| \varepsilon_j^* F^0 \right\| \left\| H \right\| + o_P((NT)^{-1}) \leq \frac{1}{T} \left( \frac{1}{N_k T} \sum_{i \in G_k} \sum_{j=1}^{N} \lambda_i^0 \left\| X'_i \varepsilon_j \right\| \left\| \varepsilon_j^* F^0 \right\| \left\| H \right\| + o_P((NT)^{-1}) \right\|^2 \left\| H \right\| + o_P((NT)^{-1}) \leq O_P(T^{-1})O_P(1)O_P(1)O_P(1) + o_P((NT)^{-1}) = O_P(T^{-1})$$

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

(iii) Let $\hat{V}_{k,NT} = \frac{1}{N_k T} \sum_{i \in G_k} X'_i M_{\bar{F}} (\varepsilon_i - \frac{1}{N} \sum_{j=1}^{N} \nu_{j,i} \varepsilon_j)$, following the proof of Lemma A.6(ii), we can show that $\hat{V}_{k,NT} = -\hat{V}_{k,NT} = \hat{V}_{k,NT} = \hat{V}_{k,NT}$, say. By (A.1), $v_{1,k,NT} = \sum_{i=1}^{4} -\tilde{A}_{k,i} \sum_{i \in G_k} X_i d_{i,1}$, say. By Lemma A.7(iv), $\left\| v_{1,k,NT,1} \right\| = \frac{1}{N_k T} \sum_{i \in G_k} X'_i (\hat{F} - F^0 H) (\hat{F} - F^0 H)' \varepsilon_i$.

By Lemma A.2(ii) and Assumption A.1(ix)

$$|\omega_{P}^T \hat{v}_{1,k,NT,1}| \leq \frac{1}{N_k T^2} \left\| \left( \hat{F} - F^0 H \right)' H' F^0 \varepsilon_i \right\| \leq \frac{1}{T} \left\| \left( \hat{F} - F^0 H \right)' H' F^0 \right\| \leq \frac{1}{N_k T} \sum_{i \in G_k} \varepsilon_i \omega_{P}^T X_i \leq O_P(T^{-1/2} + \delta_{NT})O_P(1)O_P(1)O_P(1) + o_P((NT)^{-1}) = O_P(T^{-1/2} + \delta_{NT})$$

43
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

$$
\|v_{1kNT,4}\| = \left\| \frac{1}{N_kT} \sum_{i \in G_k^0} X_i^0 F^0 H \left[ I_{R^0} - (T^{-1} H' F^0 F^0 H)^{-1} \right] H' F^0 \varepsilon_i \right\|
$$

$$\leq \left\{ \frac{1}{T} \left\| F^0 \right\| H^2\right\} \left\| I_{R^0} - (T^{-1} H' F^0 F^0 H)^{-1} \right\| \left\{ \frac{1}{N_kT} \sum_{i \in G_k^0} \|X_i\| \|F^0\varepsilon_i\| \right\}
$$

$$= O_P(T^{-1/2})O_P(T^{-1/2} + \delta_{NT}^{-2})O_P(1) = O_P(T^{-1}).$$

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

Recall that $X_i^0 = \frac{1}{N} \sum_{j=1}^N \nu_{ij} X_j$. Using (A.1), we have $v_{2kNT} = \sum_{i=1}^4 \frac{1}{N_kT} \sum_{i \in G_k^0} \lambda^0_j X_i^0 d_1 \varepsilon_i \equiv \sum_{i=1}^4 v_{2kNT,i}$, say. By Cauchy-Schwarz inequality, the fact that $\frac{1}{N_kT} \sum_{i \in G_k^0} \|X_i^0\|^2 = O_P(1)$, Lemmas A.2(i) and A.7(iii), and Theorem 3.2,

$$|\omega' v_{2kNT,1}| = \left| \frac{1}{N_kT^2} \sum_{i \in G_k^0} \omega_p' X_i^0 (\hat{F} - F^0) H' F^0 \varepsilon_i \right|
$$

$$\leq \left\{ \frac{1}{N_kT^2} \sum_{i \in G_k^0} \omega_p' X_i^0 (\hat{F} - F^0) H' F^0 \varepsilon_i \right\}^{1/2} \left\{ \frac{1}{N_kT} \sum_{i \in G_k^0} \|X_i^0\|^2 \right\}^{1/2} \left\{ \frac{1}{N_kT^2} \sum_{i \in G_k^0} \varepsilon_i' (\hat{F} - F^0) H' (\hat{F} - F^0) \varepsilon_i \right\}^{1/2}
$$

$$= O_P\left(\delta_{NT}^{-1}\right) O_P(1) O_P(\delta_{NT}^{-2}) = O_P\left(\delta_{NT}^{-3}\right).
$$

By Lemma A.2(i) and Theorem 3.2

$$|\omega' v_{2kNT,2}| = \left| \frac{1}{N_kT^2} \sum_{i \in G_k^0} \text{tr} \left[ \omega_p' X_i^0 (\hat{F} - F^0) H' F^0 \varepsilon_i \right] \right|
$$

$$= \left| \frac{1}{N_kT^2} \sum_{i \in G_k^0} \text{tr} \left[ \omega_p' \sum_{j=1}^N \nu_{ji} X_j^0 (\hat{F} - F^0) H' F^0 \varepsilon_i \right] \right|
$$

$$= \left| \frac{1}{N_kT^2} \sum_{i \in G_k^0} \text{tr} \left[ \omega_p' \sum_{j=1}^N \nu_{ji} X_j^0 (\hat{F} - F^0) H' F^0 \varepsilon_i \right] \right|
$$

$$\leq \left\| \lambda^{0,i}_j \right\| \left\| X_j \right\| \left\{ \frac{1}{T^{1/2}} \left\| \hat{F} - F^0 H \right\| \right\} \left\| H \right\| \left\| \frac{1}{N_kT} \sum_{i \in G_k^0} F^0 \varepsilon_i \right\| \left\| \lambda^{0,i} \right\|
$$

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

$$
\|v_{2k NT, 4}\| = \left\| \frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i^t F^0 H \left[ I_{R_0} - (T^{-1} H' F^0 F^0 H)^{-1} \right] H' F^0 \varepsilon_i \right\|
$$

$$
\leq \left\{ \frac{1}{T} \| F^0 \| \| H \| \right\} \left\| I_{R_0} - (T^{-1} H' F^0 F^0 H)^{-1} \right\| \left\{ \frac{1}{N_k T} \sum_{i \in G_k^0} \| X_i^t \|_2^2 \right\}^{1/2} \left\{ \frac{1}{N_k T} \sum_{i \in G_k^0} \| F^0 \varepsilon_i \|_2 \right\}^{1/2}
$$

$$
= O_P(T^{-1/2}) O_P(T^{-1/2}) O_P (1) O_P (1) = O_P(T^{-1}).
$$

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

(iii) Note that $V_{k NT} = \frac{1}{N_k T} \sum_{i \in G_k^0} X_i^t M_{P0} \varepsilon_i - \frac{1}{N_k T} \sum_{i \in G_k^0} X_i^t M_{P0} \sum_{j=1}^N \nu_{ji} \varepsilon_j$. The conclusion follows from Assumption A.4(ii).

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

$$
\frac{1}{N_k T^2} \left\| \frac{1}{NT} \sum_{j=1}^N \sum_{i \in G_k^0} \hat{F}^t \varepsilon_j \varepsilon_j' \varepsilon_i \omega_p X_i^t \right\|
$$

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

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

$$
\left\| \hat{F} - F^0 H \right\| \left\| \frac{1}{NT} \sum_{j=1}^N \varepsilon_j \varepsilon_j' \right\| \leq O_P \left( T^{1/2} \delta_{NT}^{-1} \right) O_P \left( \delta_{NT}^{-1} \right) O_P \left( \delta_{NT}^{-1} \right) = O_P \left( T^{1/2} \delta_{NT}^{-3} \right).
$$

So (i) follow.

(ii) By Lemma A.8,

$$
\frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i^t F^0 \left( T^{-1} F^0 F^0 \right)^{-1} (\hat{F} H^{-1} - F^0) \varepsilon_i
$$

$$
= \sum_{i=1}^8 \frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i^t F^0 \left( T^{-1} F^0 F^0 \right)^{-1} (V_{NT}^{-1} H^{-1})\varepsilon_i'
$$

$$
= \sum_{i=1}^8 \frac{1}{N_k T^2} \sum_{i \in G_k^0} X_i^t F^0 \left( T^{-1} F^0 F^0 \right)^{-1} G' \varepsilon_i \equiv \sum_{i=1}^8 L_i, \text{ say.}
$$

Noting that $\|a_1\| = O_P \left( T^{1/2} \delta_{NT}^{-1} \right), O_P \left( T^{-1/2} \delta_{NT}^{-2} \right)$, 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 $a_2 = -\frac{1}{NT} \sum_{i=1}^N X_i \hat{b}_i$
\[ x_l^{(y)} F_0^0 \tilde{F} = -\frac{1}{NT} \sum_{l=1}^{K_O} \sum_{i \in G_l^0} X_i (\hat{\alpha}_l - \alpha_l^0) \lambda_l^{(y)} F_0^0 \tilde{F} + o_P((NT)^{-1/2}). \] It follows that

\[
\|\omega_p L_2\| = \frac{1}{N T_k T^2} \sum_{l=1}^{K_O} \sum_{j \in G_l^0} \sum_{i \in G_l^0} \omega_p^j X_i^j F_0^0 (T^{-1} F_0^0 F_0^0)^{-1} G' F \hat{F} \hat{F}^t \varepsilon_j (\hat{\alpha}_l - \alpha_l^0) X_j^t \varepsilon_i + o_P((NT)^{-1/2})
\]

\[
= \frac{1}{N T_k T^2} \sum_{l=1}^{K_O} \sum_{j \in G_l^0} \sum_{i \in G_l^0} \varepsilon_j X_i \left\| (F_0^0 (T^{-1} F_0^0 F_0^0)^{-1} (N^{-1} \Lambda_0^0 \Lambda_0^0)^{-1} \lambda_l^{(y)} (\hat{\alpha}_l - \alpha_l^0) X_j^t \left( \sum_{i \in G_l^0} \varepsilon_i \omega_p^i X_i^i \right) \right\| + o_P((NT)^{-1/2})
\]

\[
\leq c_{1NT} \sum_{l=1}^{K_O} \left\| \hat{\alpha}_l - \alpha_l^0 \right\| \left\| F_0^0 (T^{-1} F_0^0 F_0^0)^{-1} \right\| \left\| (N^{-1} \Lambda_0^0 \Lambda_0^0)^{-1} \right\| \left( F_0^0 (T^{-1} F_0^0 F_0^0)^{-1} \right) + o_P((NT)^{-1/2})
\]

where \( c_{1NT} = T^{-1/2} \left\| F_0^0 (T^{-1} F_0^0 F_0^0)^{-1} \right\| \left\| (N^{-1} \Lambda_0^0 \Lambda_0^0)^{-1} \right\|. \) Similarly, using \( a_3 = \frac{1}{NT} \sum_{i=1}^{N} X_i \hat{b}_i \hat{c}_i \tilde{F} = -\frac{1}{NT} \sum_{i=1}^{K_O} \sum_{j \in G_l^1} X_i (\hat{\alpha}_l - \alpha_l^0) \varepsilon_i \hat{F} + o_P((NT)^{-1/2}) \), we have

\[
\|\omega_p L_3\| = \frac{1}{N T_k T^3} \left\| \sum_{l=1}^{K_O} \sum_{j \in G_l^0} \sum_{i \in G_l^0} \omega_p^j X_i^j F_0^0 (T^{-1} F_0^0 F_0^0)^{-1} G' \hat{F}^t \varepsilon_j \left( \hat{\alpha}_l - \alpha_l^0 \right)^t \varepsilon_i \right\|
\]

\[
= \frac{1}{N T_k T^3} \sum_{l=1}^{K_O} \sum_{j \in G_l^0} \sum_{i \in G_l^0} \varepsilon_j \left\| \hat{F}^t \varepsilon_j \left( \hat{\alpha}_l - \alpha_l^0 \right)^t \varepsilon_i \right\|
\]

\[
\leq c_{2NT} \sum_{l=1}^{K_O} \left\| \hat{\alpha}_l - \alpha_l^0 \right\| \left\| F_0^0 (T^{-1} F_0^0 F_0^0)^{-1} \right\| \left\| (N^{-1} \Lambda_0^0 \Lambda_0^0)^{-1} \right\| \left( F_0^0 (T^{-1} F_0^0 F_0^0)^{-1} \right) + o_P((NT)^{-1/2})
\]

where \( c_{2NT} = T^{-1/2} \left\| F_0^0 (T^{-1} F_0^0 F_0^0)^{-1} \right\| \left\| (N^{-1} \Lambda_0^0 \Lambda_0^0)^{-1} \right\|. \) and we use the fact that

\[
\frac{1}{NT^{3/2}} \sum_{j \in G_l^0} \|X_j\| \left\| \varepsilon_j \hat{F} \right\| \leq \frac{1}{NT^{3/2}} \sum_{j \in G_l^0} \|X_j\| \left\| X_j^t \right\| \leq \frac{1}{NT^{3/2}} \sum_{j \in G_l^0} \|X_j\| \left\| \varepsilon_j \hat{F} \right\|
\]

\[
= O_P(T^{-1/2}) + \frac{1}{NT^{3/2}} \sum_{j \in G_l^0} \|X_j\| \left\| \varepsilon_j \hat{F} \right\|
\]

\[
= O_P(T^{-1/2} + \frac{1}{NT^{3/2}} \sum_{j \in G_l^0} \|X_j\| \left\| \varepsilon_j \hat{F} \right\|
\]

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

\[
\|\omega_p L_6\| = \frac{1}{NT_k T^3} \sum_{i \in G_l^0} \sum_{j=1}^{N} \omega_p^j X_i^j F_0^0 (T^{-1} F_0^0 F_0^0)^{-1} G' \hat{F} \varepsilon_j \lambda_l^{(y)} F_0^0 \varepsilon_i
\]

\[
= \frac{1}{NT_k T^3} \sum_{i \in G_l^0} \sum_{j=1}^{N} \varepsilon_j X_i \left\| \hat{F} \varepsilon_j \lambda_l^{(y)} \right\| \left\| F_0^0 \varepsilon_i \omega_p^i X_i^i \right\|
\]

\[
\leq c_{2NT} T^{-1/2} \sum_{i \in G_l^0} \left\| \hat{F} \varepsilon_j \lambda_l^{(y)} \right\| \left\| F_0^0 \varepsilon_i \omega_p^i X_i^i \right\|
\]

\[
= T^{-1/2} O_P(\delta_{NT} \right) O_P(\delta_{NT}^{-1}) = o_P((NT)^{-1/2}).
\]
In addition, \( L_7 = \frac{1}{N T^2} \sum_{i \in G_k^0} \sum_{j=1}^{N} X_i' F_0 (T^{-1} F_0' F_0)^{-1} G' a_1^i \varepsilon_i = \frac{1}{N N_k T^2} \sum_{i \in G_k^0} \sum_{j=1}^{N} X_i' F_0 (T^{-1} F_0' F_0)^{-1} \times (N^{-1} A_0' A_0^{-1})^{-1} \lambda_j^0 \varepsilon_j \varepsilon_i \), and

\[
\| \omega_0' L_2 \| = \frac{1}{N_k T} \left\| \sum_{i \in G_k^0} \omega_0' X_i' F_0 (T^{-1} F_0' F_0)^{-1} G' a_2^i \varepsilon_i \right\| \\
= \frac{1}{N N_k T^3} \left\| \sum_{j=1}^{N} \sum_{i \in G_k^0} \omega_0' X_i' F_0 (T^{-1} F_0' F_0)^{-1} G' \hat{F'} \varepsilon_j \varepsilon_i \right\| \\
= \frac{1}{N N_k T^3} \left\| \text{tr} \left( (T^{-1} F_0' F_0)^{-1} G' \sum_{j=1}^{N} \sum_{i \in G_k^0} \hat{F'} \varepsilon_j \varepsilon_i \omega_0' X_i' F_0 \right) \right\| \\
\leq c_{NT} \frac{1}{T^{1/2}} \left\| \frac{1}{N N_k T^2} \sum_{j=1}^{N} \sum_{i \in G_k^0} \hat{F'} \varepsilon_j \varepsilon_i \omega_0' X_i' \right\| = T^{-1/2} O_P((T^{1/2} \delta_{NT}^{-1/2}) = o_P((NT)^{-1/2}).
\]

In sum, we have shown that \( \frac{1}{N T} \sum_{i=1}^{N} (Y_i - X_i \beta_i)' M_{F_i} (Y_i - X_i \beta_i) + F_0 \lambda_i^0 + \varepsilon_i \), we have

\[
0 \geq Q_{1,NT} (\beta, F_{(r)}) - Q_{1,NT} (\beta^0, F_{(r)}) \\
= \frac{1}{N T} \sum_{i=1}^{N} \left[ (Y_i - X_i \beta_i)' M_{F_i} (Y_i - X_i \beta_i) - (Y_i - X_i \beta_i)' M_{F_i} (Y_i - X_i \beta_i) \right] \\
= \frac{1}{N T} \sum_{i=1}^{N} \left[ (\beta_i - \beta_i)' X_i' M_{F_i} X_i (\beta_i - \beta_i) - 2 (\beta_i - \beta_i)' X_i' M_{F_i} X_i \lambda_i \right] \\
- \frac{2}{N T} \sum_{i=1}^{N} \left( \beta_i - \beta_i \right)' X_i' M_{F_i} (\beta_i - \beta_i). \\
\]

By Lemma A.1 (with \( R_0 \) and \( F \) replaced by \( R \) and \( F_{(r)} \)), we can show that \( \frac{1}{N T} \sum_{i=1}^{N} (\beta_i - \beta_i)' X_i' M_{F_i} (\beta_i - \beta_i) = o_P(1) \) provided that \( \frac{1}{N T} \sum_{i=1}^{N} \| \beta_i - \beta_i \|^2 = O_P(1) \), which can be shown by using similar arguments as used in the proof of Theorem 3.1. Let \( \tilde{d}_{\beta, R} = \text{vec}(\beta_{(r)} - \beta^0) \) and \( \tilde{d}_{F, R} = T^{-1/2} \text{vec}(M_{F_i} F_0) \). Define

\[
\hat{A}_R = \frac{1}{T} \text{diag} \left( X_1' M_{F_i} X_1, ..., X_N' M_{F_i} X_N \right) \quad \text{and} \quad \hat{C}_R = \frac{1}{T^{1/2}} \text{diag} \left( \lambda_1^0 \otimes M_{F_i} X_1, ..., \lambda_N^0 \otimes M_{F_i} X_N \right).
\]

Then \( \frac{1}{N} \sum_{i=1}^{N} (\hat{\beta}_{i, R} - \beta_i)' X_i' M_{F_i} (\hat{\beta}_{i, R} - \beta_i) - 2 (\hat{\beta}_{i, R} - \beta_i)' X_i' M_{F_i} F_0 \lambda_i^0 = \frac{1}{N} \tilde{d}_{\beta, R} \hat{A}_R \tilde{d}_{\beta, R} - \frac{2}{N} \tilde{d}_{F, R} \times \hat{C}_R \tilde{d}_{\beta, R} \).

It follows that

\[
0 \geq \frac{1}{N} \tilde{d}_{\beta, R} \hat{A}_R \tilde{d}_{\beta, R} - \frac{2}{N} \tilde{d}_{F, R} \hat{C}_R \tilde{d}_{\beta, R} + o_P(1) \\
\geq \frac{1}{N} \tilde{d}_{\beta, R} \hat{A}_R \tilde{d}_{\beta, R} - 2 \left( \tilde{d}_{F, R} \hat{d}_{F, R} \right)^{1/2} + \frac{1}{N} \left( \frac{1}{N} \tilde{d}_{\beta, R} \hat{C}_R \tilde{d}_{\beta, R} \right)^{1/2} + o_P(1).
\]

Proof of Lemma A.10. (i) Noting that \( Q_{1,NT} (\beta, F_{(r)}) = \frac{1}{N T} \sum_{i=1}^{N} (Y_i - X_i \beta_i)' M_{F_i} (Y_i - X_i \beta_i) \) and \( Y_i - X_i \beta_i, R = X_i (\beta_i - \beta_i, R) + F_0 \lambda_i^0 + \varepsilon_i \), we have

\[
0 \geq Q_{1,NT} (\beta, F_{(r)}) - Q_{1,NT} (\beta^0, F_{(r)}) \\
= \frac{1}{N T} \sum_{i=1}^{N} \left[ (Y_i - X_i \beta_i)' M_{F_i} (Y_i - X_i \beta_i) - (Y_i - X_i \beta_i)' M_{F_i} (Y_i - X_i \beta_i) \right] \\
= \frac{1}{N T} \sum_{i=1}^{N} \left[ (\beta_i - \beta_i)' X_i' M_{F_i} X_i (\beta_i - \beta_i) - 2 (\beta_i - \beta_i)' X_i' M_{F_i} F_0 \lambda_i \right] \\
- \frac{2}{N T} \sum_{i=1}^{N} \left( \beta_i - \beta_i \right)' X_i' M_{F_i} (\beta_i - \beta_i). \\
\]

By Lemma A.1 (with \( R_0 \) and \( F \) replaced by \( R \) and \( F_{(r)} \)), we can show that \( \frac{1}{N T} \sum_{i=1}^{N} (\beta_i - \beta_i)' X_i' M_{F_i} (\beta_i - \beta_i) = o_P(1) \) provided that \( \frac{1}{N T} \sum_{i=1}^{N} \| \beta_i - \beta_i \|^2 = O_P(1) \), which can be shown by using similar arguments as used in the proof of Theorem 3.1. Let \( \tilde{d}_{\beta, R} = \text{vec}(\beta_{(r)} - \beta^0) \) and \( \tilde{d}_{F, R} = T^{-1/2} \text{vec}(M_{F_i} F_0) \). Define

\[
\hat{A}_R = \frac{1}{T} \text{diag} \left( X_1' M_{F_i} X_1, ..., X_N' M_{F_i} X_N \right) \quad \text{and} \quad \hat{C}_R = \frac{1}{T^{1/2}} \text{diag} \left( \lambda_1^0 \otimes M_{F_i} X_1, ..., \lambda_N^0 \otimes M_{F_i} X_N \right).
\]

Then \( \frac{1}{N} \sum_{i=1}^{N} (\beta_{i, R} - \beta_i)' X_i' M_{F_i} (\beta_{i, R} - \beta_i) - 2 (\beta_{i, R} - \beta_i)' X_i' M_{F_i} F_0 \lambda_i^0 = \frac{1}{N} \tilde{d}_{\beta, R} \hat{A}_R \tilde{d}_{\beta, R} - \frac{2}{N} \tilde{d}_{F, R} \times \hat{C}_R \tilde{d}_{\beta, R} \).

It follows that

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

\[
\frac{1}{N} \hat{d}_{\beta,R} \hat{A}_R \hat{d}_\beta R = \frac{1}{NT} \sum_{i=1}^{N} \left( \hat{\beta}_{i,R} - \beta_i^0 \right)' X_i' M_{F(R)} \left( Y_i - X_i \hat{\beta}_{i,R0} \right) \]

Note that \( V(R, \hat{\beta}(R)) = \min_{\beta, F, \theta} Q_{\text{LNT}}(\beta, F(R)) \) 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

\[
V \left( R_0, \hat{\beta}(R_0) \right) = V \left( R_0, \hat{\beta}(R_0) \right)
\]

Next,

\[
\frac{1}{NT} \sum_{i=1}^{N} (\lambda_i^0)' \lambda_i^0 F_i^0 \lambda_i^0
\]

\[
= \frac{1}{T} \sum_{t=R+1}^{T} \mu_t \left( \frac{1}{N} \sum_{i=1}^{N} (\lambda_i^0)' (\lambda_i^0) \right) = \frac{1}{T} \sum_{t=R+1}^{T} \mu_t \left( \frac{F_0^0 \lambda_0^0 \lambda_0^0}{N} \right) \geq \mu_{\min} \left( \frac{\lambda_0^0 \lambda_0^0}{T} \right) \sum_{t=R+1}^{T} \mu_t \left( \frac{F_0^0 F_0^0}{T} \right)
\]

\[
\geq \left( R_0 - R \right) \mu_{\min} \left( \frac{\lambda_0^0 \lambda_0^0}{N} \right) \mu_{\min} \left( \frac{F_0^0 F_0^0}{T} \right) = (R_0 - R) \mu_{\min} (\Sigma_\Lambda) \mu_{\min} (\Sigma_F) + o_P(1).
\]
It follows that w.p.a.1, \( V(R, \hat{\beta}_R^R) - V(R_0, \hat{\beta}_R^{R_0}) \geq c_R \) with \( c_R = (R_0 - R) \mu_{\min} (\Sigma_A) \mu_{\min} (\Sigma_R) / 2 > 0 \).

(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. \( \blacksquare \)

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

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

Noting that \( \beta_i^0, \ i = 1, ..., N \), 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 \( k = 1, ..., K \). 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

\[
\sum_{i \in \hat{G}_k^0} P \left( \hat{E}_{kNT,i} \right) = o(1) \quad \text{for } k = 1, ..., K_0 \quad \text{and} \quad \sum_{i \in \hat{G}_k^0} P \left( \hat{F}_{kNT,i} \right) = o(1) \quad \text{for } k = 1, ..., K_0.
\]

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 \( 1 \{ i \in \hat{G}_k \} = 1 \{ i \in G^0_k \} + 1 \{ i \in \hat{G}_k \} \), we have

\[
\hat{\sigma}^2_{\hat{G}(K, \kappa)} = \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in G^0_k} \sum_{t=1}^{T} \left[ \hat{\epsilon}_{it} \right]^2 + \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in \hat{G}_k(K, \kappa)} \sum_{t=1}^{T} \left[ \hat{\epsilon}_{it} \right]^2
\]

\[
- \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in \hat{G}_k} \sum_{t=1}^{T} \left[ \hat{\epsilon}_{it} \right]^2 + \frac{1}{NT} \sum_{k=K_0+1}^{K} \sum_{i \in \hat{G}_k(K, \kappa)} \sum_{t=1}^{T} \left[ \hat{\epsilon}_{it} \right]^2
\]

\[
\equiv D_{1NT} + D_{2NT} - D_{3NT} + D_{4NT}.
\]

where \( \hat{\epsilon}_{it} = Y_{it} - \hat{\alpha}_{\hat{G}_k(K, \kappa)} X_{it} - \hat{\lambda}_i^{(K, \kappa)} \hat{F}_i^{(K, \kappa)} \). Following the proof of Theorem 3.7, for \( k = 1, ..., K_0 \), we have \( \hat{\alpha}_{\hat{G}_k(K, \kappa)} - \alpha_k^0 = O_p(\delta^{-2}) \). In addition, we can show that \( \frac{1}{T} \sum_{t=1}^{T} \left\| \hat{F}_t^{(K, \kappa)} - H^T \hat{F}_t \right\|^2 = O_p(\delta^{-2}) \)

and \( \frac{1}{N} \sum_{i=1}^{N} \left\| \chi_i^{(K, \kappa)} - H^{-1} \chi_i \right\|^2 = O_p(\delta^{-2}) \). With these results, we can readily show that \( D_{1NT} = \hat{\sigma}^2_{G^0} + O_p(\delta^{-2}) \). For \( D_{2NT}, D_{3NT}, \) and \( D_{4NT} \), we have that for any \( \epsilon > 0 \),

\[
P \left( D_{2NT} \geq \delta^{-2} \epsilon \right) \leq \sum_{i=1}^{K_0} P \left( \hat{F}_{kNT} \right) \rightarrow 0,
\]

\[
P \left( D_{3NT} \geq \delta^{-2} \epsilon \right) \leq \sum_{i=1}^{K_0} P \left( \hat{E}_{kNT} \right) \rightarrow 0, \quad \text{and}
\]

\[
P \left( D_{4NT} \geq \delta^{-2} \epsilon \right) \leq \sum_{i=1}^{N} P \left( i \in \cup_{K_0+1 \leq k \leq K} \hat{G}_k(K, \kappa) \right) \rightarrow 0.
\]

It follows that \( \hat{\sigma}^2_{\hat{G}(K, \kappa)} = \hat{\sigma}^2_{G^0} + O_p(\delta^{-2}) \) for all \( K_0 \leq K \leq K_{max} \). \( \blacksquare \)
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 $V_{kNT} = \frac{1}{N_k T} \sum_{i \in G_k^0} X_i' M_{F^0} \varepsilon_i - \frac{1}{N_k N_T} \sum_{i \in G_k^0} \sum_{j=1}^N X_i' M_{F^0} \varepsilon_j \equiv V_{1kNT} - V_{2kNT}$, say. We further decompose $V_{1kNT}$ as follows

$V_{1kNT} = \frac{1}{N_k T} \sum_{i \in G_k^0} X_i' \varepsilon_i - \frac{1}{N_k T} \sum_{i \in G_k^0} X_i' F^0 (T^{-1} F^{0'} F^0)^{-1} F^{0'} \varepsilon_i \equiv V_{1kNT,1} - V_{1kNT,2}$. Apparently, $V_{1kNT,1} = O_P \left( (N_k T)^{-1/2} \right)$ under Assumption A.1(vii). Let $\mu_{ts} = F_{t}^0 (T^{-1} F^{0'} F^0)^{-1} F_{s}^0$. Note that $\mu_{ts}$ is measurable with respect to $D$ and $V_{1kNT,2} = \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T X_{it} \varepsilon_{is} \mu_{ts}$. 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

$$
|E_D \left( V_{1kNT,2} \right)| = \left| \frac{1}{N_k T^2} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T E_D (X_i t \varepsilon_{is}) \mu_{ts} \right|
$$

$$
\leq \frac{c_{NT}}{2 N_k T^2} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \|X_i t\|_{2+\sigma, D} \| \varepsilon_{is} \|_{2+\sigma, D} \| F_0 \| \| F_0^0 \| \alpha_{NT}^\sigma (t-s)^{(2+\sigma)/\sigma}
$$

$$
\leq \frac{c_{NT}}{2 N_k T^2} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \| \varepsilon_{is} \|_{2+\sigma, D} \| F_0 \| \| F_0^0 \| \alpha_{NT}^\sigma (t-s)^{(2+\sigma)/\sigma}
$$

$$
\equiv v_{1,1} + v_{1,2}, \text{ say},
$$

where $c_{NT} = \left\| (T^{-1} F^{0'} F^0)^{-1} \right\|$, 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), $\sum_{r=1}^{\infty} \alpha_{NT}^\sigma (\tau)^{(2+\sigma)/(1+\sigma)} = O_P (1)$. Hence

$$
v_{1,1} \leq \frac{c_{NT}}{2 T} \left\{ \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^T \|X_i t\|_{2+\sigma, D} \| F_0 \| \| F_0^0 \| \right\} \sum_{r=1}^{\infty} \alpha_{NT}^\sigma (\tau)^{(2+\sigma)/(1+\sigma)} = T^{-1} O_P (1) \leq O_P (T^{-1}) .
$$

By the same token, $v_{1,2} = O_P \left( T^{-1} \right)$. It follows that $E_D \left( V_{1kNT,2} \right) = O_P \left( T^{-1} \right)$. Next,

$$
E_D \left( V_{1kNT,2}^2 \right) = \frac{1}{N_k T^4} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T \sum_{q=1}^T E_D \left( X_{it} \varepsilon_{is} X_{ir} \varepsilon_{iq} \right) \mu_{ts} \mu_{tq}
$$

$$
+ \frac{1}{N_k T^4} \sum_{i \in G_k^0, j \in G_k^0, j \neq i} \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T \sum_{q=1}^T E_D \left( X_{it} \varepsilon_{is} X_{ir} \varepsilon_{iq} \right) E_D \left( X_{jt} \varepsilon_{js} X_{jr} \varepsilon_{jq} \right) \mu_{ts} \mu_{tq}
$$

$$
= \frac{1}{N_k T^4} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T \sum_{q=1}^T E_D \left( X_{it} \varepsilon_{is} X_{ir} \varepsilon_{iq} \right) \mu_{ts} \mu_{tq} + [E_D \left( V_{1kNT,2} \right)]^2
$$

$$
- \frac{1}{N_k T^4} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T \sum_{q=1}^T E_D \left( X_{it} \varepsilon_{is} \right) E_D \left( X_{ir} \varepsilon_{iq} \right) \mu_{ts} \mu_{tq}
$$

$$
\equiv v_{2,1} + v_{2,2} - v_{2,3}, \text{ say}.
$$
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(ii) and Davydov and H"{o}lder 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 $E_D \left( X_{it} \varepsilon_{is} X_{ir} \varepsilon_{iq} \right) = 0$ 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 $v_{2,1}(s)$ analogously to $v_{2,1}$ but with the time indices restricted to subcase $s$ for $s = a_1, a_2, a_3$.

$$
|v_{2,1}(a1)| \leq \frac{1}{N_k^2 T^4} \sum_{i \in G_k^0} \sum_{T \geq t > s > r > q \geq 1} |E_D \left( X_{it} \varepsilon_{is} X_{ir} \varepsilon_{iq} \right) | \left| \mu_{is} \mu_{rq} \right|
$$

$$
\leq \frac{c_1^2}{N_k^2 T^4} \sum_{i \in G_k^0} \sum_{T \geq t > s > r > q \geq 1} \|X_{it} \varepsilon_{is} X_{ir}\|_{\delta + 4\sigma, D} \|\varepsilon_{iq}\|_{\delta + 4\sigma, D} \alpha_N^D (r - q)^{(2\sigma)/(1 + \sigma)}
\times \|F_0^0\| \|F_0^0\| \|F_0^0\| |
$$

$$
\leq \frac{c_2 NT}{2 N_k^2 T^4} \sum_{i \in G_k^0} \sum_{T \geq t > s > r > q \geq 1} \|X_{it} \varepsilon_{is} X_{ir}\|_{\delta + 4\sigma, D} \|\varepsilon_{iq}\|_{\delta + 4\sigma, D} \|F_0^0\|^2 \alpha_N^D (r - q)^{(2\sigma)/(1 + \sigma)}
\leq \frac{c_3}{N_k^2 T^2} \sum_{i \in G_k^0} \sum_{T \geq t > s > r > q \geq 1} \|X_{it} \varepsilon_{is} X_{ir}\|_{\delta + 4\sigma, D} \|\varepsilon_{iq}\|_{\delta + 4\sigma, D} \|F_0^0\|^2 \alpha_N^D (r - q)^{(2\sigma)/(1 + \sigma)}
$$

$$
\equiv \ v_{2,11}(a1) + v_{2,12}(a1), \text{ say.}
$$

By H"{o}lder inequality

$$
\|X_{it} \varepsilon_{is} X_{ir}\|_{\delta + 4\sigma, D} = \left| E_D \left( X_{it} \varepsilon_{is} X_{ir} \right) \right|^{\delta + 4\sigma} \leq \left( \left| E_D \left( X_{it} \varepsilon_{is} X_{ir} \right) \right|^{1/3} \left| E_D \left( \varepsilon_{is} X_{ir} \right) \right|^{4/3} \left| E_D \left( X_{ir} \right) \right|^{4/3} \right)^{\delta + 4\sigma}
$$

$$
\leq \left( \left| E_D \left( X_{it} \varepsilon_{is} X_{ir} \right) \right|^{1/3} \left| E_D \left( \varepsilon_{is} \right) \right|^{4/3} \left| E_D \left( X_{ir} \right) \right|^{4/3} \right)^{\delta + 4\sigma}
$$

$$
= \left( \left| E_D \left( X_{it} \varepsilon_{is} X_{ir} \right) \right|^{1/3} \left| E_D \left( \varepsilon_{is} \right) \right|^{4/3} \left| E_D \left( X_{ir} \right) \right|^{4/3} \right)^{\delta + 4\sigma}
$$

It follows that

$$
v_{2,11}(a1) \leq \frac{c_4 NT}{N_k T} \sum_{i \in G_k^0} \sum_{T = 1}^{T} \sum_{s = 1}^{T} \sum_{r = 1}^{T} \|X_{it} \varepsilon_{is} X_{ir}\|_{\delta + 4\sigma, D} \|\varepsilon_{iq}\|_{\delta + 4\sigma, D} \|F_0^0\|^2 \sum_{\tau = 1}^{\infty} \alpha_N^D (\tau)^{(2\sigma)/(1 + \sigma)}
$$

$$
\leq \frac{c_5}{N_k T} \sum_{i \in G_k^0} \sum_{q = 1}^{T} \sum_{s = 1}^{T} \|\varepsilon_{iq}\|_{\delta + 4\sigma, D} \|F_0^0\|^2 \sum_{\tau = 1}^{\infty} \alpha_N^D (\tau)^{(2\sigma)/(1 + \sigma)}
$$

$$
= O_P \left( N^{-1}T^{-1} \right).
$$

In addition

$$
v_{2,11}(a1) \leq \frac{c_6}{N_k T} \sum_{i \in G_k^0} \sum_{q = 1}^{T} \sum_{s = 1}^{T} \|\varepsilon_{iq}\|_{\delta + 4\sigma, D} \|F_0^0\|^2 \sum_{\tau = 1}^{\infty} \alpha_N^D (\tau)^{(2\sigma)/(1 + \sigma)}
$$

$$
\leq \frac{c_7}{N_k T} \sum_{i \in G_k^0} \sum_{q = 1}^{T} \sum_{s = 1}^{T} \|\varepsilon_{iq}\|_{\delta + 4\sigma, D} \|F_0^0\|^2 = O_P \left( N^{-1}T^{-1} \right).
$$
So \(v_{2,1}(a1) = O_P(N^{-1}T^{-1})\). Similarly we can show that \(v_{2,1}(a2) = O_P(N^{-1}T^{-1})\) and \(v_{2,1}(a3) = O_P(N^{-1}T^{-1})\). So \(v_{2,1} = O_P(N^{-1}T^{-1})\). By the same token, we can show that \(v_{2,3} = O_P(N^{-1}T^{-1})\). By the analysis of \(E_D(V_{1kNT,2})\), \(v_{2,2} = O_P(T^{-2})\). It follows that \(E_D(V_{1kNT,2}^2) = O_P(N^{-1}T^{-1} + T^{-2})\) 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 \(V_{2kNT} = \frac{1}{N^2 T^2} N \sum_{i \in G_k^0} X_i^T M F_0 e_i^2 = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)\). Write \(V_{2kNT} = \frac{1}{N N_k T} \sum_{i \in G_k^0} X_i^T \sum_{j=1}^N \nu_{ji} \varepsilon_j - \frac{1}{N N_k T} \sum_{i \in G_k^0} X_i^T P D \sum_{j=1}^N \nu_{ji} \varepsilon_j = V_{2kNT,1} - V_{2kNT,2}\), say. Noting that

\[
E_D(\nu_{ji}) = \frac{1}{N N_k T} \sum_{i \in G_k^0} \sum_{j=1}^N \sum_{t=1}^T E_D(X_{it} \varepsilon_{jt} X_{ks} \varepsilon_{js}) \nu_{ji} \nu_{kl}
\]
and

\[
E_D(V_{2kNT,1}) = \frac{1}{N^2 N_k T^2} \sum_{i \in G_k^0} \sum_{j=1}^N \sum_{k \in G_k^0} \sum_{l=1}^N \sum_{s=1}^T E_D(X_{it} \varepsilon_{jt} X_{ks} \varepsilon_{ls}) \nu_{ji} \nu_{kl}
\]

\[
= \frac{1}{N^2 N_k T^2} \sum_{i \in G_k^0} \sum_{j=1}^N \sum_{k \neq j} \sum_{k \in G_k^0} \sum_{l=k}^N \sum_{s=1}^T E_D(X_{it} \varepsilon_{jt} X_{ks} \varepsilon_{ls}) \nu_{ji} \nu_{kl}
\]

\[
= \frac{1}{N^2 N_k T^2} \sum_{i \in G_k^0} \sum_{j=1}^N \sum_{k \neq j} \sum_{k \in G_k^0} \sum_{l=k}^N \sum_{s=1}^T E_D(X_{it} \varepsilon_{jt} X_{ks} \varepsilon_{js}) \nu_{ji} \nu_{kl} + O_P\left(N^{-1}T^{-1}\right)
\]

\[
= O_P\left(N^{-1}T^{-1}\right)
\]

we have \(V_{2kNT,1} = O_P((NT)^{-1/2})\). For \(V_{2kNT,2}\), we have

\[
|E_D(V_{2kNT,2})| = \frac{1}{N N_k T^2} \sum_{i \in G_k^0} \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T E_D(X_{it} \varepsilon_{jt} \mu_{ts}) \nu_{ji}
\]

\[
= \frac{1}{N N_k T^2} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T E_D(X_{it} \varepsilon_{is} \mu_{ts}) \nu_{ii}
\]

\[
\leq \frac{C_{NT}}{2 N N_k T^2} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \|X_{it}\| \|X_{is}\| \|\varepsilon_{is}\| \|F_t\| \|F_s\| \|\lambda_{it}\|^2 \alpha_{NT}^2 (t-s)^{(2+\sigma)/(1+\sigma)}
\]

\[
\leq \frac{C_{NT}}{2 N N_k T^2} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \|X_{it}\| \|\varepsilon_{is}\| \|\varepsilon_{is}\| \|F_t\| \|F_s\| \|\lambda_{it}\|^2 \alpha_{NT}^2 (t-s)^{(2+\sigma)/(1+\sigma)}
\]

\[
= O_P\left(N^{-1}\right) + O_P\left(N^{-1}\right) = O_P\left(N^{-1}\right)
\]
where $c_{NT} = \left\| (T^{-1}F^0F^0)^{-1} \right\| \left\| (T^{-1}\Lambda^0\Lambda^0)^{-1} \right\|$. In addition,

$$E_D \left( V_{2kNT,2} \right) = \frac{1}{N^2N_k^2T^4} \sum_{i \in G_k^0} \sum_{j=1}^N \sum_{k \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T \sum_{q=1}^T E_D \left( \epsilon_{ij} \epsilon_{ij} \epsilon_{ij} \epsilon_{ij} \right) \mu_{ts}^{\nu_i} \mu_{ts}^{\nu_j} \mu_{ts}^{\nu_k} \mu_{ts}^{\nu_k}$$

$$= \frac{1}{N^2N_k^2T^4} \sum_{i \in G_k^0} \sum_{j=1}^N \sum_{k \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T \sum_{q=1}^T E_D \left( \epsilon_{ij} \epsilon_{ij} \epsilon_{ij} \epsilon_{ij} \right) \mu_{ts}^{\nu_i} \mu_{ts}^{\nu_j} \mu_{ts}^{\nu_k} \mu_{ts}^{\nu_k}$$

$$+ \frac{1}{N^2N_k^2T^4} \sum_{i \in G_k^0} \sum_{j=1}^N \sum_{k \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T \sum_{q=1}^T E_D \left( \epsilon_{ij} \epsilon_{ij} \epsilon_{ij} \epsilon_{ij} \right) \mu_{ts}^{\nu_i} \mu_{ts}^{\nu_j} \mu_{ts}^{\nu_k} \mu_{ts}^{\nu_k}$$

$$+ \frac{1}{N^2N_k^2T^4} \sum_{i \in G_k^0} \sum_{j=1}^N \sum_{k \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T \sum_{q=1}^T E_D \left( \epsilon_{ij} \epsilon_{ij} \epsilon_{ij} \epsilon_{ij} \right) \mu_{ts}^{\nu_i} \mu_{ts}^{\nu_j} \mu_{ts}^{\nu_k} \mu_{ts}^{\nu_k}$$

$$+ O_P \left( N^{-3} \right)$$

$$= O_P \left( N^{-1}T^{-1} \right) + O_P \left( N^{-2}T^{-1} \right) + O_P \left( N^{-2}T^{-1} \right) + O_P \left( N^{-2}T^{-1} \right) + O_P \left( N^{-3} \right)$$

$$= O_P \left( N^{-1}T^{-1} \right).$$

It follows that $V_{2kNT} = O_P \left( N^{-1/2}T^{-1/2} \right)$. In sum, we have $V_{2kNT} = O_P \left( N^{-1/2}T^{-1/2} \right)$. This completes the proof. ■

## D Numerical algorithm

In this appendix, we present the numerical algorithm to obtain the PPC estimates $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_K_0)$ and $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_N)$. The algorithm is iterative and goes as follows:

1. Start with arbitrary initial values $\hat{\alpha}^{(0)} = (\hat{\alpha}_1^{(0)}, \ldots, \hat{\alpha}_K_0^{(0)})$ and $\hat{\beta}^{(0)} = (\hat{\beta}_1^{(0)}, \ldots, \hat{\beta}_N^{(0)})$ such that $\sum_{i=1}^N \left\| \hat{\beta}_i^{(0)} \right\|_{\beta_k^{(0)}} \neq 0$ for each $k = 2, \ldots, K_0$.\(^3\)

2. Suppose that we have obtained $\hat{\alpha}_i^{(r-1)} = (\hat{\alpha}_1^{(r-1)}, \ldots, \hat{\alpha}_K_0^{(r-1)})$ and $\hat{\beta}_i^{(r-1)} = (\hat{\beta}_1^{(r-1)}, \ldots, \hat{\beta}_N^{(r-1)})$. In Step $r \geq 1$, we first choose $(\beta, \alpha_1)$ to minimize

$$Q_{K_0NT}^{(r-1)}(\beta, \alpha_1) = Q_{1NT}(\beta) + \frac{\kappa}{N} \sum_{i=1}^N \left\| \beta_i - \alpha_1 \right\| \left\| \hat{\beta}_i^{(r-1)} - \hat{\alpha}_k^{(r-1)} \right\|$$

\(^3\)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 $\left\{ \hat{\beta}_i^{(0)} \right\}$. 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 \((\hat{\beta}_1^{(r,1)}, \hat{\alpha}_1^{(r)})\) of \((\beta, \alpha_1)\). Then we choose \((\beta, \alpha_2)\) to minimize

\[
Q_{K_0,NT}^{(r,2)}(\beta, \alpha_2) = Q_{1,NT}(\beta) + \frac{K}{N} \sum_{i=1}^{N} \|\beta_i - \alpha_2\| \|\hat{\gamma}_{i}^{(r,1)} - \hat{\alpha}_1^{(r)}\| \prod_{k \neq 1, 2} \|\hat{\beta}_i^{(r,1)} - \hat{\alpha}_k^{(r-1)}\|
\]

to obtain the updated estimate \((\hat{\beta}_1^{(r,2)}, \hat{\alpha}_2^{(r)})\) of \((\beta, \alpha_2)\). Repeat this procedure until we choose \((\beta, \alpha_{K_0})\) to minimize

\[
Q_{K_0,NT}^{(r,K_0)}(\beta, \alpha_{K_0}) = Q_{1,NT}(\beta) + \frac{K}{N} \sum_{i=1}^{N} \|\beta_i - \alpha_{K}\| \prod_{k = 1}^{K_0} \|\hat{\beta}_i^{(r,K_0-1)} - \hat{\alpha}_k^{(r)}\|
\]

to obtain the updated estimate \((\hat{\beta}_1^{(r,K_0)}, \hat{\alpha}_{K_0}^{(r)})\) of \((\beta, \alpha_{K_0})\). Let \(\hat{\beta}^{(r)} = (\hat{\beta}_1^{(r)}, ..., \hat{\beta}_{K_0}^{(r)})\) and \(\hat{\alpha}^{(r)} = (\hat{\alpha}_1^{(r)}, ..., \hat{\alpha}_{K_0}^{(r)})\).

3. Repeat the above step until certain convergence criterion is met, say, when

\[
\frac{\sum_{i=1}^{N} \|\hat{\beta}_i^{(r)} - \hat{\beta}_i^{(r-1)}\|^2}{\sum_{i=1}^{N} \|\hat{\beta}_i^{(r-1)}\|^2} + \frac{0.0001}{\sum_{k=1}^{K_0} \|\hat{\alpha}_k^{(r)} - \hat{\alpha}_k^{(r-1)}\|^2} < \epsilon_{tol},
\]

where \(\epsilon_{tol}\) is some prescribed tolerance level (e.g., 0.0001). Define the final iterative estimate of \(\alpha\) as \(\hat{\alpha} = (\hat{\alpha}_1^{(R)}, ..., \hat{\alpha}_{K_0}^{(R)})\) for sufficiently large \(R\) such that the convergence criterion is met. The final iterative estimate of \(\beta\) is defined as \(\hat{\beta} = (\hat{\beta}_1, ..., \hat{\beta}_N)\) where

\[
\hat{\beta}_i = \sum_{k=1}^{K_0} \hat{\alpha}_k^{(R)} 1 \left\{ \hat{\gamma}_i^{(R,l)} = \hat{\alpha}_k^{(R)} \text{ for some } l = 1, ..., K_0 \right\}
+ \hat{\beta}_i^{(R,K_0)} \left[ 1 - \sum_{k=1}^{K_0} 1 \left\{ \hat{\gamma}_i^{(R,l)} = \hat{\alpha}_k^{(R)} \text{ for some } l = 1, ..., K_0 \right\} \right]
\]

where \(\hat{\gamma}_i^{(R,l)}\) denotes the \(i\)th column of \(\hat{\beta}^{(R,l)}\) for \(l = 1, 2, ..., K\). Intuitively speaking, we classify individual \(i\) to group \(\hat{G}_k\) if \(\hat{\gamma}_i^{(R,l)} = \hat{\alpha}_k^{(R)}\) 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., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.