1-2019

Nonstationary panel models with latent group structures and cross-section dependence

Wenxin HUANG
Shanghai Jiaotong University

Sainan JIN
Singapore Management University, snjin@smu.edu.sg

Peter C. B. PHILLIPS
Singapore Management University, peterphillips@smu.edu.sg

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

SMU ECONOMICS & STATISTICS

Nonstationary Panel Models with Latent Group Structures and Cross-Section Dependence

Wenxin Huang, Sainan Jin, Peter C.B. Phillips, Liangjun Su

March 2020

Paper No. 07-2020

Nonstationary Panel Models with Latent Group Structures and Cross-Section Dependence*

Wenxin Huang[a], Sainan Jin[b], Peter C.B. Phillips[c], and Liangjun Su[b]

[a] Antai College of Economics and Management, Shanghai Jiao Tong University

[b] School of Economics, Singapore Management University

[c] Yale University, University of Auckland

University of Southampton & Singapore Management University

March 16, 2020

Abstract

This paper proposes a novel Lasso-based approach to handle unobserved parameter heterogeneity and cross-section dependence in nonstationary panel models. In particular, a penalized principal component (PPC) method is developed to estimate group-specific long-run relationships and unobserved common factors and jointly to identify the unknown group membership. The PPC estimators are shown to be consistent under weakly dependent innovation processes. But they suffer an asymptotically non-negligible bias from correlations between the nonstationary regressors and unobserved stationary common factors and/or the equation errors. To remedy these shortcomings we provide three bias-correction procedures under which the estimators are re-centered about zero as both dimensions (N and T) of the panel tend to infinity. We establish a mixed normal limit theory for the estimators of the group-specific long-run coefficients, which permits inference using standard test statistics. Simulations suggest good finite sample performance. An empirical application applies the methodology to study international R&D spillovers and the results offer a convincing explanation for the growth convergence puzzle through the heterogeneous impact of R&D spillovers.

JEL Classification: C13; C33; C38; C51; F43; O32; O40.

Keywords: Nonstationarity; Parameter heterogeneity; Latent group patterns; Penalized principal component; Cross-section dependence; Classifier Lasso; R&D spillovers

*The authors thank Serena Ng, an associate editor, and two anonymous referees for their many helpful comments on the paper. They also thank the participants in the 2017 Asian Meeting of the Econometric Society at CUHK for their valuable comments. Su gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under Grant MOE2012-T2-2-021 and funding support provided by the Lee Kong Chian Fund for Excellence. Phillips acknowledges research support from the NSF under Grant No. SES 18-50860 and the Kelly Fund at the University of Auckland. Huang gratefully acknowledges funding support provided by the Shanghai Institute of International Finance and Economics. 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

Nonstationary panel models have been extensively used in empirical analyses. Their asymptotic properties are well explored in classical settings when assumptions of common coefficients and independence across individuals are in place. Although these assumptions offer efficient estimation and simplify asymptotic theory, they are often hard to meet in real-world economic problems. On the one hand, researchers often face the issue of unobserved parameter heterogeneity in empirical models; see the study of the “convergence clubs” (e.g., Durlauf and Johnson (1995), Quah (1997), Phillips and Sul (2009)), the relation between income and democracy (e.g., Acemoglu et al. (2008) and Lu and Su (2017)), and the “resource curse” (e.g., Van der Ploeg (2011)). On the other hand, globalization and international spillovers give rise to a new challenge – the presence of cross-section dependence. In general, ignoring these two features may lead to biased or even inconsistent estimators in nonstationary panels, which can severely distort the reliability of classical methods. The goal of this paper is to study efficient estimation (in terms of convergence rates) and inference in nonstationary panel data models by allowing for the presence of both unobserved parameter heterogeneity and cross-section dependence.

Specifically, we consider a nonstationary panel data model with latent group structures and unobserved common factors. First, we assume that the long-run cointegration relationships associated with the observables are heterogeneous across different groups and homogeneous within a group. The latent grouped patterns offer flexible parameter settings by allowing for different slope coefficients across groups and remain parsimonious and efficient by pooling the cross-section observations within a group in the estimation procedure. Moreover, there is often economic intuition for considering grouped patterns in long-run relationships. For example, long-run equilibria in the growth regressions typically share some common features within a subsample, such as developing or developed countries, but reveal distinct patterns across subsamples. We also allow for stationary regressors and their parameters are completely heterogeneous. Second, we employ factor structures to model cross-section dependence. In our nonstationary panel model we consider both unobserved stationary and nonstationary common factors. For example, both oil price shocks and global technology innovations affect GDP levels in all countries. Similarly, both stock market shocks and macro-economic news affect security prices. But it is hard to tell whether these shock processes are stationary or not. In general, our framework allows us to fit more complex features to the data in empirical applications and offers flexibility so that the methods encourage the data to reveal latent features that may not be immediately apparent.

We take advantage of a growing literature on *Classifier-Lasso* (C-Lasso) techniques and models with interactive fixed effects (IFEs); see, e.g., Bai (2009), Su, Shi, and Phillips (2016a, SSP hereafter), Qian and Su (2016), Moon and Weidner (2017), Su and Ju (2018), Miao et al. (2020), among others. We propose a penalized principal component (PPC) method, which can be regarded as an iterative procedure between penalized regression and principal component analysis (PCA). In the first step, we introduce the unobserved nonstationary common factors into the PPC-based objective function and iteratively solve a regularized least-squares problem and an eigen-decomposition problem to obtain the *C-Lasso* estimators of the group-specific long-run coefficients and the nonstationary factors and factor loadings. We can do this simply because the presence of unobserved stationary common factors will not affect the consistency of the long-run coefficient estimators while neglecting the unobserved nonstationary factors would lead to inconsistency of such estimators due to the induced spurious regression. Note that the individual's group membership is also estimated at this stage. In the second step, we can explore the first-stage residuals to estimate the unobserved stationary factors and factor loadings. In the third step, we introduce three bias-correction procedures to obtain the bias-corrected estimators of the group-specific coefficients.

Our theoretical results are concerned with developing a limit theory for Lasso-type estimators in the present model setting which allows for stationary, nonstationary variates, and various cointegrating linkages. The presence of unobserved common factors complicates the asymptotic analysis in several ways. First, we establish preliminary rates of convergence for the estimators of the group-specific long-run coefficients and the unobserved nonstationary common factors. To show classification consistency, we also prove several uniform convergence results with the involvement of unobserved common factors. Given these uniform results, we show that all individuals are classified into the correct group with probability approaching one (w.p.a.1). The group-specific estimators enjoy the oracle property in the latent group literature, so that the three bias-corrected estimators are asymptotically equivalent to the corresponding infeasible ones that are obtained with full knowledge of the individual group identities.

Since our model allows for both contemporaneous and serial correlation in the errors, nonstationary regressors, and unobserved common factors, the usual endogeneity bias in nonstationary panels is present, originating in two primary sources. The first bias is commonly noted in nonstationary panels due to the weak dependence between the errors and nonstationary regressors (e.g., Phillips and Moon (1999)). As expected, the unobserved nonstationary common factors enter into the bias formula. The second bias arises from the presence of unobserved stationary common factors that can be correlated with the nonstationary regressors. We show that stationary common factors complicate the asymptotic biases and covariance structures but do not affect the consistency of the long-run coefficient estimators. Based on the bias formula we can employ the Phillips and Hansen (1990) fully-modified OLS (FM-OLS) procedure to achieve bias correction. Further, we explore a continuous-updating mechanism to obtain continuously updated Lasso (Cup-Lasso) estimators of the group-specific parameters, in which procedure we update the estimators of the individual's group membership, and the unobserved nonstationary and stationary common factor components. With these modifications our estimators are centered on zero and achieve the $\sqrt{N}T$ consistency rate that usually applies in homogeneous nonstationary panel models. Lastly, we establish a mixed normal limit theory for the bias-corrected group-specific long-run estimators, which validates the use of t, Wald, and F statistics for inference.

In the above analyses we assume that the numbers of groups and common factors are known. For practical work we propose three information criteria to determine the number of groups, the number of nonstationary common factors and stationary common factors, respectively. These information criteria are shown to select the correct numbers of groups and common factors w.p.a.1.

We illustrate the use of our methods by studying potentially heterogeneous behavior in the international R&D spillover model using a sample of OECD countries for the period 1971-2004. As in earlier work by Coe and Helpman (1995) we regress total factor productivity (TFP) on domestic R&D capital stock and foreign R&D capital stock. Coe and Helpman assume all countries obey a common linear specification and ignore the presence of common shocks across countries. In seeking greater flexibility, our methods allow parameters to vary across countries but with certain latent group structures and model the common shocks through the use of IFEs. Our latent group structural model is consistent with the fact that cross-country productivity may exhibit multiple long-run steady states. As a result, our methods reveal different spillover patterns than those discovered in Coe and Helpman (1995).

Specifically, our empirical analysis yields two key findings. First, we confirm positive technology spillovers in the pooled sample by allowing for the presence of common factors. This finding implies overall convergence behavior in technology growth through direct R&D spillovers when controlling for the unobserved global technology trend. Second, the group-specific estimates identify heterogeneous spillover patterns across countries and indicate the existence of two types of R&D spillovers – positive technology spillovers and negative market rivalry effects in the country-level data. This corroborates the findings of Bloom et al. (2013) who also found two types of R&D spillovers from firm-level data. Based on the empirically determined group patterns, we classify the OECD countries into three groups designated as *Convergence*, *Divergence*, and *Balance*. The major sources of technology change in the Convergence group come from positive technology diffusion and, as a result, the catch-up effects through technology diffusion favor the growth convergence hypothesis. Conversely, when market rivalry effects dominate technology spillovers, we observe overall negative R&D spillovers. For these countries, technology growth relies on domestic innovations and exhibits divergence behavior. Our findings therefore explain the growth convergence puzzle through heterogeneous behavior in R&D spillovers.

A major contribution of this paper is to offer a practical approach that accommodates both unobserved heterogeneity and cross-section dependence in nonstationary panels. We provide consistent and efficient estimators of group-specific long-run relationships for the observables even when individual group membership is unknown. The penalization method borrows from the *C-Lasso* formulation in SSP (2016a), but is modified here by using the principal component method to account for cross-section dependence simultaneously. Various papers account for unobserved heterogeneity in large dimensional panel models by clustering and grouping; see, e.g., Bonhomme and Manresa (2015) on grouped fixed effects, Qian and Su (2016) on structural changes, and Ando and Bai (2016) on grouped factor models. But almost all the literature focuses on stationary panel data models.

Recently, Huang et al. (2020) have considered latent group patterns in cointegrated panels but they do not allow for cross-section dependence.

Our theoretical results also contribute to two strands of the literature on cointegrated panels and factor models. First, it is noted that the average and common long-run estimators permit normal asymptotic distributions, whereas the heterogeneous and time-series long-run estimators have a non-standard limit theory; see, e.g., Phillips and Moon (1999), Kao and Chiang (2001), and Pedroni (2004). In our context, due to the presence of the common components, we maintain the simplicity of asymptotic mixed normality under grouped parameter heterogeneity. Second, there is a growing literature using factor models to capture cross-section dependence under the large N and large T settings; see, e.g., Bai and Ng (2002, 2004), Phillips and Sul (2003), Pesaran (2006), Bai (2009), and Moon and Weidner (2017). Compared with existing work, our approach accommodates both stationary and nonstationary common factors and provides a corresponding limit theory for inference. Our asymptotic theory therefore applies to more general forms of nonstationary panel data models with internally grouped but unknown patterns of behavior and to models of this type with both stationary and nonstationary common factors.

The rest of the paper is structured as follows. Section 2 introduces a nonstationary panel model with latent group structures and cross-section dependence and proposes a penalized principal component method for estimation. Section 3 explains the main assumptions and establishes the asymptotic properties of the three Lasso-type estimators. Section 4 reports the Monte Carlo simulation results. Section 5 applies the methodology to study heterogeneous cross country behavior in R&D spillovers. Section 6 concludes. The proofs of the main results are given in the online supplement that also contains some additional discussions and simulation results.

NOTATION. We write integrals such as $\int_0^1 W(s)ds$ simply as $\int W$ and define $\Omega^{1/2}$ to be any matrix such that $\Omega = (\Omega^{1/2})(\Omega^{1/2})'$. $BM(\Omega)$ denotes Brownian motion with covariance matrix Ω. For any $m \times n$ real matrix A, we write its Frobenius norm, spectral norm and transpose as $\|A\|$, $\|A\|_{sp}$, and A', respectively. When A is symmetric, we use $\mu_{\max}(A)$ and $\mu_{\min}(A)$ to denote its largest and smallest eigenvalues, respectively. Let $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$, where $A'A$ is of full rank, and I is an identity matrix. Let $0_{p\times 1}$ denote a $p \times 1$ vector of zeros, I_b a $b \times b$ identity matrix, and $\mathbf{1}\{\cdot\}$ an indicator function. Let M denote a generic positive constant whose values can vary in different locations. We use "p.d." and "p.s.d." to abbreviate "positive definite" and "positive semidefinite," respectively. The operator $\xrightarrow{p}$ denotes convergence in probability, $\Rightarrow$ weak convergence, $a.s.$ almost surely, and the floor function $\lfloor x \rfloor$ to denote the largest integer less than or equal to x. Unless indicated otherwise, we use $(N, T) \to \infty$ to signify that N and T pass to infinity jointly.

2 Model and Estimation

This section introduces a nonstationary panel model with latent group structures and unobserved common factors. A penalized principal component method is then proposed to estimate the parameters of the model and the unobserved group structure.

2.1 Model setup

We start by considering a panel cointegration model with both nonstationary and stationary regressors. Assume that for individuals $i = 1, ..., N$, we observe $\{y_{it}, x_{1,it}, x_{2,it}\}_{t=1}^{T}$ where $x_{1,it}$ denotes nonstationary regressors of order one ($I(1)$ process) and $x_{2,it}$ denotes stationary ones ($I(0)$ process), such that

$$\begin{cases} y_{it} = \beta_{1,i}^{0\prime} x_{1,it} + \beta_{2,i}^{0\prime} x_{2,it} + e_{it} \\ x_{1,it} = x_{1,it-1} + \varepsilon_{it}, \end{cases} , \tag{2.1}$$

where y_{it} is a scalar, $\beta_{1,i}^{0}$ is a $p_1 \times 1$ vector of parameters that is associated with the long-run cointegration relationship, $\beta_{2,i}$ is a $p_2 \times 1$ vector of parameters that may capture the short-run dynamics, and ε_{it} has zero mean and finite long-run variance. We assume that the error terms e_{it} are cross-sectionally dependent due to the presence of some unobserved common factors, specified as

$$e_{it} = \lambda_i^{0\prime} f_t^0 + u_{it} = \lambda_{1i}^{0\prime} f_{1t}^0 + \lambda_{2i}^{0\prime} f_{2t}^0 + u_{it}, \tag{2.2}$$

where f_t^0 is an $r \times 1$ vector of unobserved common factors that contains an $r_1 \times 1$ vector of nonstationary factors f_{1t}^0 of order one (I(1) process) and an $r_2 \times 1$ vector of stationary factors f_{2t}^0 (I(0) process), $\lambda_i^0 = (\lambda_{1i}^{0\prime}, \lambda_{2i}^{0\prime})'$ is an $r \times 1$ vector of factor loadings, and u_{it} is the idiosyncratic component of e_{it} with zero mean and finite long-run variance. For simplicity, u_{it} is assumed to be cross-sectionally independent so that the cross-section dependence among the e_{it} only arises from the common factors f_t^0, and $\mathbb{E}(e_{it}e_{jt}) = \mathbb{E}(\lambda_i^{0\prime} f_t^0 f_t^{0\prime} \lambda_j^0) \neq 0$ in general.

In addition, we introduce latent group structures in $\beta_{1,i}^0$, which are heterogeneous across different groups and homogeneous within a group:

$$\beta_{1,i}^0 = \begin{cases} \alpha_1^0 & \text{if } i \in G_1^0 \\ \vdots & \qquad \vdots \\ \alpha_K^0 & \text{if } i \in G_K^0 \end{cases} , \tag{2.3}$$

where $\alpha_j^0 \neq \alpha_k^0$ for any $j \neq k$, $\bigcup_{k=1}^{K} G_k^0 = \{1, 2, \ldots, N\}$, and $G_k^0 \bigcap G_j^0 = \varnothing$ for any $j \neq k$. Let $N_k = \#G_k$ denote the cardinality of the set G_k^0. For the moment in this section, we assume that the number of groups, K, is known and fixed, but each individual's group membership is unknown. In Section 3.7, we propose an information criterion to determine the number of groups.

There are three main complications in this panel cointegration model. First, least-squares estimators that ignore the factor component are inconsistent due to the presence of nonstationary common factors. Noting that the components $\left\{\lambda_{1i}^{0\prime} f_{1t}^0 + \lambda_{2i}^{0\prime} f_{2t}^0,\ t \geq 1\right\}$ are still $I(1)$ processes in general when $r_1 \geq 1$, the least-squares estimators of $\beta_{1,i}^0$ and $\beta_{2,i}^0$ from the time series regression of y_{it} on $x_{1,it}$

and $x_{2,it}$ suffer from spurious regression. For this reason, we must take account of the nonstationary factor component to obtain consistent estimators of the slope coefficients. Therefore, our panel latent factor cointegration model is more general than the traditional panel cointegration model: the cointegration vector here is $(1, -\beta^{0\prime}_{1,i}, \lambda^{0\prime}_{1,i})$ and the equilibrium errors $\{y_{it} - \beta^{0\prime}_{1,i}x_{1,it} - \lambda^{0\prime}_{1i}f^0_{1t},\, t \geq 1\}$ are stationary whereas standard cointegrating equilibrium errors do not involve unobserved factors such as f^0_{1t}. Second, even though Bai et al. (2009) study a homogeneous panel cointegration model with nonstationary common factors, it is a big further step to establish desirable asymptotic properties of the group-specific long-run coefficient estimators and to recover unobserved group identities. Due to the presence of common factors, the grouping C-Lasso algorithm and derivation of the oracle property are considerably more difficult than that those of SSP (2016a).[1] Third, both unobserved group structures and common factors complicate the non-negligible asymptotic bias in the long-run estimators arising from endogeneity and serial correlations. An effective new bias-correction procedure is then needed to re-center the limit distributions around zero to facilitate inference. All these complications call for a new estimation methodology and asymptotic theory.

In the next subsection, we introduce the estimation procedure based on the level equations in (2.1). A natural question (raised by a referee) is why not proceed to first difference the data and use an estimation procedure based on the first-differenced equation

$$\Delta y_{it} = \beta^{0\prime}_{1,i}\Delta x_{1,it} + \beta^{0\prime}_{2,i}\Delta x_{2,it} + \lambda^{0\prime}_{1i}\Delta f^0_{1t} + \lambda^{0\prime}_{2i}\Delta f^0_{2t} + \Delta u_{it}, \tag{2.4}$$

where, e.g., $\Delta y_{it} = y_{it} - y_{i,t-1}$. To appreciate the importance of working on the level equations in (2.1), we make two remarks.

Remark 2.1. Let $e_i = (e'_{i1}, ..., e'_{iT})'$, $x_{l,i} = (x'_{l,i1}, ..., x'_{l,iT})'$ and $M_{l,i} = I_T - x_{l,i}\left(x'_{l,i}x_{l,i}\right)x'_{l,i}$ for $l = 1, 2$. If the error terms e_{it} are independent across individuals such that the common components are absent in (2.2), we can run time series OLS estimation of y_{it} on $(x'_{1,it},\, x'_{2,it})$ for each i to obtain the OLS estimators $(\tilde{\beta}'_{1,i}, \tilde{\beta}'_{2,i})$ of $(\beta^{0\prime}_{1,i}, \beta^{0\prime}_{2,i})$. It is well known that the OLS estimator $\tilde{\beta}_{1,i}$ is super-consistent and robust to problems such as omitted (stationary) regressors, serial correlations, and endogeneity (see Phillips (1995), which also allowed for cointegrated regressors in a VAR setting). For simplicity, we review the asymptotic properties of $\tilde{\beta}_{1,i}$ and $\tilde{\beta}_{2,i}$ by assuming $\mathbb{E}(x_{2,it}) = 0$. Then, under some standard conditions that ensure proper behavior of $\frac{1}{T^2}x'_{1,i}x_{1,i}$, $\frac{1}{T}x'_{2,i}x_{2,i}$, and $\frac{1}{T}x'_{1,i}x_{2,i}$,

[1] The oracle property in the latent group literature is that the group-specific estimators are asymptotically equivalent to the corresponding infeasible estimators that are obtained by knowing all individual group identities.

inter alia, we have

$$
\begin{aligned}
T(\tilde{\beta}_{1,i}-\beta_{1,i}^{0}) &= \left(\frac{1}{T^2}x_{1,i}'M_{2,i}x_{1,i}\right)^{-1}\frac{1}{T}x_{1,i}'M_{2,i}e_i = \left(\frac{1}{T^2}x_{1,i}'x_{1,i}\right)^{-1}\frac{1}{T}x_{1,i}'M_{2,i}e_i + o_P(1),\\
\sqrt{T}\left(\tilde{\beta}_{2,i}-\beta_{2,i}^{0}\right) &= \left(\frac{1}{T}x_{2,i}'M_{1,i}x_{2,i}\right)^{-1}\frac{1}{\sqrt{T}}x_{2,i}'M_{1,i}e_i = \left(\frac{1}{T}x_{2,i}'x_{2,i}\right)^{-1}\frac{1}{\sqrt{T}}x_{2,i}'e_i + o_P(1),\\
\frac{1}{T}x_{1,i}'M_{2,i}e_i &= O_p(1) \text{ and } \frac{1}{\sqrt{T}}x_{2,i}'M_{1,i}e_i = O_p(1) \text{ as } T\to\infty,
\end{aligned}
$$

where we use the facts that

$$
\begin{aligned}
\frac{1}{T^2}x_{1,i}'M_{2,i}x_{1,i} &= \frac{1}{T^2}x_{1,i}'x_{1,i} - \frac{1}{T}\left(\frac{1}{T}x_{1,i}'x_{2,i}\right)\left(\frac{1}{T}x_{2,i}'x_{2,i}\right)^{-1}\left(\frac{1}{T}x_{2,i}'x_{1,i}\right) = \frac{1}{T^2}x_{1,i}'x_{1,i} + O_P(T^{-1}),\\
\frac{1}{T}x_{2,i}'M_{1,i}x_{2,i} &= \frac{1}{T}x_{2,i}'x_{2,i} - \frac{1}{T}\left(\frac{1}{T}x_{2,i}'x_{1,i}\right)\left(\frac{1}{T^2}x_{1,i}'x_{1,i}\right)^{-1}\left(\frac{1}{T}x_{1,i}'x_{2,i}\right) = \frac{1}{T}x_{2,i}'x_{2,i} + O_P(T^{-1}),
\end{aligned}
$$

and similarly $\frac{1}{T}x_{2,i}M_{1,i}e_i = \frac{1}{T}x_{2,i}'e_i - \frac{1}{T}\left(\frac{1}{T}x_{2,i}'x_{1,i}\right)\left(\frac{1}{T^2}x_{1,i}'x_{1,i}\right)^{-1}\left(\frac{1}{T}x_{1,i}'e_i\right) = \frac{1}{T}x_{2,i}'e_i + O_P(T^{-1})$. The above results imply different convergence rates for $\tilde{\beta}_{1,i}$ and $\tilde{\beta}_{2,i}$. In particular, $\tilde{\beta}_{1,i}$ is super-consistent regardless of the properties of $I(0)$ regressors or the endogeneity caused by the correlation between $\{\Delta x_{1,it}\}$ and $\{e_{it}\}$. If one further assumes orthogonality conditions on the stationary regressors that ensure $\frac{1}{T}x_{2,i}'e_i = o_P(1)$, then we also have

$$
T(\tilde{\beta}_{1,i}-\beta_{1,i}^{0}) = \left(\frac{1}{T^2}x_{1,i}'x_{1,i}\right)^{-1}\frac{1}{T}x_{1,i}'e_i + o_P(1).
$$

In this case, we have the asymptotic independence between $\tilde{\beta}_{1,i}$ and $\tilde{\beta}_{2,i}$. In the presence of the factor structure in (2.2), we can continue to obtain super-consistent estimators of $\beta_{1,i}^{0}$ and consistent estimators of $\beta_{2,i}^{0}$ even if $\{\Delta x_{1,it}\}$ and $\{e_{it}\}$ are contemporaneously correlated. These appealing properties are completely lost if one works on the first-differenced data. See the next remark.

Remark 2.2. In the absence of the factor structure in (2.2), we have the following first-differenced equation:

$$
\Delta y_{it} = \beta_{1,i}^{0\prime}\Delta x_{1,it} + \beta_{2,i}^{0\prime}\Delta x_{2,it} + \Delta u_{it}. \tag{2.5}
$$

Apparently, the OLS estimator of $(\beta_{1,i}^{0\prime}, \beta_{2,i}^{0\prime})$ based on the time series regression of Δy_{it} on $(\Delta x_{1,it}', \Delta x_{2,it}')$ is inconsistent if $\mathbb{E}\left[\Delta x_{1,it}\Delta u_{it}\right] \neq 0$ (or $\mathbb{E}\left[\Delta x_{2,it}\Delta u_{it}\right] \neq 0$), not to mention the super-consistency of the estimator of $\beta_{1,i}^{0}$. Since we allow for correlation between $\{\Delta x_{1,it}\}$ and $\{\Delta u_{it}\}$, estimation based on (2.5) inevitably leads to inconsistency. This inconsistency of OLS-type estimators of $(\beta_{1,i}^{0\prime}, \beta_{2,i}^{0\prime})$ continues to hold in (2.4) even when PCA is used to handle the factor components.

To proceed with the development of level equation estimation in (2.1), let

$$\begin{aligned} \boldsymbol{\alpha} &\equiv (\alpha_1, ..., \alpha_K), \quad \boldsymbol{\beta} \equiv (\beta_1, ..., \beta_N), \quad \boldsymbol{\beta_l} \equiv (\beta_{l,1}, ..., \beta_{l,N}), \quad \Lambda \equiv (\lambda_1, ..., \lambda_N)' \\ \Lambda_l &\equiv (\lambda_{l1}, ..., \lambda_{lN})', \quad F \equiv (f_1, ..., f_T)', \text{ and } F_l \equiv (F_{l1}, ..., F_{lT})' \text{ where } l = 1, 2. \end{aligned}$$

The true values of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\beta}_l$, Λ, Λ_l, F, and F_l are denoted $\boldsymbol{\alpha}^0, \boldsymbol{\beta}$, $\boldsymbol{\beta}_l^0$, Λ^0, Λ_l^0, F^0, and F_l^0. We also use α_k^0, β_i^0, $\beta_{1,i}^0$, $\beta_{2,i}^0$, $\lambda_i^0 = (\lambda_{1i}^{0\prime}, \lambda_{2i}^{0\prime})'$, and $f_t^0 = (f_{1t}^{0\prime}, f_{2t}^{0\prime})'$ to denote the true values of $\alpha_k, \beta_i, \beta_{1,i}, \beta_{2,i}, \lambda_i = (\lambda_{1i}', \lambda_{2i}')'$, and $f_t = (f_{1t}', f_{2t}')'$. Interest focuses primarily on establishing each individual's group identity and on consistent estimation of the group-specific long-run relationships α_k in the presence of stationary regressors and both unobserved stationary and nonstationary common factors.

2.2 Penalized principal component estimation

In this subsection we propose an iterative PPC-based procedure to jointly estimate the long-run cointegrating coefficients $\beta_{1,i}$, the short-run parameters $\beta_{2,i}$ and unobserved common factors f_t, and to identify the group structure in these long-run relationships. Combining (2.1)-(2.2) yields

$$y_{it} = \beta_{1,i}^{0\prime} x_{1,it} + \beta_{2,i}^{0\prime} x_{2,it} + \lambda_{1i}^{0\prime} f_{1t}^0 + \lambda_{2i}^{0\prime} f_{2t}^0 + u_{it}, \tag{2.6}$$

or in vector observation form:

$$y_i = x_i \beta_i^0 + F_1^0 \lambda_{1i}^0 + F_2^0 \lambda_{2i}^0 + u_i = x_{1,i} \beta_{1,i}^0 + x_{2,i} \beta_{2,i}^0 + F_1^0 \lambda_{1i}^0 + F_2^0 \lambda_{2i}^0 + u_i, \tag{2.7}$$

where $y_i = (y_{i1}, ..., y_{iT})'$, $x_{1,i}$, $x_{2,i}$, F_1^0, F_2^0 and u_i are similarly defined, and $x_i = (x_{1,i}, x_{2,i})$.

Ideally, one might attempt to estimate both the stationary and nonstationary common components along with the parameters of interest, $\beta_{1,i}$ and $\beta_{2,i}$. But due to the fact that the stationary components and nonstationary components behave differently and require different normalization rules, it is difficult to study the asymptotic properties of the resulting joint estimators. Nevertheless, as mentioned above, one can obtain least square estimators of $\beta_{1,i}$ and $\beta_{2,i}$ by taking into account the nonstationary factor component and ignoring the stationary factor component. As we discussed in the model setup, the estimators of the coefficients of the nonstationary regressors still achieve consistency regardless of endogeneity, serial correlation or the presence of stationary regressors, and the estimators of the coefficients of the stationary regressors are consistent under some orthogonality conditions.[2] Lastly, we estimate the stationary common component from the resultant residuals. This motivates the following sequential approach to estimate the unknown parameters in the model. We first estimate the nonstationary factor component along with $\beta_{1,i}$ and $\beta_{2,i}$, then estimate the stationary factor component along with $\beta_{2,i}$ from the resultant residuals. The step-wise procedure is as follows.

[2] We require the stationary regressors to be uncorrelated with the stationary common factors, factor loadings and error terms (c.f., Phillips (1995) and Bai et al.(2009)).

Step 1. We estimate $(\boldsymbol{\beta}, F_1, \Lambda_1)$ by minimizing the following least squares (LS) objective function:

$$\mathrm{SSR}(\boldsymbol{\beta}, F_1, \Lambda_1) = \sum_{i=1}^{N} (y_i - x_i\beta_i - F_1\lambda_{1i})'(y_i - x_i\beta_i - F_1\lambda_{1i}) \tag{2.8}$$

under the constraints that $\frac{1}{T^2}F_1'F_1 = I_{r_1}$ and $\Lambda_1'\Lambda_1$ is diagonal. It is well known that the LS estimator $(\tilde{\beta}_i, \tilde{F}_1)$ is the solution to the following set of nonlinear equations:

$$\tilde{\beta}_i = \left(\tilde{\beta}_{1,i}', \tilde{\beta}_{2,i}'\right)' = \left(x_i' M_{\tilde{F}_1} x_i\right)^{-1} x_i' M_{\tilde{F}_1} y_i, \tag{2.9}$$

$$\tilde{F}_1 \tilde{V}_{1,NT} = \left[\frac{1}{NT^2}\sum_{i=1}^{N}(y_i - x_i\tilde{\beta}_i)(y_i - x_i\tilde{\beta}_i)\right]\tilde{F}_1, \tag{2.10}$$

where $M_{\tilde{F}_1} = I_T - \frac{1}{T^2}\tilde{F}_1\tilde{F}_1'$, $\frac{1}{T^2}\tilde{F}_1'\tilde{F}_1 = I_{r_1}$, and $\tilde{V}_{1,NT}$ is a diagonal matrix consisting of the r_1 largest eigenvalues of the matrix inside the square brackets in (2.10), arranged in decreasing order. The LS estimator of $\Lambda_1 = (\lambda_{11}, ..., \lambda_{1N})'$ is given by $\tilde{\Lambda}_1 = (\tilde{\lambda}_{11}, ..., \tilde{\lambda}_{1N})'$ where $\tilde{\lambda}_{1i}' = \frac{1}{T^2}(y_i - x_i\tilde{\beta}_i)'\tilde{F}_1$. It is easy to verify that $\frac{1}{N}\tilde{\Lambda}_1'\tilde{\Lambda}_1 = T^{-2}\tilde{F}_1'[\frac{1}{NT^2}\sum_{i=1}^{N}(y_i - x_i\tilde{\beta}_i)(y_i - x_i\tilde{\beta}_i)'\tilde{F}_1] = T^{-2}\tilde{F}_1'\tilde{F}_1\tilde{V}_{1,NT} = \tilde{V}_{1,NT}$.

Step 2. Using the initial estimates of $\tilde{\beta}_i$ and $\tilde{F}_1$ as starting values, we employ the methodology of SSP (2016a) by minimizing the following PPC criterion function to obtain estimates of $(\boldsymbol{\beta}, \boldsymbol{\alpha}, F_1)$:

$$Q_{NT}^{\lambda,K}(\boldsymbol{\beta}, \boldsymbol{\alpha}, F_1) = Q_{NT}(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, F_1) + \frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\left\|\beta_{1,i} - \alpha_k\right\|, \tag{2.11}$$

where $Q_{NT}(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, F_1) = \frac{1}{NT^2}\sum_{i=1}^{N}\left(y_i - x_{1,i}\beta_{1,i} - x_{2,i}\beta_{2,i}\right)' M_{F_1}\left(y_i - x_{1,i}\beta_{1,i} - x_{2,i}\beta_{2,i}\right)$, and $\lambda = \lambda(N,T)$ is a tuning parameter. Minimizing the PPC criterion function in (2.11) produces the C-Lasso estimators $(\hat{\beta}_i, \hat{\alpha}_k, \hat{F}_1)$ of (β_i, α_k, F_1) where $\hat{F}_1 = (\hat{f}_{11}, ..., \hat{f}_{1T})'$ and $\hat{\beta}_i = (\hat{\beta}_{1,i}', \hat{\beta}_{2,i}')'$. Note that

$$\hat{F}_1 V_{1,NT} = \left[\frac{1}{NT^2}\sum_{i=1}^{N}(y_i - x_{1,i}\hat{\beta}_{1,i} - x_{2,i}\hat{\beta}_{2,i})(y_i - x_{1,i}\hat{\beta}_{1,i} - x_{2,i}\hat{\beta}_{2,i})'\right]\hat{F}_1, \tag{2.12}$$

where $\frac{1}{T^2}\hat{F}_1'\hat{F}_1 = I_{r_1}$ and $V_{1,NT}$ is a diagonal matrix consisting of the r_1 largest eigenvalues of the matrix inside the square brackets in (2.12), arranged in decreasing order. The PPC estimator of $\Lambda_1 = (\lambda_{11}, ..., \lambda_{1N})'$ is given by $\hat{\Lambda}_1 = (\hat{\lambda}_{11}, ..., \hat{\lambda}_{1N})'$ where $\hat{\lambda}_{1i}' = \frac{1}{T^2}(y_i - x_{1,i}\hat{\beta}_{1,i} - x_{2,i}\hat{\beta}_{2,i})'\hat{F}_1$. Define the resulting estimated groups

$$\hat{G}_k = \{i \in \{1, 2, ..., N\} : \hat{\beta}_{1,i} = \hat{\alpha}_k\} \text{ for } k = 1, ..., K. \tag{2.13}$$

Step 3. Given the estimates $\hat{\beta}_{1,i}$, $\hat{\alpha}_k$, and $\hat{F}_1$, we obtain the cointegration residuals $\hat{r}_{it} = y_{it} - \hat{\beta}_{1,i}'x_{1,it} -$

$\hat{\lambda}_{1i}'\hat{f}_{1t}$. Based on the consistency in estimation of the nonstationary part, we have $\hat{r}_{it} = \lambda_{2i}^{0\prime} f_{2t}^{0} + \beta_{2,i}^{0\prime} x_{2,it} + u_{it} + v_{it}$ where v_{it} signifies the estimation error from the early stages. Then we can employ the standard procedure in stationary panel models with interactive fixed effects, see Bai (2009), Moon and Weidner (2017). The LS estimator of $(\breve{\beta}_{2,i}, \hat{F}_2)$ is the solution to the following set of nonlinear equations:

$$\breve{\beta}_{2,i} = \left(x_{2,i}' M_{\hat{F}_2} x_{2,i}\right)^{-1} x_{2,i}' M_{\hat{F}_2} \hat{r}_i, \tag{2.14}$$

$$\hat{F}_2 \tilde{V}_{2,NT} = \left[\frac{1}{NT}\sum_{i=1}^{N} (\hat{r}_i - x_{2,i}\breve{\beta}_{2,i})(\hat{r}_i - x_{2,i}\breve{\beta}_{2,i})\right] \hat{F}_2, \tag{2.15}$$

where $\frac{1}{T}\hat{F}_2'\hat{F}_2 = I_{r_2}$ and $V_{2,NT}$ is a diagonal matrix consisting of the r_2 largest eigenvalues of the matrix inside the square brackets in (2.15), arranged in decreasing order.

Let $\hat{\boldsymbol{\beta}}_l \equiv (\hat{\beta}_{l,1}, ..., \hat{\beta}_{l,N})$ and $\hat{\boldsymbol{\alpha}} \equiv (\hat{\alpha}_1, ..., \hat{\alpha}_K)$ for $l = 1, 2$. We will study the asymptotic properties of $\hat{\beta}_{1,i}, \hat{\alpha}_k$, and $\hat{F}_1$ in Section 3.2 and the classification consistency of the group structure in Section 3.3. Noting that $\hat{\alpha}_k$ has an asymptotic bias, we will propose various methods to correct its bias in Section 3.4. The asymptotic properties of $\breve{\beta}_{2,i}$ and $\hat{F}_2$ may also be studied but they are not the focus of the present paper.

3 Asymptotic Theory

3.1 Main assumptions

We introduce the main assumptions used to study the asymptotic properties of the estimators $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$, $\hat{\boldsymbol{\alpha}}$, and $\hat{F}_1$. Let $Q_{i,xx}(F_1) = \frac{1}{T^2} x_{1,i}' M_{F_1} x_{1,i}$, $Q_1(F_1) = \mathrm{diag}(Q_{1,xx}(F_1), ..., Q_{N,xx}(F_1))$, and

$$Q_2(F_1) = \begin{pmatrix} \frac{1}{NT^2} x_{1,1}' M_{F_1} x_{1,1} a_{11} & \frac{1}{NT^2} x_{1,1}' M_{F_1} x_{1,2} a_{12} & \cdots & \frac{1}{NT^2} x_{1,1}' M_{F_1} x_{1,N} a_{1N} \\ \frac{1}{NT^2} x_{1,2}' M_{F_1} x_{1,1} a_{21} & \frac{1}{NT^2} x_{1,2}' M_{F_1} x_{1,2} a_{22} & \cdots & \frac{1}{NT^2} x_{1,1}' M_{F_1} x_{1,N} a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{NT^2} x_{1,N}' M_{F_1} x_{1.1} a_{N1} & \frac{1}{NT^2} x_{1,N}' M_{F_1} x_{1,2} a_{N2} & \cdots & \frac{1}{NT^2} x_{1,N}' M_{F_1} x_{1,N} a_{NN} \end{pmatrix},$$

where F_1 satisfies $\frac{1}{T^2} F_1' F_1 = I_{r_1}$ and $a_{ij} = \lambda_{1i}^{0\prime}(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}\lambda_{1j}^0$. Note that $Q_2(F_1)$ is an $Np_1 \times Np_1$ matrix. Let $\mathcal{C} = \sigma(\Lambda^0, F^0)$, the sigma algebra generated by the common factors and factor loadings. Let M denote a generic constant that may vary across occurrences. Define $w_{it} = (u_{it}, \varepsilon_{it}', \Delta f_{1t}^{0\prime}, f_{2t}^{0\prime}, x_{2,it}')'$ and let $\Omega_i = \sum_{j=-\infty}^{\infty} \mathbb{E}(w_{ij} w_{i0}')$, be the long-run covariance matrix of $\{w_{it}\}$. We also define the contemporaneous variance matrix $\Sigma_i = \mathbb{E}(w_{i0} w_{i0}')$ and the one-sided long-run covariance matrix $\Delta_i = \sum_{j=0}^{\infty} \mathbb{E}(w_{i0} w_{ij}') = \Gamma_i + \Sigma_i$. Note that $\Omega_i = \Gamma_i' + \Gamma_i + \Sigma_i$. Conformably with w_{it}, Ω_i and Δ_i are partitioned as follows

$$\Omega_i = \begin{pmatrix} \Omega_{11,i} & \Omega_{12,i} & \Omega_{13,i} & \Omega_{14,i} & \Omega_{15,i} \\ \Omega_{21,i} & \Omega_{22,i} & \Omega_{23,i} & \Omega_{24,i} & \Omega_{25,i} \\ \Omega_{31,i} & \Omega_{32,i} & \Omega_{33} & \Omega_{34} & \Omega_{35,i} \\ \Omega_{41,i} & \Omega_{42,i} & \Omega_{43} & \Omega_{44} & \Omega_{45,i} \\ \Omega_{51,i} & \Omega_{52,i} & \Omega_{53,i} & \Omega_{54,i} & \Omega_{55,i} \end{pmatrix} \text{ and } \Delta_i = \begin{pmatrix} \Delta_{11,i} & \Delta_{12,i} & \Delta_{13,i} & \Delta_{14,i} & \Delta_{15,i} \\ \Delta_{21,i} & \Delta_{22,i} & \Delta_{23,i} & \Delta_{24,i} & \Delta_{25,i} \\ \Delta_{31,i} & \Delta_{32,i} & \Delta_{33} & \Delta_{34} & \Delta_{35,i} \\ \Delta_{41,i} & \Delta_{42,i} & \Delta_{43} & \Delta_{44} & \Delta_{45,i} \\ \Delta_{51,i} & \Delta_{52,i} & \Delta_{53,i} & \Delta_{54,i} & \Delta_{55,i} \end{pmatrix}.$$

Partition Σ_i correspondingly. Let $p = p_1 + p_2$. Let $S_1, S_2, S_3,\ S_4$, and S_5 denote, respectively, the $1 \times (1+p+r)$, $p_1 \times (1+p+r)$, $r_1 \times (1+p+r)$, $r_2 \times (1+p+r)$ and $p_2 \times (1+p+r)$ selection matrices for which $S_1 w_{it} = u_{it}$, $S_2 w_{it} = \varepsilon_{it}$, $S_3 w_{it} = \Delta f_{1t}^0$, $S_4 w_{it} = f_{2t}^0$ and $S_5 w_{it} = x_{2,it}$. Let $S_{23} = (S_2', S_3')'$, a $(p_1 + r_1) \times (1 + p + r)$ selection matrix. We assume without loss of generality that $x_{2,it}$ has zero mean.[3]

We make the following assumptions on $\{w_{it}\}$ and $\{\lambda_i\}$.

Assumption 3.1 *(i) For each i , $\{w_{it}, t \geq 1\}$ is a linear process: $w_{it} = \phi_i(L) v_{it} = \sum_{j=0}^{\infty} \phi_{ij} v_{i,t-j}$, where $v_{it} = (v_{it}^u, v_{it}^{\varepsilon\prime}, v_t^{f_1\prime}, v_t^{f_2\prime}, v_{it}^{x_2\prime})'$ is a $(1+p+r) \times 1$ random vector that is i.i.d. over t with zero mean and variance matrix I_{1+p+r}; $\sup_{N\geq 1} \max_{1\leq i\leq N} E(\|v_{it}\|^{2q+\epsilon}) < M$, where $q > 4$ and ϵ is an arbitrarily small positive constant; v_{it}^u, v_{it}^{ε}, $v_t^{f_1}$, $v_t^{f_2}$, and $v_{it}^{x_2}$ are mutually independent; and $(v_{it}^u, v_{it}^{\varepsilon\prime}, v_{it}^{x_2\prime})'$ are independent across i.*

(ii) $\sup_{N\geq 1} \max_{1\leq i\leq N} \sum_{j=0}^{\infty} j^k \|\phi_{ij}\| < \infty$ for some $k \geq 2$, and $S_{23}\Omega_i S_{23}'$ has full rank uniformly in i.

(iii) $\left(u_{it}, \varepsilon_{it}', x_{2,it}'\right)$ are independent across i conditional on $\mathcal{C}$.

(iv) $\mathbb{E}(x_{2,it} u_{is}) = 0$ and $\mathbb{E}(x_{2,it} f_{2s}^{0\prime}) = 0$ for $s \geq t$.

(v) λ_i^0 is independent of v_{jt} for all $i, j,$and t.

Following Phillips and Solo (1992, PS hereafter), we assume that $\{w_{it}, t \geq 1\}$ is a linear process in Assumption 3.1(i). For later reference, we partition the matrix operator $\phi_i(L)$ conformably with w_{it} as follows:

$$\phi_i(L) = \begin{pmatrix} \phi_i^{uu}(L) & \phi_i^{u\varepsilon}(L) & \phi_i^{uf_1}(L) & \phi_i^{uf_2}(L) & \phi_i^{ux_2}(L) \\ \phi_i^{\varepsilon u}(L) & \phi_i^{\varepsilon\varepsilon}(L) & \phi_i^{\varepsilon f_1}(L) & \phi_i^{\varepsilon f_2}(L) & \phi_i^{\varepsilon x_2}(L) \\ \phi^{f_1 u}(L) & \phi^{f_1\varepsilon}(L) & \phi^{f_1 f_1}(L) & \phi^{f_1 f_2}(L) & \phi^{f_1 x_2}(L) \\ \phi^{f_2 u}(L) & \phi^{f_2\varepsilon}(L) & \phi^{f_2 f_1}(L) & \phi^{f_2 f_2}(L) & \phi^{f_2 x_2}(L) \\ \phi_i^{x_2 u}(L) & \phi_i^{x_2\varepsilon}(L) & \phi_i^{x_2 f_1}(L) & \phi_i^{x_2 f_2}(L) & \phi_i^{x_2 x_2}(L) \end{pmatrix} = \begin{pmatrix} \phi_i^{uu}(L) & \phi_i^{u\varepsilon}(L) & 0 & 0 & \phi_i^{ux_2}(L) \\ \phi_i^{\varepsilon u}(L) & \phi_i^{\varepsilon\varepsilon}(L) & \phi_i^{\varepsilon f_1}(L) & \phi_i^{\varepsilon f_2}(L) & \phi_i^{\varepsilon x_2}(L) \\ 0 & 0 & \phi^{f_1 f_1}(L) & \phi^{f_1 f_2}(L) & 0 \\ 0 & 0 & \phi^{f_2 f_1}(L) & \phi^{f_2 f_2}(L) & 0 \\ \phi_i^{x_2 u}(L) & \phi_i^{x_2\varepsilon}(L) & \phi_i^{x_2 f_1}(L) & \phi_i^{x_2 f_2}(L) & \phi_i^{x_2 x_2}(L) \end{pmatrix}. \tag{3.1}$$

Since nonstationary and stationary common factors do not depend on i, $\phi^{f_1 u}(L)$, $\phi^{f_1\varepsilon}(L)$, $\phi^{f_1 x_2}(L)$, $\phi^{f_2 u}(L)$, $\phi^{f_2\varepsilon}(L)$ and $\phi^{f_2 x_2}(L)$ are all matrices of zeros. Moreover, we assume that $\phi_i^{uf_l}(L) = 0$ for $l = 1, 2$. This assumption indicates that there exists no serial or contemporaneous correlation

[3] If $E(x_{2,it}) = v_{2i} \neq 0$, we can rewrite the model (2.6) with the inclusion of an intercept, such that $y_{it} = \mu_i + \beta_{1,i}^{0\prime} x_{1,it} + \beta_{2,i}^{0\prime} x_{2,it}^* + \lambda_{1i}^{0\prime} f_{1t}^0 + \lambda_{2i}^{0\prime} f_{2t}^0 + u_{it}$, where $x_{2,it}^* = x_{2,it} - v_{2i}$ has zero mean and $\mu_i = \beta_{2,i}^{0\prime} v_{2i}$.

between the regression error u_{it} and $\left(\Delta f_{1t}^{0\prime},\ f_{2t}^{0\prime}\right)$. In Assumption 3.1(iv), we also require the stationary regressors to be sequentially exogenous to simplify the asymptotic analysis. These conditions ensure the consistency of the initial estimators of $\beta_{1,i}^0$'s and impose some restrictions on $\phi_i^{ux_2}(L)$, $\phi_i^{x_2u}(L)$ and $\phi_i^{x_2f_2}(L)$. For the consistency of the estimators of $\beta_{2,i}^0$'s, we further require that the stationary regressors are uncorrelated with the stationary common factors as in Assumption 3.1(iv).

The moment condition in Assumption 3.1(i) is needed to ensure the validity of the functional central limit theorem for the weakly dependent linear process $\{w_{it}\}$. We apply the Beveridge and Nelson (1981, BN hereafter) decomposition

$$w_{it} = \phi_i(1)v_{it} + \tilde{w}_{it-1} - \tilde{w}_{it}, \tag{3.2}$$

where $\tilde{w}_{it} = \sum_{j=0}^{\infty} \tilde{\phi}_{ij} v_{i,t-j}$ and $\tilde{\phi}_{ij} = \sum_{s=j+1}^{\infty} \phi_{is}$. Assumption 3.1(ii) imposes a uniform k-summability condition on the coefficient matrix ϕ_{ij} that ensures $\sum_{j=0}^{\infty} \|\tilde{\phi}_{ij}\|^k < \infty$ by Lemma 2.1 in PS, thereby assuring the validity of (3.2). This condition further implies that $\tilde{w}_{it}$ behaves like a stationary process with a finite kth moment. The second part of Assumption 3.1(ii) rules out potential cointegration relationships among the variables in $(x_{1,it}', f_{1t}^{0\prime})'$. Assumption 3.1(iii) allows $(u_{it}, \varepsilon_{it}', x_{2,it}')$ to be cross-sectionally dependent, but they become independent across i given $\mathcal{C}$. By saying that "$(u_{it}, \varepsilon_{it})$ are cross-sectionally dependent but they become independent across i given $\mathcal{C}$," we mean that cross-section dependence among $\{(u_{it}, \varepsilon_{it})\}$, if it exists, only comes from the sigma algebra generated by the common factors and factor loadings, $\mathcal{C} = \sigma(\Lambda^0, F^0)$. Unconditionally, we allow for cross-section dependence among $\{(u_{it}, \varepsilon_{it})\}$. Assumption 3.1(v) ensures that the factor loadings are independent of the generalization of the error processes over t and across i. Assumption 3.1 validates the following multivariate invariance principle for partial sums of w_{it}

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\cdot \rfloor} w_{it} \Rightarrow B_i(\cdot) \equiv BM_i(\Omega_i) \text{ as } T \to \infty \text{ for all } i,$$

where $B_i = (B_{1i}, B_{2i}', B_3', B_4', B_{5i}')'$ is a $(1+p+r) \times 1$ vector Brownian motion with a covariance matrix Ω_i.

Assumption 3.2 *(i) As* $N \to \infty$, $\frac{1}{N}\Lambda^{0\prime}\Lambda^0 \xrightarrow{p} \Sigma_\lambda > 0$ *and* $\Lambda_1^{0\prime}\Lambda_2^0 = O_P\left(N^{1/2}\right)$. $\sup_{N\geq 1} \max_{1\leq i\leq N} \|\lambda_i^0\| \leq \bar{c}_\lambda < \infty$.

(ii) $\mathbb{E}\|\Delta f_{1t}^0\|^{2q+\epsilon} \leq M$ *and* $\mathbb{E}\|f_{2t}^0\|^{2q+\epsilon} \leq M$ *for some* $\epsilon > 0$, $q \geq 4$ *and for all* t. *As* $T \to \infty$, $\frac{1}{T^2}\sum_{t=1}^T f_{1t}^0 f_{1t}^{0\prime} \xrightarrow{d} \int B_3 B_3'$ *and* $\frac{1}{T}\sum_{t=1}^T f_{2t}^0 f_{2t}^{0\prime} \xrightarrow{p} \Sigma_{44} > 0$, *where* B_3 *is an* r_1*-vector of Brownian motions with a long-run covariance matrix* $\Omega_{33} > 0$.

(iii) Let $\gamma_N(s,t) = \frac{1}{N}\sum_{i=1}^N \mathbb{E}(u_{it}u_{is})$ *and* $\xi_{st} = \frac{1}{N}\sum_{i=1}^N [u_{it}u_{is} - \mathbb{E}(u_{it}u_{is})]$. *Then* $\sup_{N\geq 1}\sup_{T\geq 1} \max_{1\leq s,t\leq T} N^2\mathbb{E}|\xi_{st}|^4 \leq M$ *and* $\sup_{N\geq 1}\sup_{T\geq 1} T^{-1}\sum_{s=1}^T\sum_{t=1}^T \|\gamma_N(s,t)\|^2 \leq M$.

(iv) We consider the linear combinations of the nonstationary regressors $x_{1,i}b_{1,i}$ *where* $b_{1,i}$ *is a* $p_1 \times 1$ *vector. Let* $\mathbf{b}_1 = (b_{1,1}, ..., b_{1,N})$, $F_1 \in \mathcal{F}_1$ *and* $\boldsymbol{\pi}_1 = (\pi_{1i}, ..., \pi_{1N})'$, *where* $\mathcal{F}_1 = \{F_1 \in \mathbb{R}^{T\times r_1}: \frac{1}{T^2}F_1'F_1 = I_{r_1}\}$ *and* π_{1i} *is an* $r_1 \times 1$ *real vector. Let* $\chi_{i,\mathbf{b}_1} = x_{1,i}b_{1,i} - F_1^0\lambda_{1i}^0$. *We assume*

(a) There does not exist $(\mathbf{b}_1, F_1, \boldsymbol{\pi}_1) \in \mathbb{R}^{p_1 \times N} \times \mathcal{F}_1 \times \mathbb{R}^{N \times r_1}$ *with* $\mathbf{b}_1 = (b_{1,1}, ..., b_{1,N}) \neq 0$ *such that we can write*

$$x_{1,i} b_{1,i} = (F_1^0, F_1) \begin{pmatrix} \lambda_{1i}^0 \\ \pi_{1i} \end{pmatrix} \text{ a.s. } \forall i; \tag{3.3}$$

(b) There exists a constant $c_1 > 0$ *such that*

$$\min_{\left\{\mathbf{b}_1 \in \mathbb{R}^{p_1 \times N}: \frac{1}{N}\|\mathbf{b}_1\|^2 = c_{NT}\right\}} \sum_{l=r_1+1}^{N} \mu_l \left(\frac{1}{NT^2} \sum_{i=1}^{N} \chi_{i,\mathbf{b}_1} \chi_{i,\mathbf{b}_1}' \right) \geq c_1 c_{NT} \quad \text{w.p.a.1.}; \tag{3.4}$$

(c) There exists a constant $\rho_{\min} > 0$ *such that* $P\left(\mu_{\min}\left(Q_1(F_1^0) - Q_2(F_1^0)\right) \geq c\rho_{\min}\right) = 1 - o(N^{-1})$.

(v) There exist constant bounds $\{\underline{c}_2, \bar{c}_2\}$ *such that* $0 < \underline{c}_2 \leq \min_{1 \leq i \leq N} \mu_{\min}\left(\mathbb{E}(x_{2,it} x_{2,it}')\right) \leq \max_{1 \leq i \leq N} \mu_{\max}\left(\mathbb{E}(x_{2,it} x_{2,it}')\right) \leq \bar{c}_2 < \infty$.

Assumption 3.2(i)-(iii) imposes some standard moment conditions in the factor literature; see, e.g., Bai and Ng (2002, 2004). Assumption 3.2(i) indicates that the stationary factor loadings and the nonstationary factor loadings can only be weakly correlated, which facilitates derivations. Assumption 3.2(iii) imposes conditions on the error process $\{u_{it}\}$, which are adapted from Bai (2003) and allow for weak forms of cross-section and serial dependence in the error processes. Assumption 3.2(iv.a) is the key identification condition that will be satisfied provided no linear combinations of $x_{1,it}$ can be written as a pure factor structure with $2r_1$ factors for all i. In particular, if there exists a combination $(\mathbf{b}_1, F_1, \boldsymbol{\pi}_1)$ such that

$$b_{1,i}' x_{1,it} = \lambda_{1i}^{0\prime} f_{1t}^0 + \pi_{1i}' f_{1t} \text{ for all } (i,t),$$

then we must have $\mathbf{b}_1 = 0$ and $\pi_{1i}' f_{1t} = -\lambda_{1i}^{0\prime} f_{1t}^0$ for all (i,t). This condition does not rule out common regressors in the model. For example, we can consider the simplest case where $r_1 = 1$ and $x_{1,it} = x_{1,t}$ is I(1). As long as λ_{1i}^0 varies across i and $x_{1,t}$ is not proportional to f_{1t}^0 ($x_{1,t}$ and f_{1t}^0 are not collinear in the general case), Assumption 3.2(iv.a) can still hold. See the Online Appendix C for more details. Assumption 3.2(iv.b) is used to establish the preliminary consistent rates in Theorem 3.1(i) below and it is in the same spirit as Assumption 4(ii.a) in Moon and Weidner (2017). Assumption 3.2(iv.c) is used to establish the uniform classification consistency in Theorem 3.3 below. It assumes $Q_1(F_1^0) - Q_2(F_1^0)$ is positive definite in the limit. Assumption 3.2(v) is required for the identification of $\beta_{2,i}^0$ and apparently it allows for the presence of both common stationary regressors and time-invariant regressors in $x_{2,it}$.

Assumption 3.3 *(i) For each* $k = 1, ..., K_0$, $N_k/N \to \tau_k \in (0,1)$ *as* $N \to \infty$.

(ii) $\min_{1 \leq k \neq j \leq K} \left\|\alpha_k^0 - \alpha_j^0\right\| \geq \underline{c}_\alpha$ *for some fixed* $\underline{c}_\alpha > 0$.

(iii) As $(N,T) \to \infty$, $N/T^2 \to c_1 \in [0, \infty)$ *and* $T/N^2 \to c_2 \in [0, \infty)$.

(iv) Let $d_T = \log\log T$. *As* $(N,T) \to \infty$, $\lambda\sqrt{T} \to 0$, $\lambda T N^{-1/q} d_T^{-2} / (\log T)^{1+\epsilon} \to \infty$, *and*

$d_T^2 N^{1/q} T^{-1} \times (\log T)^{1+\epsilon} \to 0.$

Assumptions 3.3(i)-(ii) were used in SSP (2016a). Assumption 3.3(i) implies that each group has an asymptotically non-negligible number of individuals as $N \to \infty$ and Assumption 3.3(ii) requires the separability of group-specific parameters. Similar conditions are assumed in the panel literature with latent group patterns, e.g., Bonhomme and Manresa (2015), Ando and Bai (2016), SSP (2016a), and Su and Ju (2018). Assumption 3.3(iii)-(iv) imposes conditions to control the relative rates at which N and T pass to infinity. They require that N pass to infinity at a rate faster than $T^{1/2}$ but slower than T^2. The involvement of the factor d_T is due to the law of iterated logarithm. One can verify that the permissible range of values for λ that satisfy Assumption 3.3(iv) is $\lambda \propto T^{-\alpha}$ for $\alpha \in (\frac{1}{2}, \frac{q-1}{q})$ for $q \geq 4$.

3.2 Preliminary rates of convergence

Let $\hat{b}_{l.i} = \hat{\beta}_{l,i} - \beta_{l,i}^0$ for $l = 1, 2$, $\delta_{NT} = \min(\sqrt{N}, T)$, $C_{NT} = \min(\sqrt{N}, \sqrt{T})$, $\eta_{1NT}^2 = \frac{1}{N}\sum_{i=1}^N \left\|\hat{b}_{1,i}\right\|^2$, and $H_1 = (\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1)V_{1,NT}^{-1}$. The following theorem establishes consistency of $\hat{\beta}_{1,i}$, $\hat{\beta}_{2,i}$, and $\hat{F}_1$.

Theorem 3.1 *Suppose that Assumptions 3.1-3.2 hold. Then*
(i) $\frac{1}{N}\sum_{i=1}^N \|\hat{\beta}_{1,i} - \beta_{1,i}^0\|^2 = O_P((T/d_T^3)^{-1/2})$,
(ii) $\left\|P_{\hat{F}_1} - P_{F_1^0}\right\|^2 = O_P((T/d_T^3)^{-1/2})$,
(iii) $\frac{1}{T}\|\hat{F}_1 - F_1^0 H_1\| = O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})$.

Theorem 3.1(i) establishes the *preliminary* mean-square consistency of $\{\hat{\beta}_{1,i}\}$. Theorem 3.1(ii) shows that the spaces spanned by the columns of $\hat{F}_1$ and F_1^0 are asymptotically the same. Theorem 3.1(iii) indicates that the true factor F_1^0 can only be identified up to a nonsingular rotation matrix H_1. Compared with Bai and Ng (2004) and Bai et al. (2009), our results allow for heterogeneous slope coefficients, stationary regressors and unobserved stationary and nonstationary common factors.

The following theorem establishes the rate of convergence for the individual and group-specific estimators, as well as for the estimated factors up to rotation.

Theorem 3.2 *Suppose that Assumptions 3.1-3.2 hold. Then*
(i) $\frac{1}{N}\sum_{i=1}^N \|\hat{\beta}_{1,i} - \beta_{1,i}^0\|^2 = O_P(d_T T^{-2})$,
(ii) $\hat{\beta}_{1,i} - \beta_{1,i}^0 = O_P(d_T^{1/2}T^{-1} + \lambda)$ *and* $\hat{\beta}_{2,i} - \beta_{2,i}^0 = O_P(d_T^{1/2}T^{-1/2} + N^{-1/2})$ *for* $i = 1, ..., N$,
(iii) $(\hat{\alpha}_{(1)}, ..., \hat{\alpha}_{(K)}) - (\alpha_1^0, ..., \alpha_K^0) = O_P(d_T^{1/2}T^{-1})$ *for some suitable permutation* $(\hat{\alpha}_{(1)}, ..., \hat{\alpha}_{(K)})$ *of* $(\hat{\alpha}_1, ..., \hat{\alpha}_K)$,
(iv) $T^{-1}\|\hat{F}_1 - F_1^0 H_1\| = O_P(d_T^{1/2}T^{-1} + (NT)^{-1/2})$.

Theorem 3.2(i) establishes the mean-square convergence for the estimators of $\beta_{1,i}^0$ while Theorem 3.2(ii) studies the preliminary point-wise convergence of $\hat{\beta}_{1,i}$ and $\hat{\beta}_{2,i}$. The usual super consistency of nonstationary estimators $\hat{\beta}_{1,i}$ is preserved if $\lambda = O(T^{-1})$ despite the fact that we ignore unobserved stationary common factors and allow for correlation between u_{it} and $\left(x_{it}', f_{1t}^{0\prime}\right)$. Theorem 3.2(iii) indicates that the group-specific parameters, $\alpha_1^0, ..., \alpha_K^0$, can be consistently estimated. Theorem 3.2(iv) updates the convergence rate of the unobserved nonstationary factors in Theorem 3.1(iii). For notational simplicity, hereafter we simply write $\hat{\alpha}_k$ for $\hat{\alpha}_{(k)}$ as the consistent estimator of α_k^0.

3.3 Classification consistency

We now study classification consistency. Define

$$\hat{E}_{kNT,i} = \{i \notin \hat{G}_k | i \in G_k^0\} \quad \text{and} \quad \hat{F}_{kNT,i} = \{i \notin G_k^0 | i \in \hat{G}_k\},$$

where $i = 1, ..., N$ and $k = 1, ...K$. Let $\hat{E}_{kNT} = \cup_{i \in \hat{G}_k} \hat{E}_{kNTi}$ and $\hat{F}_{kNT} = \cup_{i \in \hat{G}_k} \hat{F}_{kNTi}$. The events $\hat{E}_{kNT}$ and $\hat{F}_{kNT}$ mimic type I and type II errors in statistical tests. Following SSP (2016a), we say that a classification method is individually consistent if $P(\hat{E}_{kNT,i}) \to 0$ as $(N,T) \to \infty$ for each $i \in G_k^0$ and $k = 1, ..., K$, and $P(\hat{F}_{kNT,i}) \to 0$ as $(N,T) \to \infty$ for each $i \in G_k^0$ and $k = 1, ..., K$. It is uniformly consistent if $P(\cup_{k=1}^K \hat{E}_{kNT}) \to 0$ and $P(\cup_{k=1}^K \hat{F}_{kNT}) \to 0$ as $(N,T) \to \infty$.

The following theorem establishes uniform classification consistency.

Theorem 3.3 *Suppose that Assumptions 3.1-3.3 hold. Then*
(i) $P(\cup_{k=1}^{K_0} \hat{E}_{kNT}) \leq \sum_{k=1}^{K_0} P(\hat{E}_{kNT}) \to 0$ *as* $(N,T) \to \infty$,
(ii) $P(\cup_{k=1}^{K_0} \hat{F}_{kNT}) \leq \sum_{k=1}^{K_0} P(\hat{F}_{kNT}) \to 0$ *as* $(N,T) \to \infty$.

Theorem 3.3 implies uniform classification consistency – all individuals within a certain group, say G_k^0, can be simultaneously and correctly classified into the same group (denoted $\hat{G}_k$) w.p.a.1. Conversely, all individuals that are classified into the same group, say $\hat{G}_k$, simultaneously belong to the same group (G_k^0) w.p.a.1. Let $\hat{N}_k = \#\hat{G}_k$. One can easily show that $P(\hat{G}_k = G_k^0) \to 1$ so that $P(\hat{N}_k = N_k) \to 1$.

Note that Theorem 3.3 is an asymptotic result. It does not ensure that all individuals can be classified into one of the estimated groups when T is not large or λ is not sufficiently big if we stick to the classification rule in (2.13). In practice, we classify $i \in \hat{G}_k$ if $\hat{\beta}_{1,i} = \hat{\alpha}_k$ for some $k = 1, ..., K$, and $i \in \hat{G}_l$ for some $l = 1, ..., K$ if $||\hat{\beta}_{1,i} - \hat{\alpha}_l|| = \min\{||\hat{\beta}_{1,i} - \hat{\alpha}_1||, ..., ||\hat{\beta}_{1,i} - \hat{\alpha}_K||\}$ and $\sum_{k=1}^K \mathbf{1}\{\hat{\beta}_{1,i} = \hat{\alpha}_k\} = 0$. Since Theorem 3.3 ensures $\sum_{k=1}^K P(\hat{\beta}_{1,i} = \hat{\alpha}_k) \to 1$ as $(N,T) \to \infty$ uniformly in i, we can ignore such a modification in large samples in subsequent theoretical analyses and restrict our attention to the classification rule in (2.13) to avoid confusion.

3.4 Oracle properties and post-Lasso and Cup-Lasso estimators

We examine the oracle properties of the three Lasso-type estimators. To proceed, we add some notation. For $k = 1, ..., K$, we define

$$
\begin{aligned}
U_{kNT} &= \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} x_{1,i}' M_{F_1^0} \left(\left(u_i + F_2^0 \lambda_{2i}^0\right) - \frac{1}{N} \sum_{j=1}^{N} \left(u_j + F_2^0 \lambda_{2j}^0\right) a_{ij} \right), \\
B_{kNT,1} &= \sum_{i=1}^{N} B_{k,iNT,1} = \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} \left(\sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{1}\{t = s\} - \varkappa_{ts} \mathbf{1}\{s \le t\} \right) \Delta_{21,i}, \\
B_{kNT,2} &= \sum_{i=1}^{N} B_{k,iNT,2} = \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} \mathbb{E}_{\mathcal{C}} \left(x_{1,i}\right)' M_{F_1^0} F_2^0 \left(\lambda_{2i}^0 - \frac{1}{N} \sum_{j=1}^{N} \lambda_{2j}^0 a_{ij} \right), \\
V_{kNT} &= \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} S^{\varepsilon} \phi_i^{\dagger}(1) \sum_{t=1}^{T} \sum_{s=1}^{T} \left\{ \bar{\varkappa}_{ts} \left(V_{it}^{ux} v_{is}^{ux\prime} \right) - \left[\mathbf{1}\{t = s\} - \varkappa_{ts} \mathbf{1}\{s \le t\}\right] I_{1+p} \right\} \phi_i^{\dagger}(1)' S^{u\prime} \\
&\quad + \frac{1}{\sqrt{N_k}T} \sum_{i=1}^{N} \left\{ \mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right) \mathbf{1}\{i \in G_k^0\} - \frac{1}{N} \sum_{j \in G_k^0} a_{ij} \mathbb{E}_{\mathcal{C}}(x_{1,j}') \right\} M_{F_1^0} u_i \\
&\quad + \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} [x_{1,i} - \mathbb{E}_{\mathcal{C}}(x_{1,i})]' M_{F_1^0} F_2^0 \lambda_{2i}^0,
\end{aligned}
$$

where $\varkappa_{ts} = f_{1t}^{0\prime}(F_1^{0\prime} F_1^0)^{-1} f_{1s}^0$, $\bar{\varkappa}_{ts} = \mathbf{1}\{t = s\} - \varkappa_{ts}$, $v_{is}^{u\varepsilon} = (v_{is}^{u}, v_{is}^{\varepsilon\prime}, v_{is}^{x_2\prime})'$, $V_{it}^{u\varepsilon} = \sum_{s=1}^{t} v_{is}^{ux}$, $\mathbb{E}_{\mathcal{C}}(\cdot) = \mathbb{E}(\cdot|\mathcal{C})$, $\phi_i^{\dagger}(L) = \begin{pmatrix} \phi_i^{u\dagger}(L) \\ \phi_i^{\varepsilon\dagger}(L) \end{pmatrix} = \begin{pmatrix} \phi_i^{uu}(L) & \phi_i^{u\varepsilon}(L) & \phi_i^{ux_2}(L) \\ \phi_i^{\varepsilon u}(L) & \phi_i^{\varepsilon\varepsilon}(L) & \phi_i^{\varepsilon x_2}(L) \end{pmatrix}$, $S^u = (1, 0_{1\times p})$, $S^{\varepsilon} = (0_{p_1 \times 1}, \iota_{p_1 \times p})$ and ι is a vector of ones. Let $Q_{1NT} = \text{diag}\left(\frac{1}{N_1 T^2} \sum_{i \in G_1^0} x_{1,i}' M_{F_1^0} x_{1,i}, \ldots, \frac{1}{N_K T^2} \sum_{i \in G_K^0} x_{1,i}' M_{F_1^0} x_{1,i} \right)$ and $Q_{2NT,kl} = \frac{1}{N_k N T^2} \sum_{i \in G_k^0} \sum_{j \in G_l^0} x_{1,i}' M_{F_1^0} x_{1,j} a_{ij}$ for $k, l = 1, ..., K$. Let $Q_{NT} = Q_{1NT} - Q_{2NT}$,

$$
Q_{2NT} = \begin{pmatrix} Q_{2NT,11} & \cdots & Q_{2NT,1K} \\ \vdots & \ddots & \vdots \\ Q_{2NT,K1} & \cdots & Q_{2NT,KK} \end{pmatrix} \text{ and } Q_0 = \begin{pmatrix} Q_{1,1} - Q_{2,11} & -Q_{2,12} & \cdots & -Q_{2,1K} \\ -Q_{2,21} & Q_{1,2} - Q_{2,22} & \cdots & -Q_{2,2K} \\ \vdots & \vdots & \ddots & \vdots \\ -Q_{2,K1} & -Q_{2,K2} & \cdots & Q_{1,K} - Q_{2,KK} \end{pmatrix},
$$

where $Q_{1,k} = \lim_{N\to\infty} \frac{1}{N_k} \sum_{i \in G_k^0} \mathbb{E}_{\mathcal{C}} \left(\int \tilde{B}_{2i} \tilde{B}_{2i}' \right)$, $Q_{2,kl} = \lim_{N\to\infty} \frac{1}{NN_k} \sum_{i \in G_k^0} \sum_{j \in G_l^0} a_{ij} \mathbb{E}_{\mathcal{C}} \left(\int \tilde{B}_{2,i} \tilde{B}_{2,j}' \right)$, and $\tilde{B}_{2i} = B_{2,i} - \int B_{2,i} B_3' \left(\int B_3 B_3' \right)^{-1} B_3$.

Let $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, ..., \hat{\alpha}_K)$. Let $U_{NT} = (U_{1NT}', \ldots, U_{KNT}')'$, $B_{NT} = (B_{1NT}', \ldots, B_{KNT}')'$, $V_{NT} = (V_{1NT}', \ldots, V_{KNT}')'$ and $B_{kNT} = B_{kNT,1} + B_{kNT,2}$. The following theorem reports the Bahadur-type representation and asymptotic distribution of vec$(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0)$.

Theorem 3.4 *Suppose that assumptions 3.1-3.3 hold. Let $\hat{\alpha}_k$ be obtained by solving (2.11). Then*

(i) $\sqrt{N}T vec(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) = \sqrt{D_{NK}} Q_{NT}^{-1} U_{NT} + o_P(1) = \sqrt{D_{NK}} Q_{NT}^{-1} (V_{NT} + B_{NT}) + o_P(1)$,

(ii) $\sqrt{N}Tvec(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)-\sqrt{D_{NK}}Q_{NT}^{-1}B_{NT}\Rightarrow\mathcal{MN}(0,\ D_0Q_0^{-1}\Omega_0Q_0^{-1})$ *as* $(N,T)\to\infty$, *where* $D_{NK}=diag\Big(\frac{N}{N_1},...,\frac{N}{N_K}\Big)\otimes I_{p_1}$, $D_0=diag\Big(\frac{1}{\tau_1},...,\frac{1}{\tau_K}\Big)\otimes I_{p_1}$, $\Omega_0=\lim_{(N,T)\to\infty}\Omega_{NT}$, *and* $\Omega_{NT}=Var(V_{NT}|\mathcal{C})$.

Theorem 3.4 indicates that V_{NT} and B_{NT} are associated with the asymptotic variance and bias of $\hat{\alpha}_k$. The decomposition $B_{kNT}=B_{kNT,1}+B_{kNT,2}$ indicates two sources of the bias. The first bias term $B_{kNT,1}$ results from the contemporaneous correlation between $(x_{1,it},f_{1t})$ and u_{it} and the serial correlation among the innovation processes $\{w_{it}\}$. Apparently, the presence of unobserved nonstationary factors f_{1t}^0 complicates the formula for $B_{kNT,1}$ through the term $\varkappa_{ts}$. The second bias term $B_{kNT,2}$ is due to the presence of the unobserved stationary factors f_{2t}^0. In the special case where neither f_{1t}^0 nor f_{2t}^0 is present in the model, we have $B_{kNT}=B_{kNT,1}=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\Delta_{21,i}$. This is the usual asymptotic bias term for panel cointegration regression that is associated with the effects of the one-sided long-run covariance (c.f., Phillips (1995) and Phillips and Moon (1999)). The ith element of V_{NT} is independent across i conditional on $\mathcal{C}$ and $\mathbb{E}_{\mathcal{C}}(V_{NT})=0$. This makes it possible for us to derive a version of the conditional central limit theorem for V_{NT} and establish the limiting mixed normal ($\mathcal{MN}$) distribution of the estimators $\hat{\boldsymbol{\alpha}}$ in Theorem 3.4(ii).

As shown in the proof of Theorem 3.4, the asymptotic bias term B_{NT} is $O_P(\sqrt{N_k})$, which implies the T-consistency of the C-Lasso estimators $\hat{\alpha}_k$. To obtain the $\sqrt{N}T$-rate of convergence, we need to remove the asymptotic bias by constructing consistent estimates of B_{NT}.

3.4.1 Bias correction, fully modified and continuous updating procedures

Three types of bias-corrected estimators are considered: the bias-corrected post-Lasso estimator $\hat{\alpha}_{\hat{G}_k}^{bc}$, the fully-modified post-Lasso estimator $\hat{\alpha}_{\hat{G}_k}^{fm}$, and the fully-modified continuously updated post-Lasso (Cup-Lasso) estimator $\hat{\alpha}_{\hat{G}_k}^{cup}$, whose definitions are given below.

Following Phillips and Hansen (1990) and Phillips (1995), we first construct consistent time series estimators of the long-run covariance matrix Ω_i and the one-sided long-run covariance matrix Δ_i by

$$\hat{\Omega}_i=\sum_{j=-T+1}^{T-1}\omega\left(\frac{j}{J}\right)\hat{\Gamma}_i(j),\ \text{and}\ \hat{\Delta}_i=\sum_{j=0}^{T-1}\omega\left(\frac{j}{J}\right)\hat{\Gamma}_i(j),$$

where $\omega(\cdot)$ is a kernel function, J is a bandwidth parameter, and $\hat{\Gamma}_i(j)=\frac{1}{T}\sum_{t=1}^{T-j}\hat{w}_{i,t+j}\hat{w}_{it}'$ with $\hat{w}_{it}=(\hat{u}_{it},\Delta x_{1,it}',\Delta\hat{f}_{1t}',\hat{f}_{2t}',x_{2,it}')'$. We partition $\hat{\Omega}_i$ and $\hat{\Delta}_i$ conformably with Ω_i. For example, $\hat{\Delta}_{jl,i}$ denotes a submatrix of $\hat{\Delta}_i$ given by $S_j\hat{\Delta}_iS_l'$ for $j,l=1,...,5$.

We make the following assumption on the kernel function and bandwidth.

Assumption 3.4 *(i) The kernel function* $\omega(\cdot)$*:* $R\to[-1,1]$ *is a twice continuously differentiable symmetric function such that* $\int_{-\infty}^{\infty}\omega(x)^2dx\leq\infty$, $\omega(0)=1$, $\omega(x)=0$ *for* $|x|\geq1$, *and* $\lim_{|x|\to1}\omega(x)/(1-|x|)^q=c>0$ *for some* $q\in(0,\infty)$.

(ii) As $(N,T)\to\infty$, $N/J^{2q}\to0$ and $J/T\to0$.

We modify the variable y_{it} with the following transformation to correct for endogeneity:

$$\hat{y}_{it}^{+} = y_{it} - \hat{\Omega}_{12,i}\hat{\Omega}_{22,i}^{-1}\Delta x_{1,it}. \tag{3.5}$$

This would lead to the modified equation $\hat{y}_{it}^{+} = \beta_{1,i}^{0\prime}x_{1,it} + \beta_{2,i}^{0\prime}x_{2,it} + \lambda_{1i}^{0\prime}f_{1t}^{0} + \lambda_{2i}^{0\prime}f_{2t}^{0} + \hat{u}_{it}^{+}$, where $\hat{u}_{it}^{+} = u_{it} - \hat{\Omega}_{12,i}\hat{\Omega}_{22i}^{-1}\Delta x_{1,it}$. Define

$$\hat{\Delta}_{12,i}^{+} = \hat{\Delta}_{12,i} - \hat{\Omega}_{12,i}\hat{\Omega}_{22i}^{-1}\hat{\Delta}_{22,i}. \tag{3.6}$$

Note that (3.5) and (3.6) help to correct for endogeneity and for serial correlation, respectively. Let $\hat{y}_{i}^{+} = (\hat{y}_{i1}^{+}, ..., \hat{y}_{iT}^{+})'$ and $\hat{\Delta}_{21,i}^{+} = \hat{\Delta}_{12,i}^{+\prime}$.

We can obtain the bias-corrected post-Lasso estimator $\hat{\boldsymbol{\alpha}}_{\hat{G}_k}^{bc}$, the fully modified post-Lasso estimator $\hat{\alpha}_{\hat{G}_k}^{fm}$, $\hat{F}_1$ and $\hat{F}_2$ by iteratively solving the following equations (3.8) to (3.10)

$$\mathrm{vec}\left(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc}\right) = \mathrm{vec}\left(\hat{\boldsymbol{\alpha}}\right) - \frac{1}{\sqrt{NT}}\sqrt{D_{NK}}\hat{Q}_{NT}^{-1}\left(\hat{B}_{NT,1} + \hat{B}_{NT,2}\right), \tag{3.7}$$

$$\hat{\alpha}_{\hat{G}_k}^{fm} = \left(\sum_{i\in\hat{G}_k} x_{1,i}'M_{\hat{F}_1}x_{1,i}\right)^{-1}\left\{\sum_{i\in\hat{G}_k} x_{1,i}'M_{\hat{F}_1}\hat{y}_i^{+} - T\sqrt{N_k}\left(\hat{B}_{kNT,1}^{+} + \hat{B}_{kNT,2}\right)\right\}, \tag{3.8}$$

$$\hat{F}_1 V_{1,NT} = \left[\frac{1}{NT^2}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}(\hat{y}_i - x_{1,i}\hat{\alpha}_{\hat{G}_k}^{fm} - x_{2,i}\hat{\beta}_{2,i})(\hat{y}_i - x_{1,i}\hat{\alpha}_{\hat{G}_k}^{fm} - x_{2,i}\hat{\beta}_{2,i})'\right]\hat{F}_1, \tag{3.9}$$

$$\hat{F}_2 V_{2,NT} = \left[\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}(\hat{y}_i - x_{1.i}\hat{\alpha}_{\hat{G}_k}^{fm} - x_{2,i}\hat{\beta}_{2,i} - \hat{F}_1\hat{\lambda}_{1i})(\hat{y}_i - x_{1,i}\hat{\alpha}_{\hat{G}_k}^{fm} - x_{2,i}\hat{\beta}_{2,i} - \hat{F}_1\hat{\lambda}_{1i})'\right]\hat{F}_2, \tag{3.10}$$

where $\hat{B}_{NT,l} = (\hat{B}_{1NT,l}', ..., \hat{B}_{KNT,l}')'$ for $l = 1,\ 2$, $\hat{B}_{kNT,1} = \frac{1}{\sqrt{\hat{N}_k T}}\sum_{i\in\hat{G}_k}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\hat{\varkappa}}_{ts}\right)\hat{\Delta}_{21,i}$, $\hat{B}_{kNT,2} = \frac{1}{\sqrt{\hat{N}_k T}}\sum_{i\in\hat{G}_k}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\hat{\varkappa}}_{ts}\right)\hat{\Delta}_{24,i}\hat{\hat{\lambda}}_{2i}$, $\hat{B}_{kNT,1}^{+} = \frac{1}{\sqrt{\hat{N}_k T}}\sum_{i\in\hat{G}_k}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\hat{\varkappa}}_{ts}\right)\hat{\Delta}_{21,i}^{+}$, $\hat{\hat{\varkappa}}_{ts} = \mathbf{1}\{t = s\} - \hat{\varkappa}_{ts}$, $\hat{\varkappa}_{ts} = \hat{f}_{1t}'(\hat{F}_1'\hat{F}_1)^{-1}\hat{f}_{1s} = \hat{f}_{1t}'\hat{f}_{1s}/T^2$, $\hat{\hat{\lambda}}_{2i} = \hat{\lambda}_{2i} - \frac{1}{N}\sum_{j=1}^{N}\hat{\lambda}_{2j}\hat{a}_{ij}$, and $\hat{a}_{ij} = \hat{\lambda}_{1i}'(\frac{1}{N}\hat{\Lambda}_1'\hat{\Lambda}_1)^{-1}\hat{\lambda}_{1j}$. Here the definitions of $\hat{F}_1$, $V_{1,NT}$, $\hat{F}_2$, and $V_{2,NT}$ are similar to those defined above.

We obtain the fully modified Cup-Lasso estimators $\hat{\alpha}_{\hat{G}_k}^{cup}$ by iteratively solving (2.11), and (3.8) to (3.10), where we also update the group structure estimates $\{\hat{G}_k\}$. Note that $\hat{F}_1$, $V_{1,NT}$, $\hat{F}_2$, $V_{2,NT}$, and the factor loading estimates $\{\hat{\lambda}_{1i}, \hat{\lambda}_{2i}\}$ are also updated continuously in the procedure to obtain $\hat{\alpha}_{\hat{G}_k}^{cup}$.

Let $\hat{\boldsymbol{\alpha}}_{\hat{G}}^{fm} = (\hat{\alpha}_{\hat{G}_1}^{fm}, ..., \hat{\alpha}_{\hat{G}_K}^{fm})$ and $\hat{\boldsymbol{\alpha}}_{\hat{G}}^{cup} = (\hat{\alpha}_{\hat{G}_1}^{cup}, ..., \hat{\alpha}_{\hat{G}_K}^{cup})$. We establish the limiting distribution of the bias-corrected post-Lasso estimators $\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc}$, the fully modified post-Lasso estimators $\hat{\boldsymbol{\alpha}}_{\hat{G}}^{fm}$, and the Cup-Lasso estimators $\hat{\boldsymbol{\alpha}}_{\hat{G}}^{cup}$ in the following theorem.

Theorem 3.5 *Suppose that assumptions 3.1-3.4 hold. Let* $\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc}$ *be obtained by iteratively solving (3.7), (3.9)-(3.10); let* $\hat{\boldsymbol{\alpha}}_{\hat{G}}^{fm}$ *be obtained by iteratively solving (3.8)-(3.10); and let* $\hat{\boldsymbol{\alpha}}_{\hat{G}}^{cup}$ *be obtained by iteratively solving (2.11) and (3.8)-(3.10). Then as* $(N,T) \to \infty$,

(i) $\sqrt{N}T vec(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc} - \boldsymbol{\alpha}^0) \Rightarrow \mathcal{MN}(0,\ D_0 Q_0^{-1} \Omega_0 Q_0^{-1})$,

(ii) $\sqrt{N}T vec(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{fm} - \boldsymbol{\alpha}^0) \Rightarrow \mathcal{MN}(0,\ D_0 Q_0^{-1} \Omega_0^+ Q_0^{-1})$,

(iii) $\sqrt{N}T vec(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{cup} - \boldsymbol{\alpha}^0) \Rightarrow \mathcal{MN}(0,\ D_0 Q_0^{-1} \Omega_0^+ Q_0^{-1})$,

where $\Omega_0^+ = \lim_{N,T\to\infty} \Omega_{NT}^+$, $\Omega_{NT}^+ = Var\left(V_{NT}^+ | \mathcal{C}\right)$, V_{NT}^+ *is defined in the proof of Theorem 3.5, and* D_0 *and* Ω_0 *are as defined in Theorem 3.4.*

Theorem 3.5 indicates that all three types of estimators achieve the $\sqrt{N}T$-rate of convergence and have a mixed normal limit distribution. Asymptotic t-tests and Wald tests may be constructed as usual, provided that one can obtain suitable estimates of Q_0, Ω_{NT}, and Ω_{NT}^+. We can estimate Q_0 by $\hat{Q}_0 = \hat{Q}_{1NT} - \hat{Q}_{2NT}$ where $\hat{Q}_{1NT}$ and $\hat{Q}_{2NT}$ are analogously defined as Q_{1NT} and Q_{2NT} with N_k, G_k^0, F_1^0, and Λ_1^0 replaced by $\hat{N}_k$, $\hat{G}_k$, $\hat{F}_1$, and $\hat{\Lambda}_1$, respectively. We can also show that Ω_{NT} and Ω_{NT}^+ can be consistently estimated by

$$\hat{\Omega}_{NT} = \frac{\hat{D}_{NK}}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{\mathbf{X}}_{it} \hat{\mathbf{X}}_{is}' \hat{u}_{it}^* \hat{u}_{is}^* - \sum_{i=1}^{N} \hat{B}_{iNT} \hat{B}_{iNT}',$$
$$\hat{\Omega}_{NT}^+ = \frac{\hat{D}_{NK}}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{\mathbf{X}}_{it} \hat{\mathbf{X}}_{is}' \hat{u}_{it}^{*+} \hat{u}_{is}^{*+} - \sum_{i=1}^{N} \hat{B}_{iNT}^+ \hat{B}_{iNT}^{+\prime},$$

where $\hat{\mathbf{X}}_{it} = (\hat{\mathbf{X}}_{1,it}', ..., \hat{\mathbf{X}}_{K,it}')'$, $\hat{\mathbf{X}}'_{k,it}$ is the tth row of $\hat{\mathbf{X}}_{k,i}$, $\hat{\mathbf{X}}_{k,i} = M_{\hat{F}_1} x_{1,i} \mathbf{1}\{i \in \hat{G}_k\} - \frac{1}{N} \sum_{j \in \hat{G}_k} \hat{a}_{ij} M_{\hat{F}_1} x_{1,j}$, $\hat{D}_{NK} = \text{diag}\,(\frac{N}{\hat{N}_1}, ..., \frac{N}{\hat{N}_K}) \otimes I_{p_1}$, $\hat{B}_{iNT} = (\hat{B}_{1,iNT}', ..., \hat{B}_{K,iNT}')'$, $\hat{B}_{k,iNT} = \hat{B}_{k,iNT,1} + \hat{B}_{k,iNT,2}$, $\hat{B}_{k,iNT,1} = \frac{1}{\sqrt{\hat{N}_k} T} \left(\sum_{t=1}^{T} \sum_{s=1}^{t} \hat{\bar{\varkappa}}_{ts} \right) \hat{\Delta}_{21,i} \mathbf{1}\{i \in \hat{G}_k\}$, $\hat{B}_{k,iNT,2} = \frac{1}{\sqrt{\hat{N}_k} T} \left(\sum_{t=1}^{T} \sum_{s=1}^{t} \hat{\bar{\varkappa}}_{ts} \right) \hat{\Delta}_{24,i} \hat{\bar{\lambda}}_{2i} \mathbf{1}\{i \in \hat{G}_k\}$, $\hat{u}_{it}^* = y_{it} - \hat{\alpha}_k^{fm\prime} x_{1,it} - \hat{\beta}_{2,i}' x_{2,it} - \hat{\lambda}_{1i}' \hat{f}_{1t}$ for $i \in \hat{G}_k$, $\hat{B}_{iNT}^+ = (\hat{B}_{1,iNT}^{+\prime}, ..., \hat{B}_{K,iNT}^{+\prime})'$, $\hat{B}_{k,iNT}^+ = \hat{B}_{k,iNT,1}^+ + \hat{B}_{k,iNT,2}$, $\hat{B}_{k,iNT,1}^+ = \frac{1}{\sqrt{\hat{N}_k} T} \left(\sum_{t=1}^{T} \sum_{s=1}^{t} \hat{\bar{\varkappa}}_{ts} \right) \hat{\Delta}_{21,i}^+ \mathbf{1}\{i \in \hat{G}_k\}$, and $\hat{u}_{it}^{*+} = \hat{y}_{it}^+ - \hat{\alpha}_k^{fm\prime} x_{1,it} - \hat{\beta}_{2,i}' x_{2,it} - \hat{\lambda}_{1i}' \hat{f}_{1t}$ for $i \in \hat{G}_k$. See the proof of Lemma A.11(ix) in the Online Supplement. Given these estimates, it is standard to conduct inference on elements of $\boldsymbol{\alpha}^0$.

3.5 The case of incidental time trends

In the above analysis, we assume that there are no deterministic linear time trends in the y-equation and the nonstationary regressors and common factors are pure unit root processes without drifts. This subsection relaxes these restrictions to incorporate the deterministic components into our panel latent factor cointegration model. Here we consider the following model:

$$\begin{cases} y_{it} = \mu_i + \rho_i t + \beta'_{1,i}x_{1,it} + \beta'_{2,i}x_{2,it} + \lambda'^0_{1,i}f^0_{1t} + \lambda'^0_{2,i}f^0_{2t} + u_{it} \\ x_{1,it} = \mu_{1,i} + x_{1,it-1} + \varepsilon_{1,it}, \\ f_{1t} = \mu^{f_1} + f_{1t-1} + \varepsilon^{f_1}_t, \end{cases} , \tag{3.11}$$

where $i = 1, ..., N$, $t = 1, ..., T$, μ_i denotes the intercept or individual fixed effects and $\rho_i t$ denotes the incidental linear time trends. We allow for the presence of drifts $\mu_{1,i}$ in the I(1) regressors $\{x_{1,it}\}$ and drift μ^{f_1} in the I(1) common factor $\{f_{1t}\}$. The remaining variables are defined as before.

We first discuss the presence of an intercept μ_i alone in the $y-$equation. In this case, as discussed in Section 3.1, μ_i could be related to the non-zero means of the stationary regressors and stationary common factors. For example, if $E(x_{2,it}) = v_{2i} \neq 0$, we can rewritten the model (2.7) with the inclusion of an intercept, such that $y_{it} = \mu_i + \beta^{0\prime}_{1,i}x_{1,it} + \beta^{0\prime}_{2,i}x^*_{2,it} + \lambda^{0\prime}_{1i}f^0_{1t} + \lambda^{0\prime}_{2i}f^0_{2t} + u_{it}$, where $x^*_{2,it} = x_{2,it} - v_{2i}$ has zero mean and $\mu_i = \beta^{0\prime}_{2,i}v_{2i}$. In this case, we can employ the within-group *demeaned* transformation to eliminate the individual fixed effects to obtain

$$\tilde{y}_{it} = \beta'_{1,i}\tilde{x}_{1,it} + \beta'_{2,i}\tilde{x}_{2,it} + \lambda'^0_{1,i}\tilde{f}^0_{1t} + \lambda'^0_{2,i}\tilde{f}^0_{2t} + \tilde{u}_{it},$$

where $\tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^T y_{it}$, and $\tilde{x}_{1,it}$, $\tilde{x}_{2,it}$, $\tilde{f}_{1t}$, $\tilde{f}_{2t}$, and $\tilde{u}_{it}$ are analogously defined. The PPC-based estimation procedure is identical to that of Section 2.2 and implemented on the demeaned data.

Second, when we have both individual effects and incidental time trends, we can similarly employ the within-group *detrended* data to eliminate both individual fixed effects and incidental time trends. Specifically, we consider the detrended model:

$$\dot{y}_{it} = \beta'_{1,i}\dot{x}_{1,it} + \beta'_{2,i}\dot{x}_{2,it} + \lambda'^0_{1,i}\dot{f}^0_{1t} + \lambda'^0_{2,i}\dot{f}^0_{2t} + \dot{u}_{it},$$

where $\dot{y}_{it}$, $\dot{x}_{1,it}$, $\dot{x}_{2,it}$, $\dot{f}_{1t}$, $\dot{f}_{2t}$, and $\dot{u}_{it}$ are linearly detrended versions of y_{it}, $x_{1,it}$, $x_{2,it}$, f_{1t}, f_{2t}, and u_{it}. We can then apply the estimation procedure used in Section 2.2 with the dotted variables replacing the original variables.

To gain a better understanding of the incidental linear time trends in (3.11), we observe that

$$x_{1,it} = x_{1,i0} + \mu_{1,i}t + \sum_{s=1}^{t}\varepsilon_{1,is} = x_{1,i0} + \mu_{1,i}t + x^0_{1i,t}, \tag{3.12}$$

where $x^0_{1,it} \equiv \sum_{s=1}^t \varepsilon_{1,is}$ is a pure unit root process. In nonstationary time series, the reformation in (3.12) reveals that nonstationary panel data with incidental parameters are composed of two components: (1) stochastic trends represented by $x^0_{1i,t}$; and (2) incidental time trends $\mu_{1,i}t$. The incidental parameters $\mu_{1,i}$ can be interpreted as the individual-specific components of the linear deterministic trend. Similarly, the nonstationary common factors f^0_{1t} can be decomposed into the stochastic trend component and the deterministic trend component, such that $f_{1t}^0 = f_{10} + \mu^{f_1} t + \sum_{s=1}^t \varepsilon_s^{f_1}$.

In general, the asymptotic properties of the resulting Lasso-type estimators will be modified by changing the Brownian motion to the corresponding demeaned or detrended version in the respective limit distributions. Specifically, for the detrended case we can define κ_T =diag$(1, T^{-1})$, $g_t = (1, t)'$, and $g(r) = (1, r)'$. Let $t = \lfloor Tr \rfloor$, the integer part of Tr for $r \in [0, 1]$. Then as $T \to \infty$, $\kappa_T g_t \to g(r)$ uniformly in $r \in [0, 1]$. By the functional central limit theorem and continuous mapping theorem, we have

$$\begin{aligned}
\frac{1}{\sqrt{T}} \dot{x}_{1,i\lfloor Tr \rfloor} =& \frac{1}{\sqrt{T}} \left[x_{1,i\lfloor Tr \rfloor} - \sum_{s=1}^T x_{1,is} g_s' \left(\sum_{s=1}^T g_s g_s' \right)^{-1} g_t \right] \\
&= \frac{1}{\sqrt{T}} \left[x_{1,i\lfloor Tr \rfloor}^0 - \sum_{s=1}^T x_{1,is}^0 g_s' \left(\sum_{s=1}^T g_s g_s' \right)^{-1} g_t \right] \\
&= \frac{x_{1,i\lfloor Tr \rfloor}^0}{\sqrt{T}} - \frac{1}{T} \sum_{s=1}^T \frac{x_{1,is}^0}{\sqrt{T}} \kappa_T g_s' \left(\frac{1}{T} \sum_{s=1}^T \kappa_T g_s g_s' \kappa_T \right)^{-1} \kappa_T g_t \\
&\Rightarrow B_{1i}(r) - \int_0^1 B_{1i}(u) g(u)' du \left(\int_0^1 g(u) g(u)' du \right)^{-1} g(r) \equiv B_{1i}^\tau(r),
\end{aligned}$$

where $B_{1i}(\cdot)$ is as defined above Assumption 3.2, and $B_{1i}^\tau(\cdot)$ is a detrended Brownian motion obtained by the $L_2\,[0, 1]$ projection residual of $B_{1i}(r)$ on $g(r)$. Following the analysis in Sections 3.1-3.4, we can show the demeaned or detrended residuals, such as $(\dot{u}_{it}, \Delta\dot{x}_{1,it}, \Delta\dot{f}_{1t}^0, \dot{f}_{2t}^0, \dot{x}_{2,it})$, satisfy Assumption 3.1-3.2 and Theorems 3.1-3.3 continue to hold with the demeaned data and detrended data. The limiting distributions in Theorem 3.4-3.5 are modified by replacing the random processes B_{1i}, B_{2i} and B_3 by the demeaned or detrended Brownian motions. The asymptotic bias and variance can be estimated from the detrended or demeaned data. In short, the mixed normal limit theory is preserved for the group-specific long-run estimators, which permits inference using standard test statistics.

3.6 Estimating the number of unobserved factors

Our analysis has so far assumed that the numbers of nonstationary and stationary factors, r_1 and r_2, are known. We also note the nonstationary factors play a key role in the PPC estimation. We notice that the presence of stationary factors does not affect the consistency of nonstationary coefficients estimates despite its introduction of a second-order endogeneity bias. Thus, we consider a two-step approach to determine r_1 and r_2. In the first step, we introduce an information criterion to determine the number of unobserved nonstationary factors, r_1, without any information about the unobserved stationary factors. In the second step, we propose another information criterion to the resultant residuals to obtain the number of stationary factors, r_2. Below, we use r_1 and r_2 to denote a generic number of nonstationary factors and stationary factors, respectively. Their true values are denoted as r_1^0 and r_2^0, which are assumed to be bounded above by a finite integer $r_{\max}$.

In the first step, we estimate the number of unobserved nonstationary factors, r_1^0, consistently based on the level data. Let $F_1^{r_1}$ be a matrix of $T \times r_1$ nonstationary factors and $\lambda_{1i}^{r_1}$ be an $r_1 \times 1$ vector of nonstationary factor loadings. Let $\Lambda_1^{r_1\prime} = (\lambda_{11}^{r_1}, ..., \lambda_{1N}^{r_1})$. Given the preliminary consistent estimators of $\hat{\beta}_{1,i}$ and $\hat{\beta}_{2,i}$ based on $r_{\max}$ nonstationary factors, we consider the following minimization problem:

$$\left\{\hat{F}_1^{r_1}, \hat{\Lambda}^{r_1}\right\} = \arg\min_{\Lambda^{r_1}, F_1^{r_1}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{\beta}_{1,i}' x_{1,it} - \hat{\beta}_{2,i}' x_{2,it} - \lambda_{1i}^{r_1\prime} f_{1t}^{r_1})^2,$$
$$\text{s.t. } F_1^{r_1\prime} F_1^{r_1}/T^2 = I_{r_1} \text{ and } \Lambda_1^{r_1\prime}\Lambda_1^{r_1} \text{ is diagonal.}$$

Given $\hat{F}_1^{r_1} = (\hat{f}_{11}^{r_1}, ..., \hat{f}_{1T}^{r_1})'$, we can solve for $\hat{\Lambda}_1^{r_1} = (\hat{\lambda}_{11}^{r_1}, ..., \hat{\lambda}_{1N}^{r_1})'$ as a function of $\hat{F}_1^{r_1}$ by least squares regression. We suppress the dependence of $\hat{\Lambda}_1^{r_1}$ on $\hat{F}_1^{r_1}$ and define $V_1(r_1, \hat{F}_1^{r_1}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{\beta}_{1,i}' x_{1,it} - \hat{\beta}_{2,i}' x_{2,it} - \hat{\lambda}_{1i}^{r_1\prime} \hat{f}_{1t}^{r_1})^2$. Then we consider the information criterion:

$$IC_1(r_1) = \log V_1(r_1, \hat{F}_1^{r_1}) + r_1 g_1(N, T), \tag{3.13}$$

where $g_1(N,T)$ is a penalty function. Let $\hat{r}_1 = \arg\min_{0 \le r_1 \le r_{\max}} IC_1(r_1)$. We add the following condition.

Assumption 3.5 *As* $(N,T) \to \infty$, $g_1(N,T)\frac{\log\log(T)}{T} \to 0$ *and* $g_1(N,T) \to \infty$.

The conditions on $g_1(N,T)$ differ from the conventional conditions for the penalty function used in information criteria in the stationary framework (e.g., $g_2(N,T)$ in Assumption 3.6 below). In particular, we now require that $g_1(N,T)$ diverge to infinity rather than converge to zero. The intuition for this requirement is that the mean squared residual, $V_1(r_1, \hat{F}_1^{r_1})$, does not have a finite probability limit when the number of nonstationary common factors is under-specified. We can show that $\frac{\log\log T}{T} V_1(r_1, \hat{F}_1^{r_1})$ converges in probability to a positive constant when $0 \le r_1 < r_1^0$. By contrast, we have $V_1(r_1, \hat{F}_1^{r_1}) - V_1(r_1^0, \hat{F}_1^{r_1^0}) = O_P(1)$ when $r_1 > r_1^0$.

The following theorem shows that the use of $IC_1(r_1)$ determines r_1^0 consistently.

Theorem 3.6 *If Assumptions 3.1-3.3 and 3.5 hold, then* $P(\hat{r}_1 = r_1^0) \to 1$ *as* $(N,T) \to \infty$.

Once we obtain a consistent estimate of r_1^0, we can also obtain a consistent estimator of the number of unobserved stationary factors, r_2^0, from the resultant residuals based on standard methods in Bai and Ng (2002). In the second step, the resultant residual takes the form:

$$\hat{r}_{it} = y_{it} - \hat{\beta}_{1,i}' x_{1,it} - \hat{\beta}_{2,i}' x_{2,it} - \hat{\lambda}_{1i}' \hat{f}_{1t},\ t = 1, ..., T, \tag{3.14}$$

where $\hat{r}_{it} = \lambda_{2i}^{0\prime} f_{2t}^0 + u_{it} + v_{it}$ and v_{it} accounts for the asymptotically negligible estimation error from the early stages. Since the true dimension r_2^0 is unknown, we start with a model with $r_{\max}$ unobserved common factors. Let $F_2^{r_2}$ be a matrix of $T \times r_2$ nonstationary factors and $\lambda_{2i}^{r_2}$ be an $r_2 \times 1$ vector of nonstationary factor loadings. Let $\Lambda_2^{r_2\prime} = (\lambda_{21}^{r_2}, ..., \lambda_{2N}^{r_2})$.We consider the following minimization problem:

$$\left\{\hat{F}_2^{r_2}, \hat{\Lambda}^{r_2}\right\} = \arg \min_{\Lambda^{r_2}, F_2^{r_2}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{r}_{it} - \lambda_i^{r_2\prime} f_{2t}^{r_2})^2,$$
$$\text{s.t. } F_2^{r_2\prime} F_2^{r_2}/T^2 = I_{r_2} \text{ and } \Lambda_2^{r_2\prime}\Lambda_2^{r_2} \text{ is diagonal,}$$

where $\hat{F}_2^{r_2} = (\hat{f}_{21}^{r_2}, ..., \hat{f}_{2T}^{r_2})'$ and $\hat{\Lambda}_2^{r_2} = (\hat{\lambda}_{21}^{r_2}, ..., \hat{\lambda}_{2N}^{r_2})'$, and $\hat{\beta}_{1,i}$, $\hat{\beta}_{2,i}$, $\hat{\lambda}_{1i}'\hat{f}_{1t}$ are consistently estimated based on $\hat{r}_1$ nonstationary factors from the first step. It is easy to show that the $\hat{\beta}_{1,i}$ are T -consistent and $\hat{\beta}_{2,i}$ are $\sqrt{T}$ -consistent under appropriate orthogonality conditions, which suffices for our purpose. It is well known that given $\hat{F}_2^{r_2}$, we can solve $\hat{\Lambda}^{r_2} = \hat{\Lambda}^{r_2}(\hat{F}_2^{r_2})$ from the least squares regression as a function of $\hat{F}_2^{r_2}$. Then we can define $V_2(r_2, \hat{F}_2^{r_2}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{r}_{it} - \hat{\lambda}_{2i}^{r_2\prime}\hat{f}_{2t}^{r_2})^2$. Following Bai and Ng (2002) we consider the information criterion

$$IC_2(r) = \log V_2(r_2, \hat{F}_2^{r_2}) + r_2 g_2(N,T), \tag{3.15}$$

where $g_2(N,T)$ is a penalty function. Let $\hat{r}_2 = \arg\min_{0 \le r \le r_{\max}} IC_2(r)$. We add the next assumption.

Assumption 3.6 *As $(N,T) \to \infty$, $g_2(N,T) \to 0$ and $C_{NT}^2 g_2(N,T) \to \infty$, where $C_{NT} = \min(\sqrt{N}, \sqrt{T})$.*

Assumption 3.6 is common in the literature. It requires that $g_2(N,T)$ pass to zero at a certain rate so that both over- and under-fitted models can be eliminated asymptotically. The following theorem demonstrates that we can apply $IC_2(r_2)$ to estimate r_2^0 consistently.

Theorem 3.7 *If Assumptions 3.1-3.3 and 3.6 hold, then $P(\hat{r}_2 = r_2^0) \to 1$ as $(N,T) \to \infty$.*

In the simulations and applications, we simply follow Bai and Ng (2002) and Bai (2004) and set

$$g_1(N,T) = \alpha_T g_2(N,T) \text{ and } g_2(N,T) = \frac{N+T}{NT}\log\left(C_{NT}^2\right) \text{ or } \frac{N+T}{NT}\log\left(\frac{NT}{N+T}\right),$$

where $\alpha_T = \frac{T}{4\log\log T}$. We first estimate the number of unobserved nonstationary factors by $\hat{r}_1$ based on level data, and next estimate the number of unobserved stationary factors by $\hat{r}_2$ based on the resultant residuals from the first step.

3.7 Determination of the number of groups

We propose a BIC-type information criterion to determine the number of groups, K. We assume that the true number of groups, K_0, is bounded from above by a finite integer $K_{\max}$.

By minimizing the criterion function in (2.11), we obtain estimates $\hat{\beta}_{1,i}(K,\lambda)$, $\hat{\beta}_{2,i}(K,\lambda)$, $\hat{\alpha}_k(K,\lambda)$, $\hat{\lambda}_{1i}(K,\lambda)$, and $\hat{f}_{1t}(K,\lambda)$ of $\beta_{1,i}^0$, $\beta_{2,i}^0$, α_k^0, λ_{1i}^0, and f_{1t}^0, in which we make the dependence of the estimates $\hat{\beta}_{1,i}$, $\hat{\beta}_{2,i}$, $\hat{\alpha}_k$, $\hat{\lambda}_{1i}$, and $\hat{f}_{1t}$ on (K,λ) explicit. Let $\hat{G}_k(K,\lambda) = \{i \in \{1,2,...,N\} : \hat{\beta}_{1,i}(K,\lambda) =$

$\hat{\alpha}_k(K,\lambda)\}$ for $k=1,...,K$, and $\hat{G}(K,\lambda)=\{\hat{G}_1(K,\lambda),...,\hat{G}_K(K,\lambda)\}$. Let $\hat{\alpha}^{cup}_{\hat{G}_k(K,\lambda)}$ denote the Cup-Lasso estimate of α^0_k. Define

$$V_3(K)=\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k(K,\lambda)}\sum_{t=1}^{T}\left[y_{it}-\hat{\alpha}^{cup\prime}_{\hat{G}_k(K,\lambda)}x_{1,it}-\hat{\beta}^{\prime}_{2,i}x_{2,it}-\hat{\lambda}_i(K,\lambda)^{\prime}\hat{f}_t(K,\lambda)\right]^2.$$

Following SSP (2016a) and Lu and Su (2016), we consider the following information criterion:

$$IC_3(K,\lambda)=\log V_3(K)+pKg_3(N,T), \tag{3.16}$$

where $g_3(N,T)$ is a penalty function. Let $\hat{K}(\lambda)=\arg\min_{1\leq K\leq K_{\max}}IC_3(K,\lambda)$.

Let $\mathcal{G}^{(K)}=(G_{K,1},...,G_{K,K})$ be any K-partition of the set of individual index $\{1,2,...,N\}$. Define $\hat{\sigma}^2_{\mathcal{G}^{(K)}}=\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in G_{K,k}}\sum_{t=1}^{T}[y_{it}-\hat{\alpha}^{cup\prime}_{\hat{G}_k(K,\lambda)}x_{1,it}-\hat{\beta}^{\prime}_{2,i}x_{2,it}-\hat{\lambda}_{1i}(\ K,\lambda)^{\prime}\hat{f}_{1t}(K,\lambda)]^2$, where $\{\hat{\alpha}^{cup}_{G_{K,k}},\hat{\beta}_{2,i}(\mathcal{G}^{(K)}),\hat{\lambda}_{1i}(\mathcal{G}^{(K)}),\ \hat{f}_{1t}(\mathcal{G}^{(K)})\}$ is analogously defined as $\{\hat{\alpha}^{cup}_{\hat{G}_k(K,\lambda)},\ \hat{\beta}_{2,i}(K,\lambda),\hat{\lambda}_{1i}(K,\lambda),$ $\hat{f}_{1t}(K,\lambda)\}$ with $\{\hat{G}_k(K,\lambda)\}$ being replaced by $\{G_{K,k}\}$. Let $\sigma^2_0=\text{plim}_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{i\in G^0_k}\sum_{t=1}^{T}[y_{it}-\alpha^{0\prime}_kx_{1,it}-\beta^{0\prime}_{2,i}x_{2,it}-\lambda^{0\prime}_{1i}f^0_{1t}]^2$. Define

$$\nu_{NT}=\begin{cases}(NT)^{-1/2}\text{ when there are neither stationary regressors nor unobserved common factors,}\\ T^{-1/2}\text{ when there are stationary regressors but no unobserved common factors,}\\ N^{-1/2}\text{ when there are common nonstationary factors but no stationary factors or regressors,}\\ C^{-1}_{NT}\text{ in other cases.}\end{cases},$$

and note that ν_{NT} indicates the effect of estimating the nonstationary panel on the use of $IC_3(K,\lambda)$ under four different scenarios.

We add the following assumption.

Assumption 3.7 (i) *As* $(N,T)\to\infty$, $\min_{1\leq K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\hat{\sigma}^2_{\mathcal{G}^{(K)}}\xrightarrow{p}\underline{\sigma}^2>\sigma^2_0$.

(ii) As $(N,T)\to\infty$, $g_3(N,T)\to0$ and $g_3(N,T)/\nu^2_{NT}\to\infty$.

Assumption 3.7(i) requires that all under-fitted models yield asymptotic mean square errors larger than σ^2_0, which is delivered by the true model. Assumption 3.7(ii) imposes typical conditions on the penalty function $g_3(N,T)$, requiring that it cannot shrink to zero too fast or too slowly.

The following theorem justifies the validity of using IC_3 to determine the number of groups.

Theorem 3.8 *Suppose that Assumptions 3.1-3.4 and 3.7 hold. Then* $P(\hat{K}(\lambda)=K_0)\to1$ *as* $(N,T)\to\infty$.

Theorem 3.8 indicates that as long as λ satisfies Assumption 3.3(iv) and $g_3(N,T)$ satisfies Assumption 3.7(ii), we have $\inf_{1\leq K\leq K_{\max},K\neq K_0}IC_3(K,\lambda)>IC_3(K_0,\lambda)$ as $(N,T)\to\infty$. Consequently, the minimizer of $IC_3(K,\lambda)$ with respect to K equals K_0 w.p.a.1 for a variety of choices of λ. In practice, we can further choose λ over a finite grid of values to minimize $IC_3(\hat{K}(\lambda), \lambda)$. The next section provides details.

4 Monte Carlo Simulations

In this section we conduct simulations to evaluate the finite sample performance of the C-Lasso procedure, the bias-corrected post-Lasso, the fully-modified post-Lasso regression, and the Cup-Lasso estimators with and without unobserved factors, stationary regressors and incidental time trends. For comparison, we also consider the Lasso-type estimators using first-differenced data, which is proposed for stationary panels with interactive fixed effects. Note that the method proposed by Su and Ju (2018) requires the regressors to be predetermined. In general, their method is not suitable for the first-differenced data in panel cointegration models with both contemporaneous and serial correlations. Before estimation, we evaluate the performance of the information criteria for determining the number of unobserved common factors and groups.

4.1 Data generating processes

We consider five data generating processes (DGPs) with stationary and/or nonstationary unobserved common factors. The observations in each of these DGPs are drawn from three groups with $N_1 : N_2 : N_3 = 0.3 : 0.4 : 0.3$. There are four combinations of sample sizes, with $N = 50, 100$ and $T = 40, 80$. Data are generated based on the following design. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$,

$$\begin{cases} y_{it} = \mu_i + \rho_i t + \beta_{1,i}' x_{1,it} + \beta_{2,i}' x_{2,it} + c_1 \lambda_{1i}' f_{1t} + c_2 \lambda_{2i}' f_{2t} + u_{it} \\ x_{1,it} = \mu_{1,i} + x_{1,it-1} + \varepsilon_{it} \\ f_{1t} = \mu^{f_1} + f_{1,t-1} + \nu_t \end{cases} . \tag{4.1}$$

For DGPs 1-4 below, we do not allow for stationary regressors so that $p_2 = 0$ and $w_{it} \equiv (u_{it}, \varepsilon_{it}', \Delta f_{1t}', f_{2t}')'$ are generated from the linear process: $w_{it} = \sum_{j=0}^{\infty} \phi_{ij} v_{i,t-j}$, where $\phi_{ij} = L(j)\Omega^{1/2}$, $L(j) = 1$ or $j^{-3.5}$,
$$\Omega = \begin{pmatrix} 0.25 & \Omega_{12} & \Omega_{13} & 0_{1\times r_2} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} & \Omega_{24} \\ 0_{r_1\times 1} & 0_{r_1\times p_1} & \Omega_{33} & \Omega_{34} \\ 0_{r_2\times 1} & 0_{r_2\times p_1} & \Omega_{43} & \Omega_{44} \end{pmatrix},$$ $v_{it} = (v_{it}^{u\varepsilon\prime}, v_t^{f_1\prime}, v_t^{f_2\prime})'$, $v_{it}^{u\varepsilon} \sim$ i.i.d. $N(0, I_{p_1+1})$ for $i = 1, ..., N$, and $(v_t^{f_1\prime}, v_t^{f_2\prime})' \sim$ i.i.d. $N(0, I_{r_1+r_2})$. The factor loadings $\lambda_i = (\lambda_{1i}', \lambda_{2i}')'$ are i.i.d. $\lambda_i \sim N(\mu_\lambda, I_{r_1+r_2})$ and $\mu_\lambda = 0.1 \cdot \iota_{(r_1+r_2)\times 1}$ with ι_a an $a \times 1$ vector of ones. The long-run slope coefficients $\beta_{1,i}$ exhibit the group structure in (2.3) for $K = 3$ and the true values for the group-specific parameters are

$$(\alpha_1^0, \alpha_2^0, \alpha_3^0) = \left(\begin{pmatrix} 0.4 \\ 1.6 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1.6 \\ 0.4 \end{pmatrix} \right).$$

We allow for stationary regressors in DGP 5 and incidental linear time trends in DGP 6 below.

Endogeneity and serial correlations in the system are controlled by ϕ_{ij} and the non-zero block matrices in Ω. The parameters c_1 and c_2 control the importance of unobserved common factors. The estimates of long-run covariance matrices are obtained by using the Fejér kernel with the bandwidth set at 10.[4] The maximum number of iterations for Cup-Lasso regression is set to 20. All simulation results are obtained from 500 replications.

DGP 1. We consider a panel cointegration model with nonstationary regressors and unobserved stationary common factors such that $p_1 = 2$, $p_2 = 0$, $r_1 = 0$, and $r_2 = 2$. Let $c_2 = 0.5$, $\mu_i = \rho_i = c_1 = 0$, and $\mu_{1,i} = 0_{2\times1}$. There is neither contemporaneous correlation nor serial correlation among the errors where $L(j) = 1$ and $\Omega = \begin{pmatrix} 0.25 & 0_{1\times4} \\ 0_{4\times1} & I_4 \end{pmatrix}$.

DGP 2. The DGP is similar to DGP 1 except that we now introduce contemporaneous correlations among the errors by setting $\phi_{ij} = \Omega^{1/2}$ with $\Omega_{12} = \Omega_{21}' = \begin{pmatrix} 0.2, 0.2 \end{pmatrix}$, $\Omega_{24} = \begin{pmatrix} 0.2, 0.2 \\ 0.2, 0.2 \end{pmatrix}$ and $\Omega_{22} = \Omega_{44} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$.

DGP 3. We consider a panel latent factor cointegration model with both nonstationary regressors and unobserved nonstationary common factors, such that $p_1 = 2$, $p_2 = 0$, $r_1 = 2$, and $r_2 = 0$. Let $c_1 = 1$, $\mu_i = \rho_i = 0$, and $\mu_{1,i} = \mu^{f_1} = 0_{2\times1}$. We allow for general forms of weak dependence among the errors where $\phi_{ij} = j^{-3.5}\Omega^{1/2}$, $\Omega_{12} = \Omega_{21}' = \begin{pmatrix} 0.2, 0.2 \end{pmatrix}$, $\Omega_{23} = \begin{pmatrix} 0.2, 0.2 \\ 0.2, 0.2 \end{pmatrix}$ and $\Omega_{22} = \Omega_{33} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$.

DGP 4. We consider a panel latent factor cointegration model with both nonstationary regressors and mixed unobserved common factors such that $p_1 = 2$, $p_2 = 0$, $r_1 = 2$, and $r_2 = 1$. Let $c_1 = 1$, $c_2 = 0.5$, $\mu_i = \rho_i = 0$, and $\mu_{1,i} = \mu^{f_1} = 0_{2\times1}$. We allow for general forms of weak dependence among the errors where $\phi_{ij} = j^{-3.5}\Omega^{1/2}$, $\Omega_{12} = \Omega_{21}' = \Omega_{24}' = \Omega_{43} = \Omega_{43}' = \begin{pmatrix} 0.2, 0.2 \end{pmatrix}$, $\Omega_{23} = \begin{pmatrix} 0.2, 0.2 \\ 0.2, 0.2 \end{pmatrix}$, $\Omega_{22} = \Omega_{33} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$ and $\Omega_{44} = 1$. In addition, we allow for weak correlation among the factor loadings with $\lambda_i = (\lambda_{1i}', \lambda_{2i}')' \sim$ i.i.d. $N(0.1 \cdot \iota_3, \Omega_\lambda)$, where $\Omega_\lambda = \begin{pmatrix} 1 & 0 & 2/\sqrt{N} \\ 0 & 1 & 2/\sqrt{N} \\ 2/\sqrt{N} & 2/\sqrt{N} & 1 \end{pmatrix}$.

DGP 5. We consider a panel latent factor cointegration model with mixed regressors and mixed unobserved common factors such that $p_1 = 2$, $p_2 = 1$, $r_1 = 2$, and $r_2 = 1$. Let $c_1 = 1$, $c_2 = 0.5$, $\mu_i = \rho_i = 0$, and $\mu_{1,i} = \mu^{f_1} = 0_{2\times1}$. The settings of the errors are the same as in DGP 4. For the stationary regressors and associated coefficients, we generate $x_{2,it} \sim$ i.i.d. $N(0, 1)$ and

[4] Findings based on other kernels (the quadratic spectral kernel and Parzen kernel) and other choices of bandwidth are similar and are not reported

$\beta_{2,i} \sim N(0.5, 1)$.

DGP 6. We consider a panel latent factor cointegration model with unobserved nonstationary common factors and incidental deterministic trends such that $p_1 = r_1 = 2$ and $p_2 = r_2 = 0$. We set $c_1 = 1$. For the incidental time trends, we generate $(\mu_i, \rho_i, \mu_{1,i}', \mu^{f_1\prime})' \sim$ i.i.d. $N(0, I_6)$. The errors are generated as in DGP 3.

4.2 Estimating the number of unobserved factors

We assess the performance of the two information criteria proposed in Section 3.6 before determining the number of groups and running the PPC-based estimation procedure. We first obtain the preliminary time-series estimates of both nonstationary and stationary slope coefficients $\beta_{1,i}$ and $\beta_{2,i}$ by setting the number of nonstationary factors $r_1 = r_{\max}$. We choose the BIC-type penalty function $g_1(N,T) = \frac{T}{4 \log\log T} g_2(N,T)$ to determine the number (r_1) of unobserved nonstationary factors and $g_2(N,T) = \frac{N+T}{NT} \log(\frac{NT}{N+T})$ to determine the number (r_2) of unobserved stationary factors. Note that $r_1^0 = 0, 0,$ 2, 2, 2, and 2 for DGPs 1-6, respectively and $r_2^0 =$ 2, 2, 0, 1, 1, and 0 for DGPs 1-6, respectively.

Table 1 displays the probability that a particular factor number from 0 to 4 is selected according to the information criteria proposed for the level data and the resultant residual data based on 500 replications. For the level data, the precision for selecting the number of nonstationary factors generally increases and approaches 1 in all DGPs as both N and T become larger. For DGPs 3-6, the performance in the case of $N = 50$ and $T = 80$ slightly deteriorates in comparison with the case $N = 50$ and $T = 40$. Similar phenomenon may occur in the use of information criteria for stationary factor models.

For the resultant residual data, the probabilities for selecting the number of stationary factors are influenced by the results in nonstationary factors. In general, it preserves similar finite sample performance as the level data. As both N and T increase, the probabilities of selecting the number of stationary factors approach 1 in all DGPs. In general, the simulation results show that the two information criteria work fairly well in finite samples.

4.3 Determination of the number of groups

The results above show that the information criteria ($IC_1(r_1)$ and $IC_2(r_2)$) in Section 3.6 are useful in determining the number of nonstationary and stationary factors. We emphasize that these information criteria do not require the knowledge of the latent group structure or even the number of groups.

Next, we focus on the performance of the information criterion ($IC_3(K, \lambda)$) for determining the number of groups by assuming that the number of unobserved factors is known. We follow SSP (2016a) and set $g_3(N,T) = \frac{2}{3} \log(\min(N,T))/\min(N,T)$ and $\lambda = c_\lambda T^{-3/4}$ with $c_\lambda = 0.05$, 0.1, 0.2, 0.4. Note that $g_3(N,T)$ satisfies the two restrictions in Assumption 3.7. Due to space limitations, we

Table 1: Frequency for selecting $r_1, r_2 = 0, 1, 2, 3, 4$ nonstationary and stationary factors

			Level Data					Resultant Residual Data				
	N	T	$r_1 = 0$	$r_1 = 1$	$r_1 = 2$	$r_1 = 3$	$r_1 = 4$	$r_2 = 0$	$r_2 = 1$	$r_2 = 2$	$r_2 = 3$	$r_2 = 4$
DGP 1	50	40	1	0	0	0	0	0	0	1	0	0
	50	80	1	0	0	0	0	0	0	1	0	0
	100	40	1	0	0	0	0	0	0	1	0	0
	100	80	1	0	0	0	0	0	0	1	0	0
	1000	1000	1	0	0	0	0	0	0	1	0	0
DGP 2	50	40	1	0	0	0	0	0	0	1	0	0
	50	80	1	0	0	0	0	0	0	1	0	0
	100	40	1	0	0	0	0	0	0	1	0	0
	100	80	1	0	0	0	0	0	0	1	0	0
	1000	1000	1	0	0	0	0	0	0	1	0	0
DGP 3	50	40	0	0.014	0.93	0.054	0.002	0.984	0.014	0	0.002	0
	50	80	0.016	0.048	0.92	0.016	0	0.932	0.004	0.002	0	0.012
	100	40	0	0	0.998	0.002	0	1	0	0	0	0
	100	80	0	0	0.988	0.012	0	1	0	0	0	0
	1000	1000	0	0	1	0	0	1	0	0	0	0
DGP 4	50	40	0.004	0.074	0.908	0.014	0	0.002	0.920	0.042	0.014	0.014
	50	80	0.042	0.114	0.836	0.008	0	0	0.844	0	0.012	0.016
	100	40	0	0.008	0.988	0.004	0	0.002	0.990	0.006	0.002	0
	100	80	0	0.004	0.996	0	0	0	0.996	0.002	0.002	0
	1000	1000	0	0	1	0	0	0	1	0	0	0
DGP 5	50	40	0.010	0.086	0.892	0.012	0	0	0.900	0.052	0.016	0.012
	50	80	0.044	0.134	0.818	0.004	0	0	0.822	0.004	0.014	0.014
	100	40	0	0.008	0.984	0.008	0	0.004	0.988	0.008	0	0
	100	80	0	0.004	0.996	0	0	0	0.996	0.002	0	0.002
	1000	1000	0	0	1	0	0	0	1	0	0	0
DGP 6	50	40	0.004	0.022	0.974	0	0	0.974	0.02	0.004	0	0.002
	50	80	0.082	0.036	0.882	0	0	0.882	0.014	0.006	0.006	0.008
	100	40	0	0.002	0.998	0	0	0.998	0.002	0	0	0
	100	80	0	0	1	0	0	1	0	0	0	0
	1000	1000	0	0	1	0	0	1	0	0	0	0

Table 2: Frequency for selecting $K = 1, 2, ..., 6$ groups

	N	T	1	2	3	4	5	6
DGP 1	50	40	0	0	1	0	0	0
	50	80	0	0	1	0	0	0
	100	40	0	0	1	0	0	0
	100	80	0	0	1	0	0	0
DGP 2	50	40	0	0	0.992	0.008	0	0
	50	80	0	0	1	0	0	0
	100	40	0	0	0.996	0.004	0	0
	100	80	0	0	1	0	0	0
DGP 3	50	40	0	0	0.996	0.002	0.002	0
	50	80	0	0	0.996	0.002	0.002	0
	100	40	0	0	0.996	0.004	0	0
	100	80	0	0	1	0	0	0
DGP 4	50	40	0	0	0.99	0.01	0	0
	50	80	0	0	0.992	0.008	0	0
	100	40	0	0	0.996	0.004	0	0
	100	80	0	0	1	0	0	0
DGP 5	50	40	0	0	0.998	0.002	0	0
	50	80	0	0	1	0	0	0
	100	40	0	0	1	0	0	0
	100	80	0	0	0.996	0	0	0.004
DGP 6	50	40	0	0	1	0	0	0
	50	80	0	0	1	0	0	0
	100	40	0	0	1	0	0	0
	100	80	0	0	1	0	0	0

only report the outcomes for $c_\lambda = 0.1$ based on 500 replications for each DGP in Table 2 as the other choices of c_λ produce similar results. Recall that the true number of groups is 3 in all DGPs. Table 2 displays the probability that a particular group number from 1 to 6 is selected according to IC_3. The probabilities are higher than 99% in all cases and tend to unity when T increases to 80. This indicates good finite sample performance of the criterion IC_3 in determining the number of groups.

4.4 Classification and point estimation

We now examine the performance of classification and estimation when we have a priori knowledge of the numbers of groups and unobserved common factors. Table 3 compares finite sample performance between our estimators obtained from the level data and the estimators obtained from the first-differenced data for DGPs 1-2. The latter are obtained by implementing the method of Su and Ju (2018) for stationary models. Tables 4-5 report classification and point estimation results for DGPs 3-6 and check the sensitivity of classification and estimation performance for different λ's. Here, we set $\lambda = c_\lambda T^{-3/4}$ where $c_\lambda = \{0.05, 0.1, 0.2, 0.4\}$. Due to space constraints, we only report results for $c_\lambda = 0.1$ in DGPs 1-2 and $c_\lambda = 0.1, 0.2$ in DGPs 3-6. The focus of our analysis is the latent group patterns in nonstationary slope coefficients. For $\alpha_k = (\alpha_{1,k}, \alpha_{2,k})'$ we only report results for the estimation of the first nonstationary slope coefficient $\alpha_{1,k}$ in each DGP.

For comparison, Table 3 summarizes group classification and estimation results from both the level data and first-differenced data. Tables 4-5 only report the corresponding results for the level data. Columns 4 and 9 in Table 3 and Columns 4 and 8 in Tables 4-5 report the percentage of correct classification over the N cross-section units, calculated as $\frac{1}{N}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k}\mathbf{1}\{\beta_{1,i}^0 = \alpha_k^0\}$, averaged over the 500 replications. Columns 5-7 and 10-11 in Table 3 and Columns 5-7 and 9-11 in Tables 4-5 summarize estimation performance in terms of root-mean-squared error (RMSE), bias (Bias), and 95% coverage probability (% coverage). For simplicity, we define the weighted average RMSE as $\frac{1}{N}\sum_{k=1}^{K} N_k \mathrm{RMSE}(\hat{\alpha}_{1,k})$ with $\hat{\alpha}_{1,k}$ being the estimate of $\alpha_{1,k}$. We define the weighted average bias and 95% coverage probability analogously. For comparison, we report the estimation and inference results based on the estimates of the C-Lasso, bias-corrected post-Lasso, fully-modified post-Lasso and Cup-Lasso methods defined in Section 3.4. We also report estimation and inference results for the oracle estimates that are obtained by utilizing the true group structures $\{G_k^0\}$.

For brevity, we only summarize the main findings in Tables 3. First, when there is no endogeneity issue in DGP 1, both level data and first-differenced data lead to consistent estimation and there is no bias in the C-Lasso estimation. In terms of RMSEs, there is a considerable convergence rate advantage to use level data, where the estimators of the nonstationary slope coefficients enjoy super-consistency—$\sqrt{N}T$-consistency, which is in contrast with the $\sqrt{NT}$-consistency of the estimators in the first-differenced model. The correct classification results generally approach 100% in both cases. Second, when there is endogeneity in DGP 2, the first-differencing approach does not lead to consistent estimation. For the first-differenced data, there is no evidence of consistency in terms of RMSE and Bias. However, the PPC-based estimators obtained from the level data generally show good finite sample performance with the bias of the C-Lasso estimator being approximately halved as T doubles.

The classification and estimation are reported in Tables 4-5 below and will now be discussed. In these tables, we first notice that the results with different c_λ's are similar, indicating some robustness in our algorithm to the choice of the tuning parameter λ. Second, the correct classification percentage approaches 100% when T increases. As expected, the correct classification percentages for the Cup-Lasso estimates are higher than those of the C-Lasso and post-Lasso estimates in all cases. This outcome suggests that iterations do help in finite samples to achieve better classification. Third, regarding parameter estimation Tables 3-5 show that the fully-modified procedure works slightly better than the direct bias-correction procedure. For DGP 2, the endogeneity bias issue is not very serious in the C-Lasso estimate since we only introduce contemporaneous correlation among the errors, nonstationary regressors, and stationary common factors. The two post-Lasso estimates and the Cup-Lasso estimates are found to perform as well as the oracle estimates in terms of RMSE, bias

Table 3: Classification and point estimation of α_1 for DGPs 1-2

			Level Data				First-Differenced Data			
N	T		% Correct classification	RMSE	Bias	%Coverage		% Correct classification	RMSE	Bias
DGP 1										
50	40	C-Lasso	99.98	0.0085	0.0002	90.90	C-Lasso	99.92	0.0340	-0.0007
		post-Lasso^{bc}	99.98	0.0083	0.0002	90.42	C-Lasso BC	99.92	0.0340	-0.0007
		post-Lasso^{fm}	99.98	0.0083	0.0002	90.24	Post-Lasso	99.92	0.0309	-0.0006
		Cup-Lasso	99.98	0.0083	0.0002	90.24	Post-Lasso BC	99.92	0.0308	-0.0006
		Oracle	-	0.0082	0.0002	90.18	Oracle	-	0.0309	-0.0004
50	80	C-Lasso	100.00	0.0040	0.0001	91.90	C-Lasso	100.00	0.0223	0.0002
		post-Lasso^{bc}	100.00	0.0040	0.0001	91.60	C-Lasso BC	100.00	0.0223	0.0002
		post-Lasso^{fm}	100.00	0.0040	0.0001	91.52	Post-Lasso	100.00	0.0207	0.0004
		Cup-Lasso	100.00	0.0040	0.0001	91.52	Post-Lasso BC	100.00	0.0207	0.0004
		Oracle	-	0.0040	0.0001	90.68	Oracle	-	0.0207	0.0004
100	40	C-Lasso	99.99	0.0057	-0.0001	93.14	C-Lasso	99.95	0.0234	0.0004
		post-Lasso^{bc}	99.99	0.0056	0.0000	92.82	C-Lasso BC	99.95	0.0234	0.0004
		post-Lasso^{fm}	99.99	0.0056	0.0000	93.06	Post-Lasso	99.95	0.0202	0.0004
		Cup-Lasso	99.99	0.0056	0.0000	93.06	Post-Lasso BC	99.95	0.0202	0.0004
		Oracle	-	0.0056	0.0000	93.66	Oracle	-	0.0202	0.0004
100	80	C-Lasso	100.00	0.0029	-0.0001	92.04	C-Lasso	100.00	0.0162	0.0011
		post-Lasso^{bc}	100.00	0.0028	0.0000	93.08	C-Lasso BC	100.00	0.0162	0.0011
		post-Lasso^{fm}	100.00	0.0028	0.0000	93.08	Post-Lasso	100.00	0.0142	0.0010
		Cup-Lasso	100.00	0.0028	0.0000	93.08	Post-Lasso BC	100.00	0.0142	0.0010
		Oracle	-	0.0028	0.0000	93.08	Oracle	-	0.0142	0.0010
DGP 2										
50	40	C-Lasso	99.98	0.0098	0.0054	83.42	C-Lasso	99.75	0.0981	0.0918
		post-Lasso^{bc}	99.98	0.0081	0.0004	91.12	C-Lasso BC	99.75	0.0980	0.0918
		post-Lasso^{fm}	99.98	0.0080	0.0005	91.00	Post-Lasso	99.75	0.0974	0.0922
		Cup-Lasso	99.98	0.0080	0.0005	91.00	Post-Lasso BC	99.75	0.0974	0.0922
		Oracle	-	0.0079	0.0005	91.00	Oracle	-	0.0976	0.0924
50	80	C-Lasso	100.00	0.0048	0.0026	84.12	C-Lasso	99.99	0.0974	0.0947
		post-Lasso^{bc}	100.00	0.0039	0.0001	91.32	C-Lasso BC	99.99	0.0974	0.0947
		post-Lasso^{fm}	100.00	0.0038	0.0002	92.04	Post-Lasso	99.99	0.0972	0.0949
		Cup-Lasso	100.00	0.0038	0.0002	92.04	Post-Lasso BC	99.99	0.0972	0.0949
		Oracle	-	0.0038	0.0002	92.04	Oracle	-	0.0972	0.0948
100	40	C-Lasso	99.97	0.0075	0.0050	79.48	C-Lasso	99.79	0.0960	0.0931
		post-Lasso^{bc}	99.97	0.0056	0.0002	92.30	C-Lasso BC	99.79	0.0960	0.0931
		post-Lasso^{fm}	99.97	0.0055	0.0003	92.60	Post-Lasso	99.79	0.0962	0.0940
		Cup-Lasso	99.97	0.0055	0.0003	92.60	Post-Lasso BC	99.79	0.0962	0.0940
		Oracle	-	0.0054	0.0002	92.60	Oracle	-	0.0961	0.0938
100	80	C-Lasso	100.00	0.0037	0.0024	80.04	C-Lasso	100.00	0.0969	0.0955
		post-Lasso^{bc}	100.00	0.0028	0.0000	92.24	C-Lasso BC	100.00	0.0969	0.0955
		post-Lasso^{fm}	100.00	0.0027	0.0001	92.60	Post-Lasso	100.00	0.0969	0.0958
		Cup-Lasso	100.00	0.0027	0.0001	92.60	Post-Lasso BC	100.00	0.0969	0.0958
		Oracle	-	0.0027	0.0001	92.60	Oracle	-	0.0969	0.0958

and coverage probability. For DGPs 3-6, the performance of the C-Lasso estimates is poorer due to the presence of unobserved nonstationary common factors. In addition, the Cup-Lasso estimates generally outperform the two post-Lasso estimates due to the updated group classification results. In DGP 5, we show that the presence of stationary regressors does not affect the finite sample performance of our estimates for nonstationary slope coefficients. We introduce incidental time trends in DGP 6 and show that our PPC-based estimation procedure work fairly well with the detrended data. In addition, the finite sample performance of the long-run estimates preserves similar patterns. In general, the finite sample performance of the Cup-Lasso estimators is close to that of the oracle estimates, which corroborates the oracle efficiency of the Cup-Lasso estimates. Accordingly, we recommend for practical implementation the use of Cup-Lasso estimates for both estimation and inference.

5 An Empirical Application to the Growth Convergence Puzzle

A longstanding leading question in the economic growth literature is whether national economies exhibit convergence across countries over time. A benchmark model in the literature is the international R&D spillover model proposed by Coe and Helpman (1995) who empirically identified positive technology spillover effects. Since technological progress is a primary source of economic growth, positive R&D spillovers are regarded as a force of convergence that activates through the channel of technology catch-up. Notwithstanding the strength and relevance of this argument, two potential problems have been identified in the Coe and Helpman study. First, the study fails to distinguish two distinct types of spillover effects: positive technology spillovers and negative market rivalry effects (Bloom et al., 2013). Second, the research does not account for unobserved common patterns across countries, such as financial crisis shocks and technological progress. These two issues may lead to biased or even inconsistent estimates for the parameters of interest – see, e.g., Griffith and Reenen (2004), Coe et al. (2009, CHH hereafter), and Ertur and Musolesi (2017).

In this section we apply our model and methodology to re-investigate this issue by allowing for heterogeneous convergence behavior through the channel of technology diffusion and unobserved common patterns across countries. In particular, we impose latent group structures on the long-run relationships between technological change, domestic R&D stock, foreign R&D stock, and human capital, at the same time capturing any common patterns of behavior via the use of unobserved factors. Interestingly, we find two directions of R&D spillover – positive technology spillovers and negative market rivalry effects, which help to explain the economic convergence puzzle through the channel of technology growth.

5.1 International R&D spillover model

We introduce two linear specifications for the international R&D spillover model. Following the standard growth literature, we define the total factor productivity (TFP) as the Solow residual,

Table 4: Classification and point estimation of α_1 for DGPs 3-4

		c_λ	0.1				0.2			
N	T		% Correct classification	RMSE	Bias	% Coverage	% Correct classification	RMSE	Bias	% Coverage
DGP 3										
50	40	C-Lasso	98.42	0.0420	0.0155	65.36	98.26	0.0443	0.0143	65.88
		post-Lasso^{bc}	98.42	0.0305	0.0028	91.62	98.26	0.0311	0.0029	91.74
		post-Lasso^{fm}	98.42	0.0305	0.0028	92.20	98.26	0.0311	0.0030	92.14
		Cup-Lasso	100.00	0.0112	0.0021	90.28	99.98	0.0112	0.0021	90.28
		Oracle	-	0.0110	0.0021	90.28	-	0.0110	0.0021	90.28
50	80	C-Lasso	99.34	0.0283	0.0072	60.60	99.31	0.0285	0.0073	60.44
		post-Lasso^{bc}	99.34	0.0188	0.0009	91.34	99.31	0.0173	0.0014	91.74
		post-Lasso^{fm}	99.34	0.0188	0.0014	91.28	99.31	0.0172	0.0018	91.62
		Cup-Lasso	100.00	0.0050	0.0009	90.44	100.00	0.0050	0.0009	90.44
		Oracle	-	0.0050	0.0009	90.44	-	0.0050	0.0009	90.44
100	40	C-Lasso	98.66	0.0281	0.0135	52.88	98.49	0.0300	0.0125	54.64
		post-Lasso^{bc}	98.66	0.0225	0.0027	89.72	98.49	0.0222	0.0033	89.86
		post-Lasso^{fm}	98.66	0.0226	0.0027	90.10	98.49	0.0223	0.0034	90.26
		Cup-Lasso	100.00	0.0073	0.0025	89.78	99.98	0.0073	0.0025	89.78
		Oracle	-	0.0073	0.0025	89.78	-	0.0073	0.0025	89.78
100	80	C-Lasso	99.41	0.0184	0.0069	49.68	99.38	0.0194	0.0064	48.78
		post-Lasso^{bc}	99.41	0.0188	0.0009	92.72	99.38	0.0190	0.0009	92.84
		post-Lasso^{fm}	99.41	0.0188	0.0014	93.08	99.38	0.0190	0.0013	93.20
		Cup-Lasso	100.00	0.0035	0.0010	93.12	100.00	0.0035	0.0010	93.12
		Oracle	-	0.0035	0.0010	93.12	-	0.0035	0.0010	93.12
DGP 4										
50	40	C-Lasso	98.22	0.0479	0.0145	70.70	98.07	0.0511	0.0133	71.44
		post-Lasso^{bc}	98.22	0.0337	0.0022	91.64	98.07	0.0335	0.0020	91.48
		post-Lasso^{fm}	98.22	0.0338	0.0024	91.44	98.07	0.0335	0.0022	91.18
		Cup-Lasso	99.97	0.0137	0.0015	89.98	99.93	0.0137	0.0015	90.10
		Oracle	-	0.0136	0.0015	89.96	-	0.0136	0.0015	89.96
50	80	C-Lasso	99.10	0.0454	0.0089	67.04	99.09	0.0451	0.0082	65.94
		post-Lasso^{bc}	99.10	0.0310	0.0008	91.52	99.09	0.0313	0.0007	91.40
		post-Lasso^{fm}	99.10	0.0310	0.0012	91.14	99.09	0.0313	0.0012	91.02
		Cup-Lasso	100.00	0.0065	0.0007	90.58	100.00	0.0065	0.0007	90.58
		Oracle	-	0.0065	0.0007	90.58	-	0.0065	0.0007	90.58
100	40	C-Lasso	98.44	0.0319	0.0140	62.60	98.28	0.0355	0.0130	62.82
		post-Lasso^{bc}	98.44	0.0277	0.0024	91.16	98.28	0.0282	0.0021	90.92
		post-Lasso^{fm}	98.44	0.0279	0.0026	90.94	98.28	0.0283	0.0023	90.72
		Cup-Lasso	99.97	0.0095	0.0021	91.12	99.94	0.0096	0.0021	91.22
		Oracle	-	0.0095	0.0021	91.12	-	0.0095	0.0021	91.12
100	80	C-Lasso	99.45	0.0198	0.0073	56.66	99.43	0.0216	0.0070	56.32
		post-Lasso^{bc}	99.45	0.0167	0.0007	92.62	99.43	0.0165	0.0006	92.66
		post-Lasso^{fm}	99.45	0.0167	0.0011	92.70	99.43	0.0165	0.0011	92.88
		Cup-Lasso	100.00	0.0047	0.0006	93.00	100.00	0.0047	0.0006	93.00
		Oracle	-	0.0047	0.0006	93.00	-	0.0047	0.0006	93.00

Table 5: Classification and point estimation of α_1 for DGPs 5-6

		c_λ	0.1				0.2			
N	T		% Correct classification	RMSE	Bias	% Coverage	% Correct classification	RMSE	Bias	% Coverage
DGP 5										
50	40	C-Lasso	98.01	0.0538	0.0165	63.74	97.78	0.0585	0.0150	63.60
		post-Lasso^{bc}	98.01	0.0365	0.0028	91.46	97.78	0.0391	0.0033	91.28
		post-Lasso^{fm}	98.01	0.0364	0.0029	91.96	97.78	0.0390	0.0034	91.78
		Cup-Lasso	99.98	0.0112	0.0026	90.80	99.96	0.0114	0.0025	90.80
		Oracle		0.0111	0.0026	90.72		0.0111	0.0026	90.72
50	80	C-Lasso	99.33	0.0254	0.0074	60.48	99.31	0.0278	0.0071	60.28
		post-Lasso^{bc}	99.33	0.0223	0.0009	91.66	99.31	0.0220	0.0009	91.86
		post-Lasso^{fm}	99.33	0.0223	0.0014	92.12	99.31	0.0219	0.0014	92.32
		Cup-Lasso	100.00	0.0051	0.0010	91.92	100.00	0.0051	0.0010	91.92
		Oracle		0.0051	0.0010	90.98		0.0051	0.0010	90.98
100	40	C-Lasso	98.72	0.0292	0.0133	53.10	98.57	0.0309	0.0126	54.30
		post-Lasso^{bc}	98.72	0.0245	0.0029	89.00	98.57	0.0252	0.0032	89.32
		post-Lasso^{fm}	98.72	0.0246	0.0031	89.44	98.57	0.0252	0.0034	89.80
		Cup-Lasso	100.00	0.0076	0.0027	89.64	99.99	0.0076	0.0027	89.64
		Oracle		0.0076	0.0027	90.68		0.0076	0.0027	90.68
100	80	C-Lasso	99.34	0.0184	0.0075	48.18	99.29	0.0203	0.0068	49.24
		post-Lasso^{bc}	99.34	0.0177	0.0008	91.04	99.29	0.0187	0.0008	91.06
		post-Lasso^{fm}	99.34	0.0178	0.0013	91.44	99.29	0.0187	0.0013	91.40
		Cup-Lasso	100.00	0.0036	0.0011	91.54	100.00	0.0036	0.0011	91.54
		Oracle		0.0036	0.0011	91.54		0.0036	0.0011	91.54
DGP 6										
50	40	C-Lasso	99.90	0.0322	0.0244	61.72	99.90	0.0308	0.0227	64.06
		post-Lasso^{bc}	99.90	0.0233	-0.0100	87.76	99.90	0.0233	-0.0100	87.76
		post-Lasso^{fm}	99.90	0.0176	0.0014	91.42	99.90	0.0177	0.0014	91.42
		Cup-Lasso	99.99	0.0172	0.0015	91.40	99.99	0.0172	0.0014	91.40
		Oracle		0.0172	0.0014	89.08		0.0172	0.0014	89.08
50	80	C-Lasso	99.98	0.0167	0.0128	62.20	99.98	0.0164	0.0119	63.64
		post-Lasso^{bc}	99.98	0.0125	-0.0079	86.18	99.98	0.0125	-0.0079	86.18
		post-Lasso^{fm}	99.98	0.0082	0.0008	93.58	99.98	0.0082	0.0008	93.58
		Cup-Lasso	100.00	0.0081	0.0009	93.58	100.00	0.0081	0.0009	93.58
		Oracle		0.0081	0.0009	91.54		0.0081	0.0009	91.54
100	40	C-Lasso	99.94	0.0277	0.0236	41.88	99.94	0.0264	0.0222	45.64
		post-Lasso^{bc}	99.94	0.0176	-0.0103	82.98	99.94	0.0175	-0.0102	83.10
		post-Lasso^{fm}	99.94	0.0122	0.0013	93.42	99.94	0.0122	0.0014	93.42
		Cup-Lasso	100.00	0.0120	0.0013	93.30	99.99	0.0120	0.0013	93.24
		Oracle		0.0120	0.0013	91.68		0.0120	0.0013	91.68
100	80	C-Lasso	99.94	0.0125	0.0097	50.38	99.94	0.0124	0.0093	51.56
		post-Lasso^{bc}	99.94	0.0104	-0.0077	73.66	99.94	0.0104	-0.0077	73.66
		post-Lasso^{fm}	99.94	0.0061	0.0004	93.50	99.94	0.0060	0.0004	93.44
		Cup-Lasso	100.00	0.0050	0.0004	93.44	100.00	0.0050	0.0004	93.44
		Oracle		0.0050	0.0004	95.06		0.0050	0.0004	95.06

which is often regarded as a measure of technology change. That is, $\log(TFP) = \log(Y) - \theta \log(K) - (1-\theta)\log(L)$, where Y, L, and K denotes final output, labor force, capital stock, respectively, and θ is the share of capital in GDP. In the first place, domestic R&D investment is a major source of technology change that stimulates innovation. Second, trade in intermediate goods enables a country to gain access to inputs available throughout the rest of the world. In this respect, foreign R&D stocks from a country's trading partners affect TFP by directly enhancing the transfer of R&D. Coe and Helpman (1995) empirically identify two sources of technology growth – innovation and catch-up effects – by running the following regression:

$$\log(F_{it}) = \mu_i + \beta^d \log(s_{it}^d) + \beta^f \log(s_{it}^f) + u_{it},$$

where i is the country index, t is the year index, μ_i are the unobserved individual fixed effects, F is total factor productivity, s^d is real domestic R&D capital stock, and s^f is real foreign R&D capital stock. We follow their specification on the international R&D spillover model and introduce unobserved common patterns to obtain

$$\log(F_{it}) = \beta_i^d \log(s_{it}^d) + \beta_i^f \log(s_{it}^f) + \lambda_i' f_t + u_{it}, \tag{5.1}$$

where f_t denotes the unobserved technology trends or global financial shocks, and the fixed effects μ_i are absorbed into the factor structure. We shall assume that the slope vector $\beta_i = (\beta_i^d, \beta_i^f)'$ exhibits the latent group structures studied in this paper. This specification is important because the latent group structures on β_i^f allow us to study the two types of spillover effects discussed above – positive technology spillovers and negative market rivalry effects, respectively.

In addition, we consider the following specification

$$\log(F_{it}) = \beta_i^d \log(s_{it}^d) + \beta_i^f \log(s_{it}^f) + \beta_i^h \log(h_{it}) + \lambda_i' f_t + u_{it}. \tag{5.2}$$

where h_{it} denotes human capital for country i in year t. Human capital accounts for innovation outside the R&D sector and other aspects of human capital not captured by formal R&D. Engelbrecht (1997) finds that human capital affects TFP directly as a factor of production and as a channel for international technology diffusion associated with catch-up effects across countries. As above, we allow the slope vector $\beta_i = (\beta_i^d, \beta_i^f, \beta_i^h)'$ to exhibit latent group structures.

CHH further extend the analysis to include institutional variables. In particular, they use various proxies for institutions to test if the estimated parameters on domestic and foreign R&D capital and on human capital vary among countries. For example, they first define the dummy variables (*high* and *low*) for some institutional variables and then consider their interaction with $\log(s_{it}^d)$ in order to provide sub-sample regression results for the above two specifications. Their results suggest that institutional differences introduce heterogeneous impacts on both innovation effects and R&D spillovers. In general, CHH employ observed institution variables to group countries into different subsamples and reveal heterogeneous degrees of R&D spillover effects from institutional differences.

Instead of using observed characteristics, such as institution variables, our PPC-based method allows us to analyze parameter heterogeneity empirically by encouraging the data to reveal latent features that may not be immediately apparent. In particular, the latent group structures on slope coefficients allow us to study potentially different impacts of innovation and catch-up effects. We can also analyze two opposite spillover effects — positive technology spillovers and negative market rivalry effects, respectively. These features of the methodology help us explain the growth convergence puzzle by means of different aspects of technological diffusion.

5.2 Data

We use the same dataset as CHH. This dataset is similar to that used in Coe and Helpman (1995) and is expanded to include two more countries and annual observations. It contains observations for $\log(F_{it})$, $\log(s^d_{it})$, $\log(s^f_{it})$, and $\log(h_{it})$ for 24 OECD countries from 1971-2004. The bilateral import-weighted R&D variable S^{f-biw} from trading partners is a measure of foreign R&D stock. Human capital is measured by years of schooling. In CHH, the relevant variables are pre-tested for unit roots and cointegration. All variables we consider have a unit root, i.e., all are non-stationary. We refer the readers directly to CHH for details on the definition and construction of these variables, and for summary statistics of the data.

5.3 Empirical results

We first determine the number of unobserved factors and the number of groups as was done in the simulation exercises. Then we report the results for the estimation of the group structures and group-specific parameters.

5.3.1 Estimation of the number of factors

Before running the PPC-based estimation procedure, we employ the information criteria IC_1 and IC_2 in Section 3.6 to estimate the number of unobserved factors. Following the simulation design, we set $g_1(N,T) = \frac{T}{4\log\log T} g_2(N,T)$ and $g_2(N,T) = c\frac{N+T}{NT}\log(\frac{NT}{N+T})$. Based on the results for level data and resultant residuals, we obtain the estimates $\hat{r}_1 = 1$ and $\hat{r}_2 = 0$. That is, we find a single nonstationary common factor and zero stationary common factors in the data. We fix $r_1 = 1$ and $r_2 = 0$ in the following empirical analysis.

5.3.2 Determination of the number of groups

As in the simulations, we set $g_3(N,T) = \frac{2}{3}\log(\min(N,T))/\min(N,T)$ and $\lambda = c_\lambda T^{-3/4}$. We use the following tuning parameter settings: $c_\lambda = 0.1, 0.2, 0.4, 0.6, 0.8$. Table 7 reports the information criterion IC_3 as a function of the number of groups under these tuning parameters. Following the majority rule, we find that the information criterion suggests three groups for both model (5.1) and

Table 6: Information criterion for the determination of the number of groups

	Model (5.1)					Model (5.2)				
$K \setminus c_\lambda$	0.1	0.2	0.4	0.6	0.8	0.1	0.2	0.4	0.6	0.8
1	-4.830	-4.807	-4.790	-4.776	-4.773	-4.680	-4.668	-4.671	-4.671	-4.669
2	**-6.387**	-5.545	-5.366	-5.234	-5.210	-4.671	-4.655	-4.430	-4.430	-4.429
3	-6.259	**-6.235**	**-6.229**	**-6.206**	**-6.213**	**-4.871**	**-5.058**	**-4.869**	**-4.835**	-4.218
4	-6.072	-6.099	-6.090	-6.177	-6.116	-4.865	-4.759	-4.783	-4.572	**-4.784**
5	-5.957	-5.974	-5.896	-5.951	-5.861	-4.528	-4.631	-4.526	-4.720	-4.137
6	-5.785	-5.706	-5.757	-5.814	-5.807	-4.255	-4.398	-4.261	-4.158	-3.701

model (5.2). Note that IC_3 achieves the minimal values for both model specifications when $c_\lambda = 0.2$. Therefore, we set $K = 3$ and $c_\lambda = 0.2$ in subsequent analyses.

5.3.3 Estimation results

For both model specifications, we employ the pooled fully modified OLS (FM-OLS) estimates under the homogeneity assumption and the Cup-Lasso estimates with one unobserved nonstationary common factor. Note that we also allow for one unobserved nonstationary factor to obtain the FM-OLS estimates. Table 6 reports the main results for these two estimates along with the fixed effects estimates of CHH.

In model (5.1), we have two explanatory variables ($\log(s^d)$ and $\log(s^f)$). We summarize some of the more interesting findings from Table 7. First, a comparison between the estimates in CHH and those obtained by pooled FM-OLS suggests that the estimate of the coefficient of $\log(s^d)$ in CHH is similar to our pooled FM-OLS estimate, whereas the estimate of the coefficient of $\log(s^f)$ decreases substantially after introducing one unobserved nonstationary factor in the model. This seems to suggest that direct spillover effects are partially offset by unobserved global technology patterns. Noting that our asymptotic variance estimation allows for both serial correlation and heteroskedasticity and appears more conservative than that of CHH, this difference explains why the standard errors (s.e.) of our estimates are much larger than those in CHH. Second, once we allow for latent group structures among the slope coefficients, our PPC estimation helps to identify quite different behavior in the estimates of the effects of both domestic R&D stock and foreign R&D stock: for Group 1, we observe the largest effect of domestic R&D stock, but the estimate on foreign R&D is negative; for Groups 2 and 3, the coefficient estimates on both domestic and foreign R&D stocks are positive. In addition, both estimates for Group 2 are larger than those for Group 3, but the estimates of the coefficient of foreign R&D stocks in Groups 2 and 3 are not statistically significant even at the 10% level.

The above findings from our PPC estimate have some interesting implications. First, the negative estimate on foreign R&D in Group 1 indicates that negative market rivalry effects dominate the technology spillovers for countries inside Group 1. Therefore, technology change in those countries relies mainly on innovations from domestic R&D stock. Moreover, this result implies that countries

Table 7: PPC estimation results

	Model (5.1)				
Slope coefficients	Pooled CHH2009	Pooled FM-OLS	Group 1 Cup-Lasso	Group 2 Cup-Lasso	Group 3 Cup-Lasso
$\log(s^d)$	0.095***	0.099***	0.289***	0.101***	0.058**
	(0.005)	(0.027)	(0.046)	(0.023)	(0.028)
$\log(s^f)$	0.213***	0.121***	-0.147***	0.120	0.086
	(0.014)	(0.044)	(0.057)	(0.099)	(0.068)
	Model (5.2)				
Slope coefficients	Pooled CHH2009	Pooled FM-OLS	Group 1 Cup-Lasso	Group 2 Cup-Lasso	Group 3 Cup-Lasso
$\log(s^d)$	0.098***	0.054**	0.464***	0.055***	-0.104***
	(0.016)	(0.023)	(0.064)	(0.021)	(0.027)
$\log(s^f)$	0.035***	0.121**	-0.413**	0.022	0.219***
	(0.011)	(0.048)	(0.138)	(0.061)	(0.063)
$\log(h)$	0.725***	0.615***	1.405**	0.550***	0.567***
	(0.087)	(0.138)	(0.564)	(0.158)	(0.130)

Note: Standard errors are in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

in Group 1 do not favor convergence through the technological change channel. We call this the "*Divergence*" group. Second, technology change for countries in Group 2 comes from balanced sources – the innovation effects from domestic R&D stock and the catch-up effects from technology spillovers, and interestingly, the magnitudes of those estimates are similar. From this perspective, countries in Group 2 favor the growth convergence hypothesis. We refer to this group as the "*Balance*" group. Last, the technology change in Group 3 is mainly determined by foreign R&D stock and we refer to Group 3 as the "*Convergence*" group, which also favors the growth convergence hypothesis.

In model (5.2), we introduce an additional regressor – human capital, which is regarded as another source of technology change. Our results from the pooled FM-OLS estimates confirm that human capital is one of the main sources of productivity growth and there exist direct technology spillovers in the full sample. When using our PPC estimation methods, we find similar heterogeneous behavior for model (5.2) as that for model (5.1). We can still classify countries into three groups and define them as groups of *Divergence*, *Balance-Human capital*, and *Convergence*, respectively. For the Divergence group (Group 1), technology growth relies on innovations and human capital and countries in Group 1 suffer from strong negative market rivalry effects. For Group 2, referred to as Balance-Human capital, the estimates of the effect of foreign R&D are not significant at the 10% level, and technology growth still benefits from the innovations and indirect catch-up effects from human capital. For Group 3, referred to as Convergence, countries benefit directly from the dominating technology spillovers. In general, the divergence behavior is more statistically significant than the convergence behavior.

Table 8: Group classification results

Model (5.1)				
Group 1 "Divergence" ($N_1 = 7$)				
Austria	Denmark	France	Germany	New Zealand
Norway	United States			
Group 2 "Balance" ($N_2 = 7$)				
Canada	Ireland	Israel	South Korea	Netherlands
Portugal	United Kingdom			
Group 3 "Convergence" ($N_3 = 10$)				
Australia	Belgium	Finland	Greece	Iceland
Italy	Japan	Spain	Sweden	Switzerland
Model (5.2)				
Group 1 "Divergence " ($N_1 = 2$)				
Ireland	United States			
Group 2 "Balance–Human capital " ($N_2 = 16$)				
Austria	Belgium	Denmark	Finland	Iceland
Israel	Italy	Japan	South Korea	Netherlands
New Zealand	Norway	Portugal	Spain	Sweden
Switzerland				
Group 3 "Convergence" ($N_3 = 6$)				
Australia	Canada	France	Germany	Greece
United Kingdom				

5.3.4 Classification results

Table 8 reports the group classification results. We summarize several interesting findings. First, based on the results for model (5.1), there are typically two types of countries in the Divergence group – "Leaders" and "Losers". Countries like France, Germany, the United States are already at the global technology frontiers, and they own 61.1% of R&D stock in our sample. By contrast, the remaining countries in Group 1 account for only 1.5% of R&D stock in our sample. Second, most OECD countries are classified into Groups 2 and 3 when model (5.2) is used. We also notice that four of the seven countries in the G7 are classified in the convergence group, viz., Canada, France, Germany and the United Kingdom. These findings confirm those in Keller (2004) who finds that the major sources of technical change leading to productivity growth in OECD countries are not domestic but come from aboard through the channel of international technology diffusion.

In summary, we re-estimate Coe and Helpman's model by using the pooled FM-OLS and the PPC-based method with one unobserved global nonstationary factor. The pooled FM-OLS estimates confirm the international R&D spillovers after allowing for an unobserved global factor. In addition, our Cup-Lasso estimates show heterogeneous behavior in innovations and catch-up effects. To the best of our knowledge, this finding is the first to empirically identify two types of technology spillovers at the country level. Further, these results build an empirical connection between the "Club convergence" theory (Quah (1996, 1997)) and the conditional convergence model (Barro and Sala-i-Martin (1997)). Consequently, economic growth patterns do vary across countries– some exhibit convergence while others do not.

6 Conclusion

The primary theoretical contribution of this paper is to develop a novel approach that handles unobserved parameter heterogeneity and cross-section dependence in nonstationary panel models with latent cointegrating structures. We assume that cross-section dependence is captured by unobserved common factors which may be stationary and nonstationary. In general, penalized least squares estimators are inconsistent due to variable omission and the induced spurious regression problem from the presence of unobserved nonstationary factors. We propose an iterative procedure based on the penalized principal component method, which provides consistent and efficient estimators for long-run cointegration relationships under cross-section dependence. Lasso-type estimators are shown to have a mixed normal asymptotic distribution after bias correction. This property facilitates the use of conventional testing procedures using t, Wald, and F statistics for inference. A secondary contribution of the paper is to employ these methods in an empirical application that provides new findings to explain the growth convergence puzzle through the heterogeneous behavior of R&D spillover effects.

Several interesting topics for future research emerge. First, we do not allow the regressors to share a similar factor structure as the dependent variable in our model. If the regressors are assumed to exhibit factor structures, it seems possible to control for the unobserved common factors via the cross-sectional averages of the dependent and independent variables and then one can extend the common correlated effects (CCE) estimation of Pesaran (2006) to our framework. Second, as a referee remarked, the factor loadings (especially those of the nonstationary factors) may also exhibit a latent group structure, which may or may not be identical to those among the slope coefficients $\{\beta^0_{1,i}\}$. If the factor loadings are not required to share the same latent group structure as $\{\beta^0_{1,i}\}$, we can estimate the model as in the current paper and then estimate the latent group structure in the estimated factor loadings, say by applying the sequential binary segmentation algorithm of Wang and Su (2020). Formal analysis of these topics is left for future research.

REFERENCES

Acemoglu, D., Johnson, S., Robinson, J.A., Yared, P., 2008. Income and democracy. *American Economic Review* 98, 808-842.

Ando, T., Bai, J., 2016. Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics* 31, 163-191.

Bai, J., 2003. Inferential theory for factor models of large dimensions. *Econometric*a 71, 135-171.

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

Bai, J., Kao, C., Ng, S., 2009. Panel cointegration with global stochastic trends. *Journal of Econometrics* 149, 82-99.

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

Bai, J., Ng, S., 2004. A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127-1177.

Barro, R. J., Sala-i-Martin, X., 1997. Technological diffusion, convergence, and growth. *Journal of Economic Growth* 2, 1-26.

Beveridge, S., Nelson, C.R., 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', *Journal of Monetary Economics* 7, 151-174.

Bloom, N., Schankerman, M., Van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. *Econometrica* 81, 1347-1393.

Bonhomme, S., Manresa, E., 2015. Grouped patterns of heterogeneity in panel data. *Econometrica* 83, 1147-1184.

Coe, D.T., Helpman, E., 1995. International R&D spillovers. *European Economic Review* 39, 859-887.

Coe, D.T., Helpman, E., Hoffmaister, A.W., 2009. International R&D spillovers and institutions. *European Economic Review* 53, 723-741 .

Durlauf, S.N., Johnson, P.A., 1995. Multiple regimes and cross-country growth behaviour. *Journal of Applied Econometrics* 10, 365-384.

Engelbrecht, H.J., 1997. International R&D spillovers, human capital and productivity in OECD economies: An empirical investigation. *European Economic Review 4*1, 1479-1488.

Ertur, C., Musolesi, A., 2017. Weak and strong cross-sectional dependence: a panel data analysis of international technology diffusion. *Journal of Applied Econometrics* 32, 477-503.

Griffith, R., Redding, S., Reenen, J. V., 2004. Mapping the two faces of R&D: Productivity growth in a panel of OECD industries. *Review of Economics and Statistics* 86, 883-895.

Huang, W., Jin, S., Su, L., 2020. Identifying latent group patterns in cointegrated panels. *Econometric Theory*, forthcoming.

Kao, C., Chiang, M. H., 2001. On the estimation and inference of a cointegrated regression in panel data. In *Nonstationary panels, panel cointegration, and dynamic panels* 179-222. Emerald Group Publishing Limited.

Keller, W., 2004. International technology diffusion. *Journal of Economic Literature* 42, 752-782.

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

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

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

Miao, K., Li, K., Su, L., 2020. Panel threshold regressions with interactive fixed effects. *Working Paper*, Singapore Management University.

Moon, H. R., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica, 83(4), 1543-1579.

Moon, H. R., Weidner, M., 2017. Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory* 33, 158-195.

Pedroni, P., 2004. Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric Theory* 20, 597-625.

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

Phillips, P.C.B., 1995. Fully modified least squares and vector autoregression. *Econometrica* 63, 1023-1078.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. *The Review of Economic Studies* 57, 99-125.

Phillips, P.C.B., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057-1111.

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. *The Annals of Statistics* 20, 971-1001.

Phillips, P.C.B., Sul, D., 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal* 6, 217-259.

Phillips, P.C.B., Sul, D., 2009. Economic transition and growth. *Journal of Applied Econometrics* 24, 1153-1185.

Qian, J., Su, L., 2016. Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics* 191, 86-109.

Quah, D.T., 1996. Twin peaks: growth and convergence in models of distribution dynamics. *The Economic Journal* 106, 1045-1055.

Quah, D.T., 1997. Empirics for growth and distribution: stratification, polarization, and convergence clubs. *Journal of Economic Growth* 2, 27-59.

Su, L., Ju, G., 2018. Identifying latent group patterns in panel data models with interactive fixed effects. *Journal of Econometrics* 206, 554-573.

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

Su, L., Shi, Z., Phillips, P.C., 2016a. Identifying latent structures in panel data. *Econometrica* 84, 2215-2264.

Su, L., Shi, Z., Phillips, P.C., 2016b. Supplement to Identifying latent structures in panel data. *Econometrica Supplemental Material* 84, http://dx.doi.org/10.3982/ECTA12560.

Van der Ploeg, F., 2011. Natural resources: Curse or blessing? *Journal of Economic Literature* 49, 366-420.

Online Supplement to
"Nonstationary Panel Models with Latent Group Structures and Cross-Section Dependence"

Wenxin Huang[a], Sainan Jin[b], Peter C.B. Phillips[c], and Liangjun Su[b],

[a] Antai College of Economics and Management, Shanghai Jiao Tong University
[b] School of Economics, Singapore Management University
[c] Yale University, University of Auckland
University of Southampton & Singapore Management University

This online supplement is composed of five parts. Appendix A contains the proofs of the main results in the paper. Appendix B contains the proofs of the technical lemmas stated and used in Appendix A. Appendix C discusses the identification of $\beta^0_{1,i}$. Appendix D contains the detailed procedure for the proposed method in the paper. Appendix E reports some additional simulation results. Let $\max_i = \max_{1\le i\le N}$ and $\min_i = \min_{1\le i\le N}$.

A Proof of the Main Results in Section 3

This Appendix provides the proofs of Theorems 3.1-3.8 in the paper. These results rely on some subsidiary technical lemmas whose proofs are provided in the Additional Online Supplement (Appendix B).

To proceed, we define some notation.

(i) Let $H_1 = \left(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0\right)\left(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1\right)V_{1,NT}^{-1}$ and $H_2 = \left(\frac{1}{N}\Lambda_2^{0\prime}\Lambda_2^0\right)\left(\frac{1}{T}F_2^{0\prime}\hat{F}_2\right)V_{2,NT}^{-1}$.

(ii) Let $\mathbf{b}_l = (b_{l,1},...,b_{l,N})$ and $\hat{\boldsymbol{b}}_l = (\hat{b}_{l,1},...,\hat{b}_{l,N})$, where $b_{l,i} = \beta_{l,i} - \beta^0_{l,i}$ and $\hat{b}_{l,i} = \hat{\beta}_{l,i} - \beta^0_{l,i}$ for $i = 1,...,N$ and $l = 1,2$.

(iii) Let $\eta^2_{lNT} = \frac{1}{N}\sum_{i=1}^N \|\hat{b}_{l,i}\|^2$ for $l = 1,2$, $\varrho^2_{NT} = \frac{1}{K}\sum_{k=1}^K \left\|\hat{\alpha}_k - \alpha^0_k\right\|^2$, $C_{NT} = \min(\sqrt{N},\sqrt{T})$, $\delta_{NT} = \min(\sqrt{N},T)$, and $\psi_{NT} = N^{1/q}T^{-1}(\log T)^{1+\epsilon}$ for some $\epsilon > 0$.

(iv) Let $\hat{Q}_{i,xx} = \frac{1}{T^2}x'_{1,i}M_{\hat{F}_1}x_{1,i}$, $Q_{i,xx}(F_1) = \frac{1}{T^2}x'_{1,i}M_{F_1}x_{1,i}$, and $Q^0_{i,xx} = Q_{i,xx}(F^0_1)$.

(v) Without loss of generality, we set $x_{1,i0} = 0$ and $x_{2,i0} = 0$ throughout the appendix.

To prove Theorem 3.1, we need the following four lemmas.

Lemma A.1 *Suppose that Assumption 3.1 hold. Then for each $i = 1,...,N$,*

(i) $\frac{1}{T^2}x'_{1,i}M_{F^0_1}x_{1,i} \Rightarrow \int \tilde{B}_{2i}\tilde{B}'_{2i}$,

(ii) $\frac{1}{T}x'_{1,i}M_{F^0_1}u_i \Rightarrow \int (B_{2i} - \pi'_iB_3)\,dB_{1i} + (\Delta_{21,i} - \pi'_i\Delta_{31,i})$,

where $\tilde{B}_{2i} = B_{2i} - \int B_{2i}B'_3\left(\int B_3B'_3\right)^{-1}B_3$ *and* $\pi_i = \left(\int B_3B'_3\right)^{-1}\int B_3B'_{2i}$.

Lemma A.2 *Suppose that Assumptions 3.1-3.2 hold. Let $W_i = \left(x_{1,i}, F^0_1\right)$ and $d_T = \log\log T$, as in Assumption 3. Then for any fixed small constant $c \in (0, 1/2)$,*

(i) $\limsup_{T\to\infty}\mu_{\max}\left(\frac{1}{d_TT^2}W'_iW_i\right) \le (1+c)\rho_{\max}$ *a.s.,*

(ii) $\liminf_{T\to\infty}\mu_{\min}\left(\frac{d_T}{T^2}W'_iW_i\right) \ge c\rho_{\min}$ *a.s.,*

(iii) $\limsup_{T\to\infty}\mu_{\max}\left(\frac{1}{d_TT^2}x'_{1,i}M_{F^0_1}x_{1,i}\right) \le (1+c)\rho_{\max}$ *a.s.,*

(iv) $\liminf_{T\to\infty}\mu_{\min}\left(\frac{d_T}{T^2}x'_{1,i}M_{F^0_1}x_{1,i}\right) \ge \rho_{\min}/2$ *a.s..*

Lemma A.3 *Suppose that Assumptions 3.1-3.2 hold. Then* *(i)* $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}u_i\right\|^2=O_P(d_T T^{-2})$,

(ii) $\sup_{N^{-1}\|\mathbf{b}_2\|^2\le M}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}u_i^*\right\|^2=o_P(T^{-1}N^{1/q}(\log T)^{(1+\epsilon)/2})$,

(iii) $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{NT^2}\sum_{j=1}^{N}x_{1,i}'M_{F_1^0}u_j a_{ij}\right\|=O_P(T^{-1})$,

(iv) $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{1,i}\right\|=O_P(1)$,

where $u_i^*\equiv u_i^*(b_{2,i})\equiv u_i+F_2^0\lambda_{2i}^0-x_{2,i}b_{2,i}$.

Lemma A.4 *Suppose that Assumptions 3.1-3.2 hold. Then*

(i) $\sup_{F_1\in\mathcal{F}_1}\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\|\frac{1}{NT^2}\sum_{i=1}^{N}b_{1,i}'x_{1,i}'M_{F_1}u_i^*\right\|=O_P((T/d_T)^{-1/2})$,

(ii)$\sup_{F_1\in\mathcal{F}_1}\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\lambda_{1i}^{0\prime}F_1^{0\prime}M_{F_1}u_i^*\right\|=O_P((T/d_T)^{-1/2})$,

(iii) $\sup_{F_1\in\mathcal{F}_1}\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\|\frac{1}{NT^2}\sum_{i=1}^{N}u_i^{*\prime}P_{F_1}u_i^*\right\|=O_P((T/d_T)^{-1/2})$,

where $\mathbf{b}=(\mathbf{b}_1,\mathbf{b}_2)$, $\mathcal{F}_1$ *is defined in Assumption 3.2(iv), and* u_i^* *is defined in Lemma A.3.*

Proof of Theorem 3.1. (i) Let $Q_{i,NT}(\beta_{1,i},\beta_{2,i},F_1)=\frac{1}{T^2}(y_i-x_{1,i}\beta_{1,i}-x_{2,i}\beta_{2,i})'M_{F_1}(y_i-x_{1,i}\beta_{1,i}-x_{2,i}\beta_{2,i})$ and $Q_{i,NT}^{K,\lambda}(\beta_{1,i},\beta_{2,i},\alpha,F_1)=Q_{i,NT}(\beta_{1,i},\beta_{2,i},F_1)+\lambda\prod_{k=1}^{K}\|\beta_{1,i}-\alpha_k\|$. Then $Q_{NT}^{K,\lambda}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha},F_1)$ $=\frac{1}{N}\sum_{i=1}^{N}Q_{i,NT}^{K,\lambda}(\beta_{1,i},\beta_{2,i},\alpha,F_1)$. Noting that $y_i-x_{1,i}\beta_{1,i}-x_{2,i}\beta_{2,i}=-x_{1,i}b_{1,i}+F_1^0\lambda_{1i}^0+u_i^*$, we have

$$\begin{aligned}&Q_{i,NT}(\beta_{1,i},\beta_{2,i},F_1)-Q_{i,NT}(\beta_{1,i}^0,\beta_{2,i}^0,F_1^0)\\=&\frac{1}{T^2}(b_{1,i}'x_{1,i}'M_{F_1}x_{1,i}b_{1,i}+\lambda_{1i}^{0\prime}F_1^{0\prime}M_{F_1}F_1^0\lambda_{1i}^0-2b_{1,i}'x_{1,i}'M_{F_1}F_1^0\lambda_{1i})\\&+\frac{1}{T^2}\left(2\lambda_{1i}^{0\prime}F_1^{0\prime}M_{F_1}u_i^*-2b_{1,i}'x_{1,i}'M_{F_1}u_i^*-u_i^{*\prime}(P_{F_1}-P_{F_1^0})u_i^*\right),\end{aligned}\tag{A.1}$$

where $u_i^*=u_i+F_2^0\lambda_{2i}^0-x_{2,i}b_{2,i}$.

Let $S_{i,NT}(\beta_{1,i},F_1)=\frac{1}{T^2}\left(b_{1,i}'x_{1,i}'M_{F_1}x_{1,i}b_{1,i}+\lambda_{1i}^{0\prime}F_1^{0\prime}M_{F_1}F_1^0\lambda_{1i}^0-2b_{1,i}'x_{1,i}'M_{F_1}F_1^0\lambda_{1i}^0\right)=\frac{1}{T^2}(x_{1,i}b_{1,i}$ $-F_1^0\lambda_{1i}^0)'M_{F_1}(x_{1,i}b_{1,i}-F_1^0\lambda_{1i}^0)\ge 0$. Then we have

$$\begin{aligned}&Q_{NT}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,F_1)-Q_{NT}(\boldsymbol{\beta}_1^0,\boldsymbol{\beta}_2^0,F_1^0)\\&=\frac{1}{N}\sum_{i=1}^{N}S_{i,NT}(\beta_{1,i},F_1)+\frac{1}{NT^2}\sum_{i=1}^{N}\left(2\lambda_{1i}^{0\prime}F_1^{0\prime}M_{F_1}u_i^*-2b_{1,i}'x_{1,i}'M_{F_1}u_i^*-u_i^{*\prime}(P_{F_1}-P_{F_1^0})u_i^*\right)\\&=S_{NT}(\boldsymbol{\beta}_1,F_1)+O_P((T/d_T)^{-1/2}),\end{aligned}\tag{A.2}$$

where $S_{NT}(\boldsymbol{\beta}_1,F_1)=\frac{1}{N}\sum_{i=1}^{N}S_{i,NT}(\beta_{1,i},F_1)$ and $O_P((T/d_T)^{-1/2})$ holds uniformly in $(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,F_1)$ such that $\frac{F_1'F_1}{T^2}=I_{r_1}$ and $\frac{1}{N}\|\mathbf{b}\|^2\le M$ by Lemma A.4(i)-(iii) and the fact that $\frac{1}{NT^2}\sum_{i=1}^{N}u_i^{*\prime}P_{F_1^0}u_i^*=O_P((T/d_T)^{-1/2})$. It follows that

$$\begin{aligned}&Q_{NT}^{K,\lambda}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\hat{\boldsymbol{\alpha}},F_1)-Q_{NT}^{K,\lambda}(\boldsymbol{\beta}_1^0,\boldsymbol{\beta}_2^0,\boldsymbol{\alpha}^0,F_1^0)\\=&\frac{1}{N}\sum_{i=1}^{N}[Q_{NT,i}(\beta_{1,i},\beta_{2,i},F_1)-Q_{NT,i}(\beta_{1,i}^0,\beta_{2,i}^0,F_1^0)]+\frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\beta_{1,i}-\hat{\alpha}_k\|\\\ge&S_{NT}(\boldsymbol{\beta}_1,F_1)+O_P((T/d_T)^{-1/2}).\end{aligned}\tag{A.3}$$

Then by (A.2) and (A.3) and the fact that $Q_{NT}^{K,\lambda}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\alpha}}, \hat{F}_1) - Q_{NT}^{K,\lambda}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, \boldsymbol{\alpha}^0, F_1^0) \leq 0$, we have

$$S_{NT}(\hat{\boldsymbol{\beta}}_1, \hat{F}_1) = \frac{1}{NT^2} \sum_{i=1}^{N} (x_{1,i}\hat{b}_{1,i} - F_1^0 \lambda_{1i}^0)' M_{\hat{F}_1} (x_{1,i}\hat{b}_{1,i} - F_1^0 \lambda_{1i}^0) = O_P((T/d_T)^{-1/2}). \tag{A.4}$$

Let $\chi_{i,\hat{\mathbf{b}}_1} = x_{1,i}\hat{b}_{1,i} - F_1^0 \lambda_{1i}^0$. Noting that $\operatorname{tr}(AB) \geq \sum_{t=1}^{T} \mu_t(A) \mu_{T-t+1}(B)$ for any two $T \times T$ symmetric p.s.d. matrices A and B where $\{\mu_t(\cdot)\}_{t=1}^{T}$ represent descending ordered eigenvalues (e.g., Bernstein (2005, p.326)) and $M_{\hat{F}_1}$ is a projection matrix with rank $T - r_1$, we have

$$S_{NT}(\hat{\boldsymbol{\beta}}_1, \hat{F}_1) = \operatorname{tr}\left\{ \left(\frac{1}{NT^2} \sum_{i=1}^{N} \chi_{i,\hat{\mathbf{b}}_1} \chi_{i,\hat{\mathbf{b}}_1}' \right) M_{\hat{F}_1} \right\} \geq \sum_{t=r_1+1}^{T} \mu_t \left(\frac{1}{NT^2} \sum_{i=1}^{N} \chi_{i,\hat{\mathbf{b}}_1} \chi_{i,\hat{\mathbf{b}}_1}' \right). \tag{A.5}$$

Therefore $\sum_{t=r_1+1}^{T} \mu_t \left(\frac{1}{NT^2} \sum_{i=1}^{N} \chi_{i,\hat{\mathbf{b}}_1} \chi_{i,\hat{\mathbf{b}}_1}' \right) = O_P((T/d_T)^{-1/2})$. Then by Assumption 3.2(v), we must have $\frac{1}{N} \left\| \hat{\mathbf{b}}_1 \right\|^2 = O_P((T/d_T)^{-1/2})$.

(ii) By the result in (i) and Lemma A.2(i),

$$\frac{1}{NT^2} \sum_{i=1}^{N} \hat{b}_{1,i}' x_{1,i}' M_{\hat{F}_1} x_{1,i} \hat{b}_{1,i} \leq d_T \max_i \mu_{\max}(\frac{1}{d_T T^2} x_{1,i}' x_{1,i}) \frac{1}{N} \left\| \hat{\mathbf{b}}_1 \right\|^2 = O_P((T/d_T^3)^{-1/2}). \tag{A.6}$$

Combining (A.4) and (A.6) and applying the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
& O_P((T/d_T^3)^{-1/2}) \\
= \; & \frac{1}{NT^2} \sum_{i=1}^{N} \lambda_{1i}^{0\prime} F_1^0 M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 - 2\hat{b}_{1,i}' x_{1,i}' M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 \\
\geq \; & \frac{1}{NT^2} \sum_{i=1}^{N} \left\{ \lambda_{1i}^{0\prime} F_1^0 M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 - 2 \left\{ \frac{1}{NT^2} \sum_{i=1}^{N} \hat{b}_{1,i}' x_{1,i}' M_{\hat{F}_1} x_{1,i} \hat{b}_{1,i} \right\}^{1/2} \left\{ \frac{1}{NT^2} \sum_{i=1}^{N} \lambda_{1i}^{0\prime} F_1^{0\prime} M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 \right\}^{1/2} \right\},
\end{aligned}$$

which, in conjunction with (A.6), further implies that $\frac{1}{NT^2} \sum_{i=1}^{N} \lambda_{1i}^{0\prime} F_1^{0\prime} M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 = O_P((T/d_T^3)^{-1/2})$. Then $O_P((T/d_T^3)^{-1/2}) = \operatorname{tr}[(\frac{1}{T^2} F_1^{0\prime} M_{\hat{F}_1} F_1^0)(\frac{1}{N} \Lambda_1^{0\prime} \Lambda_1^0)] \geq \operatorname{tr}(\frac{1}{T^2} F_1^{0\prime} M_{\hat{F}_1} F_1^0) \mu_{\min}(\frac{1}{N} \Lambda_1^{0\prime} \Lambda_1^0)$. It follows that $\operatorname{tr}(\frac{1}{T^2} F_1^{0\prime} M_{\hat{F}_1} F_1^0) = O_P((T/d_T^3)^{-1/2})$ as $\mu_{\min}(\frac{1}{N} \Lambda_1^{0\prime} \Lambda_1^0)$ is bounded away from zero in probability by Assumption 3.2(i). As in Bai (2009, p.1265), this implies that

$$\frac{F_1^{0\prime} M_{\hat{F}_1} F_1^0}{T^2} = \frac{F_1^{0\prime} F_1^0}{T^2} - \frac{F_1^{0\prime} \hat{F}_1}{T^2} \frac{\hat{F}_1' F_1^0}{T^2} = O_P((T/d_T^3)^{-1/2}), \tag{A.7}$$

and $\frac{1}{T^2} F_1^{0\prime} \hat{F}_1$ is asymptotically invertible by the fact that $\frac{1}{T^2} F_1^{0\prime} F_1^0$ is asymptotically invertible from Assumption 3.2(ii). (A.7) implies that $\frac{1}{T^2} \hat{F}_1' P_{F_1^0} \hat{F}_1 - I_{r_1} = O_P((T/d_T^3)^{-1/2})$, which further implies that $\left\| P_{\hat{F}_1} - P_{F_1^0} \right\|^2 = 2\operatorname{tr}(I_{r_1} - \frac{1}{T^2} \hat{F}_1' P_{F_1^0} \hat{F}_1) = O_P((T/d_T^3)^{-1/2})$.

(iii) We want to establish the consistency of the estimated factor space $\hat{F}_1$, which extends the results of Bai and Ng (2004) and Bai (2009). Our model allows for heterogeneous slope coefficients in both nonstationary and stationary regressors and unobserved stationary common factors. Here we don't need the consistency of $\hat{\beta}_{2,i}$ but require that $\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{2,i}\|^2 \leq M$ for some sufficiently large constant M w.p.a.1 (which can be proved as in Su and Ju (2018)). Note that $\hat{F}_1$ satisfies

$$\left[\frac{1}{NT^2}\sum_{i=1}^{N}(y_i - x_i\hat{\beta}_i)(y_i - x_i\hat{\beta}_i)'\right]\hat{F}_1 = \hat{F}_1 V_{1,NT}. \tag{A.8}$$

Using (A.8) and the fact that $y_i - x_i\hat{\beta}_i = -x_{1,i}\hat{b}_{1,i} + F^0\lambda_i^0 + \tilde{u}_i = -x_{1,i}\hat{b}_{1,i} + F_1^0\lambda_{1i}^0 + F_2^0\lambda_{2i}^0 + \tilde{u}_i$ and $\tilde{u}_i = u_i - x_{2,i}\hat{b}_{2,i}$. we have

$$\begin{aligned}
\hat{F}_1 V_{1,NT} =& \frac{1}{NT^2}\sum_{i=1}^{N} x_{1,i}\hat{b}_{1,i}\hat{b}_{1,i}'x_{1,i}'\hat{F}_1 - \frac{1}{NT^2}\sum_{i=1}^{N} x_{1,i}\hat{b}_{1,i}\lambda_i^{0\prime}F^{0\prime}\hat{F}_1 - \frac{1}{NT^2}\sum_{i=1}^{N} x_{1,i}\hat{b}_{1,i}\tilde{u}_i'\hat{F}_1 \\
& - \frac{1}{NT^2}\sum_{i=1}^{N} F^0\lambda_i^0\hat{b}_{1,i}'x_{1,i}'\hat{F}_1 - \frac{1}{NT^2}\sum_{i=1}^{N}\tilde{u}_i\hat{b}_{1,i}'x_{1,i}'\hat{F}_1 + \frac{1}{NT^2}\sum_{i=1}^{N} F^0\lambda_i^0\tilde{u}_i'\hat{F}_1 \\
& + \frac{1}{NT^2}\sum_{i=1}^{N}\tilde{u}_i\lambda_i^{0\prime}F^{0\prime}\hat{F}_1 + \frac{1}{NT^2}\sum_{i=1}^{N}\tilde{u}_i\tilde{u}_i'\hat{F}_1 + \frac{1}{NT^2}\sum_{i=1}^{N} F_2^0\lambda_{2i}^0\lambda_{2i}^{0\prime}F_2^{0\prime}\hat{F}_1 \\
& + \frac{1}{NT^2}\sum_{i=1}^{N} F_1^0\lambda_{1i}^0\lambda_{2i}^{0\prime}F_2^{0\prime}\hat{F}_1 + \frac{1}{NT^2}\sum_{i=1}^{N} F_2^0\lambda_{2i}^0\lambda_{1i}^{0\prime}F_1^{0\prime}\hat{F}_1 + \frac{1}{NT^2}\sum_{i=1}^{N} F_1^0\lambda_{1i}^0\lambda_{1i}^{0\prime}F_1^{0\prime}\hat{F}_1 \\
\equiv& I_1 + ... + I_{11} + \frac{1}{NT^2}\sum_{i=1}^{N} F_1^0\lambda_{1i}^0\lambda_{1i}^{0\prime}F_1^{0\prime}\hat{F}_1, \text{ say.}
\end{aligned}$$

It follows that $\hat{F}_1 V_{1,NT} - F_1^0(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1) = I_1 + ... + I_{11}$. Let $H_1 = (\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1)V_{1,NT}^{-1}$. It is easy to show that $H_1 = O_P(1)$ and is asymptotically nonsingular. Then $\hat{F}_1 H_1^{-1} - F_1^0 = [I_1 + ... + I_{11}]\,(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1)^{-1}(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}$ and $\frac{1}{T}\left\|\hat{F}_1 H^{-1} - F_1^0\right\| \leq \frac{1}{T}(\|I_1\| + ... + \|I_{11}\|)\left\|(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1)^{-1}\right\| \times \left\|(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}\right\|$. It remains to analyze $\|I_l\|$ for $l = 1, 2, ..., 11$. For I_1, we have that by the result in (iii),

$$\frac{1}{T}\|I_1\| \leq \frac{1}{N}\sum_{i=1}^{N}\frac{\|x_{1,i}\|}{T}\|\hat{b}_{1,i}\|^2\frac{\|x_{1,i}'\hat{F}_1\|}{T^2} \leq \max_{1\leq i\leq N}\frac{\|x_{1,i}'\|^2}{T^2}\frac{\|\hat{F}_1\|}{T}\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2 = O_P(d_T\eta_{1NT}^2) = o_P(\eta_{1NT}),$$

where we use the fact that $\max_{1\leq i\leq N}\frac{\|x_{1,i}\|^2}{T^2} \leq \max_{1\leq i\leq N} p_1 d_T \mu_{\max}\left(\frac{x_i'x_i}{d_T T^2}\right) = O_P(d_T)$ by Lemma A.2(i) and $\frac{\|\hat{F}_1\|}{T} \leq \sqrt{r_1}$. For I_2, we have

$$\frac{1}{T}\|I_2\| \leq \frac{\|F^{0\prime}\hat{F}_1\|}{T^2}\max_{1\leq i\leq N}\frac{\|\lambda_i^0\|}{T}\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\|x_{1,i}\|^2}{T^2}\right\}^{1/2} = O_P(\eta_{1NT}),$$

where we use the fact that $\frac{\|F^{0\prime}\hat{F}_1\|}{T^2} = O_P(1)$ and $\|\lambda_i^0\| = O_P(1)$ by Assumption 3.2(i). For I_3,

$$\frac{1}{T}\|I_3\| \leq \frac{1}{\sqrt{T}}\frac{\|\hat{F}_1\|}{T}\max_{1\leq i\leq N}\frac{\|x_{1,i}\|}{T}\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\|\tilde{u}_i\|^2}{T}\right\}^{1/2} = O_P\left(\sqrt{\frac{d_T}{T}}\eta_{1NT}\right),$$

where $\frac{1}{N}\sum_{i=1}^{N}\frac{\|\tilde{u}_i\|^2}{T} \leq 2\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\|u_i\|^2}{T} + \max_{1\leq i\leq N}\frac{\|x_{2,i}\|^2}{T}\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{2,i}\|^2\right) = O_P(1)$ by Assumption 3.1(i) and (iv) and the fact that $\max_{1\leq i\leq N}\frac{\|x_{2,i}\|^2}{T} = \max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}[\|x_{2,it}\|^2 - E(\|x_{2,it}\|^2)] + \max_{i,t}E(\|x_{2,it}\|^2) = o_P(1) + O(1) = O_P(1)$ by Lemma S1.2(iii) in Su, Shi and Phillips (2016, SSPb hereafter). Similarly, for I_4 and I_5,

$$\begin{aligned}
\frac{1}{T}\|I_4\| &\leq \frac{\|F^0\|}{T}\frac{\|\hat{F}_1\|}{T}\max_{1\leq i\leq N}\|\lambda_i^0\|\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\|x_{1,i}\|^2}{T^2}\right\}^{1/2} = O_P(\eta_{1NT}), \text{ and} \\
\frac{1}{T}\|I_5\| &\leq \frac{1}{\sqrt{T}}\frac{\|\hat{F}_1\|}{T}\max_{1\leq i\leq N}\frac{\|x_{1,i}\|}{T}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\|\tilde{u}_i\|^2}{T}\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2} = O_P\left(\sqrt{\frac{d_T}{T}}\eta_{1NT}\right),
\end{aligned}$$

where we use the fact that $\frac{\|F^0\|}{T} \leq \frac{\|F_1^0\|}{T} + \frac{1}{\sqrt{T}}\frac{\|F_2^0\|}{\sqrt{T}} = O_P(1)$. For I_6, we have

$$\frac{1}{T}\|I_6\| = \frac{1}{T}\left\|\frac{1}{NT^2}F^0\Lambda^{0\prime}\tilde{u}\hat{F}_1\right\| \leq \frac{1}{\sqrt{NT}}\left(\frac{1}{T}\left\|\hat{F}_1\right\|\right)\left(\frac{1}{T}\left\|F^0\right\|\right)\frac{1}{\sqrt{NT}}\|\Lambda^{0\prime}\tilde{u}\| = O_P(T^{-1/2}N^{-1/2}),$$

where $\tilde{u} = (\tilde{u}_1, ..., \tilde{u}_N)'$ and we have used the fact that $\frac{1}{NT}\left\|\Lambda^{0\prime}\tilde{u}\right\|^2 = O_P(1)$ and $\frac{1}{NT}\left\|\sum_{i=1}^{N}\lambda_i^0\hat{b}_{2,i}'x_{2,i}\right\|^2 = O_P(1)$ by Assumption 3.1(iv) and straightforward calculations. Analogously, we can show that $\frac{1}{T}\|I_7\| = O_P(T^{-1/2}N^{-1/2})$. For I_8, $I_8 = \frac{1}{NT^2}\sum_{i=1}^{N}\left(u_iu_i' - u_i\hat{b}_{2,i}'x_{2,i}' - x_{2,i}\hat{b}_{2,i}u_i' + x_{2,i}\hat{b}_{2,i}\hat{b}_{2,i}'x_{2,i}'\right)\hat{F}_1 = I_{8,1} + I_{8,2} + I_{8,3} + I_{8,4}$. For $I_{8,1}$,

$$\begin{aligned}
\frac{1}{T^2}\|I_{8,1}\|^2 &= \frac{1}{T^2}\left\|\frac{1}{NT^2}u'u\hat{F}_1\right\|^2 \leq 2\sum_{t=1}^{T}\left\|T^{-3}\sum_{s=1}^{T}\gamma_N(s,t)\hat{f}_{1s}'\right\|^2 + 2\sum_{t=1}^{T}\left\|T^{-3}\sum_{s=1}^{T}\xi_{st}\hat{f}_{1s}'\right\|^2 \\
&\equiv 2\left(\|I_8(a)\| + \|I_8(b)\|\right),
\end{aligned}$$

where $\gamma_N(s,t)$ and ξ_{st} are defined in Assumption 3.2(iii). Note that $\|I_8(a)\| = T^{-3}(T^{-2}\sum_{s=1}^{T}\|\hat{f}_{1s}\|^2)$ $\times(T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\|\gamma_N(s,t)\|^2) = O_P(T^{-3})$ and $\|I_8(b)\| = T^{-2}N^{-1}(T^{-2}\sum_{s=1}^{T}\left\|\hat{f}_{1s}\right\|^2)(T^{-2}N\sum_{t=1}^{T}\sum_{s=1}^{T}$ $\|\xi_{st}\|^2) = O_P(T^{-2}N^{-1})$ by the fact that $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\gamma_N(s,t)\|^2 \leq M$ by Assumption 3.2(iii) (see also Lemma 1(i) in Bai and Ng (2002)) and that $\mathbb{E}(\|\xi_{st}\|^2) \leq N^{-2}M$ under Assumption 3.2(iii). Then $\frac{1}{T}\|I_{8,1}\| = O_P(N^{-1/2}T^{-1} + T^{-3/2})$. For $I_{8,2}$, we have

$$\frac{1}{T}\|I_{8,2}\| \leq \frac{1}{T}\max_{1\leq i\leq N}\frac{\|x_{2,i}\|}{\sqrt{T}}\frac{\|\hat{F}_1\|}{T}\left(\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{2,i}\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\|u_i\|^2}{T}\right)^{1/2} = O_P(T^{-1})$$

Similarly, we have $\frac{1}{T}\|I_{8,3}\| = O_P(T^{-1})$ and $\frac{1}{T}\|I_{8,4}\| \leq \frac{1}{T}\max_{1\leq i\leq N}\frac{\|x_{2,i}\|^2}{T}\frac{\|\hat{F}_1\|}{T}\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{2,i}\|^2 = O_P(T^{-1})$. Then, we have $\frac{1}{T}\|I_8\| = O_P(T^{-1})$. For I_9 and I_{10}, we have

$$\begin{aligned}
\frac{1}{T}\|I_9\| &= \frac{1}{T}\left\|\frac{1}{NT^2}F_2^0\Lambda_2^{0\prime}\Lambda_2^0F_2^{0\prime}\hat{F}_1\right\| \leq \frac{1}{T}\frac{\|F_2^0\|^2}{T}\frac{\|\hat{F}_1\|}{T}\left\|\frac{\Lambda_2^{0\prime}\Lambda_2^0}{N}\right\| = O_P(T^{-1}), \text{ and} \\
\frac{1}{T}\|I_{10}\| &= \frac{1}{T}\left\|\frac{1}{NT^2}F_1^0\Lambda_1^{0\prime}\Lambda_2^0F_2^{0\prime}\hat{F}_1\right\| \leq \frac{1}{\sqrt{NT}}\frac{\|F_1^0\|}{T}\frac{\|F_2^0\|}{\sqrt{T}}\frac{\|\hat{F}_1\|}{T}\frac{\|\Lambda_1^{0\prime}\Lambda_2^0\|}{\sqrt{N}} = O_P((NT)^{-1/2}),
\end{aligned}$$

where $\frac{\Lambda_1^{0\prime}\Lambda_2^0}{\sqrt{N}} = O_P(1)$ by Assumption 3.2(i). Analogously, we have $\frac{1}{T}\|I_{11}\| = O_P((NT)^{-1/2})$. In sum, we have shown that $\frac{1}{T}\left\|\hat{F}_1 H_1^{-1} - F_1^0\right\| = O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})$. Then (iv) follows. ■

To prove Theorem 3.2 we need the following two lemmas.

Lemma A.5 *Suppose that Assumptions 3.1-3.2 hold. Let $\hat{u}_i^* = u_i + F_2^0\lambda_{2i}^0 - x_{2,i}\hat{b}_{2,i}$. Then*

(i) $\left\|P_{\hat{F}_1} - P_{F_1^0}\right\| = O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})$,

(ii) $\eta_{2NT}^2 \equiv \frac{1}{N}\sum_{i=1}^N \|\hat{b}_{2,i}\|^2 = O_P(T^{-1/2}C_{NT}^{-1} + Td_T\eta_{1NT}^2)$,

(iii) $\frac{1}{T}F_1^{0\prime}(\hat{F}_1 - F_1^0 H_1) = O_P(T\eta_{1NT} + \delta_{NT}^{-1} + T^{-1/4}C_{NT}^{-1/2})$,

(iv) $\frac{1}{T}\hat{F}_1'(\hat{F}_1 - F_1^0 H_1) = O_P(T\eta_{1NT} + \delta_{NT}^{-1} + T^{-1/4}C_{NT}^{-1/2})$,

(v) $\frac{1}{T}\hat{u}_i^{*\prime}(\hat{F}_1 H_1^{-1} - F_1^0) = O_P(\sqrt{T}\eta_{1NT} + \delta_{NT}^{-1})$ for each $i = 1, ..., N$.

Lemma A.6 *Suppose that Assumptions 3.1-3.2 hold. Let $R_{1i} = \frac{1}{T^2}x_{1,i}'(P_{F_1^0} - P_{\hat{F}_1})\hat{u}_i^*$, $R_{2i} = \frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 - \frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_{1,j}a_{ij}\hat{b}_{1,j} + \frac{1}{NT^2}\sum_{j=1}^N a_{ij}x_{1,i}'M_{\hat{F}_1}u_j$, $R_{3i} = \frac{1}{NT^2}\sum_{j=1}^N a_{ij}x_{1,i}'(P_{F_1^0} - P_{\hat{F}_1})u_j$, and $R_{4i} = \frac{1}{T^2}x_{1,i}'M_{F_1^0}\hat{u}_i^* - \frac{1}{NT^2}\sum_{j=1}^N a_{ij}x_{1,i}'M_{F_1^0}u_j$. Then*

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

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

(iii) $R_{3i} = O_P(\varsigma_{1NT})$ *for each* $i = 1, ..., N$, *and* $N^{-1}\sum_{i=1}^N \|R_{3i}\|^2 = O_P(d_T\varsigma_{1NT}^2)$,

(iv) $R_{4i} = O_P(T^{-1})$ *for each* $i = 1, ..., N$, *and* $N^{-1}\sum_{i=1}^N \|R_{4i}\|^2 = O_P(d_T T^{-2})$,

where $\varsigma_{1NT} = T^{-1/2}\eta_{1NT} + T^{-1}C_{NT}^{-1}$ and $\varsigma_{2NT} = T^{-5/4}C_{NT}^{-1/2} + T^{-1/2}\eta_{1NT} + d_T\eta_{1NT}^2 + T^{-1}\delta_{NT}^{-1}$.

Proof of Theorem 3.2. (i) Based on the sub-differential calculus, a necessary condition for $\hat{\beta}_{1,i}, \hat{\beta}_{2,i}$ $\hat{\alpha}_k$, and $\hat{F}_1$ to minimize the objective function (2.11) is, for each $i = 1, ..., N$, that $0_{p_1\times 1}$ belongs to the sub-differential of $Q_{NT}^{\lambda,K}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}, F_1)$ with respect to $\beta_{1,i}$ (resp. α_k) evaluated at $\{\hat{\beta}_{1,i}\}$, $\{\hat{\beta}_{2,i}\}$, $\{\hat{\alpha}_k\}$ and $\hat{F}_1$. That is, for each $i = 1, ..., N$ and $k = 1, ..., K$, we have

$$0_{p_1\times 1} = -\frac{2}{T^2}x_{1,i}'M_{\hat{F}_1}(y_i - x_{1,i}\hat{\beta}_{1,i} - x_{2,i}\hat{\beta}_{2,i}) + \lambda\sum_{j=1}^K \hat{e}_{ij}\prod_{l=1,l\neq j}^K \|\hat{\beta}_{1,i} - \hat{\alpha}_l\|, \tag{A.9}$$

where $\hat{e}_{ij} = \frac{\hat{\beta}_{1,i} - \hat{\alpha}_j}{\|\hat{\beta}_{1,i} - \hat{\alpha}_j\|}$ if $\|\hat{\beta}_{1,i} - \hat{\alpha}_j\| \neq 0$ and $\|\hat{e}_{ij}\| \leq 1$ if $\|\hat{\beta}_{1,i} - \hat{\alpha}_j\| = 0$. Noting that $y_i - x_{1,i}\hat{\beta}_{1,i} - x_{2,i}\hat{\beta}_{2,i} = -x_{1,i}\hat{b}_{1,i} + \hat{F}_1 H_1^{-1}\lambda_{1i}^0 + \hat{u}_i^* + (F_1^0 - \hat{F}_1 H_1^{-1})\lambda_{1i}^0$, $\hat{u}_i^* = u_i + F_2^0\lambda_{2i}^0 - x_{2,i}\hat{b}_{2,i}$ (A.9) implies that

$$\hat{Q}_{i,xx}\hat{b}_{1,i} = \frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}\hat{u}_i^* + \frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 - \frac{\lambda}{2}\sum_{j=1}^{K_0}\hat{e}_{ij}\prod_{l=1,l\neq j}^K \|\hat{\beta}_{1,i} - \hat{\alpha}_l\|, \tag{A.10}$$

which can be rewritten as

$$\hat{Q}_{i,xx}\hat{b}_{1,i} = \frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_{1,j}a_{ij}\hat{b}_{1,j} + R_i, \tag{A.11}$$

where $R_i = R_{1i} + R_{2i} - R_{3i} + R_{4i} - R_{5i}$, R_{1i}, R_{2i}, R_{3i} and R_{4i} are defined in the statement of Lemma A.6, and $R_{5i} = \frac{\lambda}{2}\sum_{j=1}^K e_{ij}\prod_{l=1,l\neq j}^K \|\hat{\beta}_{1,i} - \hat{\alpha}_l\|$. By Lemma A.6(i)-(iv), we have that $\sum_{l=1}^4 \frac{1}{N}\sum_{i=1}^N \|R_{li}\|^2 =$

$O_P(T^{-1}d_T\eta_{1NT}^2 + d_T^2\eta_{1NT}^4 + d_T T^{-2})$. In addition, we can show that $\frac{1}{N}\sum_{i=1}^N \|R_{5i}\|^2 = O_P\left(\lambda^2\right)$. It follows that $\frac{1}{N}\sum_{i=1}^N \|R_i\|^2 = O_P(T^{-1}d_T\eta_{1NT}^2 + d_T^2\eta_{1NT}^4 + d_T T^{-2} + \lambda^2)$.

Let $\hat{Q}_1$ =diag$(\hat{Q}_{1,xx}, ..., \hat{Q}_{N,xx})$ and $\hat{Q}_2$ as an $Np_1 \times Np_1$ matrix with typical blocks $\frac{1}{NT^2}x_{1,i}' M_{\hat{F}_1} x_{1,j} a_{ij}$, such that

$$\hat{Q}_2 = \begin{pmatrix} \frac{1}{NT^2}x_{1,1}'M_{\hat{F}_1}x_{1,1}a_{11} & \frac{1}{NT^2}x_{1,1}'M_{\hat{F}_1}x_{1,2}a_{12} & \cdots & \frac{1}{NT^2}x_{1,1}'M_{\hat{F}_1}x_{1,N}a_{1N} \\ \frac{1}{NT^2}x_{1,2}'M_{\hat{F}_1}x_{1,1}a_{21} & \frac{1}{NT^2}x_{1,2}'M_{\hat{F}_1}x_{1,2}a_{22} & \cdots & \frac{1}{NT^2}x_{1,2}'M_{\hat{F}_1}x_{1,N}a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{NT^2}x_{1,N}'M_{\hat{F}_1}x_{1,1}a_{N1} & \frac{1}{NT^2}x_{1,N}'M_{\hat{F}_1}x_{1,2}a_{N2} & \cdots & \frac{1}{NT^2}x_{1,N}'M_{\hat{F}_1}x_{1,N}a_{NN} \end{pmatrix}.$$

Let $R = (R_1', ..., R_N')'$. Then (A.11) implies that $(\hat{Q}_1 - \hat{Q}_2)\text{vec}(\hat{\mathbf{b}}_1) = R$. It follows that

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

By Assumption 3.2(iv) and Lemma A.5(i), we can readily show that $\mu_{\min}(\hat{Q}_1 - \hat{Q}_2) \geq \rho_{\min}/2 > 0$ w.p.a.1. Then $\eta_{1NT}^2 \equiv \frac{1}{N}\|\hat{\mathbf{b}}_1\|^2 \leq \frac{\rho_{\min}^2}{4N}\sum_{i=1}^N \|R_i\|^2 = O_P(T^{-1}d_T\eta_{1NT}^2 + d_T^2\eta_{1NT}^4 + d_T T^{-2} + \lambda^2)$. This implies that $\frac{1}{N}\|\hat{\mathbf{b}}_1\|^2 = \frac{1}{N}\sum_{i=1}^N \|\hat{b}_{1,i}\|^2 = O_P\left(d_T T^{-2} + \lambda^2\right)$.

Next, we want to strengthen the last result to the stronger version: $\frac{1}{N}\sum_{i=1}^N \|\hat{b}_{1,i}\|^2 = O_P(d_T T^{-2})$. Let $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0 + d_T^{1/2}T^{-1}\boldsymbol{v}$, where $\boldsymbol{v} = (v_1, ..., v_N)$ is a $p_1 \times N$ matrix. Let v =vec$(\boldsymbol{v})$. 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

$$P\left\{\inf_{\frac{1}{N}\sum_{i=1}^N \|v_i\|^2 = L} Q_{NT}^{\lambda,K}(\boldsymbol{\beta}_1 + d_T^{1/2}T^{-1}v, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\alpha}}, \hat{F}_1) > Q_{NT}^{\lambda,K}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, \boldsymbol{\alpha}^0, \hat{F}_1)\right\} \geq 1 - \epsilon^*,$$

regardless of the property of $\hat{\boldsymbol{\beta}}_2$, $\hat{F}_1$ and $\hat{\alpha}$. This implies that w.p.a.1 there is a local minimum $\hat{\boldsymbol{\beta}}_1 = (\hat{\beta}_1, ..., \hat{\beta}_N)$ such that $\frac{1}{N}\sum_{i=1}^N \|\hat{b}_{1,i}\|^2 = O_P(T^{-2})$. Note that

$$\begin{aligned}
& T^2\left[Q_{NT}^{\lambda,K}(\boldsymbol{\beta}_1 + d_T^{1/2}T^{-1}v, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\alpha}}, \hat{F}_1) - Q_{NT}^{\lambda,K}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, \boldsymbol{\alpha}^0, \hat{F}_1)\right] \\
& \geq \frac{d_T^{1/2}}{N}\sum_{i=1}^N \left(\frac{d_T^{1/2}}{T^2}v_i'x_{1,i}'M_{\hat{F}_1}x_{1,i}'v_i - \frac{2}{T}v_i'x_{1,i}'M_{\hat{F}_1}(F_1^0 - \hat{F}_1 H_1)\lambda_{1i}^0 - \frac{2}{T}v_i'x_{1,i}'M_{\hat{F}_1}\hat{u}_i^*\right) \\
& = \frac{d_T}{N}\sum_{i=1}^N \frac{1}{T^2}v_i'x_{1,i}'M_{\hat{F}_1}x_{1,i}'v_i \\
& \quad - \frac{2d_T^{1/2}}{N}\sum_{i=1}^N v_i'\left\{T \cdot R_{2i} + \frac{1}{T}x_{1,i}'M_{\hat{F}_1}\hat{u}_i^* + \frac{1}{NT}\sum_{j=1}^N a_{ij}x_{1,i}'M_{\hat{F}_1}x_{1,j}\hat{b}_{1,j} - \frac{1}{NT}\sum_{j=1}^N a_{ij}x_{1,i}'M_{\hat{F}_1}u_j\right\} \\
& \equiv D_{1NT} - 2D_{2NT},
\end{aligned}$$

where $R_{2i} = \frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 - \frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_{1,j}a_{ij}\hat{b}_j + \frac{1}{NT^2}\sum_{j=1}^N a_{ij}x_{1,i}'M_{\hat{F}_1}u_j$ as defined in Lemma A.6. By Assumption 3.2(iv) and Lemma A.5(iv), $D_{1NT} = \frac{d_T}{N}v'\hat{Q}_1 v \geq d_T\mu_{\min}(\hat{Q}_1)N^{-1}\|\boldsymbol{v}\|^2 \geq d_T\rho_{\min}N^{-1}\|\boldsymbol{v}\|^2/2$ w.p.a.1. Note that $|D_{2NT}| \leq \{\frac{1}{N}\sum_{i=1}^N \|v_i\|^2\}^{1/2}\sum_{l=1}^4 (D_{2NT,l})^{1/2}$, where $D_{2NT,1} =$

$\frac{T^2}{d_T N}\sum_{i=1}^N \|R_{2i}\|^2$, $D_{2NT,2}=\frac{1}{d_T NT^2}\sum_{i=1}^N \|x_{1,i}' M_{\hat{F}_1}\hat{u}_i^*\|^2$, $D_{2NT,3}=\frac{1}{d_T N^3T^2}\sum_{i=1}^N\sum_{j=1}^N \|a_{ij}x_{1,i}'M_{\hat{F}_1}x_{1,j}\hat{b}_{1,j}\|^2$, and $D_{2NT,4}=\frac{1}{d_T N^3T^2}\sum_{i=1}^N\sum_{j=1}^N \|a_{ij}x_{1,i}'M_{\hat{F}_1}u_j\|^2$. By Lemmas A.6(i)-(ii) and A.5(i), we can show that $D_{2NT,1}=T^2 O_P(T^{-5/2}C_{NT}^{-1}+T^{-1}\eta_{1NT}^2+d_T^2\eta_{1NT}^4+T^{-2}\delta_{NT}^{-2})=o_P(1)$, and $D_{2NT,2}\leq \frac{2T^2}{d_T N}\sum_{i=1}^N ||\frac{1}{T^2}x_{1,i}'(M_{\hat{F}_1}-M_{F_1^0})\hat{u}_i^*||^2+\frac{2}{d_T N}\sum_{i=1}^N ||\frac{1}{T}x_i' M_{F_1^0}\hat{u}_i^*||^2 = TO_P(\eta_{1NT}^2+T^{-1}C_{NT}^{-2})+o_P(1)=o_P(1)$. Next,

$$
\begin{aligned}
D_{2NT,3} &\leq \frac{1}{d_T}\frac{1}{N^3T^2}\sum_{i=1}^N\sum_{j=1}^N \|a_{ij}\|^2 \left\|x_{1,i}'M_{\hat{F}_1}x_{1,j}\hat{b}_{1,j}\right\|^2 \\
&\leq \frac{T^2}{N}\left[\mu_{\min}\left(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0\right)\right]^{-2}\left\{\max_{1\leq j\leq N}\frac{\|x_{1,j}\|^2}{d_T T^2}\right\}\max_{1\leq j\leq N}\|\lambda_{1j}^0\|^2\left\{\frac{1}{NT^2}\sum_{i=1}^N \|\lambda_{1i}^0\|^2\|x_{1,i}\|^2\right\}\frac{1}{N}\sum_{j=1}^N\left\|\hat{b}_{1,j}\right\|^2 \\
&= \frac{T^2}{N}O_P(1)\,O_P(1)\,O_P(1)O_P(1)\,O_P\left(d_T T^{-2}+\lambda^2\right)=o_P(1),
\end{aligned}
$$

where we use the fact that $\max_{1\leq j\leq N}\frac{\|x_{1,j}\|^2}{d_T T^2}=O_P(1)$ by Lemma A.2(i), $\max_{1\leq j\leq N}\left\|\lambda_{1j}^0\right\|^2=O_P(1)$ by Assumption 3.2(i), and $\frac{1}{NT^2}\sum_{i=1}^N\left\|\lambda_{1i}^0\right\|^2\|x_{1,i}\|^2=O_P(1)$ by Markov inequality and $\frac{1}{N}||\hat{\mathbf{b}}_1||^2=O_P\left(d_T T^{-2}+\lambda^2\right)$. Similarly, we have by Lemma A.5(i),

$$
\begin{aligned}
D_{2NT,4} &\leq \frac{1}{d_T}\frac{1}{N^3T^2}\sum_{i=1}^N\sum_{j=1}^N \|a_{ij}\|^2\left\|x_{1,i}'M_{\hat{F}_1}u_j\right\|^2 \\
&\leq \frac{1}{d_T}\left[\mu_{\min}\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)\right]^{-2}\frac{2}{N^3T^2}\sum_{i=1}^N\sum_{j=1}^N\|\lambda_{1i}^0\|^2\|\lambda_{1j}^0\|^2\left\{\left\|x_{1,i}'(M_{\hat{F}_1}-M_{F_1^0})u_j\right\|^2+\left\|x_{1,i}'M_{F_1^0}u_j\right\|^2\right\} \\
&= \frac{1}{d_T}O_P(N^{-1}T(\eta_{1NT}^2+T^{-1}C_{NT}^{-2})+1=o_P(1).
\end{aligned}
$$

It follows that $|D_{2NT}|=d_T N^{-1/2}\|\boldsymbol{v}\|\,o_P(1)$. Then D_{1NT} dominates D_{2NT} for sufficiently large L. That is, $T^2\left[Q_{NT}^{\lambda,K}(\boldsymbol{\beta}_1+d_T^{1/2}T^{-1}v,\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\alpha}},\hat{F}_1)-Q_{NT}^{\lambda,K}(\boldsymbol{\beta}_1^0,\boldsymbol{\beta}_2^0,\boldsymbol{\alpha}^0,\hat{F}_1)\right]>0$ for sufficiently large L. Consequently, the result in (i) follows.

(ii) We study the probability bound for each term on the right side of (A.10). For the first term, we have by Lemma A.6(i) and straightforward calculations

$$
\begin{aligned}
\left\|\frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}\hat{u}_i^*\right\| &\leq \left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}\hat{u}_i^*\right\|+\left\|\frac{1}{T^2}x_{1,i}'(M_{\hat{F}_1}-M_{F_1^0})\hat{u}_i^*\right\| \\
&= O_P(T^{-1})+O_P(T^{-1/2}\eta_{1NT}+T^{-1}C_{NT}^{-1})=O_P(T^{-1}). \qquad \text{(A.12)}
\end{aligned}
$$

For the second term, we can readily apply Lemmas A.6(ii), A.5(i) and A.3(iii), and Theorem 3.2(i) to obtain

$$
\begin{aligned}
\left\|\frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right\| &\leq \|R_{2i}\|+\left\|\frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_j\hat{b}_j a_{ij}\right\|+\left\|\frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}u_j a_{ij}\right\| \\
&= O_P(T^{-5/4}C_{NT}^{-1/2}+T^{-1/2}\eta_{1NT}+d_T\eta_{1NT}^2+T^{-1}\delta_{NT})+O_P(\eta_{1NT})+O_P(d_T T^{-1}) \\
&= O_P(d_T T^{-1}). \qquad \text{(A.13)}
\end{aligned}
$$

The third term is $O_P(\lambda)$. By Lemma A.5(i), $\mu_{\min}(\frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}x_{1,i}) = \mu_{\min}(\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{1,i}) + o_P(1)$. Noting that $(\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{1,i})^{-1}$ is the principal $p\times p$ submatrix of $(\frac{1}{T^2}W_i'W_i)^{-1}$, $\mu_{\min}(\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{1,i}) \geq \mu_{\min}(\frac{1}{T^2}W_i'W_i)$, and the last object is bounded away from zero w.p.a.1. It follows that $\hat{b}_{1,i} = O_P(d_T T^{-1} + \lambda)$ for $i = 1, 2, ..., N$.

Note that $\hat{\beta}_{2,i} = \left(x_{2,i}'M_{\hat{F}_1}x_{2,i}\right)^{-1} x_{2,i}'M_{\hat{F}_1}(y_i - x_{1,i}\hat{\beta}_{1,i})$ and

$$\|\hat{b}_{2,i}\| = \left\|\left(\frac{1}{T}x_{2,i}'M_{\hat{F}_1}x_{2,i}\right)^{-1}\right\|_{sp}\left\{\left\|\frac{1}{T}x_{2,i}'M_{\hat{F}_1}\left(u_i + F_2^0\lambda_{2i}^0\right)\right\| + \left\|\frac{1}{T}x_{2,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right\| + \left\|\frac{1}{T}x_{2,i}'M_{\hat{F}_1}x_{1,i}\hat{b}_{1,i}\right\|\right\}.$$

By the proof of Lemma A.5(ii) and Assumption 3.2(v), we can show that $\left\|(\frac{1}{T}x_{2,i}'M_{\hat{F}_1}x_{2,i})^{-1}\right\|_{sp} \leq M$ uniformly in i w.p.a.1. Note that $\|P_{\hat{F}_1} - P_{F_1^0}\| = O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1}) = O_P(d_T^{1/2}T^{-1} + (NT)^{-1/2})$ and similarly $\frac{1}{T}\left\|\hat{F}_1H_1^{-1} - F_1^0\right\| = O_P(d_T^{1/2}T^{-1} + (NT)^{-1/2})$ by Theorem 3.1(iii) and Lemma A.5(i). Thus

$$\begin{aligned}
\left\|\frac{1}{T}x_{2,i}'M_{\hat{F}_1}(u_i + F_2^0\lambda_{2i}^0)\right\| &\leq \left\|\frac{1}{T}x_{2,i}'M_{F_1^0}(u_i + F_2^0\lambda_{2i}^0)\right\| + \frac{\|x_{2,i}\|}{\sqrt{T}}\left\|P_{\hat{F}_1} - P_{F_1^0}\right\|\frac{\|u_i + F_2^0\lambda_{2i}^0\|}{\sqrt{T}} \\
&= O_P(T^{-1/2}) + O_P(d_T^{1/2}T^{-1} + (NT)^{-1/2}) = O_P(T^{-1/2}), \\
\left\|\frac{1}{T}x_{2,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right\| &= \left\|\frac{1}{T}x_{2,i}'M_{\hat{F}_1}(\hat{F}_1H_1^{-1} - F_1^0)\lambda_{1i}^0\right\| \leq T^{1/2}\frac{\|x_{2,i}\|}{T^{1/2}}\frac{1}{T}\left\|\hat{F}_1H_1^{-1} - F_1^0\right\|\|\lambda_{1i}^0\| \\
&= T^{1/2}O_P(d_T^{1/2}T^{-1} + (NT)^{-1/2}) = O_P(d_T^{1/2}T^{-1/2} + N^{-1/2}), \\
\left\|\frac{1}{T}x_{2,i}'M_{\hat{F}_1}x_{1,i}\hat{b}_{1,i}\right\| &\leq \left\|\frac{1}{T}x_{2,i}'M_{F_1^0}x_{1,i}\right\|\|\hat{b}_{1,i}\| + \frac{\|x_{2,i}\|}{\sqrt{T}}\left\|P_{\hat{F}_1} - P_{F_1^0}\right\|\frac{\|x_{1,i}\|}{T}\sqrt{T}\|\hat{b}_{1,i}\| \\
&= O_P(d_T T^{-1} + \lambda) + O_P(d_T T^{-1} + (NT)^{-1/2})o_P(1) = O_P(d_T T^{-1} + \lambda),
\end{aligned}$$

where $\frac{1}{T}x_{2,i}'M_{F_1^0}(u_i+F_2^0\lambda_{2i}^0) = \frac{1}{T}x_{2,i}'(u_i+F_2^0\lambda_{2i}^0)+\frac{1}{T}\frac{x_{2,i}'F_1^0}{T}\left(\frac{F_1^{0\prime}F_1^0}{T}\right)^{-1}\frac{F_1^{0\prime}(u_i+F_2^0\lambda_{2i}^0)}{T} = \frac{1}{T}x_{2,i}'\left(u_i + F_2^0\lambda_{2i}^0\right) + O_P(T^{-1}) = O_P(T^{-1/2})$ by Assumption 3.1(v). It follows that $\left\|\hat{b}_{2,i}\right\| = O_P(d_T^{1/2}T^{-1/2} + N^{-1/2})$ for $i = 1, 2, ..., N$.

(iii) Let $P_{NT}(\boldsymbol{\beta}_1, \boldsymbol{\alpha}) = \frac{1}{N}\sum_{i=1}^N\prod_{k=1}^K\|\beta_{1,i} - \alpha_k\|$ and $\hat{c}_{iNT}(\alpha) = \prod_{k=1}^{K-1}\|\hat{\beta}_{1,i} - \alpha_k\| + \prod_{k=1}^{K-2}\|\hat{\beta}_{1,i} - \alpha_k\| \times \|\beta_{1,i}^0 - \alpha_K\| + ... + \prod_{k=2}^K\|\beta_{1,i}^0 - \alpha_k\|$. By SSP (2016a), we have that as $(N,T) \to \infty$,

$$\left|\prod_{k=1}^K\|\hat{\beta}_{1,i} - \alpha_k\| - \prod_{k=1}^K\|\beta_{1,i}^0 - \alpha_k\|\right| \leq \hat{c}_{iNT}(\alpha)\|\hat{\beta}_{1,i} - \beta_i^0\|,$$

where $\hat{c}_{iNT}(\alpha) \leq C_{KNT}(\alpha)(1+2\|\hat{\beta}_{1,i}-\beta_i^0\|)$ and $C_{KNT}(\alpha) = \max_{1\leq i\leq N}\max_{1\leq s\leq k\leq K-1}\prod_{k=1}^s c_{ks}\|\beta_{1,i}^0 - \alpha_k\|^{K-1-s} = \max_{1\leq l\leq K}\max_{1\leq s\leq k\leq K_0-1}\prod_{k=1}^s c_{ks}\|\alpha_l^0 - \alpha_k\|^{K-1-s} = O(1)$ with c_{ks} being finite integers. It follows that as $(N,T)\to\infty$

$$
\begin{aligned}
|P_{NT}(\hat{\boldsymbol{\beta}}_1,\boldsymbol{\alpha})-P_{NT}(\boldsymbol{\beta}_1^0,\boldsymbol{\alpha})| &\le C_{KNT}(\alpha)\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|+2C_{KNT}(\alpha)\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\\
&\le C_{KNT}(\alpha)\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2}+O_P(d_T T^{-2})=O_P(d_T^{1/2}T^{-1}). \quad \text{(A.14)}
\end{aligned}
$$

By (A.14) and the fact that $P_{NT}(\boldsymbol{\beta}_1^0,\boldsymbol{\alpha}_1^0)=0$ and that $P_{NT}(\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\alpha}}_1)-P_{NT}(\hat{\boldsymbol{\beta}}_1,\boldsymbol{\alpha}_1^0)\le 0$. we have

$$
\begin{aligned}
0 &\ge P_{NT}(\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\alpha}}_1)-P_{NT}(\hat{\boldsymbol{\beta}}_1,\boldsymbol{\alpha}_1^0)=P_{NT}(\boldsymbol{\beta}_1^0,\hat{\boldsymbol{\alpha}})-P_{NT}(\boldsymbol{\beta}_1^0,\boldsymbol{\alpha}^0)+O_P(d_T^{1/2}T^{-1})\\
&=\frac{1}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\beta_{1,i}^0-\hat{\alpha}_k\|+O_P(d_T^{1/2}T^{-1})\\
&=\frac{N_1}{N}\prod_{k=1}^{K}\|\hat{\alpha}_k-\alpha_1^0\|+\frac{N_2}{N}\prod_{k=1}^{K}\|\hat{\alpha}_k-\alpha_2^0\|+\dots+\frac{N_K}{N}\prod_{k=1}^{K}\|\hat{\alpha}_k-\alpha_K^0\|+O_P(d_T^{1/2}T^{-1}). \quad \text{(A.15)}
\end{aligned}
$$

By Assumption 3.3(i), $N_k/N\to\tau_k\in(0,1)$ for each $k=1,...K$. So (A.15) implies that $\prod_{k=1}^{K}\|\hat{\alpha}_k-\alpha_l^0\|=O_P(d_T^{1/2}T^{-1})$ for $l=1,...K$. It follows that $(\hat{\alpha}_{(1)},...,\hat{\alpha}_{(K)})-(\alpha_1^0,...,\alpha_K^0)=O_P(d_T^{1/2}T^{-1})$.

(iv) By Theorem 3.1(iii) and Theorem 3.2(i), we have $\frac{1}{T}\|\hat{F}_1-F_1^0H_1\|=O_P(d_T^{1/2}\eta_{1NT}+T^{-1/2}C_{NT}^{-1})=O_P(d_T^{1/2}T^{-1}+(NT)^{-1/2})$. ■

To prove Theorem 3.3 we use the following two lemmas.

Lemma A.7 *Suppose that Assumptions 3.1-3.3 hold. Then for any $c>0$,*
(i) $P\left(\max_{1\le i\le N}\left\|\frac{1}{T^2}x_{1,i}'\hat{u}_i^*\right\|>c\psi_{NT}\right)=o(N^{-1})$,
(ii) $P\left(\max_{1\le i\le N}\left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}\hat{u}_i^*\right\|>cd_T\psi_{NT}\right)=o(N^{-1})$.

Lemma A.8 *Suppose that Assumptions 3.1-3.3 hold. Then for any $c>0$,*
(i) $P\left(\max_{1\le i\le N}\|R_{1i}\|>cd_T^{1/2}(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})\left(\psi_{NT}+T^{-1/2}(\log T)^3\right)\right)=o(N^{-1})$,
(ii) $P\left(\max_{1\le i\le N}\|R_{2i}\|>cd_T^{1/2}\varsigma_{2NT}\right)=o(N^{-1})$,
(iii) $P\left(\max_{1\le i\le N}\|R_{3i}\|>cd_T^{1/2}\varsigma_{1NT}\right)=o(N^{-1})$,
(iv) $P\left(\max_{1\le i\le N}\|R_{4i}\|>cd_T\psi_{NT}\right)=o(N^{-1})$,
(v) $P\left(\max_{1\le i\le N}\left\|\hat{\beta}_{1,i}-\beta_{1,i}^0\right\|>c\left(\psi_{NT}+\lambda(\log T)^{\epsilon/2}\right)\right)=o(N^{-1})$ *for any* $\epsilon>0$,
(vi) $P\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\beta}_{1,i}-\beta_{1,i}^0\right\|^2>cd_T^2\psi_{NT}^2\right)=o(N^{-1})$ *for any* $\epsilon>0$,
(vii) $P\left(\max_{1\le i\le N}\left\|\frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right\|>cd_T^{1/2}(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})\right)=o(N^{-1})$,
(viii) $P\left(\max_{1\le i\le N}\left\|\hat{\beta}_{2,i}-\beta_{2,i}^0\right\|>c(\log T)^3(d_T^{1/2}T^{-1/2}+C_{NT}^{-1})\right)=o(N^{-1})$.

Proof of Theorem 3.3. (i) Fix $k\in\{1,...,K\}$. By the consistency of $\hat{\alpha}_k$ and $\hat{\beta}_{1,i}$, we have $\hat{\beta}_{1,i}-\hat{\alpha}_l\xrightarrow{p}\alpha_k^0-\alpha_l^0\neq 0$ for all $i\in G_k^0$ and $l\neq k$. Now, suppose that $\|\hat{\beta}_{1,i}-\hat{\alpha}_k\|\neq 0$ for some $i\in G_k^0$. Then the first order condition (with respect to $\beta_{1,i}$) for the minimization of the objective function

(2.8) implies that

$$0_{p_1\times 1} = -\frac{2}{T}x_{1,i}'M_{F_1^0}\hat{u}_i^* + \frac{2}{T}x_{1,i}'(M_{F_1^0}-M_{\hat{F}_1})\hat{u}_i^* - \frac{2}{T}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 + \frac{2}{T^2}x_{1,i}'M_{\hat{F}_1}x_{1,i}'T(\hat{\alpha}_k-\alpha_k^0)$$
$$+\left(\frac{2}{T^2}x_{1,i}'M_{\hat{F}_1}x_{1,i}' + \frac{\lambda\hat{c}_{ki}}{\|\hat{\beta}_{1,i}-\hat{\alpha}_k\|}I_{p_1}\right)T(\hat{\beta}_{1,i}-\hat{\alpha}_k) + T\lambda\sum_{j=1,j\neq k}^{K}\hat{e}_{ij}\prod_{l=1,l\neq j}^{K}\|\hat{\beta}_{1,i}-\hat{\alpha}_l\|$$
$$\equiv -\hat{A}_{1i}+\hat{A}_{2i}-\hat{A}_{3i}+\hat{A}_{4i}+\hat{A}_{5i}+\hat{A}_{6i},\ \text{say},$$

where $\hat{e}_{ij}$ are defined in the proof of Theorem 3.2(i), $\hat{c}_{ki}=\prod_{l=1,l\neq k}^{K}\|\hat{\beta}_{1,i}-\hat{\alpha}_l\| \xrightarrow{p} c_k^0 \equiv \prod_{l=1,l\neq k}^{K}\|\alpha_k^0-\alpha_l^0\|>0$ for $i\in G_k^0$ by Assumption 3.3(ii). Let $\Psi_{NT}=\psi_{NT}+\lambda(\log T)^{\epsilon/2}$. Let c denote a generic constant that may vary across lines. By Lemma A.8(v)-(vi), we have

$$P\left(\max_{i\in G_k^0}\left\|\hat{\beta}_{1,i}-\beta_{1,i}^0\right\|>c\Psi_{NT}\right)=o(N^{-1}) \text{ and } P\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\beta}_{1,i}-\beta_{1,i}^0\right\|^2>cd_T^2\psi_{NT}^2\right)=o(N^{-1}). \tag{A.16}$$

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

$$P(\|\hat{\alpha}_k-\alpha_k^0\|>cd_T\psi_{NT})=o(N^{-1}) \text{ and } P(\max_{i\in G_k^0}\left|\hat{c}_{ki}-c_k^0\right|\geq c_k^0/2)=o(N^{-1}). \tag{A.17}$$

By (A.16)-(A.17) and the fact that $\max_{i\in G_k^0}\frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}x_{1,i}'\leq cd_T\rho_{\max}$ a.s., $P\left(\max_{i\in G_k^0}\left\|\hat{A}_{4i}\right\|>cd_T^2T\psi_{NT}\right)$ $=o(N^{-1})$ and $P\left(\max_{i\in G_k^0}\left\|\hat{A}_{6i}\right\|>c\lambda T\Psi_{NT}\right)=o(N^{-1})$. By Lemmas A.7(ii), A.8(i), and A.8(vii), we have $P\left(\max_{i\in G_k^0}\|\hat{A}_{1i}\|>cTd_T\psi_{NT}\right)=o(N^{-1})$, $P\left(\max_{i\in G_k^0}\|\hat{A}_{3i}\|>cd_T^{1/2}(T\eta_{1NT}+T^{1/2}C_{NT}^{-1})\right)$ $=o(N^{-1})$, and $P\left(\max_{i\in G_k^0}\|\hat{A}_{2i}\|>cd_T^{1/2}(T\eta_{1NT}+T^{1/2}C_{NT}^{-1})\left(\psi_{NT}+T^{-1/2}(\log T)^3\right)\right)=o(N^{-1})$. For $\hat{A}_{5i}$, we have

$$(\hat{\beta}_{1,i}-\hat{\alpha}_k)'\hat{A}_{5i}=(\hat{\beta}_{1,i}-\hat{\alpha}_k)'\left(\frac{2}{T^2}x_{1,i}'M_{\hat{F}_1}x_{1,i}'+\frac{\lambda\hat{c}_{ki}}{\|\hat{\beta}_{1,i}-\hat{\alpha}_k\|}I_{p_1}\right)T(\hat{\beta}_{1,i}-\hat{\alpha}_k)$$
$$\geq 2\hat{Q}_{i,xx}T\|\hat{\beta}_{1,i}-\hat{\alpha}_k\|^2+T\lambda\hat{c}_{ki}\|\hat{\beta}_{1,i}-\hat{\alpha}_k\|\geq cT\lambda c_k^0\|\hat{\beta}_{1,i}-\hat{\alpha}_k\|.$$

Combining the above results yields $P(\Xi_{k,NT})=1-o(N^{-1})$, where

$$\Xi_{k,NT}=\left\{\max_{i\in G_k^0}\|\hat{A}_{2i}\|<cd_T^{1/2}\left(T\eta_{1NT}+T^{1/2}C_{NT}^{-1}\right)\left(\psi_{NT}+T^{-1/2}(\log T)^3\right)\right\}$$
$$\cap\left\{\max_{i\in G_k^0}\|\hat{A}_{3i}\|<cd_T^{1/2}(T\eta_{1NT}+T^{1/2}C_{NT}^{-1})\right\}\cap\left\{\max_{i\in G_k^0}\left|\hat{c}_{ki}-c_k^0\right|<c_k^0/2\right\}$$
$$\cap\left\{\max_{i\in G_k^0}\left\|\hat{A}_{4i}\right\|<cd_T^2T\psi_{NT}\right\}\cap\left\{\max_{i\in G_k^0}\left\|\hat{A}_{6i}\right\|<c\lambda T\Psi_{NT}\right\}.$$

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

$$
\begin{aligned}
&\left|(\hat{\beta}_{1,i} - \hat{\alpha}_k)'(\hat{A}_{2i} + \hat{A}_{3i} + \hat{A}_{4i} + \hat{A}_{5i} + \hat{A}_{6i})\right| \\
\geq &\left|(\hat{\beta}_{1,i} - \hat{\alpha}_k)'\hat{A}_{5i}\right| - \left|(\hat{\beta}_i - \hat{\alpha}_k)'(\hat{A}_{2i} + \hat{A}_{3i} + \hat{A}_{4i} + \hat{A}_{6i})\right| \\
\geq &\left\{cT\lambda c_k^0 - c\left(Td_T^{1/2}\eta_{1NT} + T^{1/2}d_T^{1/2}C_{NT}^{-1} + Td_T^2\psi_{NT} + \lambda T\Psi_{NT}\right)\right\}\|\hat{\beta}_{1,i} - \hat{\alpha}_k\| \\
\geq &cT\lambda c_k^0\|\hat{\beta}_{1,i} - \hat{\alpha}_k\|/2,
\end{aligned}
$$

where the last inequality follows by the fact that $Td_T^{1/2}\eta_{1NT} + T^{1/2}d_T^{1/2}C_{NT}^{-1} + Td_T^2\psi_{NT} + \lambda T\Psi_{NT}$ $= o(T\lambda)$ for sufficiently large (N, T) by Assumption 3.3(iv). It follows that

$$
\begin{aligned}
P(\hat{E}_{kNT,i}) &= P(i \notin \hat{G}_k | i \in G_k^0) = P(\hat{A}_{1i} = \hat{A}_{2i} + \hat{A}_{3i} + \hat{A}_{4i} + \hat{A}_{5i} + \hat{A}_{6i}) \\
&\leq P\left(|(\hat{\beta}_{1,i} - \hat{\alpha}_k)'\hat{A}_{1i}| \geq |(\hat{\beta}_{1,i} - \hat{\alpha}_k)'\hat{A}_{5i} - (\hat{\beta}_{1,i} - \hat{\alpha}_k)'(\hat{A}_{2i} + \hat{A}_{3i} + \hat{A}_{4i} + \hat{A}_{6i})\right) \\
&\leq P(\|\hat{A}_{i1}\| \geq cT\lambda c_k^0/4, \Xi_{kNT}) + o(N^{-1}) \to 0 \quad \text{as} \quad (N, T) \to \infty,
\end{aligned}
$$

where the last inequality follows because $T\lambda \gg Td_T\psi_{NT}$ by Assumption 3.3(iv). Consequently, we can conclude that w.p.a.1, $\hat{\beta}_{1,i} - \hat{\alpha}_k$ must be in a position where $\|\beta_{1,i} - \alpha_k\|$ is not differentiable with respect to β_i for any $i \in G_k^0$. That is, $P(\|\hat{\beta}_{1,i} - \hat{\alpha}_k\| = 0 | i \in G_k^0) = 1 - o(N^{-1})$ as $(N, T) \to \infty$.

For uniform consistency, we have that $P(\cup_{k=1}^K \hat{E}_{kNT}) \leq \sum_{k=1}^K P(\hat{E}_{kNT}) \leq \sum_{k=1}^K \sum_{i \in G_k^0} P(\hat{E}_{kNT,i}) \leq$ $N \max_{1 \leq i \leq N} P(\|\hat{A}_{i1}\| \geq cT\lambda c_k^0/4) + o(1) \to 0$ as $(N, T) \to \infty$.This completes the proof of (i). Then the proof of (ii) directly follows SSP (2016a) and is therefore omitted. ■

To prove Theorem 3.4, we use the following two lemmas.

Lemma A.9 *Suppose that Assumptions 3.1-3.3 hold. Then for any $k = 1, ..., K$,*

(i) $\frac{1}{N_k T^2}\sum_{i \in \hat{G}_k} x_{1,i}' M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 = \frac{1}{N_k T^2}\sum_{i \in \hat{G}_k} \frac{1}{N}\sum_{j=1}^N x_{1,i}' M_{\hat{F}_1} x_{1,j} a_{ij} \hat{b}_{1,j} - \frac{1}{N_k T^2}\sum_{i \in \hat{G}_k} \frac{1}{N}\sum_{j=1}^N a_{ij} x_{1,i}' M_{\hat{F}_1} u_j - \frac{1}{N_k T^2}\sum_{i \in \hat{G}_k} \frac{1}{N}\sum_{j=1}^N a_{ij} x_{1,i}' M_{\hat{F}_1} F_2^0 \lambda_{2j}^0 + o_P(N^{-1/2}T^{-1})$,

(ii) $\frac{1}{N_k T^2}\sum_{i \in \hat{G}_k} x_{1,i}' M_{\hat{F}_1} x_{1,i} = \frac{1}{N_k T^2}\sum_{i \in G_k^0} x_{1,i}' M_{F_1^0} x_{1,i} + o_P(1)$,

(iii) $\frac{1}{\sqrt{N_k}T}\sum_{i \in \hat{G}_k} x_i' M_{\hat{F}_1}\left[(u_i + F_2^0\lambda_{2i}^0) - \frac{1}{N}\sum_{j=1}^N (u_j + F_2^0\lambda_{2j}^0) a_{ij}\right] = U_{kNT} + o_P(1)$,

(iv) $\frac{1}{N_k T^2}\sum_{i \in \hat{G}_k} \frac{1}{N}\sum_{j \in \hat{G}_l} x_{1,i}' M_{\hat{F}_1} x_{1,j} a_{ij} = \frac{1}{N_k T^2}\sum_{i \in G_k^0} \frac{1}{N}\sum_{j \in G_l^0} x_{1,i}' M_{F_1^0} x_{1,j} a_{ij} + o_P(1)$,

(v) $\frac{1}{N_k T^2}\sum_{i \in \hat{G}_k} x_{1,i}' M_{\hat{F}_1} x_{2,i} \hat{b}_{2,i} = o_P(N^{-1/2}T^{-1})$.

Lemma A.10 *Suppose that Assumptions 3.1-3.3 hold. Then*

(i) $Q_{NT} \xrightarrow{d} Q_0$,

(ii) $U_{kNT} = V_{kNT} + B_{kNT} + o_P(1)$ *for* $k = 1, ..., K$,

(iii) $V_{NT} \xrightarrow{d} N(0, \Omega_0)$ *conditional on* $\mathcal{C}$ *where* $\Omega_0 = \lim_{N,T \to \infty} \Omega_{NT}$.

Proof of Theorem 3.4. (i) To study of the oracle property of the C-Lasso estimator, we invoke the sub-differential calculus. A necessary and sufficient condition for $\{\hat{\beta}_{1,i}\}$ and $\{\hat{\alpha}_k\}$ to minimize the objective function in (2.11) is that for each $i = 1, ..., N$ (resp. $k = 1, ..., K$), the null vector $0_{p_1 \times 1}$ belongs to the sub-differential of $Q_{NT}^{\lambda,K}(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \boldsymbol{\alpha}, \hat{F}_1)$ with respect to $\beta_{1,i}$ (resp. α_k) evaluated at

$\{\hat\beta_{1,i}\}$ and $\{\hat\alpha_k\}$. That is, for each $i=1,...,N$ and $k=1,...,K$, we have

$$0_{p_1\times 1} = -\frac{2}{NT^2}x_{1,i}'M_{\hat F_1}(y_i - x_{1,i}\hat\beta_{1,i} - x_{2,i}\hat\beta_{2,i}) + \frac{\lambda}{N}\sum_{j=1}^{K}\hat e_{ij}\prod_{l=1,l\neq j}^{K}\|\hat\beta_{1,i}-\hat\alpha_l\|, \tag{A.18}$$

$$0_{p_1\times 1} = \frac{\lambda}{N}\sum_{i=1}^{N}\hat e_{ik}\prod_{l=1,l\neq k}^{K}\|\hat\beta_{1,i}-\hat\alpha_l\|, \tag{A.19}$$

where $\hat e_{ij} = \frac{\hat\beta_{1,i}-\hat\alpha_j}{\|\hat\beta_{1,i}-\hat\alpha_j\|}$ if $\|\hat\beta_i - \hat\alpha_j\| \neq 0$ and $\|\hat e_{ij}\|\leq 1$ if $\|\hat\beta_{1,i}-\hat\alpha_j\|=0$. First, we observe that $\|\hat\beta_{1,i}-\hat\alpha_k\|=0$ for any $i\in\hat G_k$ by the definition of $\hat G_k$, implying that $\hat\beta_{1,i}-\hat\alpha_l \to \alpha_k^0-\alpha_l^0\neq 0$ for any $i\in\hat G_k$ and $l\neq k$ by Assumption 3.3(ii). It follows that $\|\hat e_{ik}\|\leq 1$ for any $i\in\hat G_k$ and $\hat e_{ij} = \frac{\hat\beta_{1,i}-\hat\alpha_j}{\|\hat\beta_{1,i}-\hat\alpha_j\|} = \frac{\hat\alpha_k-\hat\alpha_j}{\|\hat\alpha_k-\hat\alpha_j\|}$ w.p.a.1 for any $i\in\hat G_k$ and $j\neq k$. This further implies that w.p.a.1 $\sum_{i\in\hat G_k}\sum_{j=1,j\neq k}^{K}\hat e_{ij}\prod_{l=1,l\neq j}^{K}\|\hat\beta_{1,i}-\hat\alpha_l\| = \sum_{i\in\hat G_k}\sum_{j=1,j\neq k}^{K}\frac{\hat\alpha_k-\hat\alpha_j}{\|\hat\alpha_k-\hat\alpha_j\|}\prod_{l=1,l\neq j}^{K}\|\hat\alpha_k-\hat\alpha_l\| = 0_{p_1\times 1}$, and

$$\begin{aligned}
0_{p_1\times 1} &= \sum_{i=1}^{N}\hat e_{ik}\prod_{l=1,l\neq k}^{K}\|\hat\beta_{1,i}-\hat\alpha_l\| \\
&= \sum_{i\in\hat G_k}\hat e_{ik}\prod_{l=1,l\neq k}^{K}\|\hat\alpha_k-\hat\alpha_l\| + \sum_{i\in\hat G_0}\hat e_{ik}\prod_{l=1,l\neq k}^{K}\|\hat\beta_{1,i}-\hat\alpha_l\| + \sum_{j=1,j\neq k}^{K}\sum_{i\in\hat G_j}\hat e_{ik}\prod_{l=1,l\neq k}^{K}\|\hat\alpha_j-\hat\alpha_l\| \\
&= \sum_{i\in\hat G_k}\hat e_{ik}\prod_{l=1,l\neq k}^{K}\|\hat\alpha_k-\hat\alpha_l\| + \sum_{i\in\hat G_0}\hat e_{ik}\prod_{l=1,l\neq k}^{K}\|\hat\beta_{1,i}-\hat\alpha_l\|.
\end{aligned} \tag{A.20}$$

Then by (A.18)–(A.20) we have

$$\frac{2}{N_kT^2}\sum_{i\in\hat G_k}x_{1,i}'M_{\hat F_1}(y_i - x_{1,i}\hat\alpha_k - x_{2,i}\hat\beta_{2,i}) + \frac{\lambda}{N}\sum_{i\in\hat G_0}\hat e_{ik}\prod_{l=1,l\neq k}^{K}\|\hat\beta_{1,i}-\hat\alpha_l\| = 0_{p_1\times 1}. \tag{A.21}$$

Noting that $\mathbf{1}\{i\in\hat G_k\} = \mathbf{1}\{i\in G_k^0\} + \mathbf{1}\{i\in\hat G_k\setminus G_k^0\} - \mathbf{1}\{i\in G_k^0\setminus\hat G_k\}$ and $y_i = x_{1,i}\alpha_k^0 + x_{2,i}\beta_{2,i}^0 + F_1^0\lambda_{1i}^0 + F_2^0\lambda_{2i}^0 + u_i$ when $i\in G_k^0$, we have

$$\begin{aligned}
&\frac{1}{N_kT^2}\sum_{i\in\hat G_k}x_{1,i}'M_{\hat F_1}y_i = \frac{1}{N_kT^2}\sum_{i\in\hat G_k}x_{1,i}'M_{\hat F_1}x_{1,i}\beta_i^0 + \frac{1}{N_kT^2}\sum_{i\in\hat G_k}x_{1,i}'M_{\hat F_1}F_1^0\lambda_{1i}^0 + \frac{1}{N_kT^2}\sum_{i\in\hat G_k}x_i'M_{\hat F_1}\hat u_i^* \\
=&\frac{1}{N_kT^2}\sum_{i\in G_k^0}x_{1,i}'M_{\hat F_1}x_{1,i}\alpha_k^0 + \frac{1}{N_kT^2}\sum_{i\in\hat G_k\setminus G_k^0}x_{1,i}'M_{\hat F_1}x_{1,i}\beta_{1,i}^0 - \frac{1}{N_kT^2}\sum_{i\in G_k^0\setminus\hat G_k}x_{1,i}'M_{\hat F_1}x_{1,i}\alpha_k^0 \\
&+\frac{1}{N_kT^2}\sum_{i\in\hat G_k}x_{1,i}'M_{\hat F_1}x_{2,i}\beta_{2,i}^0 + \frac{1}{N_kT^2}\sum_{i\in\hat G_k}x_{1,i}'M_{\hat F_1}F_1^0\lambda_{1i}^0 + \frac{1}{N_kT^2}\sum_{i\in\hat G_k}x_{1,i}'M_{\hat F_1}(u_i + F_2^0\lambda_{2i}^0).
\end{aligned} \tag{A.22}$$

Combining (A.21) and (A.22) yields

$$\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}x_{1,i}(\hat{\alpha}_k-\alpha_k^0)=\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0+\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}\left(u_i+F_2^0\lambda_{2i}^0\right)$$
$$+\hat{C}_{1k}-\hat{C}_{2k}+\hat{C}_{3k}-\hat{C}_{4k}, \tag{A.23}$$

where $\hat{C}_{1k}=\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k\setminus G_k^0}x_{1,i}'M_{\hat{F}_1}x_{1,i}\beta_{1,i}^0$, $\hat{C}_{2k}=\frac{1}{N_kT^2}\sum_{i\in G_k^0\setminus\hat{G}_k}x_{1,i}'M_{\hat{F}_1}x_{1,i}\alpha_k^0$, $\hat{C}_{3k}=\frac{\lambda}{2N_k}\sum_{i\in\hat{G}_0}\hat{e}_{ik}\times\prod_{l=1,l\neq k}^K\|\hat{\beta}_{1,i}-\hat{\alpha}_l\|$ and $\hat{C}_{4k}=\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}x_{2,i}\hat{b}_{2,i}$. By Theorem 3.3 and Lemmas S1.11-S1.12 in Su et al. (2016b), we have $P(N^{1/2}T\|\hat{C}_{1k}\|\geq\epsilon)\leq P(\hat{F}_{kNT})\to 0$, $P(N^{1/2}T\|\hat{C}_{2k}\|\geq\epsilon)\leq P(\hat{E}_{kNT})\to 0$, and $P(N^{1/2}T\|\hat{C}_{3k}\|\geq\epsilon)\leq\sum_{k=1}^K\sum_{i\in G_k^0}P(i\in\hat{G}_0|i\in G_k^0)\leq\sum_{k=1}^K\sum_{i\in G_k^0}P(\hat{E}_{kNT,i})=o(1)$. It follows that $\|\hat{C}_{1k}-\hat{C}_{2k}+\hat{C}_{3k}\|=o_P(N^{-1/2}T^{-1})$. By Lemma A.9 (v), $\|\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}x_{2,i}\hat{b}_{2,i}\|=o_P(N^{-1/2}T^{-1})$. We have By Lemma A.9(i), we have as $\frac{\sqrt{N}}{T}\to 0$

$$\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0=\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_{1,j}a_{ij}\hat{b}_{1,j}-\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N a_{ij}x_{1,i}'M_{\hat{F}_1}u_j$$
$$-\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N a_{ij}x_{1,i}'M_{\hat{F}_1}F_2^0\lambda_{2j}^0+o_P(N^{-1/2}T^{-1}). \tag{A.24}$$

In addition,

$$\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_{1,j}a_{ij}\hat{b}_{1,j}=\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{l=1}^K\sum_{j\in\hat{G}_l}x_{1,i}'M_{\hat{F}_1}x_{1,j}a_{ij}\left(\hat{\alpha}_l-\alpha_l^0\right)+o_P(N^{-1/2}T^{-1}) \tag{A.25}$$

by Theorem 3.3. Let $\hat{Q}_{1NT}=\text{diag}\Big(\frac{1}{N_1T^2}\sum_{i\in\hat{G}_1}x_{1,i}'M_{\hat{F}_1}x_{1,i},\ldots,\frac{1}{N_KT^2}\sum_{i\in\hat{G}_K}x_{1,i}'M_{\hat{F}_1}x_{1,i}\Big)$ and $\hat{Q}_{2NT}$ is a $Kp_1\times Kp_1$ matrix with typical blocks $\frac{1}{NN_kT}\sum_{i\in\hat{G}_k}\sum_{j\in\hat{G}_l}a_{ij}x_{1,i}'M_{\hat{F}_1}x_{1,j}$ such that

$$\hat{Q}_{2NT}=\begin{pmatrix}\frac{1}{NN_1T^2}\sum_{i\in\hat{G}_1}\sum_{j\in\hat{G}_1}a_{ij}x_{1,i}'M_{\hat{F}_1}x_{1,j}, & \cdots & \frac{1}{NN_1T^2}\sum_{i\in\hat{G}_1}\sum_{j\in\hat{G}_K}a_{ij}x_{1,i}'M_{\hat{F}_1}x_{1,j}\\ \frac{1}{NN_2T^2}\sum_{i\in\hat{G}_2}\sum_{j\in\hat{G}_1}a_{ij}x_{1,i}'M_{\hat{F}_1}x_{1,j}, & \cdots & \frac{1}{NN_2T^2}\sum_{i\in\hat{G}_2}\sum_{j\in\hat{G}_K}a_{ij}x_{1,i}'M_{\hat{F}_1}x_{1,j},\\ \vdots & \ddots & \vdots\\ \frac{1}{NN_KT^2}\sum_{i\in\hat{G}_K}\sum_{j\in\hat{G}_1}a_{ij}x_{1,i}'M_{\hat{F}_1}x_{1,j}, & \cdots & \frac{1}{NN_KT^2}\sum_{i\in\hat{G}_K}\sum_{j\in\hat{G}_K}a_{ij}x_{1,i}'M_{\hat{F}_1}x_{1,j}\end{pmatrix}.$$

Combining (A.23)–(A.25), we have $\sqrt{N}T\text{vec}(\ \hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)=(\hat{Q}_{1NT}-\hat{Q}_{2NT})^{-1}\sqrt{D_{NK}}\hat{U}_{NT}+o_P(1)$, where the kth element of $\hat{U}_{NT}$ is

$$\hat{U}_{kNT}=\frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}\left[\left(u_i+F_2^0\lambda_{2i}^0\right)-\frac{1}{N}\sum_{j=1}^N a_{ij}\left(u_j+F_2^0\lambda_{2j}^0\right)\right]$$

and $D_{NK}=\text{diag}(\frac{N}{N_1},...,\frac{N}{N_K})\otimes I_{p_1}$. By Lemma A.9(ii)-(iv), we have that $\hat{Q}_{1NT}-\hat{Q}_{2NT}=Q_{NT}+o_P(1)$, $\hat{U}_{NT}=U_{NT}+o_P(1)$, where U_{NT} and Q_{NT} are defined in Theorem 3.4. Then we have $\sqrt{N}T\text{vec}(\ \hat{\boldsymbol{\alpha}}-$

$\boldsymbol{\alpha}^0) = Q_{NT}^{-1}\sqrt{D_{NK}}U_{NT}+o_P(1)$. By Lemma A.10(ii), we have $U_{kNT}-B_{kNT,1}-B_{kNT,2} = V_{kNT}+o_P(1)$, where V_{kNT} and $B_{kNT} = B_{kNT,1} + B_{kNT,2}$ are defined in Theorem 3.4. Thus,

$$\sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) = Q_{NT}^{-1}\sqrt{D_{NK}}\,(V_{NT} + B_{NT}) + o_P(1), \tag{A.26}$$

where $V_{NT} = (V_{1NT}', ..., V_{KNT}')'$ and $B_{NT} = (B_{1NT}', ..., B_{KNT}')'$.

(ii) By Lemma A.10 (i) and (iii), $Q_{NT} \xrightarrow{d} Q_0$ and $V_{NT} \xrightarrow{d} N(0, \Omega_0)$ conditional $\mathcal{C}$. This result, in conjunction with (A.26), implies that $\sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)-\sqrt{D_{NK}}Q_{NT}^{-1}B_{NT} \xrightarrow{d} \mathcal{MN}(0, D_0Q_0^{-1}\Omega_0Q_0^{-1})$. ■

To prove Theorem 3.5 we use the following lemma.

Lemma A.11 *Suppose that Assumptions 3.1-3.3 hold. Then, as* $(N,T) \to \infty$,

(i) $\frac{1}{\sqrt{T}}\|\hat{F}_1\hat{\lambda}_{1i} - F_1^0\lambda_{1i}^0\| = O_P(\sqrt{T}\eta_{1NT}) + O_P(C_{NT}^{-1})$,

(ii) $\frac{1}{\sqrt{T}}\|\hat{F}_2 - F_2^0H_2\| = O_P(C_{NT}^{-1})$,

(iii) $\frac{1}{\sqrt{N_k}}\sum_{i\in\hat{G}_k}(\hat{\lambda}_{2i} - H_2^{-1}\lambda_{2i}^0) = o_P(1)$,

(iv) $\frac{1}{\sqrt{T}}\left\|\hat{F}_2\hat{\lambda}_{2i} - F_2^0\lambda_{2i}^0\right\| = O_P(C_{NT}^{-1})$,

(v) $\frac{1}{\sqrt{N_k}}\sum_{i\in\hat{G}_k}(\hat{\Delta}_{21,i} - \Delta_{21,i}) = o_P(1)$,

(vi) $\frac{\sqrt{N_k}}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{\varkappa}_{ts} - \varkappa_{ts})\,\mathbf{1}\{s \le t\} = o_P(1)$,

(vii) $\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i} - \Delta_{24,i}\bar{\lambda}_{2i}^0) = o_P(1)$,

(viii) $\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}[\hat{\varkappa}_{ts}\mathbf{1}\{s \le t\}\,\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i} - \varkappa_{ts}\mathbf{1}\{s \le t\}\,\Delta_{24,i}\bar{\lambda}_{2i}^0] = o_P(1)$,

(ix) $\hat{\Omega}_{NT} = \Omega_{NT} + o_P(1)$ and $\hat{\Omega}_{NT}^{+} = \Omega_{NT}^{+} + o_P(1)$,

where $\bar{\lambda}_{2i}^0 = \lambda_{2i}^0 - \frac{1}{N}\sum_{j=1}^{N}\lambda_{2j}^0 a_{ij}$.

Proof of Theorem 3.5. (i) We first consider the bias-corrected post-Lasso estimators $\text{vec}(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc})$. By construction and Theorem 3.4, we have

$$\begin{aligned}
&\sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc} - \boldsymbol{\alpha}^0)\\
=\;&\sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc} - \hat{\boldsymbol{\alpha}}) + \sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0)\\
=\;&\sqrt{D_{NK}}Q_{NT}^{-1}V_{NT} + \sqrt{D_{NK}}\left[Q_{NT}^{-1}(B_{NT,1} + B_{NT,2}) - \hat{Q}_{NT}^{-1}(\hat{B}_{NT,1} + \hat{B}_{NT,2})\right] + o_P(1).
\end{aligned}$$

It suffices to show that $\sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc} - \boldsymbol{\alpha}^0) = \sqrt{D_{NK}}Q_{NT}^{-1}V_{NT} + o_P(1)$ by showing that (i1) $\hat{Q}_{1NT} - \hat{Q}_{2NT} = Q_{NT} + o_P(1)$, (i2) $\hat{B}_{NT,1} = B_{NT,1} + o_P(1)$, and (i3) $\hat{B}_{NT,2} = B_{NT,2} + o_P(1)$. (i1) holds by Lemma A.9 (ii) and (iv). For (i2), it suffices to show that $\hat{B}_{kNT,1}-B_{kNT,1} = o_P(1)$ for $k = 1, ..., K$. By Theorem 3.3 and using arguments like those in the proof of Lemma A.9(ii), we can readily show that $\hat{B}_{kNT,1} = \tilde{B}_{kNT,1}+o_P(1)$, where $\tilde{B}_{kNT,1} = \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\hat{\Delta}_{21,i}-\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\varkappa}_{ts}\mathbf{1}\{s \le t\}\,\hat{\Delta}_{21,i}$.

It follows that

$$\begin{aligned}
\hat{B}_{kNT,1}-B_{kNT,1}= & \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{21,i}-\Delta_{21,i})-\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{1}\left\{s\le t\right\}\left[\hat{\varkappa}_{ts}\hat{\Delta}_{21,i}-\varkappa_{ts}\Delta_{21,i}\right]+o_P(1)\\
= & \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{21,i}-\Delta_{21,i})-\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\varkappa}_{ts}\mathbf{1}\left\{s\le t\right\}\left(\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{21,i}-\Delta_{21,i})\right)\\
& -\frac{\sqrt{N_k}}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{\varkappa}_{ts}-\varkappa_{ts})\mathbf{1}\left\{s\le t\right\}\left(\frac{1}{N_k}\sum_{i\in G_k^0}\Delta_{21,i}\right)+o_P(1)\\
\equiv & B_{kNT,1}\left(1\right)+B_{kNT,1}\left(2\right)+B_{kNT,1}\left(3\right)+o_P(1).
\end{aligned}$$

We can prove $\hat{B}_{kNT,1}=B_{kNT,1}+o_P(1)$ by showing that $B_{kNT,1}\left(l\right)=o_P\left(1\right)$ for $l=1,2,3$. Noting that $\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\varkappa}_{ts}\mathbf{1}\left\{s\le t\right\}\right|\le\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\hat{f}_{1t}\right\|\left\|\hat{f}_{1s}\right\|=O_P\left(1\right)$ and $\frac{1}{N_k}\sum_{i\in G_k^0}\Delta_{21,i}=O_P\left(1\right),$ these results would follow by Lemma A.11(v)-(vi). To show (i3), we first observe that

$$\begin{aligned}
B_{kNT,2}= & \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\mathbb{E}\left(x_{1,i}'|\mathcal{C}\right)M_{F_1^0}F_2^0\left(\lambda_{2i}^0-\frac{1}{N}\sum_{j=1}^{N}\lambda_{2j}^0a_{ij}\right)\\
= & \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\mathbb{E}\left(x_{1,i}'|\mathcal{C}\right)F_2^0\bar{\lambda}_{2i}^0-\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\mathbb{E}\left(x_{1,i}'|\mathcal{C}\right)P_{F_1^0}F_2^0\bar{\lambda}_{2i}^0\equiv B_{kNT,21}-B_{kNT,22},
\end{aligned}$$

where $\bar{\lambda}_{2i}^0=\lambda_{2i}^0-\frac{1}{N}\sum_{j=1}^{N}\lambda_{2j}^0a_{ij}$. Let $\phi^{f_2,f_1f_2}=(\phi^{f_2f_1}(L),\phi^{f_2f_2}(L)),\phi_i^{\varepsilon,f_1f_2}=(\phi_i^{\varepsilon f_1}\left(L\right),\phi_i^{\varepsilon f_2}\left(L\right))=(\phi^{\varepsilon f_1}\left(L\right),\phi^{\varepsilon f_2}\left(L\right))$, and $v_t^{f_1f_2}=(v_t^{f_1\prime},v_t^{f_1\prime})'$. Note that $\varepsilon_{it}=w_{it}^{\varepsilon}=\phi_i^{\varepsilon u}\left(L\right)v_{it}^{u}+\phi_i^{\varepsilon\varepsilon}\left(L\right)v_{it}^{\varepsilon}+\phi_i^{\varepsilon x_2}\left(L\right)v_{it}^{\varepsilon x_2}$ $+\phi^{\varepsilon f_1}\left(L\right)v_t^{f_1}+\phi^{\varepsilon f_2}\left(L\right)v_t^{f_2}$. By the BN decomposition and the independence of $\{v_{it}^{ux}\}$ and $\{v_s^{f_1f_2}\}$, we have

$$\begin{aligned}
f_{2t}^0= & S_4w_{it}=\phi^{f_2f_1}(L)v_t^{f_1}+\phi^{f_2f_2}(L)v_t^{f_2}=\phi^{f_2,f_1f_2}(L)v_t^{f_1f_2}\\
= & \phi^{f_2,f_1f_2}(1)v_t^{f_1f_2}+S_4\tilde{w}_{it-1}-S_4\tilde{w}_{it},\\
\mathbb{E}_{\mathcal{C}}\left(x_{1,it}\right)= & \mathbb{E}_{\mathcal{C}}\left(S_2\sum_{m=1}^{t}w_{im}\right)=\sum_{m=1}^{t}\left(\phi_i^{\varepsilon f_1}\left(L\right)v_m^{f_1}+\phi_i^{\varepsilon f_2}\left(L\right)v_m^{f_2}\right)=\phi^{\varepsilon,f_1f_2}\left(L\right)V_t^{f_1f_2}\\
= & \phi_i^{\varepsilon,f_1f_2}\left(1\right)V_t^{f_1f_2}+S_2\mathbb{E}_{\mathcal{C}}\left(\tilde{w}_{i0}-\tilde{w}_{it}\right).
\end{aligned}$$

where $V_t^{f_1f_2}=(V_t^{f_1\prime},V_t^{f_2\prime})'=\left(\sum_{m=1}^{t}v_m^{f_1\prime},\sum_{m=1}^{t}v_m^{f_2\prime}\right)'$, w_{it} and $\tilde{w}_{it}$ are defined in Assumption 3.1.

Let $B^*_{kNT,21} = \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} S_2 \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r} \phi'_{i,l} S'_4 \bar{\lambda}^0_{2i}$. It follows that

$$
\begin{aligned}
& B_{kNT,21} - B^*_{kNT,21} \\
= & \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} \frac{1}{T} \sum_{t=1}^{T} \phi_i^{\varepsilon, f_1 f_2}(L) V_t^{f_1 f_2} v_t^{f_1 f_2 \prime} \phi^{f_2, f_1 f_2}(L)' \bar{\lambda}^0_{2i} - \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} S_2 \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r} \phi'_{i,l} S_4 \bar{\lambda}^0_{2i} \\
= & \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} \frac{1}{T} \sum_{t=1}^{T} \phi^{\varepsilon, f_1 f_2}(1) (V_t^{f_1 f_2} v_t^{f_1 f_2 \prime} - I_r) \phi^{f_2, f_1 f_2}(1)' \bar{\lambda}^0_{2i} \\
& + \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} S_2 \left\{ \frac{1}{T} \sum_{t=1}^{T-1} \left(\mathbb{E}_{\mathcal{C}}(w_{it+1}) \tilde{w}'_{it} - \sum_{l=0}^{\infty} \phi_{i,l+1} \phi'_{i,l} \right) S'_4 \bar{\lambda}^0_{2i} - \frac{1}{T} \sum_{l=0}^{\infty} \phi_{i,l+1} \phi'_{i,l} S'_4 \bar{\lambda}^0_{2i} \right. \\
& - \frac{1}{T} \sum_{t=1}^{T} \left(\mathbb{E}_{\mathcal{C}}(\tilde{w}_{i0}) v_t^{f_1 f_2 \prime} \phi^{f_2, f_1 f_2}(1)' - \tilde{\phi}_{i,0} \phi_i(1)' S'_4 \right) \bar{\lambda}^0_{2i} + \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\mathcal{C}}(\tilde{w}_{it}) v_t^{f_1 f_2 \prime} \phi^{f_2, f_1 f_2}(1)' \bar{\lambda}^0_{2i} \\
& \left. - \frac{1}{T} \mathbb{E}_{\mathcal{C}} \left(\sum_{t=1}^{T} w_{it} \right) \tilde{w}'_{iT} S'_4 \bar{\lambda}^0_{2i} + \frac{1}{T} \mathbb{E}_{\mathcal{C}}(w_{i1}) \tilde{w}'_{i0} S'_4 \bar{\lambda}^0_{2i} \right\} \\
\equiv & \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} Q^{f_2}_{iT} + \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} S_2 \left\{ R^{f_2}_{iT,1} + R^{f_2}_{iT,2} + R^{f_2}_{iT,3} + R^{f_2}_{iT,4} + R^{f_2}_{iT,5} + R^{f_2}_{iT,6} \right\} S'_4 \bar{\lambda}^0_{2i},
\end{aligned}
$$

where we use the fact that $\phi_i^{\varepsilon, f_1 f_2}(1) \phi^{f_2, f_1 f_2}(1)' = S_2 \phi_i(1) \phi_i(1)' S'_4$ by construction and that $\sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r} \phi'_{i,l} = \phi_i(1) \phi_i(1)' - \sum_{l=0}^{\infty} \phi_{i,l+1} \phi'_{i,l} + \tilde{\phi}_{i,0} \phi_i(1)'$. Following the proof of Lemma A.7 in HJS, we can show that $\frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} S_2 R^{f_2}_{iT,l} S'_4 \bar{\lambda}^0_{2i} = o_P(1)$ for $l = 1, ..., 6$ and $\frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} \mathbb{E}(Q^{f_2}_{iT}) = 0$. It follows that $B_{kNT,21} = B^*_{kNT,21} + o_P(1) = \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} \Delta_{24,i} \bar{\lambda}^0_{2i} + o_P(1)$. Analogously, we have $B_{kNT,22} = B^*_{kNT,22} + o_P(1)$, where $B^*_{kNT,22} = \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \varkappa_{ts} \mathbf{1}\{s \le t\} S_2 \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r} \phi'_{i,l} S'_4 \bar{\lambda}^0_{2i}$. Let $B^*_{kNT,2} = B^*_{kNT,21} - B^*_{kNT,22}$. Then

$$
\begin{aligned}
B^*_{kNT,2} = & \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} (\mathbf{1}\{s = t\} - \varkappa_{ts} \mathbf{1}\{s \le t\}) S_2 \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r} \phi'_{i,l} S'_4 \bar{\lambda}^0_{2i} \\
= & \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t} \bar{\varkappa}_{ts} \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \left(\phi^{\varepsilon f_1}_{l+r} \phi^{f_2 f_1}_{l} + \phi^{\varepsilon f_2}_{l+r} \phi^{f_2 f_2}_{l} \right) \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} \bar{\lambda}^0_{2i} \\
= & \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t} \bar{\varkappa}_{ts} \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} \Delta_{24,i} \bar{\lambda}^0_{2i}.
\end{aligned}
$$

By Theorem 3.3 and using arguments as used in the proof of Lemma A.9(ii), we can readily show that $\hat{B}_{kNT,2} = \tilde{B}_{kNT,2} + o_P(1)$, where $\tilde{B}_{kNT,2} = \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t} \hat{\varkappa}_{ts} \hat{\Delta}_{24,i} \hat{\bar{\lambda}}_{2i}$. Thus we can prove that $\hat{B}_{NT,2} = B_{NT,2} + o_P(1)$ by showing $\tilde{B}_{kNT,2} = B^*_{kNT,2} + o_P(1)$ for $k = 1, ..., K$. Note that $\tilde{B}_{kNT,2} - B^*_{kNT,2} = \frac{1}{\sqrt{N_k}} \sum_{i \in G^0_k} (\hat{\Delta}_{24,i} \hat{\bar{\lambda}}_{2i} - \Delta_{24,i} \bar{\lambda}^0_{2i}) - \frac{1}{\sqrt{N_k} T} \sum_{i \in G^0_k} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{1}\{s \le t\} [\hat{\varkappa}_{ts} \hat{\Delta}_{24,i} \hat{\bar{\lambda}}_{2i} - \varkappa_{ts} \Delta_{24,i} \bar{\lambda}^0_{2i}] = o_P(1) - o_P(1) = o_P(1)$ by Lemma A.11(vii)-(viii). Consequently, $\hat{B}_{kNT,2} - B_{kNT,2} =$

$o_P(1)$. In sum, we have $\sqrt{N}T\mathrm{vec}(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc}-\boldsymbol{\alpha}^0)=\sqrt{D_{NK}}Q_{NT}^{-1}V_{NT}+o_P(1)$.

(ii) For the fully-modified post-Lasso estimators $\hat{\alpha}_{G_k}^{fm}$, we first consider the asymptotic distribution for the infeasible version of the fully modified post-Lasso estimator $\tilde{\alpha}_{G_k}^{fm}$. Noting that $y_i^+=x_{1,i}\alpha_k^0+x_{2,i}\beta_{2,i}^0+F_1^0\lambda_{1i}^0+F_2^0\lambda_{2i}^0+u_i^+$, by (A.23) and (A.24) and Theorem 3.3, we have

$$\begin{aligned}\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}x_{1,i}(\tilde{\alpha}_{G_k}^{fm}-\alpha_k^0)=&\frac{1}{N_kT^2}\sum_{i\in G_k^0}x_{1,i}'M_{\hat{F}_1}\left(u_i^++F_2^0\lambda_{2i}^0\right)+\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\\&-\frac{1}{\sqrt{N_k}T}B_{kNT,1}^+-\frac{1}{\sqrt{N_k}T}B_{kNT,2}+o_P(N^{-1/2}T^{-1}).\end{aligned}\tag{A.27}$$

Combining (A.25), (A.27) and Lemma A.9(i) yields

$$\begin{aligned}&\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}x_{1,i}(\tilde{\alpha}_{G_k}^{fm}-\alpha_k^0)-\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^{N}x_{1,i}'M_{\hat{F}_1}x_{1,j}a_{ij}\hat{b}_{1,j}\\=&\frac{1}{N_kT^2}\sum_{i\in G_k^0}x_{1,i}'M_{F_1^0}\left(u_i^+-\frac{1}{N}\sum_{j=1}^{N}u_j^+a_{ij}\right)+\frac{1}{N_kT^2}\sum_{i\in G_k^0}x_{1,i}'M_{F_1^0}F_2^0\left(\lambda_{2i}^0-\frac{1}{N}\sum_{j=1}^{N}\lambda_{2j}^0a_{ij}\right)\\&-\frac{1}{\sqrt{N_k}T}B_{kNT,1}^+-\frac{1}{\sqrt{N_k}T}B_{kNT,2}+o_P(N^{-1/2}T^{-1}).\end{aligned}$$

By (A.25) and Lemma A.10 (i)-(iii), we have $\sqrt{N}T\mathrm{vec}(\tilde{\boldsymbol{\alpha}}_G^{fm}-\boldsymbol{\alpha}^0)=(\hat{Q}_{1NT}-\hat{Q}_{2NT})^{-1}\sqrt{D_{NK}}[(U_{NT}^{u+}+U_{NT}^{f_2})-B_{NT,1}^+-B_{NT,2}]+o_P(1)=\sqrt{D_{NK}}Q_{NT}^{-1}V_{NT}^++o_P(1)$, where

$$\begin{aligned}U_{k,NT}^{u+}&=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_{1,i}'M_{F_1^0}\left(u_i^+-\frac{1}{N}\sum_{j=1}^{N}a_{ij}u_j^+\right),\\U_{k,NT}^{f_2}&=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_{1,i}'M_{F_1^0}\left(F_2^0-\frac{1}{N}\sum_{j=1}^{N}a_{ij}F_2^0\right),\\V_{kNT,1}^+&=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}S^{\varepsilon}\phi_i^{\dagger}(1)\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\bar{\varkappa}_{ts}\left(V_{it}^{ux}v_{is}^{ux,+\prime}\right)-\left[\mathbf{1}\{t=s\}-\varkappa_{ts}\mathbf{1}\{s\le t\}\right]I_{1+p}\right\}\phi_i^{\dagger}(1)'S^{w\prime},\\V_{kNT,2}^+&=\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N}\left\{\frac{1}{T}\mathbb{E}\left(x_{1,i}'|\mathcal{C}\right)\mathbf{1}\left\{i\in G_k^0\right\}-\frac{1}{N}\sum_{j\in G_k^0}a_{ij}\frac{1}{T}\mathbb{E}(x_{1,j}'|\mathcal{C})\right\}M_{F_1^0}u_i^+,\\V_{kNT,3}&=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}[x_{1,i}-\mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)]'M_{F_1^0}F_2^0\lambda_{2i}^0,\end{aligned}$$

and $U_{k,NT}^+=U_{k,NT}^{u+}+U_{k,NT}^{f_2}$ and $V_{kNT}^+=V_{kNT,1}^++V_{kNT,2}^++V_{kNT,3}$ are the kth block-elements of U_{NT}^+ and V_{NT}^+, respectively. We have a new error process $w_{it}^+=(u_{it}^+,\Delta x_{1,it}',\Delta f_{1t}',f_{2t}',x_{2,it}')'$ whose partial sum satisfies the multivariate invariance principle: $\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\cdot]}w_{it}^+\Rightarrow B_i^+=BM(\Omega_i^+)$. Following the proof of Lemma A.10(iii) (see also Theorem 9 in Phillips and Moon, 1999), we can show that

$V_{NT}^{+}\xrightarrow{d}N(0,\Omega_0^{+})$ conditional on $\mathcal{C}$ where $\Omega_0^{+}=\lim_{N,T\to\infty}\Omega_{NT}^{+}$ and $\Omega_{NT}^{+}=\mathrm{Var}\left(V_{NT}^{+}|\mathcal{C}\right)$. Then we have

$$\sqrt{N}T\mathrm{vec}(\tilde{\boldsymbol{\alpha}}_G^{fm}-\boldsymbol{\alpha}^0)\xrightarrow{d}\mathcal{MN}(0,\ D_0Q_0^{-1}\Omega_0^{+}Q_0^{-1}).$$

Next, we show that $\hat{\boldsymbol{\alpha}}_G^{fm}$ is asymptotically equivalent to $\tilde{\boldsymbol{\alpha}}_G^{fm}$ by showing that $\sqrt{N}T(\hat{\boldsymbol{\alpha}}_G^{fm}-\tilde{\boldsymbol{\alpha}}_G^{fm})=o_P(1)$. Note that

$$\sqrt{N}T(\hat{\boldsymbol{\alpha}}_G^{fm}-\tilde{\boldsymbol{\alpha}}_G^{fm})=\sqrt{D_{NK}}\left[(\hat{Q}_{1NT}-\hat{Q}_{2NT})^{-1}(\hat{U}_{NT}^{+}+\hat{B}_{NT,1}^{+}+\hat{B}_{NT,2})-Q_{NT}^{-1}\left(U_{NT}^{+}+B_{NT,1}^{+}+B_{NT,2}\right)\right].$$

Then it suffices to show (ii1) $\hat{Q}_{1NT}-\hat{Q}_{2NT}=Q_{NT}+o_P(1)$, (ii2) $\hat{B}_{NT,1}^{+}=B_{NT,1}^{+}+o_P(1)$,(ii3) $\hat{U}_{NT}^{+}=U_{NT}^{+}+o_P(1)$, and (ii4) $\hat{B}_{NT,2}=B_{NT,2}+o_P(1)$. (ii1) and (ii4) have been established in the proof of part (i) of the theorem. For (ii2), we can apply arguments analogous to those used in the proof of Lemma A.11(v) to establish that $\mathbb{E}_{\mathcal{C}}\left\|\frac{1}{\sqrt{N_k}}\sum_{i\in\hat{G}_k}(\hat{\Omega}_i-\Omega_i)\right\|=O_P(\frac{J}{T}+\frac{N}{J^{2q}})=o_P(1)$. Since $\Delta_{lm,i}^{+}=\Delta_{lm,i}-\Omega_{lm,i}\Omega_{mi}^{-1}\Delta_{m,i}$, this implies that $\left\|\frac{1}{\sqrt{N_k}}\sum_{i\in\hat{G}_k}(\hat{\Delta}_{21,i}^{+}-\Delta_{21,i}^{+})\right\|^2=o_P(1)$. The latter further implies that $\hat{B}_{NT,1}^{+}=B_{NT,1}^{+}+o_P(1)$. For (ii3) we can apply Theorem 3 to show that

$$\begin{aligned}
&\hat{U}_{kNT}^{+}-U_{kNT}^{+}\\
=&\hat{U}_{kNT}^{u+}-\tilde{U}_{kNT}^{u+}+\tilde{U}_{kNT}^{u+}-U_{kNT}^{u+}\\
=&\frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}\left(\hat{u}_i^{+}-\frac{1}{N}\sum_{j=1}^{N}a_{ij}\hat{u}_j^{+}\right)-\frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}\left(u_i^{+}-\frac{1}{N}\sum_{j=1}^{N}a_{ij}u_j^{+}\right)+o_P(1)\\
=&\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_{1,i}'M_{\hat{F}_1}\left(\hat{u}_i^{+}-u_i^{+}\right)-\frac{1}{\sqrt{N_k}NT}\sum_{i\in G_k^0}\sum_{j=1}^{N}x_{1,i}'M_{\hat{F}_1}\left(\hat{u}_j^{+}-u_j^{+}\right)a_{ij}+o_P(1)\\
=&\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_{1,i}'\Delta x_{1,i}\left(\Omega_{12,i}\Omega_{22i}^{-1}-\hat{\Omega}_{12,i}\hat{\Omega}_{22i}^{-1}\right)-\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_i'P_{\hat{F}_1}\Delta x_{1,i}\left(\Omega_{12,i}\ \Omega_{22i}^{-1}-\hat{\Omega}_{12,i}\hat{\Omega}_{22i}^{-1}\right)\\
&-\frac{1}{\sqrt{N_k}NT}\sum_{i\in G_k^0}\sum_{j=1}^{N}x_{1,i}'M_{\hat{F}_1}\Delta x_{1,j}\left(\Omega_{12,j}\Omega_{22j}^{-1}-\hat{\Omega}_{12,j}\hat{\Omega}_{22j}^{-1}\right)a_{ij}+o_P(1)\\
\equiv&UU_1+UU_2+UU_3+o_P(1),
\end{aligned}$$

where $\tilde{U}_{kNT}^{u+}=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_{1,i}'M_{\hat{F}_1}\left(u_i^{+}-\frac{1}{N}\sum_{j=1}^{N}a_{ij}u_j^{+}\right)$ and $\tilde{U}_{kNT}^{u+}-U_{kNT}^{u+}=o_P(1)$ by Lemma A.9(iii). Following the proof of Lemma A.11(v), we can show that $UU_l=o_P(1)$ for $l=1,2,3$. Then (ii3) follows. This completes the proof of (ii).

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

To prove Theorems 3.6-3.7 we use the following two lemmas.

Lemma A.12 *Suppose that Assumptions 3.1-3.3 and 3.6 hold. Then*

(i) For any $1\le r_1\le r_1^0$, $V_1(r_1,\hat{F}_1^{r_1})-V_1(r_1,F_1^0H_1^{r_1})=O_P(\sqrt{T})$,

(ii) For any $1\le r_1<r_1^0$, $\operatorname{plim\,inf}_{(N,T)\to\infty}d_TT^{-1}[V_1(r_1,F_1^0H_1^{r_1})-V_1(r_1,F_1^0)]=d_{r_1}$ *for some* $d_{r_1}>0$,

(iii) For any $r_1^0\le r_1\le r_{\max}$, $V_1(r_1,\hat{F}_1^{r_1})-V_1(r_1^0,\hat{F}_1^{r_1^0})=O_P(1)$,

where $V_1(r_1, F_1^0 H_1^{r_1})$ *is defined analogously to* $V_1(r_1, \hat{F}_1^{r_1})$ *with* $\hat{F}_1^{r_1}$ *replaced by* $F_1^0 H_1^{r_1}$*, and* $H_1^{r_1} = (N^{-1}\Lambda^{0\prime}\Lambda^0) \times (T^{-2} F_1^{0\prime} \hat{F}_1^{r_1})$.

Lemma A.13 *Suppose that Assumptions 3.1-3.3 and 3.5 hold. Then*

(i) For any $1 \le r_2 \le r_2^0$, $V_2(r_2, \hat{F}_2^{r_2}) - V_2(r_2, F_2^0 H_2^{r_2}) = O_P(C_{NT}^{-1})$,

(ii) For each r_2 *with* $0 \le r_2 < r_2^0$*, there exist a positive number* c_r *such that* $\operatorname{plim\,inf}_{(N,T)\to\infty}[V_2(r_2, F_2^0 H_2^{r_2}) - V_2(r_2^0, F_2^0)] = c_r$,

(iii) For any fixed r_2*, with* $r_2^0 \le r_2 \le r_{\max}$, $V_2(r_2, \hat{F}_2^{r_2}) - V_2(r^0, \hat{F}_2^{r_2^0}) = O_P(C_{NT}^{-2})$,

where $V_2(r_2, F_2^0 H_2^{r_2})$ *is defined analogously to* $V_2(r_2, \hat{F}_2^{r_2})$ *with* $\hat{F}_2^{r_2}$ *replaced by* $F_2^0 H_2^{r_2}$, $H_2^{r_2} = (N^{-1}\Lambda_2^{0\prime}\Lambda_2^0) \times (T^{-1} F_2^{0\prime} \hat{F}_2^{r_2})$.

Proof of Theorem 3.6. Noting that $IC_1(r_1) - IC_1(r_1^0) = V_1(r_1, \hat{F}_1^{r_1}) - V_1(r_1^0, \hat{F}_1^{r_1^0}) - (r_1^0 - r_1) g_1(N,T)$, it suffices to show that $P\left(V_1(r_1, \hat{F}_1^{r_1}) - V_1(r_1^0, \hat{F}_1^{r_1^0}) < (r_1^0 - r_1) g_1(N,T)\right) \to 0$ as $(N,T) \to \infty$ when $r_1 \neq r_1^0$. First, when $r_1 < r_1^0$, we consider the decomposition

$$
\begin{aligned}
V_1(r_1, \hat{F}_1^{r_1}) - V_1(r_1^0, \hat{F}_1^{r_1^0}) &= [V_1(r_1, \hat{F}_1^{r_1}) - V_1(r_1, F_1^0 H_1^{r_1})] + [V_1(r_1, F_1^0 H_1^{r_1}) - V_1(r_1^0, F_1^0 H_1^{r_1^0})] \\
&\quad + [V_1(r_1^0, F_1^0 H_1^{r_1^0}) - V_1(r_1^0, \hat{F}_1^{r_1^0})] \equiv DV_{1,1} + DV_{1,2} + DV_{1,3}.
\end{aligned}
$$

By Lemma A.12, $DV_{1,1} = O_P(T^{1/2})$, $DV_{1,2}$ is of exact probability order $O_P(T/\log\log T)$, and $DV_{1,3} = O_P(1)$. It follows that

$$
P(IC_1(r_1) < IC_1(r_1^0)) = P\left(V_1(r_1, \hat{F}_1^{r_1}) - V_1(r_1^0, \hat{F}_1^{r_1^0}) < (r_1^0 - r_1) g_1(N,T)\right) \to 0
$$

as $g_1(N,T)(\log\log T)/T \to 0$ under Assumption 3.6.

Next, for $r_1 > r_1^0$, we have $V_1(r_1, \hat{F}_1^{r_1}) - V_1(r_1^0, \hat{F}_1^{r_1^0}) = O_P(1)$ for $r_1 > r_1^0$ by Lemma A.12(iii), and $(r_1 - r_1^0) g_1(N,T) \to \infty$ by Assumption 3.6. This implies that $P(IC_1(r_1) - IC_1(r_1^0) < 0) = P(V_1(r_1, \hat{F}_1^{r_1}) - V_1(r_1^0, \hat{F}_1^{r_1^0}) < (r_1^0 - r_1) g_1(N,T)) \to 0$ as $N, T \to \infty$. ■

Proof of Theorem 3.7. Noting that $IC_2(r_2) - IC_2(r_2^0) = V_2(r_2, \hat{F}_2^{r_2}) - V_2(r_2^0, \hat{F}_2^{r_2^0}) - (r_2^0 - r_2) g_2(N,T)$, it suffices to show that $P\left(V_2(r_2, \hat{F}_2^{r_2}) - V_2(r_2^0, \hat{F}_2^{r_2^0}) < (r_2^0 - r_2) g_2(N,T)\right) \to 0$ as $(N,T) \to \infty$ when $r_2 \neq r_2^0$. We consider the under- and over-fitted models, respectively. When $0 \le r_2 < r_2^0$, we make the following decomposition:

$$
\begin{aligned}
V_2(r_2, \hat{F}_2^{r_2}) - V_2(r_2^0, \hat{F}_2^{r_2^0}) =& [V_1(r_2, \hat{F}_2^{r_2}) - V_1(r_2, F_2^0 H_2^{r_2})] + [V_1(r_2, F_2^0 H_2^{r_2}) - V_1(r_2^0, F_2^0 H_2^{r_2^0})] \\
&+ [V_1(r_2^0, F_2^0 H_2^{r_2^0}) - V_1(r_2^0, \hat{F}_2^{r_2^0})] \equiv DV_{2,1} + DV_{2,2} + DV_{2,3}.
\end{aligned}
$$

$DV_{1,l} = O_P(C_{NT}^{-1})$ for $l = 1, 3$ by Lemma A.13(i). Noting that $V_1(r_2, F_2^0 H_2^{r_2}) = V_1(r_2, F_2^0)$, $\operatorname{plim\,inf}_{(N,T)\to\infty}$ $DV_{1,2} = c_r$ when $r_2 < r_2^0$ by Lemma A.13(ii). It follows that $P(IC_2(r_2) < IC_2(r_2^0)) \to 0$ as $g_1(N,T) \to 0$ as $(N,T) \to \infty$ under Assumption 3.6.

Now, we consider the case where $r_2^0 < r_2 \le r_{\max}$. Note that $C_{NT}^2[V_2(r_2, \hat{F}_2^{r_2}) - V_2(r_2^0, \hat{F}_2^{r_2^0})] = O_P(1)$ and $C_{NT}^2 (r_2 - r_2^0) g_2(N,T) > C_{NT}^2 g_2(N,T) \to \infty$ by Lemma A.13(iii) and Assumption 3.6, we have $P(IC_2(r_2) < IC_2(r_2^0)) = P(V_2(r_2, \hat{F}_2^{r_2}) - V_2(r_2^0, \hat{F}_2^{r_2^0}) < (r_2^0 - r_2) g_2(N,T)) \to 0$ as $(N,T) \to \infty$. ■

To prove Theorem 3.8 we use the following lemma.

Lemma A.14 *Suppose that Assumptions 3.1-3.3 and 3.7 hold. Then* $\max_{K_0\le K\le K_{\max}}|\hat\sigma^2_{G_{(K,\lambda)}} - \hat\sigma^2_{\hat G(K_0,\lambda)}| = O_P(\nu^2_{NT})$, *where* $\hat\sigma^2_{G_{(K,\lambda)}} = \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat G_k(K,\lambda)}\sum_{t=1}^{T}[y_{it} - \hat\alpha^{cup\prime}_{\hat G_k(K,\lambda)}x_{1,it} - \hat\beta'_{2,i}x_{2,it} - \hat\lambda_{1i}(K,\lambda)'\hat f_{1t}(K,\lambda)]^2$ *and* ν_{NT} *is defined in Section 3.6.*

Proof of Theorem 3.8. First, we show that

$$\begin{aligned}IC_3(K_0,\lambda) &= \log[V_3(K_0)] + pK_0g_3(N,T)\\ &= \log\frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in\hat G_k(K_0,\lambda)}\sum_{t=1}^{T}\Big[y_{it} - \hat\alpha^{fm\prime}_{\hat G_k(K_0,\lambda)}x_{1,it} - \hat\beta'_{2,i}x_{2,it} - \hat\lambda_{1i}(K_0,\lambda)'\hat f_{1t}(K_0,\lambda)\Big]^2 + o_P(1)\\ &\xrightarrow{p}\log(\sigma_0^2).\end{aligned}$$

We consider the cases of under- and over-fitted models separately. When $1\le K<K_0$, for $G^{(K)} = (G_{K,1},...,G_{K,K})$ we have

$$\begin{aligned}V_3(K) =&\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat G_k(K,\lambda)}\sum_{t=1}^{T}\Big[y_{it} - \hat\alpha^{fm\prime}_{\hat G_k(K,\lambda)}x_{it} - \hat\beta_{2,i}(\hat G(K,\lambda))'x_{2,it} - \hat\lambda_{1i}(K,\lambda)'\hat f_{1t}(K,\lambda)\Big]^2\\ \ge&\min_{1\le K<K_0}\inf_{G^{(K)}\in\mathcal G^{(K)}}\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in G_{K,k}}\sum_{t=1}^{T}\Big[y_{it} - \hat\alpha^{fm\prime}_{G_{K,k}}x_{it} - \hat\beta_{2,i}(\hat G(K,\lambda))'x_{2,it} - \hat\lambda_{1i}(G^{(K)})'\hat f_{1t}(G^{(K)})\Big]^2\\ =&\min_{1\le K<K_0}\inf_{G^{(K)}\in\mathcal G^{(K)}}\hat\sigma^2_{G^{(K)}}.\end{aligned}$$

By Assumption 3.7 and Slutsky's lemma, we can demonstrate

$$\min_{1\le K<K_0} IC_3(K,\lambda)\ge\min_{1\le K<K_0}\inf_{G^{(K)}\in G_K}\log(\hat\sigma^2_{G^{(K)}}) + pKg_3(N,T)\xrightarrow{p}\log(\underline{\sigma}^2) > \log(\sigma_0^2).$$

It follows that $P(\min_{1\le K<K_0} IC_3(K,\lambda) > IC_3(K_0,\lambda))\to 1$.

When $K_0<K\le K_{\max}$, we can show that $NT[\hat\sigma^2_{\hat G(K,\lambda)} - \hat\sigma^2_{\hat G(K_0,\lambda)}] = O_P(1)$ when there are no stationary regressors, unobserved common factors, or endogeneity in $x_{1,it}$, $T[\hat\sigma^2_{\hat G(K,\lambda)} - \hat\sigma^2_{\hat G(K_0,\lambda)}] = O_P(1)$ when there are stationary regressors but no unobserved common factors, $N[\hat\sigma^2_{\hat G(K,\lambda)} - \hat\sigma^2_{\hat G(K_0,\lambda)}] = O_P(1)$ when there are nonstationary factors but no stationary regressors or factors, and $C^2_{NT}[\hat\sigma^2_{\hat G(K,\lambda)} - \hat\sigma^2_{\hat G(K_0,\lambda)}] = O_P(1)$ otherwise. Then by Lemma 14,

$$\begin{aligned}&P\left(\min_{K\in\mathcal K^+} IC_3(K,\lambda) > IC_3(K_0,\lambda)\right)\\ =&P\left(\min_{K\in\mathcal K^+}\nu_{NT}^{-2}\log(\hat\sigma^2_{\hat G(K,\lambda)}/\hat\sigma^2_{\hat G(K_0,\lambda)}) + \nu_{NT}^{-2}g_3(N,T)(K-K_0) > 0\right)\\ \approx&P\left(\min_{K\in\mathcal K^+}\nu_{NT}^{-2}(\hat\sigma^2_{\hat G(K,\lambda)} - \hat\sigma^2_{\hat G(K_0,\lambda)})/\hat\sigma^2_{\hat G(K_0,\lambda)} + \nu_{NT}^{-2}g_3(N,T)(K-K_0) > 0\right)\\ \to&1\quad\text{as}\quad (N,T)\to\infty\end{aligned}$$

where $\mathcal K^+ = \{K : K_0 < K\le K_{\max}\}$. ■

B Proofs of the lemmas in Appendix A

Proof of Lemma A.1. (i) By Lemma 2.1(c) in Park and Phillips (1988), we can show that

$$\begin{aligned}\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{1,i} &= \frac{1}{T^2}x_{1,i}'x_{1,i} - \frac{1}{T^2}x_{1,i}'P_{F_1^0}x_{1,i} \\ &= \frac{1}{T^2}\sum_{t=1}^{T} x_{1,it}x_{1,it}' - \frac{1}{T^2}\sum_{t=1}^{T} x_{1,it}f_{1t}^{0\prime}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_{1t}^0 f_{1t}^{0\prime}\right)^{-1}\frac{1}{T^2}\sum_{t=1}^{T} f_{1t}^0 x_{1,it}' \\ &\Rightarrow \int B_{2i}B_{2i}' - \int B_{2i}B_3'\left(\int B_3B_3'\right)^{-1}\int B_3B_{2i}' = \int \tilde{B}_{2i}\tilde{B}_{2i}',\end{aligned}$$

where $\tilde{B}_{2i} = B_{2i} - \int B_{2i}B_3'\left(\int B_3B_3'\right)^{-1}B_3$.

(ii) By Lemma 2.1(e) in Park and Phillips (1988), we can show that

$$\begin{aligned}\frac{1}{T}x_{1,i}'M_{F_1^0}u_i =& \frac{1}{T}x_{1,i}'u_i - \frac{1}{T}x_{1,i}'P_{F_1^0}u_i \\ =& \frac{1}{T}\sum_{t=1}^{T} x_{1,it}u_{it} - \frac{1}{T^2}\sum_{t=1}^{T} x_{1,it}f_{1t}^{0\prime}\left(\frac{1}{T^2}\sum_{t=1}^{T} f_{1t}^0 f_{1t}^{0\prime}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} f_{1t}^0 u_{it} \\ \Rightarrow& \left(\int B_{2i}dB_{1i} + \Delta_{21,i}\right) - \int B_{2i}B_3'\left(\int B_3B_3'\right)^{-1}\left(\int B_3dB_{1i} + \Delta_{31,i}\right) \\ =& \int \left(B_{2i} - \pi_i'B_3\right)dB_{1i} + (\Delta_{21,i} - \pi_i'\Delta_{31,i}),\end{aligned}$$

where $\pi_i = \left(\int B_3B_3'\right)^{-1}\int B_3B_{2i}'$, $\Delta_{21,i}$ and $\Delta_{31,i}$ are the one-sided long-run variances, defined above Assumption 3.1. ■

Proof of Lemma A.2. (i) This follows from Lemma A.3(i) in Huang, Jin, and Su (2020, HJS hereafter).

(ii) This follows from Donsker and Varadhan (1977, eqn (4.6) on p.751) and Lai and Wei (1982 ,eqn (3.23) on p.163).

(iii) Note that $\mu_{\max}\left(\frac{x_{1,i}'M_{F_1^0}x_{1,i}}{d_TT^2}\right) \le \mu_{\max}\left(\frac{x_{1,i}'x_{1,i}}{d_TT^2}\right) \le \mu_{\max}\left(\frac{W_i'W_i}{d_TT^2}\right)$ where $\mu_{\max}(M_{F_1^0}) = 1$. Then the result follows from Lemma A.3(i) in HJS.

(iv) Noting that $\left(\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{1,i}\right)^{-1}$ is the principal $p_1 \times p_1$ submatrix of $\left(\frac{1}{T^2}W_i'W_i\right)^{-1}$, we have by (ii)

$$\mu_{\max}\left(\left(\frac{d_T}{T^2}x_{1,i}'M_{F_1^0}x_{1,i}\right)^{-1}\right) \le \mu_{\max}\left(\left(\frac{d_T}{T^2}W_i'W_i\right)^{-1}\right) = \left[\mu_{\min}\left(\frac{d_T}{T^2}W_i'W_i\right)\right]^{-1} \le 2\rho_{\min}^{-1}$$

by the inclusion principle (see, e.g., Corollary 8.4.6 in Bernstein (2005)). It follows that

$$\mu_{\min}\left(\frac{d_T}{T^2}x_{1,i}'M_{F_1^0}x_{1,i}\right) = \left\{\left[\mu_{\max}\left(\frac{d_T}{T^2}x_{1,i}'M_{F_1^0}x_{1,i}\right)^{-1}\right]\right\}^{-1} \ge \rho_{\min}/2. \blacksquare$$

Remark. By Lemma 2.1(c) in Park and Phillips (1988), the continuous mapping theorem and the inversion formula for partitioned matrix,

$$\left(\frac{1}{T^2}W_i'W_i\right)^{-1} \Rightarrow \begin{pmatrix} \int B_{2i}B_{2i}' & \int B_{2i}B_3' \\ \int B_3B_{2i}' & \int B_3B_3' \end{pmatrix}^{-1} = \begin{pmatrix} \left(\int \tilde{B}_{2i}\tilde{B}_{2i}'\right)^{-1} & -\left(\int \tilde{B}_{2i}\tilde{B}_{2i}'\right)^{-1}\pi_i \\ -\pi_i'\left(\int \tilde{B}_{2i}\tilde{B}_{2i}'\right)^{-1} & \left(\int B_3B_3'\right)^{-1} + \pi_i'\left(\int \tilde{B}_{2i}\tilde{B}_{2i}'\right)^{-1}\pi_i \end{pmatrix},$$

where π_i is defined in the statement of Lemma A.1.

Proof of Lemma A.3. (i) Note that $\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{1,i} = \frac{1}{T^2}x_{1,i}'u_i - \frac{1}{T^2}x_{1,i}'P_{F_1^0}u_i$. It suffices to show that $\frac{1}{N}\sum_{i=1}^N \left\|\frac{x_{1,i}'u_i}{T^2}\right\|^2 = O_P(T^{-2})$ and $\frac{1}{N}\sum_{i=1}^N \left\|\frac{1}{T^2}x_{1,i}'P_{F_1^0}u_i\right\|^2 = O_P(d_TT^{-2})$. Note that

$$\begin{aligned}\frac{1}{NT^4}\sum_{i=1}^N \left\|x_{1,i}'P_{F_1^0}u_i\right\|^2 &= \frac{1}{NT^4}\sum_{i=1}^N \operatorname{tr}(x_{1,i}x_{1,i}'P_{F_1^0}u_iu_i'P_{F_1^0}) \le \frac{1}{NT^4}\sum_{i=1}^N \operatorname{tr}(x_{1,i}x_{1,i}')(u_i'P_{F_1^0}u_i)\\ &\le Cd_T\max_i \mu_{\max}\left(\frac{x_{1,i}'x_{1,i}}{d_TT^2}\right)\left[\mu_{\min}\left(\frac{F_1^{0\prime}F_1^0}{T^2}\right)\right]^{-1}\frac{1}{N}\sum_{i=1}^N \left\|\frac{F_1^{0\prime}u_i}{T^2}\right\|^2\\ &= O_P(d_T)\frac{1}{N}\sum_{i=1}^N\left\|\frac{F_1^{0\prime}u_i}{T^2}\right\|^2.\end{aligned}$$

where we use the fact that the limit of $\frac{F_1^{0\prime}F_1^0}{T^2}$ is p.d. a.s. and $\max_i \mu_{\max}\left(\frac{x_{1,i}'x_{1,i}}{d_TT^2}\right) = O_{a.s.}(1)$ by Lemma A.2(i). The result in (i) follows provided $\frac{1}{N}\sum_{i=1}^N\left\|\frac{x_{1,i}'u_i}{T^2}\right\|^2 = O_P(T^{-2})$ and $\frac{1}{N}\sum_{i=1}^N\left\|\frac{F_1^{0\prime}u_i}{T^2}\right\|^2 = O_P(T^{-2})$. Noting that $x_{1,it} = \sum_{s=1}^t \varepsilon_{it} + x_{1,i0} = S_2\sum_{s=1}^t w_{is}$ and $f_{1t}^0 = S_3\sum_{s=1}^t w_{is}$, it is sufficient to prove either of these two claims. Here we show the former one. Note that

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T^2}x_{1,i}'u_i\right\|^2 &= \frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T^2}\sum_{t=1}^T S_2(\sum_{s=1}^t w_{is})w_{it}'S_1'\right\|^2\\ &\le \frac{2}{N}\sum_{i=1}^N\left(\left\|S_2\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^{t-1} w_{is}w_{it}'S_1'\right\|^2 + \left\|S_2\frac{1}{T^2}\sum_{t=1}^T w_{it}w_{it}'S_1'\right\|^2\right) \equiv 2D_1 + 2D_2.\end{aligned}$$

By the panel BN-decomposition, we have $w_{it} = \phi_i(1)v_{it} + \tilde{w}_{it-1} - \tilde{w}_{it}$, where $\tilde{w}_{it} = \sum_{j=0}^\infty \tilde{\phi}_{ij}v_{i,t-j}$ and $\tilde{\phi}_{ij} = \sum_{s=j+1}^\infty \phi_{is}$. Then by the Cauchy-Schwarz inequality

$$\begin{aligned}D_1 &\le \frac{2}{N}\sum_{i=1}^N\left(\left\|S_2\frac{1}{T^2}\sum_{t=1}^T\left(\sum_{s=1}^{t-1}w_{is}\right)v_{it}'\phi_i(1)'S_1'\right\|^2 + \left\|S_2\frac{1}{T^2}\sum_{t=1}^T\left(\sum_{s=1}^{t-1}w_{is}\right)(\tilde{w}_{it-1}-\tilde{w}_{it})'S_1'\right\|^2\right)\\ &\equiv 2(D_{11}+D_{12}),\ \text{say.}\end{aligned}$$

Let $z_{it} = S_2\left(\sum_{s=1}^{t-1}w_{is}\right)v_{it}'\phi_i(1)'S_1'$ and $\mathcal{F}_{i,t} = \sigma(v_{it}, v_{i,t-1}, \ldots)$, the sigma-field generated by the series

$\{v_{is}, s \leq t\}$. Since $\mathbb{E}(z_{it}|\mathcal{F}_{i,t-1}) = 0$, we have

$$\mathbb{E}(D_{11}) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left\|\frac{1}{T^2}\sum_{t=1}^{T} z_{it}\right\|^2 \leq \frac{1}{NT^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\|z_{it}\|^2 \leq \frac{C}{NT^4}\sum_{i=1}^{N}\sum_{t=1}^{T} t = O(T^{-2}),$$

where the inequality follows by the fact that $\mathbb{E}\|z_{it}\|^2 \leq C\mathbb{E}\left\|\sum_{s=1}^{t-1} w_{is}\right\|^2 \mathbb{E}\|v_{it}\|^2 \leq Ct$. Then $D_{11} = O_P(T^{-2})$ by the Markov inequality. For D_{12}, we have

$$\begin{aligned} D_{12} &= \frac{1}{N}\sum_{i=1}^{N}\left\|S_2\frac{1}{T^2}\sum_{s=1}^{T-1} w_{is}\sum_{t=s+1}^{T}\left(\tilde{w}_{it-1} - \tilde{w}_{it}\right)' S_1'\right\|^2 \\ &\leq \frac{2}{N}\sum_{i=1}^{N}\left(\left\|S_2\frac{1}{T^2}\sum_{s=1}^{T-1} w_{is}\tilde{w}_{is}' S_1'\right\|^2 + \left\|S_2\frac{1}{T^2}\sum_{s=1}^{T-1} w_{is}\tilde{w}_{iT}' S_1'\right\|^2\right) \equiv 2\left(D_{12,1} + D_{12,2}\right), \text{ say.} \end{aligned}$$

Under Assumption 3.1(i)-(ii) and Phillips and Solo (1992), we have $\mathbb{E}\left\|\tilde{w}_{it}\right\|^4 \leq C < \infty$. By similar arguments in the proof of Lemma A.2. in HJS, we can show $D_{12,1} = O_P(T^{-2})$. It's easy to show $D_{12,2} = o_P(T^{-2})$. Thus $D_1 = O_P(T^{-2})$. For D_2, we have

$$\begin{aligned} \mathbb{E}(D_2) &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left\|S_2\frac{1}{T^2}\sum_{t=1}^{T} w_{it}w_{it}' S_1\right\|^2 \leq \frac{C}{NT^2}\sum_{i=1}^{N}\mathbb{E}\left(\frac{1}{T}\sum_{t=1}^{T}\|w_{it}\|^2\right)^2 \leq \frac{C}{NT^2}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\|w_{it}\|^4 \\ &= O(T^{-2}), \end{aligned}$$

where the second inequality comes from the Cauchy-Schwarz inequality. It implies that $D_2 = O_P(T^{-2})$. Consequently, $\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left\|\frac{1}{T^2}x_{1,i}'u_i\right\|^2 = O(T^{-2})$ and $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_{1,i}'u_i\right\|^2 = O_P(T^{-2})$ by the Markov inequality. This completes the proof of (i).

(ii) Note that

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}u_i^*\right\|^2 &\leq \frac{3}{N}\sum_{i=1}^{N}\left(\left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}u_i\right\|^2 + \left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}F_2^0\lambda_{2i}^0\right\|^2 + \left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{2,i}b_{2,i}\right\|^2\right) \\ &\equiv 3\left(I + II + III\right), \end{aligned}$$

where recall that $f_{2t}^0 = S_4 w_{it}$. For I, we have $I = \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_i' M_{F_1^0}u_i\right\|^2 = O_P(d_T T^{-2})$ by the result in part (i). By arguments analogous to those used in the proof of part (i) and using $F_2^0\lambda_{2i}^0$ in place of u_i, we can show that $II = \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}F_2^0\lambda_{2i}^0\right\|^2 = O_P(d_T T^{-2})$. For III, we have

$$III \leq \max_i\left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{2,i}\right\|^2\frac{1}{N}\sum_{i=1}^{N}\|b_{2,i}\|^2 = o_P(T^{-1}N^{1/q}(\log T)^{(1+\epsilon)/2})$$

uniformly in $\frac{1}{N}\left\|\mathbf{b}_2\right\|^2 \leq M$, where we use the fact that $\max_i\left\|\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{2,i}\right\|^2 = o_P(T^{-1}N^{1/q}(\log T)^{(1+\epsilon)/2})$ by similar analysis as used in the proof of Lemma A.3(i) in HJS. Then the result in (ii) follows.

(iii) Note that

$$\begin{aligned}
\left\|\frac{1}{NT^2}\sum_{j=1}^{N} x_{1,i}' M_{F_1^0} u_j a_{ij}\right\| \leq & \frac{1}{NT^2}\sum_{j=1}^{N}\left(\left\|x_{1,i}' u_j a_{ij}\right\| + \left\|x_{1,i}' P_{F_1^0} u_j a_{ij}\right\|\right) \\
\leq & \left\{\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{x_{1,i}' u_j}{T^2}\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{j=1}^{N}\|a_{ij}\|^2\right\}^{1/2} \\
& + \left[\mu_{\min}\left(\frac{F_1^{0\prime} F_1^0}{T^2}\right)\right]^{-1}\frac{\|x_{1,i}' F_1^0\|}{T^2}\left\{\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{F_1^0 u_j}{T^2}\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{j=1}^{N}\|a_{ij}\|^2\right\}^{1/2} \\
= & O_P(T^{-1}) + O_P(T^{-1}) = O_P(T^{-1}),
\end{aligned}$$

where we use the fact that $\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{x_{1,i}' u_j}{T^2}\right\|^2 = O_P(T^{-2})$, $\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{F_1^0 u_j}{T^2}\right\|^2 = O_P(T^{-2})$, $\frac{1}{T^2}x_{1,i}' F_1^0 = \int B_{2i} B_3' + o_P(1) = O_P(1)$ by Lemma A.1(i), and $\frac{1}{N}\sum_{j=1}^{N}\|a_{ij}\|^2 = O_P(1)$ by Assumption 3.2(i). Similarly, we can show that $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{NT^2}\sum_{j=1}^{N} x_{1,i}' M_{F_1^0} u_j a_{ij}\right\| = O_P(T^{-1})$.

(iv) Noting that $\left\|x_{1,i}' M_{F_1^0} x_{1,i}\right\|^2 =\text{tr}(M_{F_1^0} x_{1,i} x_{1,i}' M_{F_1^0} x_{1,i} x_{1,i}') \leq \text{tr}(x_{1,i} x_{1,i}' M_{F_1^0} x_{1,i} x_{1,i}') \leq \text{tr}(x_{1,i} x_{1,i}' x_{1,i} x_{1,i}') = \left\|x_{1,i}' x_{1,i}\right\|^2$ by the fact that $\mu_{\max}(M_{F_1^0}) = 1$, we have $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_{1,i}' M_{F_1^0} x_{1,i}\right\|^2 \leq \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_{1,i}' x_{1,i}\right\|^2$. It suffices to show that $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_{1,i}' x_{1,i}\right\|^2 = O_P(1)$. Using the panel BN decomposition $x_{1,it} = S_2(\phi_i(1)\sum_{s=1}^{t} v_{is} + \tilde{w}_{i0} - \tilde{w}_{it})$ and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left\|\frac{1}{T^2}x_{1,i}' x_{1,i}\right\|^2 \leq & \frac{3}{N}\sum_{i=1}^{N}\mathbb{E}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{l=1}^{t} S_2\phi_i(1) v_{is} v_{il}'\phi_i(1)' S_2'\right\|^2 \\
& + \frac{3}{N}\sum_{i=1}^{N}\mathbb{E}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t} S_2\phi_i(1) v_{is}(\tilde{w}_{i0} - \tilde{w}_{it})' S_2'\right\|^2 \\
& + \frac{3}{N}\sum_{i=1}^{N}\mathbb{E}\left\|\frac{1}{T^2}\sum_{t=1}^{T} S_2(\tilde{w}_{i0} - \tilde{w}_{it})(\tilde{w}_{i0} - \tilde{w}_{it})' S_2'\right\|^2 \equiv 3\left(D_1 + D_2 + D_3\right).
\end{aligned}$$

For D_1, we have $D_1 \leq \frac{C}{N}\sum_{i=1}^{N}\mathbb{E}\left\{\frac{1}{T^2}\sum_{t=1}^{T}\left\|\sum_{s=1}^{t} v_{is}\right\|^2\right\} \leq \frac{C}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T} t = O(1)$, where we use the fact $\mathbb{E}\left\|\sum_{s=1}^{t} v_{is}\right\|^2 \leq Ct$. Similarly, we can show that $D_2 = O(T^{-1})$ and $D_3 = O(T^{-2})$. It follows that $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_{1,i}' M_{F_1^0} x_{1,i}\right\|^2 = O_P(1)$. ■

Proof of Lemma A.4. (i) Note that

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N} b_{1,i}' x_{1,i}' M_{F_1} u_i^*\right\| \leq \left\|\frac{1}{NT^2}\sum_{i=1}^{N} b_{1,i}' x_{1,i}' u_i^*\right\| + \left\|\frac{1}{NT^2}\sum_{i=1}^{N} b_{1,i}' x_{1,i}' P_{F_1} u_i^*\right\|.$$

For the first term, we have

$$
\begin{aligned}
&\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\|\frac{1}{NT^2}\sum_{i=1}^N b_{1,i}'x_{1,i}'u_i^*\right\| \\
&\le \sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left(\left\|\frac{1}{NT^2}\sum_{i=1}^N b_{1,i}'x_{1,i}'u_i\right\|+\left\|\frac{1}{NT^2}\sum_{i=1}^N b_{1,i}'x_{1,i}'F_2^0\lambda_{2i}^0\right\|+\left\|\frac{1}{NT^2}\sum_{i=1}^N b_{1,i}'x_{1,i}'x_{2,i}b_{2,i}\right\|\right) \\
&\le \sup_{N^{-1}\|\mathbf{b}\|^2\le M}\frac{1}{T}\left\{\frac{1}{N}\sum_{i=1}^N\|b_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\left\|\frac{x_{1,i}'u_i}{T}\right\|^2\right\}^{1/2} \\
&+\frac{1}{\sqrt{T}}\frac{\|F_2^0\|}{\sqrt{T}}\max_i\|\lambda_{2i}^0\|\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\{\frac{1}{N}\sum_{i=1}^N\|b_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\frac{\|x_{1,i}\|^2}{T^2}\right\}^{1/2} \\
&+\frac{1}{\sqrt{T}}\max_i\frac{\|x_{1,i}\|}{T}\max_i\frac{\|x_{2,i}\|}{\sqrt{T}}\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\{\frac{1}{N}\sum_{i=1}^N\|b_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\|b_{2,i}\|^2\right\}^{1/2} \\
&=O_P(T^{-1})+O_P(T^{-1/2})+O_P(T^{-1/2}\sqrt{d_T})=O_P(T^{-1/2}\sqrt{d_T}),
\end{aligned}
$$

where we use the fact that $\max_i\|\lambda_{2i}^0\|=O_P(1)$ by Assumption 3.2(i), $\frac{1}{N}\sum_{i=1}^N\left\|\frac{x_{1,i}'u_i}{T}\right\|^2=O_P(1)$ as shown in the proof of Lemma A.3(i), $\max_i\frac{\|x_{1,i}\|}{T}=O_P(\sqrt{d_T})$ by Lemma A.2(i), and $\max_i\frac{\|x_{2,i}\|}{\sqrt{T}}=O_P(1)$ by an application of Lemma S1.2(iii) in Su, Shi and Phillips (2016b, SSPb hereafter). For the second term, we have

$$
\begin{aligned}
&\sup_{F_1\in\mathcal{F}_1}\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\|\frac{1}{NT^2}\sum_{i=1}^N b_{1,i}'x_{1,i}'P_{F_1}u_i^*\right\| \\
&\le \sup_{F_1\in\mathcal{F}_1}\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left(\left\|\frac{1}{NT^4}\sum_{i=1}^N b_{1,i}'x_{1,i}'F_1F_1'u_i\right\|+\left\|\frac{1}{NT^4}\sum_{i=1}^N b_{1,i}'x_{1,i}'F_1F_1'F_2^0\lambda_{2i}^0\right\|\right. \\
&\left.+\left\|\frac{1}{NT^4}\sum_{i=1}^N b_{1,i}'x_{1,i}'F_1F_1'x_{2,i}b_{2,i}\right\|\right) \\
&\le \frac{1}{\sqrt{T}}\sup_{F_1\in\mathcal{F}_1}\frac{\|F_1\|^2}{T^2}\max_i\frac{\|x_{1,i}\|}{T}\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\{\frac{1}{N}\sum_{i=1}^N\|b_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\frac{\|u_i\|^2}{T}\right\}^{1/2} \\
&+\frac{\bar{c}_\lambda}{\sqrt{T}}\sup_{F_1\in\mathcal{F}_1}\frac{\|F_1\|^2}{T^2}\frac{\|F_2^0\|}{\sqrt{T}}\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\{\frac{1}{N}\sum_{i=1}^N\|b_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\frac{\|x_{1,i}\|^2}{T^2}\right\}^{1/2} \\
&+\frac{1}{\sqrt{T}}\sup_{F_1\in\mathcal{F}_1}\frac{\|F_1\|^2}{T^2}\max_i\frac{\|x_{1,i}\|}{T}\max_i\frac{\|x_{2,i}\|}{\sqrt{T}}\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\{\frac{1}{N}\sum_{i=1}^N\|b_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\|b_{2,i}\|^2\right\}^{1/2} \\
&=O_P(T^{-1/2}\sqrt{d_T})+O_P(T^{-1/2})+O_P(T^{-1/2}\sqrt{d_T})=O_P(T^{-1/2}\sqrt{d_T}),
\end{aligned}
$$

This proves (i).

(ii)-(iii) The proofs of (ii) and (iii) are analogous to that of (i) and are therefore omitted. ■

Proof of Lemma A.5. (i) We make the following decomposition

$$
\begin{aligned}
P_{\hat{F}_1} - P_{F_1^0} &= P_{\hat{F}_1} - P_{F_1^0 H_1} = \frac{1}{T^2}\hat{F}_1\hat{F}_1' - F_1^0 H_1 \left(H_1' F_1^{0\prime} F_1^0 H_1\right)^{-1} H_1' F_1^{0\prime} \\
&= \frac{1}{T^2}(\hat{F}_1 - F_1^0 H_1)(\hat{F}_1 - F_1^0 H_1)' - \frac{1}{T^2}(\hat{F}_1 - F_1^0 H_1) H_1' F_1^{0\prime} \\
&\quad + \frac{1}{T^2} F_1^0 H_1 (\hat{F}_1 - F_1^0 H_1)' + \frac{1}{T^2} F_1^0 H_1 \left[I_{r_1} - \left(\frac{1}{T^2} H' F^{0\prime} F^0 H \right)^{-1} \right] H_1' F_1^{0\prime} \\
&\equiv \hat{p}_1 + \hat{p}_2 + \hat{p}_3 + \hat{p}_4, \text{ say.}
\end{aligned}
\tag{B.1}
$$

By Theorem 3.1(iii), $\|\hat{p}_1\| = O_P\left(\eta_{1NT}^2 + T^{-1}C_{NT}^{-2}\right)$ and $\|\hat{p}_2\| = \|\hat{p}_3\| = O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})$. In addition, noting that

$$
\begin{aligned}
\left\| I_{r_1} - \frac{1}{T^2} H' F^{0\prime} F^0 H \right\| &= \left\| \frac{1}{T^2}\left(\hat{F}_1'\hat{F}_1 - H' F^{0\prime} F^0 H \right) \right\| \\
&\le \left\| \frac{1}{T^2}\left(\hat{F}_1 - F^0 H \right)' \left(\hat{F}_1 - F^0 H \right) \right\| + 2 \left\| \frac{1}{T^2}\left(\hat{F}_1 - F^0 H \right)' F^0 H \right\| \\
&= O_P\left(\eta_{1NT}^2 + T^{-1}C_{NT}^{-2}\right) + O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1}),
\end{aligned}
$$

$\|\hat{p}_4\| = O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})$. It follows that $\left\| P_{\hat{F}_1} - P_{F_1^0} \right\| \le \sum_{l=1}^4 \|\hat{p}_l\| = O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})$.

(ii) By taking the sub-differential of $Q_{NT}^{\lambda,K}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}, F_1)$ with respect to β_{1i}, for each $i = 1, ..., N$ and $k = 1, ..., K$, we have $0_{p_{1\times 1}} = -\frac{2}{T^2} x_{1,i}' M_{\hat{F}_1} (y_i - x_{1,i}\hat{\beta}_{1,i} - x_{2,i}\hat{\beta}_{2,i}) + \lambda \sum_{j=1}^K \hat{e}_{ij} \prod_{l=1, l\neq j}^K \|\hat{\beta}_{1,i} - \hat{\alpha}_l\|$, where $\hat{e}_{ij}$ is as defined in the proof of Theorem 3.2. From this, we can derive that

$$
\|\hat{b}_{1,i}\| \le \left\| \left(\frac{1}{T^2} x_{1,i}' M_{\hat{F}_1} x_{1,i} \right)^{-1} \right\|_{sp} \left\{ \left\| \frac{1}{T^2} x_{1,i}' M_{\hat{F}_1} \hat{u}_i^* \right\| + \left\| \frac{1}{T^2} x_{1,i}' M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 \right\| + C\lambda \right\}.
$$

Note that $\left\| \left(\frac{1}{T^2} x_{1,i}' M_{\hat{F}_1} x_{1,i} \right)^{-1} \right\|_{sp} = [\mu_{\min}(\frac{1}{T^2} x_{1,i}' M_{\hat{F}_1} x_{1,i})]^{-1}$. By (i) and Assumption A.2(i), $\mu_{\min}(\frac{1}{T^2} x_{1,i}' M_{\hat{F}_1} x_{1,i}) = \mu_{\min}(\frac{1}{T^2} x_{1,i}' M_{F_1^0} x_{1,i}) + o_P\left(d_T^{-2}\right)$ uniformly in i. Then

$$
\left\| \left(\frac{d_T}{T^2} x_{1,i}' M_{\hat{F}_1} x_{1,i} \right)^{-1} \right\|_{sp} = \left\| \left(\frac{d_T}{T^2} x_{1,i}' M_{F_1^0} x_{1,i} \right)^{-1} \right\|_{sp} + o_P(1) \le 4\rho_{\min}^{-1}
$$

by Lemma A.2(iv) w.p.a.1. It follows that

$$
\max_i \|\hat{b}_{1,i}\| \le O_P\left(d_T\right) \max_i \left\{ \left\| \frac{1}{T^2} x_{1,i}' M_{\hat{F}_1} \hat{u}_i^* \right\| + \left\| \frac{1}{T^2} x_{1,i}' M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 \right\| + C\lambda \right\}.
$$

In addition, it is easy to show that $\max_i \left\| \frac{1}{T^2} x_{1,i}' M_{\hat{F}_1} \hat{u}_i^* \right\| \le T^{-1/2} \max_i \frac{1}{T} \|x_{1,i}\| \max_i \frac{1}{T^{1/2}} \|\hat{u}_i^*\| =$

$O_P(d_T^{1/2}T^{-1/2})$ and

$$\begin{aligned}\max_i \left\| \frac{1}{T^2} x'_{1,i} M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 \right\| &= \max_i \left\| \frac{1}{T^2} x'_{1,i} M_{\hat{F}_1} (\hat{F}_1 H_1^{-1} - F_1^0) \lambda_{1i}^0 \right\| \le C \max_i \frac{\|x_{1,i}\|}{T} \frac{1}{T} \left\| \hat{F}_1 H_1^{-1} - F_1^0 \right\| \\ &= \sqrt{d_T} O_P(\eta_{1NT} + T^{-1/2} C_{NT}^{-1}).\end{aligned}$$

Thus, we have $\max_i \|\hat{b}_{1,i}\| = d_T O_P(d_T^{1/2} T^{-1/2} + d_T^{1/2}\eta_{1NT} + \lambda)$.

Note that $\hat{\beta}_{2,i} = \left(x'_{2,i} M_{\hat{F}_1} x_{2,i} \right)^{-1} x'_{2,i} M_{\hat{F}_1} (y_i - x_{1,i}\hat{\beta}_{1,i})$ and

$$\|\hat{b}_{2,i}\| \le \left\| \left(\frac{1}{T} x'_{2,i} M_{\hat{F}_1} x_{2,i} \right)^{-1} \right\|_{sp} \left\{ \left\| \frac{1}{T} x'_{2,i} M_{\hat{F}_1} \left(u_i + F_2^0 \lambda_{2i}^0 \right) \right\| + \left\| \frac{1}{T} x'_{2,i} M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 \right\| + \left\| \frac{1}{T} x'_{2,i} M_{\hat{F}_1} x_{1,i} \hat{b}_{1,i} \right\| \right\}.$$

Note that $\left\| \left(\frac{1}{T} x'_{2,i} M_{\hat{F}_1} x_{2,i} \right)^{-1} \right\|_{sp} = [\mu_{\min}(\frac{1}{T} x'_{2,i} M_{\hat{F}_1} x_{2,i})]^{-1}$, $\mu_{\min}(\frac{1}{T} x'_{2,i} M_{\hat{F}_1} x_{2,i}) = \mu_{\min}(\frac{1}{T} x'_{2,i} M_{F_1^0} x_{2,i})$ $+ o_P(1) = \mu_{\min}(\frac{1}{T} x'_{2,i} x_{2,i}) + o_P(1)$ uniformly in i, and $\min_i \mu_{\min}(\frac{1}{T} x'_{2,i} x_{2,i})$ is bounded away from zero w.p.a.1 by Assumption 3.2(v). It follows that $\max_i ||(\frac{1}{T} x'_{2,i} M_{\hat{F}_1} x_{2,i})^{-1}||_{sp} = O_P(1)$ and

$$\begin{aligned}\frac{1}{N} \sum_{i=1}^N \|\hat{b}_{2,i}\|^2 &\le O_P(1) \frac{1}{N} \sum_{i=1}^N \left\{ \left\| \frac{1}{T} x'_{2,i} M_{\hat{F}_1} \left(u_i + F_2^0 \lambda_{2i}^0 \right) \right\|^2 + \left\| \frac{1}{T} x'_{2,i} M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 \right\|^2 + \left\| \frac{1}{T} x'_{2,i} M_{\hat{F}_1} x_{1,i} \hat{b}_{1,i} \right\|^2 \right\} \\ &\equiv O_P(1) \{ II_1 + II_2 + II_3 \}, \text{ say.}\end{aligned}$$

Then we have

$$\begin{aligned} II_1 &\le \frac{2}{N} \sum_{i=1}^N \left\| \frac{1}{T} x'_{2,i} M_{F_1^0} \left(u_i + F_2^0 \lambda_{2i}^0 \right) \right\|^2 + \frac{2}{N} \sum_{i=1}^N \left\| \frac{1}{T} x'_{2,i} (P_{\hat{F}_1} - P_{F_1^0}) \left(u_i + F_2^0 \lambda_{2i}^0 \right) \right\|^2 \\ &\le \frac{4}{N} \sum_{i=1}^N \left\| \frac{1}{T} x'_{2,i} \left(u_i + F_2^0 \lambda_{2i}^0 \right) \right\|^2 + \frac{4}{N} \sum_{i=1}^N \left\| \frac{1}{T} \frac{x'_{2,i} F_1^0}{T} \left(\frac{F_1^{0\prime} F_1^0}{T^2} \right)^{-1} \frac{F_1^{0\prime}(u_i + F_2^0 \lambda_{2i}^0)}{T} \right\|^2 \\ &+ 2 \max_i \left\| \frac{x_{2,i}}{\sqrt{T}} \right\|^2 \frac{1}{N} \sum_{i=1}^N \left\| \frac{u_i + F_2^0 \lambda_{2i}^0}{\sqrt{T}} \right\|^2 \left\| P_{\hat{F}_1} - P_{F_1^0} \right\|^2 \\ &= O_P(T^{-1}) + T^{-2} o_P(N^{2/q} (\log T)^{1+\epsilon}) + O_P(\eta_{1NT}^2 + T^{-1} C_{NT}^{-2}) = O_P(\eta_{1NT}^2 + T^{-1}), \end{aligned}$$

where we use the result in (i) and the fact that $\max_i \frac{1}{T} \|x'_{2,i} F_1^0\| = o_P(N^{1/q} (\log T)^{(1+\epsilon)/2})$ by arguments as used in the proof of Lemma A.2(i) in HJS. For II_2, we have

$$\begin{aligned} II_2 &= \frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{T} x'_{2,i} M_{\hat{F}_1} (\hat{F}_1 H_1^{-1} - F_1^0) \lambda_{1i}^0 \right\|^2 \le \max_i \left\| \lambda_{1i}^0 \right\|^2 \frac{1}{T} \left\| \hat{F}_1 H_1^{-1} - F_1^0 \right\|^2 \frac{1}{N} \sum_{i=1}^N \frac{\|x_{2,i}\|^2}{T} \\ &= O_P(T \eta_{1NT}^2 + C_{NT}^{-2}), \end{aligned}$$

By arguments as used in the proof of Lemma A.2(i) in HJS, $\max_i \left\| \frac{1}{T} x'_{2,i} M_{F_1^0} x_{1,i} \right\| = o_P(N^{1/q} (\log T)^{(1+\epsilon)/2})$.

Then

$$\begin{aligned}
\max_i \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} x_{1,i} \right\| &\leq \max_i \left\| \frac{1}{T} x_{2,i}' M_{F_1^0} x_{1,i} \right\| + \max_i \left\| \frac{1}{T} x_{2,i}' (P_{\hat{F}_1} - P_{F_1^0}) x_{1,i} \right\| \\
&= o_P(N^{1/q} (\log T)^{(1+\epsilon)/2} + T^{1/2} \max_i \frac{\|x_{1,i}\|}{T} \max_i \frac{\|x_{2,i}\|}{T^{1/2}} \left\| P_{\hat{F}_1} - P_{F_1^0} \right\| \\
&= o_P(N^{1/q} (\log T)^{(1+\epsilon)/2} + \sqrt{T d_T} O_P(\eta_{1NT} + T^{-1/2} C_{NT}^{-1}) = o_P\left(T^{1/2}\right)
\end{aligned}$$

and

$$II_3 \leq \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} x_{1,i} \hat{b}_{1,i} \right\|^2 \leq \max_i \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} x_{1,i} \right\|^2 \frac{1}{N} \sum_{i=1}^{N} \|\hat{b}_{1,i}\|^2 = o_P(T \eta_{1NT}^2).$$

[Alternatively, we can show that $II_3 \leq \max_i \|\hat{b}_{1,i}\|^2 \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} x_{1,i} \right\|^2 = d_T^2 O_P(d_T T^{-1} + d_T \eta_{1NT}^2 + \lambda^2)$.] It follows that $\eta_{2NT}^2 \equiv \frac{1}{N} \sum_{i=1}^{N} \|\hat{b}_{2,i}\|^2 = O_P(T^{-1/2} C_{NT}^{-1} + T \eta_{1NT}^2)$.

(iii) By the proof of Theorem 3.1(iii), we have $\hat{F}_1 V_{1,NT} \equiv I_1 + ... + I_{11} + \frac{1}{NT^2} \sum_{i=1}^{N} F_1^0 \lambda_{1i}^0 \lambda_{1i}^{0\prime} F_1^{0\prime} \hat{F}_1$, where the I_l are defined in the proof of Theorem 3.1(iii). It follows that $\frac{1}{T} F_1^{0\prime} (\hat{F}_1 - F_1^0 H_1) = \frac{1}{T} (F_1^{0\prime} I_1 + ... + F_1^{0\prime} I_{11}) V_{1,NT}^{-1}$, where we recall that $H_1 = (\frac{1}{N} \Lambda_1^{0\prime} \Lambda_1^0)(\frac{1}{T^2} F_1^{0\prime} \hat{F}_1) V_{1,NT}^{-1}$. It remains to study the probabilistic order of $\frac{1}{T} F_1^{0\prime} I_l$ for $l = 1, 2, ..., 11$. For $\frac{1}{T} F_1^{0\prime} I_1$ and $\frac{1}{T} F_1^{0\prime} I_2$, we have

$$\begin{aligned}
\frac{1}{T} \|F_1^{0\prime} I_1\| &\leq T \frac{\|F_1^0\|}{T} \frac{\|\hat{F}_1\|}{T} \max_i \frac{\|x_{1,i}\|^2}{T^2} \frac{1}{N} \sum_{i=1}^{N} \|\hat{b}_{1,i}\|^2 = O_P(T d_T \eta_{1NT}^2), \text{ and} \\
\frac{1}{T} \|F_1^{0\prime} I_2\| &\leq T \frac{\|F^{0\prime} \hat{F}_1\|}{T^2} \frac{\|F_1^0\|}{T} \max_i \|\lambda_i^0\| \left\{ \frac{1}{N} \sum_{i=1}^{N} \|\hat{b}_{1,i}\|^2 \right\}^{1/2} \left\{ \frac{1}{N} \sum_{i=1}^{N} \frac{\|x_{1,i}\|^2}{T^2} \right\}^{1/2} = O_P(T \eta_{1NT}).
\end{aligned}$$

For $\frac{1}{T} F_1^{0\prime} I_3$, we have

$$\frac{1}{T} \|F_1^{0\prime} I_3\| \leq \frac{\|F_1^0\|}{T} \max_i \frac{\|x_{1,i}\|}{T} \left\{ \frac{1}{N} \sum_{i=1}^{N} \|\hat{b}_{1,i}\|^2 \right\}^{1/2} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\tilde{u}_i' \hat{F}_1}{T} \right\|^2 \right\}^{1/2},$$

where $\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\tilde{u}_i' \hat{F}_1}{T} \right\|^2 \leq \frac{2}{N} \sum_{i=1}^{N} \left(\left\| \frac{\tilde{u}_i' F_1^0 H_1}{T} \right\|^2 + \left\| \frac{\tilde{u}_i' (\hat{F}_1 - F_1^0 H_1)}{T} \right\|^2 \right)$. Noting that $\tilde{u}_i = u_i - x_{2,i} \hat{b}_{2,i}$, we have

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{\tilde{u}_i' (\hat{F}_1 - F_1^0 H_1)}{T} \right\|^2 &\leq \frac{2}{T} \left\| \hat{F}_1 - F_1^0 H_1 \right\|^2 \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\|u_i\|^2}{T} + \frac{\max_i \|x_{2,i}\|^2}{T} \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{b}_{2,i} \right\|^2 \right) \\
&= O_P(T \eta_{1NT}^2 + C_{NT}^{-2})[1 + O_P(T^{-1/2} C_{NT}^{-1} + T \eta_{1NT}^2)] = O_P(T \eta_{1NT}^2 + C_{NT}^{-2}).
\end{aligned}$$

In addition,

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{\tilde{u}_i'F_1^0}{T}\right\|^2 \leq \frac{1}{N}\sum_{i=1}^{N}\left(\left\|\frac{u_i'F_1^0}{T}\right\|^2 + \|\hat{b}_{2,i}\|^2\left\|\frac{x_{2i}'F_1^0}{T}\right\|^2\right) = O_P(1) + \max_i\left\|\frac{x_{2i}'F_1^0}{T}\right\|^2\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{2,i}\|^2$$
$$= O_P(1) + o_P(N^{2/q}(\log T)^{1+\varepsilon})O_P(T^{-1/2}C_{NT}^{-1} + T\eta_{1NT}^2).$$

Then $\frac{1}{T}\|F_1^{0\prime}I_3\| = d_T^{1/2}\eta_{1NT}[O_P(T\eta_{1NT}^2 + C_{NT}^{-2}) + o_P(N^{2/q}(\log T)^{1+\varepsilon})O_P(T^{-1/2}C_{NT}^{-1} + T\eta_{1NT}^2)] = o_P(T\eta_{1NT})$. For $\frac{1}{T}F_1^{0\prime}I_4$ and $\frac{1}{T}F_1^{0\prime}I_5$, we have

$$\frac{1}{T}\left\|F_1^{0\prime}I_4\right\| \leq T\frac{\|F_1^{0\prime}F^0\|}{T^2}\frac{\|\hat{F}_1\|}{T}\max_i\|\lambda_i^0\|\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\|x_{1,i}\|^2}{T^2}\right\}^{1/2} = O_P(T\eta_{1NT}),$$
$$\frac{1}{T}\|F_1^{0\prime}I_5\| \leq \frac{\|\hat{F}_1\|}{T}\max_i\frac{\|x_{1,i}\|}{T}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\|F_1^{0\prime}\tilde{u}_i\|^2}{T^2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2} = O_P(\sqrt{d_T}\eta_{1NT}).$$

For $\frac{1}{T}F_1^{0\prime}I_6$, $\frac{1}{T}\|F_1^{0\prime}I_6\| \leq \left\|\frac{1}{NT^3}\sum_{i=1}^{N}F_1^{0\prime}F^0\lambda_i^0\tilde{u}_i'F_1^0H_1\right\| + \left\|\frac{1}{NT^3}\sum_{i=1}^{N}F_1^{0\prime}F^0\lambda_i^0\tilde{u}_i'(\hat{F}_1 - F_1^0H_1)\right\| \equiv D_1 + D_2$. For D_1, we have

$$D_1 \leq \frac{1}{T^2}\left\|F_1^{0\prime}F^0\right\|\left(\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i^0u_{it}f_{1t}^{0\prime}\right\| + \left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^0\hat{b}_{2,i}'x_{2,it}f_{1t}^{0\prime}\right\|\right)\|H_1\|$$

By analysis as used to obtain (B.8) in Bai (2004, p.172), $\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_i^0u_{it}f_{1t}^{0\prime}\right\| = O_P(N^{-1/2})$. In addition,

$$\left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^0\hat{b}_{2,i}'\sum_{t=1}^{T}x_{2,it}f_{1t}^{0\prime}\right\| \leq C\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{b}_{2,i}\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}x_{2,it}f_{1t}^{0\prime}\right\|^2\right\}^{1/2} = O_P\left(\eta_{2NT}\right).$$

So $D_1 = O_P\left(\eta_{2NT} + N^{-1/2}\right)$. For D_2, we have

$$\begin{aligned}D_2 \leq& \frac{1}{T^2}\left\|F_1^{0\prime}F^0\right\|\left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_{it}^0\tilde{u}_{it}(\hat{f}_{1t} - H_1f_{1t}^0)\right\| \\ \leq& \sqrt{\frac{T}{N}}\frac{1}{T^2}\left\|F_1^{0\prime}F^0\right\|\left\{\frac{1}{T^2}\sum_{t=1}^{T}\left\|\hat{f}_{1t} - H_1f_{1t}^0\right\|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{t=1}^{T}\left\|\sum_{i=1}^{N}\lambda_{it}^0\tilde{u}_{it}\right\|^2\right\}^{1/2} \\ =& \sqrt{\frac{T}{N}}O_P(1)O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})O_P\left(1\right) = \sqrt{\frac{1}{N}}O_P(\sqrt{T}\eta_{1NT} + C_{NT}^{-1}).\end{aligned}$$

It follows that $\frac{1}{T}\|F_1^{0\prime}I_6\| = O_P(\sqrt{T/N}\eta_{1NT} + N^{-1/2} + \eta_{2NT})$. Similarly, we can show that $\frac{1}{T}\|F_1^{0\prime}I_7\| =$

$O_P(\sqrt{T/N}\eta_{1NT} + N^{-1/2})$. Next, $\frac{1}{T}F_1^{0\prime}I_8 = \frac{1}{T}F_1^{0\prime}(I_{8,1} + I_{8,2} + I_{8,3} + I_{8,4})$. For $\frac{1}{T}F_1^{0\prime}I_{8,1}$, we have

$$
\begin{aligned}
\frac{1}{T}\|F_1^{0\prime}I_{8,1}\| &= \frac{1}{T^3}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{f}_{1s}'f_{1t}^0(\gamma_N(s,t)+\xi_{st})\right\| \\
&\leq \frac{1}{T^3}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{f}_{1s}'f_{1t}^0\gamma_N(s,t)\right\| + \frac{1}{T^3}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{f}_{1s}-H_1'f_{1s}^0)'f_{1t}^0\xi_{st}\right\| + \frac{1}{T^3}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}H_1 f_{1s}^{0\prime}f_{1t}^0\xi_{st}\right\| \\
&\equiv I + II + III.
\end{aligned}
$$

For I and II, we apply Assumption 3.2(iii) to obtain

$$
\begin{aligned}
I &\leq \frac{1}{2T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\left\|\hat{f}_{1s}\right\|^2 + \|f_{1t}^0\|^2\right)|\gamma_N(s,t)| \\
&\leq \frac{1}{2T}\left\{\frac{1}{T^2}\sum_{s=1}^{T}\left\|\hat{f}_{1s}\right\|^2 \max_{1\leq t\leq T}\sum_{t=1}^{T}|\gamma_N(s,t)| + \frac{1}{T^2}\sum_{t=1}^{T}\|f_{1t}^0\|^2 \max_{1\leq s\leq T}\sum_{s=1}^{T}|\gamma_N(s,t)|\right\} = O_P(T^{-1}),
\end{aligned}
$$

and

$$
\begin{aligned}
II &\leq \left\{\frac{1}{T^4}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|\hat{f}_{1s}-H_1'f_{1s}^0\right\|^2\|f_{1t}^0\|^2\right\}^{1/2}\left\{\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\xi_{st}\|^2\right\}^{1/2} \\
&\leq \frac{\left\|\hat{F}_1 - F_1^0H_1\right\|}{T}\frac{\|F_1^0\|}{T}O_P(N^{-1/2}) = N^{-1/2}O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1}),
\end{aligned}
$$

where we use the fact that $\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}E\|\xi_{st}\|^2 = O(N^{-1})$ under Assumption 3.2(iii) and Theorem 3.1(iii). Noting that $E\|f_{1s}^0\|^4 \leq Cs^2$ and $\max_{s,t}E|\xi_{st}|^2 \leq CN^{-1}$, we have

$$
\begin{aligned}
E\left\|\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}f_{1s}^{0\prime}f_{1t}^0\xi_{st}\right\| &\leq \frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\|f_{1s}^0\|^2|\xi_{st}|\right) \leq \frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{E\|f_{1s}^0\|^4\right\}^{1/2}\left\{E|\xi_{st}|^2\right\}^{1/2} \\
&\leq C\max_{s,t}\left\{E|\xi_{st}|^2\right\}^{1/2}\frac{1}{T^2}\sum_{s=1}^{T}s = O(N^{-1/2}).
\end{aligned}
$$

Then $III \leq \frac{1}{T^3}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}f_{1s}^{0\prime}f_{1t}^0\xi_{st}\right\|\|H_1\| = O_P\left(N^{-1/2}\right)$. It follows that $\frac{1}{T}\|F_1^{0\prime}I_{8,1}\| = O_P(N^{-1/2}\eta_{1NT} + T^{-1} + N^{-1/2})$. For $\frac{1}{T}\|F_1^{0\prime}I_{8,2}\|$, we have

$$
\begin{aligned}
\frac{1}{T}\|F_1^{0\prime}I_{8,2}\| &= \left\|\frac{1}{NT^3}\sum_{i=1}^{N}F_1^{0\prime}u_i\hat{b}_{2,i}'x_{2,i}'\hat{F}_1\right\| \leq \frac{1}{T}\max_i\frac{\|x_{2,i}'\hat{F}_1\|}{T}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\|F_1^{0\prime}u_i\|^2}{T^2}\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{2,i}\|^2\right)^{1/2} \\
&= T^{-1}o_P\left(N^{1/q}(\log T)^{(1+\epsilon)/2}\right)O_P(1)\,O_P(\eta_{2NT}) = o_P\left(T^{-1}N^{1/q}(\log T)^{(1+\epsilon)/2}\eta_{2NT}\right)
\end{aligned}
$$

where we use the fact that

$$
\begin{aligned}
\max_i \frac{1}{T}\|x_{2,i}'\hat{F}_1\| &\leq \max_i \frac{1}{T}\|x_{2,i}'F_1^0H_1\| + \max_i \frac{1}{T}\|x_{2,i}'(\hat{F}_1 - F_1^0H_1)\| \\
&\leq \max_i \frac{1}{T}\|x_{2,i}'F_1^0\| \, \|H_1\| + \sqrt{T}\max_i \frac{1}{\sqrt{T}} \left\|x_{2,i}'\right\| \frac{1}{T}\left\|\hat{F}_1 - F_1^0H_1\right\| \\
&= o_P(N^{1/q}(\log T)^{(1+\epsilon)/2}) + \sqrt{T}O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1}) = o_P(N^{1/q}(\log T)^{(1+\epsilon)/2})
\end{aligned}
$$

by arguments as used in the proof of Lemma A.2(i) in HJS and the result in Theorem 3.1(iii). Similarly, we have $\frac{1}{T}\|F_1^{0\prime}I_{8,3}\| = O_P(T^{-1}N^{1/q}(\log T)^{(1+\epsilon)/2}\,\eta_{2NT})$ and

$$
\begin{aligned}
\frac{1}{T}\|F_1^{0\prime}I_{8,4}\| &= \left\|\frac{1}{NT^3}\sum_{i=1}^N F_1^{0\prime}x_{2,i}\hat{b}_{2,i}\hat{b}_{2,i}'x_{2,i}'\hat{F}_1\right\| \leq \frac{1}{T}\max_i \frac{\|x_{2,i}'\hat{F}_1\|}{T}\max_i \frac{\|F_1^{0\prime}x_{2,i}\|}{T}\frac{1}{N}\sum_{i=1}^N \|\hat{b}_{2,i}\|^2 \\
&= o_P(T^{-1}N^{2/q}(\log T)^{(1+\epsilon)}\,\eta_{2NT}^2).
\end{aligned}
$$

It follows that $\frac{1}{T}\|F_1^{0\prime}I_8\| = O_P(T^{-1}+N^{-1/2})+o_P\left(T^{-1}[N^{1/q}(\log T)^{(1+\epsilon)/2}\,\eta_{2NT} + N^{2/q}(\log T)^{(1+\epsilon)}\,\eta_{2NT}^2]\right)$. For $\frac{1}{T}F_1^{0\prime}I_9$,

$$
\begin{aligned}
\frac{1}{T}\|F_1^{0\prime}I_9\| &\leq \frac{1}{T}\frac{\|F_1^{0\prime}F_2^0\|}{T}\left\|\frac{\Lambda_2^{0\prime}\Lambda_2^0}{N}\right\|\left(\frac{\|f_2^{0\prime}F_1^0\|\,\|H_1\|}{T} + \left\|F_2^0\right\|\frac{1}{T}\|\hat{F}_1 - F_1^0H_1\|\right) \\
&= O_P(T^{-1})[O_P(1) + O_P(\sqrt{T}\eta_{1NT} + C_{NT}^{-1})] = T^{-1}O_P(1+\sqrt{T}\eta_{1NT}),
\end{aligned}
$$

where we use Theorem 3.1(iii) and the fact that $\frac{\|F_1^{0\prime}F_2^0\|}{T} = O_P(1)$ by similar arguments as used in the proof of Lemma A.3(i). For $\frac{1}{T}F_1^{0\prime}I_{10}$,

$$
\begin{aligned}
\frac{1}{T}\|F_1^{0\prime}I_{10}\| =& \frac{1}{NT^3}\|F_1^{0\prime}F_1^0\Lambda_1^{0\prime}\Lambda_2^0 f_2^{0\prime}\hat{F}_1\| \\
\leq& O_P(N^{-1/2})\frac{\|F_1^{0\prime}F_1^0\|}{T^2}\frac{\left\|\Lambda_1^{0\prime}\Lambda_2^0\right\|}{\sqrt{N}}\left(\frac{\|F_2^{0\prime}F_1^0\|\,\|H_1\|}{T} + \left\|F_2^0\right\|\frac{1}{T}\|\hat{F}_1 - F_1^0H_1\|\right) \\
=& O_P(N^{-1/2})\left[O_P(1) + O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})\right] = O_P(N^{-1/2}).
\end{aligned}
$$

Similarly, we can show that $\frac{1}{T}\|F_1^{0\prime}I_{11}\| = O_P(N^{-1/2})$. Combining the above results and noting that $\sqrt{Td_T}\eta_{1NT} = o(1)$ by the proof of Theorem 3.1(iii), we obtain $\frac{1}{T}F_1^{0\prime}(\hat{F}_1 - F_1^0H_1) = O_P(T\eta_{1NT} + \delta_{NT}^{-1} + \eta_{2NT}) = O_P(T\eta_{1NT} + \delta_{NT}^{-1} + T^{-1/4}C_{NT}^{-1/2})$ and the conclusion in (ii) follows.

(iv) By (ii) and Theorem 3.1(iii),

$$
\begin{aligned}
&\left\|\frac{1}{T}\hat{F}_1'(\hat{F}_1 - F_1^0H_1)\right\| \leq \frac{1}{T}\left\|\hat{F}_1 - F_1^0H_1\right\|^2 + \|H_1\|\frac{1}{T}\left\|F_1^{0\prime}\left(\hat{F}_1 - F_1^0H_1\right)\right\| \\
&= O_P(T\eta_{1NT}^2 + C_{NT}^{-2}) + O_P(T\eta_{1NT} + \delta_{NT}^{-1} + T^{-1/4}C_{NT}^{-1/2}) = O_P(T\eta_{1NT} + \delta_{NT}^{-1} + T^{-1/4}C_{NT}^{-1/2}).
\end{aligned}
$$

(v) Note that $\frac{1}{T}\hat{u}_k^{*\prime}(\hat{F}_1H_1^{-1} - F_1^0) = \frac{1}{T}\hat{u}_k^{*\prime}[I_1 + ... + I_{11}]\,G_1 \equiv \frac{1}{T}(J_1 + ... + J_{11})$, where $G_1 =$

$\left(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1\right)^{-1}\left(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0\right)^{-1}$. Note that $\|G_1\| = O_P(1)$. For J_1,

$$
\begin{aligned}
\frac{1}{T}\|J_1\| &= \frac{1}{NT^3}\sum_{i=1}^{N}\left\|\hat{u}_k^{*\prime}x_{1,i}\hat{b}_{1,i}\hat{b}_{1,i}^{\prime}x_{1,i}^{\prime}\hat{F}_1G_1\right\| \\
&\leq \sqrt{T}\|G_1\|\frac{\|\hat{u}_k^*\|}{\sqrt{T}}\frac{\|\hat{F}_1\|}{T}\max_i\frac{\|x_{1,i}\|^2}{T^2}\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2 = O_P(\sqrt{T}d_T\eta_{1NT}^2),
\end{aligned}
$$

where $\frac{\|\hat{u}_k^*\|}{\sqrt{T}} \leq \frac{\|\hat{u}_k\|}{\sqrt{T}} + \frac{\|F_2^0\|}{\sqrt{T}}\|\lambda_{2k}^0\| + \frac{\|x_{2,k}\|}{\sqrt{T}}\|\hat{b}_{2,k}\| = O_P(1)$. For J_2 and J_3, we have

$$
\begin{aligned}
\frac{1}{T}\|J_2\| &\leq \frac{1}{NT^3}\sum_{i=1}^{N}\left\|\hat{u}_k^{*\prime}x_{1,i}\hat{b}_{1,i}\lambda_i^{0\prime}F_1^{0\prime}\hat{F}_1G_1\right\| \\
&\leq \sqrt{T}\|G_1\|\frac{\|F_1^{0\prime}\hat{F}_1\|}{T^2}\frac{\|\hat{u}_k^*\|}{\sqrt{T}}\max_i\|\lambda_i^0\|\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\|x_{1,i}\|^2}{T^2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2} = O_P(\sqrt{T}\eta_{1NT}),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{T}\|J_3\| \leq& \frac{1}{NT^3}\sum_{i=1}^{N}\left\|\hat{u}_k^{*\prime}x_{1,i}\hat{b}_{1,i}\tilde{u}_i^{\prime}\hat{F}_1G_1\right\| \leq \frac{1}{\sqrt{T}}\|G_1\|\frac{\|\hat{u}_k^*\|}{\sqrt{T}}\max_i\frac{\|x_{1,i}\|}{T}\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{\tilde{u}_i^{\prime}\hat{F}_1}{T}\right\|^2\right\}^{1/2} \\
=& O_P\left(\sqrt{\frac{d_T}{T}}\eta_{1NT}\right)O_P(1+Td_T\eta_{1NT}^2) = o_P(\sqrt{T}\eta_{1NT}),
\end{aligned}
$$

where we use the fact that $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{\tilde{u}_i^{\prime}\hat{F}_1}{T}\right\|^2 \leq \frac{2}{N}\sum_{i=1}^{N}\left(\left\|\frac{1}{T}\tilde{u}_i^{\prime}F_1^0H_1\right\|^2 + \frac{\|\tilde{u}_i\|^2}{T}\frac{1}{T}\left\|\hat{F}_1 - F_1^0H_1\right\|^2\right) = O_P(1) + O_P(T\eta_{1NT}^2 + C_{NT}^{-2}) = O_P(1+T\eta_{1NT}^2)$. For J_4, J_5, and J_6, we have

$$
\begin{aligned}
\frac{1}{T}\|J_4\| &\leq \frac{1}{NT^3}\sum_{i=1}^{N}\left\|\hat{u}_k^{*\prime}F^0\lambda_i^0\hat{b}_{1,i}^{\prime}x_{1,i}^{\prime}\hat{F}_1G_1\right\| \\
&\leq \|G_1\|\frac{\|\hat{u}_k^{*\prime}F^0\|}{T}\frac{\|\hat{F}_1\|}{T}\max_i\|\lambda_i^0\|\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\|x_{1,i}\|^2}{T^2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2} = O_P(\eta_{1NT}), \\
\frac{1}{T}\|J_5\| &\leq \frac{1}{NT^3}\sum_{i=1}^{N}\left\|\hat{u}_k^{*\prime}\tilde{u}_i\hat{b}_{1,i}^{\prime}x_{1,i}^{\prime}\hat{F}_1G_1\right\| \\
&\leq \|G_1\|\frac{\|\hat{u}_k^*\|}{\sqrt{T}}\frac{\|\hat{F}_1\|}{T}\max_i\frac{\|x_{1,i}\|}{T}\left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\|\tilde{u}_i\|^2}{T}\right\}^{1/2} = O_P(\sqrt{d_T}\eta_{1NT}),
\end{aligned}
$$

and

$$\begin{aligned}\frac{1}{T}\|J_6\| &\leq \frac{1}{NT^3}\left\|\sum_{i=1}^{N}\hat{u}_k^{*\prime}F^0\lambda_i^0\tilde{u}_i^{\prime}\hat{F}_1G_1\right\| \\ &\leq \|G_1\|\frac{1}{T}\left\|\frac{\hat{u}_k^{*\prime}F^0}{T}\right\|\left(\left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^0\tilde{u}_i^{\prime}F_1^0\right\|\|H_1\|+\left\|\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^0\tilde{u}_i^{\prime}(\hat{F}_1-F_1^0H_1)\right\|\right)=O_P(T^{-1}).\end{aligned}$$

For J_7,

$$\begin{aligned}\frac{1}{T}\|J_7\| \leq \frac{1}{NT^3}\left\|\sum_{i=1}^{N}\hat{u}_k^{*\prime}\tilde{u}_i\lambda_i^{0\prime}F_1^{0\prime}\hat{F}_1G_1\right\| &\leq \left\|\frac{1}{NT}\sum_{i=1}^{N}\hat{u}_k^{*\prime}\tilde{u}_i\lambda_i^{0\prime}\right\|\left\|\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)^{-1}\right\| \\ &\leq \left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i^0\tilde{u}_{it}\hat{u}_{kt}^*\right\|O_P(1)=O_P((NT)^{-1/2}+N^{-1}),\end{aligned}$$

where we use the fact that

$$\begin{aligned}\left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i^0\tilde{u}_{it}\hat{u}_{kt}^*\right\| \leq \left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i^0\tilde{u}_{it}u_{kt}\right\|+\left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i^0\tilde{u}_{it}f_{2t}^{0\prime}\lambda_{2k}^0\right\| \\ +\left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i^0\tilde{u}_{it}x_{2,ik}^{\prime}\hat{b}_{2,i}\right\|=O_P((NT)^{-1/2}+N^{-1});\end{aligned}$$

see, e.g., eq (B.1) in Bai (2003, p.164). For J_8 and J_9, we can show that

$$\begin{aligned}\frac{1}{T}\|J_8\| &\leq \left\{\left\|\frac{1}{NT^3}\sum_{i=1}^{N}\hat{u}_k^{*\prime}\tilde{u}_i\tilde{u}_i^{\prime}F_1^0H_1\right\|+\left\|\frac{1}{NT^3}\sum_{i=1}^{N}\hat{u}_k^{*\prime}\tilde{u}_i\tilde{u}_i^{\prime}(\hat{F}_1-F_1^0H_1)\right\|\right\}\|G_1\| \\ &\leq \left\{\frac{1}{NT}\sum_{i=1}^{N}\left\|\frac{\hat{u}_k^{*\prime}\tilde{u}_i}{T}\right\|\left\|\frac{\tilde{u}_i^{\prime}F_1^0}{T}\right\|\|H_1\|+\frac{1}{\sqrt{T}}\frac{1}{T}\left\|\hat{F}_1-F_1^0H_1\right\|\frac{\|\hat{u}_k^*\|}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N}\frac{\|\tilde{u}_i\|^2}{T}\right\}\|G_1\| \\ &=O_P(T^{-1})+T^{-1/2}O_P(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})=O_P(T^{-1/2}\eta_{1NT}+T^{-1}),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{T}\|J_9\| =&\frac{1}{T}\left\|\frac{1}{NT^2}\hat{u}_k^{*\prime}F_2^0\Lambda_2^{0\prime}\Lambda_2^0F_2^{0\prime}\hat{F}_1G_1\right\| \\ &\leq \frac{1}{T}\left(\frac{\|u_k\|}{\sqrt{T}}\frac{\|F_2^0\|}{\sqrt{T}}+\|\lambda_{2k}\|\frac{\|F_2^{0\prime}F_2^0\|}{T}+\|\hat{b}_{2,k}\|\frac{\|x_{2,k}\|}{\sqrt{T}}\frac{\|F_2^0\|}{\sqrt{T}}\right)\left\|\frac{\Lambda_2^{0\prime}\Lambda_2^0}{N}\right\|\left\|\frac{F_2^{0\prime}\hat{F}_1}{T}\right\|\|G_1\| \\ &=T^{-1}O_P(1+\sqrt{T}\eta_{1NT}),\end{aligned}$$

where we use the fact that $\left\|\frac{1}{T}F_2^{0\prime}\hat{F}_1\right\|=\left\|\frac{1}{T}F_2^{0\prime}F_1^0\right\|\|H_1\|+\left\|\frac{1}{T}F_2^{0\prime}(\hat{F}_1-F_1^0H_1)\right\|=O_P(1)+\sqrt{T}O_P(\eta_{1NT}$

$+T^{-1/2}C_{NT}^{-1}) = O_P(1+\sqrt{T}\eta_{1NT})$. For J_{10} and J_{11},

$$\begin{aligned}
\frac{1}{T}\|J_{10}\| &= \frac{1}{T}\left\|\frac{1}{NT^2}\hat{u}_k^{*\prime}F_1^0\Lambda_1^{0\prime}\Lambda_2^0F_2^{0\prime}\hat{F}_1G_1\right\| \\
&\leq \frac{1}{T\sqrt{N}}\left(\frac{\|u_k\|}{\sqrt{T}}\frac{\|F_2^0\|}{\sqrt{T}}+\|\lambda_{2k}\|\frac{\|F_2^{0\prime}F_2^0\|}{T}+\|\hat{b}_{2,i}\|\frac{\|x_{2,i}\|}{\sqrt{T}}\frac{\|F_2^0\|}{\sqrt{T}}\right)\left\|\frac{\Lambda_1^{0\prime}\Lambda_2^0}{\sqrt{N}}\right\|\left\|\frac{F_2^{0\prime}\hat{F}_1}{T}\right\|\|G_1\| \\
&= T^{-1}N^{-1/2}O_P(1+\sqrt{T}\eta_{1NT}), \text{ and} \\
\frac{1}{T}\|J_{11}\| &= \frac{1}{T}\left\|\frac{1}{NT^2}\hat{u}_k^{*\prime}F_2^0\Lambda_2^{0\prime}\Lambda_1^0F_1^{0\prime}\hat{F}_1G_1\right\| \\
&\leq \frac{1}{\sqrt{N}}\left(\frac{\|u_k\|}{\sqrt{T}}\frac{\|F_2^0\|}{\sqrt{T}}+\|\lambda_{2k}\|\frac{\|F_2^{0\prime}F_2^0\|}{T}+\|\hat{b}_{2,i}\|\frac{\|x_{2,i}\|}{\sqrt{T}}\frac{\|F_2^0\|}{\sqrt{T}}\right)\left\|\frac{\Lambda_2^{0\prime}\Lambda_1^0}{\sqrt{N}}\right\|\left\|\frac{F_1^{0\prime}\hat{F}_1}{T^2}\right\|\|G_1\| = O_P(N^{-1/2}).
\end{aligned}$$

Combining the above results yields the conclusion in (v). ■

Proof of Lemma A.6. We only prove the first part of (i)-(iv) as the second part can be shown analogously by the repeated use of the fact that $\max_i \frac{\|x_{1,i}\|}{T} = O_P(d_T^{1/2})$.

(i) By the decomposition in (B.1),

$$R_{1i} = \frac{1}{T^2}x_{1,i}'(P_{\hat{F}_1}-P_{F_1^0H_1})\ \hat{u}_i^* = \frac{1}{T^2}x_{1,i}'(\hat{p}_1+\hat{p}_2+\hat{p}_3+\hat{p}_4)\ \hat{u}_i^* \equiv D_{1i}+D_{2i}+D_{3i}+D_{4i}. \quad \text{(B.2)}$$

For D_{1i}, we have $\|D_{1i}\| = \frac{1}{T^4}\left\|x_{1,i}'(\hat{F}_1-F_1^0H_1)(\hat{F}_1-F_1^0H_1)'\hat{u}_i^*\right\| \leq \frac{1}{T^{1/2}}\frac{1}{T^2}\left\|\hat{F}_1-F_1^0H_1\right\|^2\frac{\|x_{1,i}\|}{T}\frac{\|\hat{u}_i^*\|}{T^{1/2}} = T^{-1/2}O_P(\eta_{1NT}^2+T^{-1}C_{NT}^{-2})$ by Theorem 3.1(iii). For D_{2i},

$$\begin{aligned}
\|D_{2i}\| &= \left\|\frac{1}{T^2}\sum_{s=1}^T(\hat{f}_{1s}-H_1'f_{1s}^0)'H_1'\frac{1}{T^2}\sum_{t=1}^T f_{1t}^0\hat{u}_{it}^*x_{1,is}'\right\| \\
&\leq \frac{1}{\sqrt{T}}\left(\frac{1}{T^2}\sum_{s=1}^T\left\|\hat{f}_{1s}-H_1'f_{1s}^0\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^T\left\|\frac{1}{T^2}\sum_{t=1}^T f_{1t}^0\hat{u}_{it}^*x_{1,is}'\right\|^2\right)^{1/2}\|H_1\| \\
&= T^{-1/2}O_P(\eta_{1NT}+T^{-1/2}C_{NT}^{-1}),
\end{aligned}$$

where we use the fact that $\frac{1}{T}\sum_{s=1}^T\left\|\frac{1}{T}(\frac{1}{T}\sum_{t=1}^T f_{1t}^0\hat{u}_{it}^*)x_{1,is}'\right\|^2 = O_P(1)$ under Assumptions 3.1-3.2. For D_{3i}, we apply Lemma A.5(v) to obtain $\|D_{3i}\| \leq \frac{1}{T}\frac{\|x_{1,i}'F_1^0\|}{T^2}\frac{1}{T}\left\|(F_1^0H_1-\hat{F}_1)'\hat{u}_i^*\right\|\|H_1\| = O_P(T^{-1/2}\eta_{1NT}+T^{-1}\delta_{NT}^{-1})$. Noting that $I_{r_1}-\frac{1}{T^2}H'F^{0\prime}F^0H = O_P(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})$ by the proof of Lemma A.5(i),

$$\begin{aligned}
\|D_{4i}\| &\leq \frac{1}{T}\left\|I_{r_1}-\left(\frac{1}{T^2}H'F^{0\prime}F^0H\right)^{-1}\right\|\frac{\left\|x_{1,i}'F_1^0\right\|}{T^2}\frac{\|F_1^{0\prime}\hat{u}_i^*\|}{T}\|H_1\|^2 \\
&= T^{-1}O_P(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})O_P(1) = O_P(T^{-1}\eta_{1NT}+T^{-3/2}C_{NT}^{-1}),
\end{aligned}$$

where we use the fact that $\left\|\frac{x_{1,i}'F_1^0}{T^2}\right\| = O_P(1)$ and $\left\|\frac{F_1^{0\prime}\ \hat{u}_i^*}{T}\right\| \leq \left\|\frac{F_1^{0\prime}u_i}{T}\right\|+\left\|\frac{F_1^{0\prime}F_2^0}{T}\right\|\|\lambda_{2i}^0\|+\left\|\frac{F_1^{0\prime}x_{2,i}}{T}\right\|\left\|\hat{b}_{2,i}\right\| =$

$O_P(1)$ by the proofs of Lemma A.5(iv) and Lemma A.3(i). Consequently, $R_{1i} = O_P(T^{-1/2}\eta_{1NT} + T^{-1}C_{NT}^{-1})$ and the first part of (i) directly follows.

(ii) By the proof of Theorem 3.1(iii), we have $\hat{F}_1 H_1^{-1} - F_1^0 = [I_1 + ... + I_{11}]\, G_1$, where $G_1 = \left(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1\right)^{-1}\left(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0\right)^{-1} = O_P(1)$. Then we have $\frac{1}{T^2}x_{1,i}' M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 = \frac{-1}{T^2}x_{1,i}' M_{\hat{F}_1}(\hat{F}_1 H_1^{-1} - F_1^0)\lambda_{1i}^0 = \frac{-1}{T^2}x_{1,i}' M_{\hat{F}_1}[I_1 + ... + I_{11}]\, G_1 \lambda_{1i}^0 \equiv L_{1i} + ... + L_{11i}$. Note that

$$
\begin{aligned}
L_{2i} &= \frac{1}{NT^2}\sum_{j=1}^{N} x_{1,i}' M_{\hat{F}_1} x_{1,j}\hat{b}_{1,j}\lambda_j^{0\prime}\frac{F^{0\prime}\hat{F}_1}{T^2}\left(\frac{F_1^{0\prime}\hat{F}_1}{T^2}\right)^{-1}\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)^{-1}\lambda_{1i}^0 \\
&= \frac{1}{NT^2}\sum_{j=1}^{N} x_{1,i}' M_{\hat{F}_1} x_{1,j}\hat{b}_{1,j}\lambda_{1j}^{0\prime}\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)^{-1}\lambda_{1i}^0 + \bar{L}_{2i} = \frac{1}{NT^2}\sum_{j=1}^{N} x_{1,i}' M_{\hat{F}_1} x_{1,j}\hat{b}_{1,j} a_{ij} + \bar{L}_{2i}, \text{ and} \\
L_{7i} &= \frac{-1}{NT^2}\sum_{j=1}^{N} x_{1,i}' M_{\hat{F}_1}\tilde{u}_j\lambda_j^{0\prime}\frac{F^{0\prime}\hat{F}_1}{T^2}\left(\frac{F_1^{0\prime}\hat{F}_1}{T^2}\right)^{-1}\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)^{-1}\lambda_{1i}^0 = \frac{-1}{NT^2}\sum_{j=1}^{N} x_i' M_{\hat{F}_1} u_j a_{ij} + \bar{L}_{7i,1} + \bar{L}_{7i,2},
\end{aligned}
$$

where $\bar{L}_{2i} = \frac{1}{NT^2}\sum_{j=1}^{N} x_{1,i}' M_{\hat{F}_1} x_{1,j}\hat{b}_{1,j}\lambda_{2j}^{0\prime}\frac{F_2^{0\prime}\hat{F}_1}{T^2}G_1\lambda_{1i}^0$, $\bar{L}_{7i,1} = \frac{-1}{NT^2}\sum_{j=1}^{N} x_{1,i}' M_{\hat{F}_1} x_{2,j}\hat{b}_{2j}' a_{ij}$, $\bar{L}_{7i,2} = -\frac{1}{NT^2}\sum_{j=1}^{N} x_{1,i}' M_{\hat{F}_1} u_j \lambda_{2j}^{0\prime}\frac{F_2^{0\prime}\hat{F}_1}{T^2}G_1\lambda_{1i}^0$, and $a_{ij} = \lambda_{1i}^{0\prime}(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}\lambda_{1j}^0$. It follows that $R_{2i} = \frac{1}{T^2}x_{1,i}' M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 - \frac{1}{NT^2}\sum_{j=1}^{N} x_{1,i}' M_{\hat{F}_1} x_{1,j}\hat{b}_{1,j} a_{ij} + \frac{1}{NT^2}\sum_{j=1}^{N} a_{ij} x_{1,i}' M_{\hat{F}_1} u_j = \bar{L}_{1i} + ... + \bar{L}_{11i}$, where $\bar{L}_{li} = L_{li}$ for $l = 1, 3, 4, 5, 6, 8, 9, 10, 11$. For $\bar{L}_{1i}$,

$$
\begin{aligned}
\left\|\bar{L}_{1i}\right\| =& \frac{1}{T^2}\left\| x_{1,i}' M_{\hat{F}_1}\frac{1}{NT^2}\sum_{j=1}^{N} x_{1,j}\hat{b}_{1,j}\hat{b}_{1,j}' x_{1,j}'\hat{F}_1 G_1 \lambda_{1i}^0\right\| \\
\leq& \|G_1\|\left\|\lambda_{1i}^0\right\|\frac{\|x_{1,i}\|}{T}\frac{\|\hat{F}_1\|}{T}\max_j\frac{\|x_{1,j}\|^2}{T^2}\frac{1}{N}\sum_{j=1}^{N}\|\hat{b}_{1,j}\|^2 = O_P(d_T\eta_{1NT}^2).
\end{aligned}
$$

where we use the fact that $\left\|M_{\hat{F}_1}\right\|_{\mathrm{sp}} = 1$. For $\bar{L}_{2i}$ and $\bar{L}_{3i}$, we have

$$
\begin{aligned}
\|\bar{L}_{2i}\| &= \frac{1}{NT^2}\left\| x_{1,i}' M_{\hat{F}_1}\sum_{j=1}^{N} x_{1,j}\hat{b}_{1,j}\lambda_{2j}^{0\prime}\frac{F_2^{0\prime}\hat{F}_1}{T^2}G_1\lambda_{1i}^0\right\| \\
&\leq \frac{1}{T}\|G_1\|\|\lambda_{1i}^0\|\frac{\|x_{1,i}\|}{T}\left\|\frac{F_2^{0\prime}\hat{F}_1}{T}\right\|\max_j\|\lambda_{2j}^0\|\left\{\frac{1}{N}\sum_{j=1}^{N}\frac{\|x_{1,j}\|^2}{T^2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{j=1}^{N}\|\hat{b}_{1,j}\|^2\right\}^{1/2} \\
&= O_P(T^{-1}\eta_{1NT})
\end{aligned}
$$

and

$$
\begin{aligned}
\left\|\bar{L}_{3i}\right\| &\leq \frac{1}{T}\|G_1\|\|\lambda_{1i}^0\|\frac{\|x_{1,i}\|}{T}\max_j\frac{\|x_{1,j}\|}{T}\left\{\frac{1}{N}\sum_{j=1}^{N}\|\hat{b}_{1,j}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{\tilde{u}_j'\hat{F}_1}{T}\right\|^2\right\}^{1/2} \\
&= T^{-1}O_P(\sqrt{d_T}\eta_{1NT})O_P(1+\sqrt{T}\eta_{1NT}) = O_P(T^{-1}\sqrt{d_T}\eta_{1NT} + T^{-1/2}d_T^{1/2}\eta_{1NT}^2),
\end{aligned}
$$

where we use the fact that $\frac{1}{N}\sum_{j=1}^N \left\|\frac{\tilde{u}_j'\hat{F}_1}{T}\right\|^2 \leq \frac{1}{N}\sum_{j=1}^N \{\left\|\frac{1}{T}\tilde{u}_j' F_1^0 H_1\right\|^2 + \left\|\frac{1}{T}\tilde{u}_j'(\hat{F}_1 - F_1^0 H_1)\right\|^2\} = O_P(1) + TO_P(\eta_{1NT}^2 + T^{-1}C_{NT}^{-1})$. For $\bar{L}_{4i}$, we have

$$
\begin{aligned}
\|\bar{L}_{4i}\| &= \frac{1}{NT^4}\left\|x_{1,i}' M_{\hat{F}_1}\sum_{j=1}^N F^0\lambda_j^0 \hat{b}_{1,j}' x_{1,j}'\hat{F}_1 G_1 \lambda_{1i}^0\right\| \\
&\leq \|G_1\|\,\|\lambda_{1i}^0\|\frac{\left\|\hat{F}_1\right\|}{T}\left\{\frac{1}{NT}\sum_{j=1}^N\left(\left\|\frac{x_{1,i}' M_{\hat{F}_1}(F_1^0 - \hat{F}_1 H_1^{-1})}{T^2}\right\|\|\lambda_{1j}^0\| + \left\|\frac{x_{1,i}' M_{\hat{F}_1} F_2^0}{T^2}\right\|\|\lambda_{2j}^0\|\right)\|x_{1,j}\hat{b}_{1,j}\|\right\} \\
&\leq \|G_1\|\,\|\lambda_{1i}^0\|\frac{\left\|\hat{F}_1\right\|}{T}\max_j \|\lambda_j^0\|\left\{\left\|\frac{x_{1,i}' M_{\hat{F}_1}(F_1^0 - \hat{F}_1 H_1^{-1})}{T^2}\right\| + \left\|\frac{x_{1,i}' M_{\hat{F}_1} F_2^0}{T^2}\right\|\right\} \\
&\times\left\{\frac{1}{N}\sum_{j=1}^N\frac{\|x_{1,j}\|^2}{T^2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{j=1}^N\|\hat{b}_{1,j}\|^2\right\}^{1/2} \\
&= [O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1}) + O_P(T^{-1/2}\eta_{1NT} + T^{-1})]O_P(\eta_{1NT}) = O_P(\eta_{1NT}^2 + T^{-1/2}C_{NT}^{-1}\eta_{1NT}),
\end{aligned}
$$

where we use the fact that $\frac{1}{T}\left\|F_1^0 - \hat{F}_1 H_1^{-1}\right\| = O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})$ and that

$$
\begin{aligned}
\left\|\frac{x_{1,i}' M_{\hat{F}_1} F_2^0}{T^2}\right\| &\leq \left\|\frac{x_{1,i}' F_2^0}{T^2}\right\| + \left\|\frac{x_{1,i}'\hat{F}_1}{T^2}\frac{\hat{F}_1' F_2^0}{T^2}\right\| \\
&\leq \left\|\frac{x_{1,i}' F_2^0}{T^2}\right\| + \left\|\frac{x_{1,i}'\hat{F}_1}{T^2}\right\|\left\{\left\|\frac{(\hat{F}_1 - F_1^0 H_1)' F_2^0}{T^2}\right\| + \|H_1\|\left\|\frac{F_1^{0\prime} F_2^0}{T^2}\right\|\right\} \\
&= O_P(T^{-1}) + \left[O_P(T^{-1/2}\eta_{1NT} + T^{-1}C_{NT}^{-1}) + O_P(T^{-1})\right] = O_P(T^{-1/2}\eta_{1NT} + T^{-1}).
\end{aligned}
$$

For $\bar{L}_{5i}$ to $\bar{L}_{10i}$, we have

$$
\begin{aligned}
\|\bar{L}_{5i}\| &= \frac{1}{NT^4}\left\|x_{1,i}' M_{\hat{F}_1}\sum_{j=1}^N \tilde{u}_j \hat{b}_{1,j}' x_{1,j}'\hat{F}_1 G_1 \lambda_{1i}^0\right\| \\
&\leq \frac{1}{T}\|G_1\|\,\|\lambda_{1i}^0\|\frac{\|x_{1,i}\|}{T}\max_j\frac{\|x_{1,j}\|}{T}\left\{\frac{1}{N}\sum_{j=1}^N\frac{\|\hat{F}_1'\tilde{u}_j\|^2}{T^2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{j=1}^N\|\hat{b}_{1,j}\|^2\right\}^{1/2} = O_P(T^{-1}\sqrt{d_T}\eta_{1NT}), \\
\|\bar{L}_{6i}\| &\leq \|G_1\|\,\|\lambda_{1i}^0\|\left\{\frac{1}{NT^2}\sum_{j=1}^N\left(\left\|\frac{x_{1,i}' M_{\hat{F}_1}(F_1^0 - \hat{F}_1 H_1^{-1})}{T^2}\right\|\|\lambda_{1j}^0\| + \left\|\frac{x_{1,i}' M_{\hat{F}_1} F_2^0}{T^2}\right\|\|\lambda_{2j}^0\|\right)\left\|\tilde{u}_j'\hat{F}_1\right\|\right\} \\
&\leq \frac{O_P(1)}{T}\left(\left\|\frac{x_{1,i}' M_{\hat{F}_1}(F_1^0 - \hat{F}_1 H_1^{-1})}{T^2}\right\| + \left\|\frac{x_i' M_{\hat{F}_1} F_2^0}{T^2}\right\|\right)\left\{\frac{1}{N}\sum_{j=1}^N\|\lambda_j^0\|^2\right\}^{1/2}\left\{\frac{1}{NT^2}\sum_{j=1}^N\|\tilde{u}_j'\hat{F}_1\|^2\right\}^{1/2} \\
&= T^{-1}[O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1}) + O_P(T^{-1/2}\eta_{1NT} + T^{-1})] = O_P(T^{-1}\eta_{1NT} + T^{-3/2}C_{NT}^{-1}),
\end{aligned}
$$

$$\begin{aligned}
\|\bar{L}_{7i,1}\| &\leq \frac{1}{NT^2}\sum_{j=1}^{N}\left\|x_{1,i}'M_{\hat{F}_1}x_{2,j}\hat{b}_{2j}'a_{ij}\right\| \leq \frac{1}{NT^2}\sum_{j=1}^{N}\left\{\left\|x_{1,i}'M_{F_1^0}x_{2,j}\hat{b}_{2j}'a_{ij}\right\| + \left\|x_{1,i}'(P_{\hat{F}_1}-P_{F_1^0})x_{2,j}\hat{b}_{2j}'a_{ij}\right\|\right\} \\
&\leq T^{-1}\left\{\frac{1}{N}\sum_{j=1}^{N}\frac{\left\|x_{1,i}'M_{F_1^0}x_{2,j}\right\|^2}{T^2}a_{ij}^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{j=1}^{N}\left\|\hat{b}_{2j}\right\|^2\right\}^{1/2} \\
&+ T^{-1/2}\frac{\|x_{1,i}\|}{T}\left\|P_{\hat{F}_1}-P_{F_1^0}\right\|\left\{\frac{1}{N}\sum_{j=1}^{N}\frac{\|x_{2,j}\|^2}{T}a_{ij}^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{j=1}^{N}\left\|\hat{b}_{2j}\right\|^2\right\}^{1/2} \\
&= T^{-1}O_P(\eta_{2NT}) + T^{-1/2}O_P(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})O_P(\eta_{2NT}) = O_P(T^{-5/4}C_{NT}^{-1/2}+T^{-1/2}\eta_{1NT}+\eta_{1NT}^2), \\
\|\bar{L}_{7i,2}\| &= \frac{1}{NT^4}\left\|x_{1,i}'M_{\hat{F}_1}\sum_{j=1}^{N}\tilde{u}_j\lambda_{2j}^{0\prime}F_2^{0\prime}\hat{F}_1G_1\lambda_{1i}^0\right\| \\
&\leq \frac{1}{NT^3}\|G_1\|\|\lambda_{1i}^0\|\left\|\frac{F_2^{0\prime}\hat{F}_1}{T}\right\|\left\{\left\|\sum_{j=1}^{N}x_{1,i}'\tilde{u}_j\lambda_{2j}^{0\prime}\right\| + \left\|\frac{x_{1,i}'\hat{F}_1}{T^2}\sum_{j=1}^{N}\hat{F}_1'\tilde{u}_j\lambda_{2j}^{0\prime}\right\|\right\} \\
&\leq \frac{O_P(1)}{T^2}\left\{\left\|\frac{1}{NT}\sum_{j=1}^{N}x_{1,i}'\tilde{u}_j\lambda_{2j}^{0\prime}\right\| + \frac{\left\|x_{1,i}'\hat{F}_1\right\|}{T^2}\left\{\frac{1}{NT^2}\sum_{j=1}^{N}\left\|\hat{F}_1'\tilde{u}_j\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{j=1}^{N}\|\lambda_{2j}^0\|^2\right\}^{1/2}\right\} \\
&= O_P(T^{-2}), \\
\|\bar{L}_{8i}\| &= \frac{1}{NT^4}\left\|x_{1,i}'M_{\hat{F}_1}\sum_{j=1}^{N}\tilde{u}_j\tilde{u}_j'\hat{F}_1G_1\lambda_{1i}^0\right\| \\
&\leq \frac{1}{NT^4}\|G_1\|\|\lambda_{1i}^0\|\left\{\left\|\sum_{j=1}^{N}x_{1,i}'\tilde{u}_j\tilde{u}_j'\hat{F}_1\right\| + \left\|\frac{x_{1,i}'\hat{F}_1}{T^2}\sum_{j=1}^{N}\hat{F}_1'\tilde{u}_j\tilde{u}_j'\hat{F}_1\right\|\right\} \\
&\leq \frac{O_P(1)}{T^2}\left\{\left(\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{x_{1,i}'\tilde{u}_j}{T}\right\|^2\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{\hat{F}_1'\tilde{u}_j}{T}\right\|^2\right)^{1/2} + \left\|\frac{x_{1,i}'\hat{F}_1}{T^2}\right\|\left(\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{\hat{F}_1'\tilde{u}_j}{T}\right\|^2\right)\right\} \\
&= O_P(T^{-2}), \\
\|\bar{L}_{9i}\| &\leq \frac{1}{T^2}\|G_1\|\|\lambda_{1i}^0\|\left\|\frac{\hat{F}_1'F_2^0}{T}\right\|\left\|\frac{\Lambda_2^{0\prime}\Lambda_2^0}{N}\right\|\left\{\left\|\frac{x_{1,i}'F_2^0}{T}\right\| + \frac{x_{1,i}'\hat{F}_1}{T^2}\frac{\hat{F}_1'F_2^0}{T}\right\} = O_P(T^{-2}), \text{ and} \\
\|\bar{L}_{10,i}\| &\leq \frac{1}{\sqrt{N}T}\|G_1\|\|\lambda_{1i}^0\|\frac{\|x_{1,i}\|}{T}\left\|\frac{F_1^0}{T}\right\|\left\|\frac{\hat{F}_1'F_2^0}{T}\right\|\left\|\frac{\Lambda_1^{0\prime}\Lambda_2^0}{\sqrt{N}}\right\| = O_P(N^{-1/2}T^{-1}).
\end{aligned}$$

Similarly, we can show that $\|\bar{L}_{11,i}\| = O_P(N^{-1/2}T^{-1})$. Then $R_{2i} = O_P(T^{-5/4}C_{NT}^{-1/2}+T^{-1/2}\eta_{1NT}+d_T\eta_{1NT}^2+T^{-1}\delta_{NT}^{-1})$ and the first part of (ii) follows.

(iii) By Lemma A.5(i),

$$
\begin{aligned}
\|R_{3i}\| &\leq \left\| \frac{1}{NT^2} \sum_{j=1}^{N} a_{ij} x_{1,i}' (P_{F_1^0} - P_{\hat{F}_1}) u_j \right\| \leq \frac{\|x_{1,i}\|}{T} \left\| P_{F_1^0} - P_{\hat{F}_1} \right\| \frac{1}{NT} \sum_{j=1}^{N} |a_{ij}| \, \|u_j\| \\
&\leq \frac{1}{T^{1/2}} \frac{\|x_{1,i}\|}{T} \left\| P_{F_1^0} - P_{\hat{F}_1} \right\| \left\{ \frac{1}{N} \sum_{j=1}^{N} a_{ij}^2 \right\}^{1/2} \left\{ \frac{1}{NT} \sum_{j=1}^{N} \|u_j\|^2 \right\}^{1/2} \\
&= O_P(T^{-1/2} \eta_{1NT} + T^{-1} C_{NT}^{-1}),
\end{aligned}
$$

where we use the fact that $\frac{1}{N}\sum_{j=1}^{N} a_{ij}^2 = \frac{1}{N}\sum_{j=1}^{N} [\lambda_{1i}^{0\prime}(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}\lambda_{1j}^0]^2 \leq \left\|\lambda_{1i}^0\right\|^2 \left\|(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}\right\|^2 \frac{1}{N}\sum_{j=1}^{N} \left\|\lambda_{1j}^0\right\|^2 = O_P(1)$ under Assumption 3.2(i). Then the first part of (iii) follows.

(iv) As in the proof of Lemma A.3(i), we can show that $R_{4i} = O_P(T^{-1})$ and $N^{-1}\sum_{i=1}^{N} \|R_{4i}\|^2 = O_P(d_T T^{-2})$. ■

Proof of Lemma A.7. (i) By the proof of Lemma A.5(ii),

$$
\begin{aligned}
\max_i \|\hat{b}_{2,i}\| &\leq O_P(1) \max_i \left\{ \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} \left(u_i + F_2^0 \lambda_{2i}^0 \right) \right\| + \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 \right\| + \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} x_{1,i} \hat{b}_{1,i} \right\| \right\} \\
&\equiv O_P(1) \{ III_1 + III_2 + III_3 \}, \text{ say.}
\end{aligned}
$$

Note that

$$
\begin{aligned}
\max_i \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} u_i \right\| &\leq \max_i \left\| \frac{1}{T} x_{2,i}' M_{F_1^0} u_i \right\| + \max_i \left\| \frac{1}{T} x_{2,i}' (P_{\hat{F}_1} - P_{F_1^0}) u_i \right\| \\
&\leq \max_i \left\| \frac{1}{T} x_{2,i}' u_i \right\| + \frac{1}{T} \max_i \left\| \frac{1}{T} x_{2,i}' F_1^0 (\frac{1}{T^2} F_1^{0\prime} F_1^0)^{-1} \frac{1}{T} F_1^{0\prime} u_i \right\| \\
&\quad + \max_i \frac{1}{\sqrt{T}} \|x_{2,i}\| \max_i \frac{1}{\sqrt{T}} \|u_i\| \left\| P_{\hat{F}_1} - P_{F_1^0} \right\| \\
&= O_P(T^{-1/2} (\log T)^3) + T^{-1} o_P(N^{2/q} (\log T)^{(1+\epsilon)}) + O_P(d_T^{1/2} T^{-1} + (NT)^{-1/2}) \\
&= o_P(1).
\end{aligned}
$$

By the same token, $\max_i \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} F_2^0 \lambda_{2i}^0 \right\| \leq O_P(T^{-1/2} (\log T)^3 + T^{-1} N^{2/q} (\log T)^{(1+\epsilon)}) = o_P(1)$,

$$
\begin{aligned}
\max_i \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 \right\| &= \max_i \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} \left(\hat{F}_1 H_1 - F_1^0 \right) \lambda_{1i}^0 \right\| \\
&\leq T^{1/2} \max_i \frac{\|x_{2,i}\|}{T^{1/2}} \max_i \left\| \lambda_{1i}^0 \right\| \frac{1}{T} \left\| \hat{F}_1 H_1 - F_1^0 \right\| \\
&= T^{1/2} N^{1/(2q)} O_P(d_T^{1/2} T^{-1} + (NT)^{-1/2}) = o_P(1), \text{ and} \\
\max_i \left\| \frac{1}{T} x_{2,i}' M_{\hat{F}_1} x_{1,i} \right\| &= \max_i \left\| \frac{1}{T} x_{2,i}' M_{F_1^0} x_{1,i} \right\| + \max_i \left\| \frac{1}{T} x_{2,i}' (P_{\hat{F}_1} - P_{F_1^0}) x_{1,i} \right\| \\
&= o_P(N^{1/q} (\log T)^{(1+\epsilon)/2}) + \sqrt{T d_T} O_P(d_T^{1/2} T^{-1} + (NT)^{-1/2}).
\end{aligned}
$$

In addition, by the proof of Lemma A.5(ii) and Theorem 3.2(i), $\max_i \|\hat{b}_{1,i}\| = d_T O_P(d_T^{1/2} T^{-1/2} +$

$d_T^{1/2}\eta_{1NT}+\lambda) = d_T O_P(d_T^{1/2}T^{-1/2}+\lambda)$. It follows that

$$\begin{aligned}\max_i \|\hat{b}_{2,i}\| &= \left[o_P(N^{1/q}(\log T)^{(1+\epsilon)/2}) + \sqrt{Td_T}O_P(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})\right]\max_i \|\hat{b}_{1,i}\| + o_P(1)\\ &= \left[o_P(N^{1/q}(\log T)^{(1+\epsilon)/2}) + \sqrt{Td_T}O_P(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})\right] d_T O_P(d_T^{1/2}T^{-1/2}+\lambda) + o_P(1)\\ &= o_P(1).\end{aligned}$$

Now, note that

$$\begin{aligned}&P\left(\max_i\left\|\frac{x_{1,i}'\hat{u}_i^*}{T^2}\right\| > c\psi_{NT}\right)\\ \le& P\left(\max_i\left\|\frac{x_{1,i}'u_i}{T^2}\right\| > \frac{c\psi_{NT}}{3}\right) + P\left(\max_i\left\|\frac{x_{1,i}'F_2^0\lambda_{2i}^0}{T^2}\right\| > \frac{c\psi_{NT}}{3}\right) + P\left(\max_i\left\|\frac{x_{1,i}'x_{2,i}}{T^2}\right\|\left\|\hat{b}_{2,i}\right\| > \frac{c\psi_{NT}}{3}\right).\end{aligned}$$

The first term on the right hand side (rhs) of the last equation is $o(N^{-1})$ by Lemma A.2(i) in HJS. Since $E\left(f_{2t}^0\lambda_{2i}^{0\prime}\right)=0$, each element of $f_{2t}^0\lambda_{2i}^{0\prime}$ can play the same role as u_{it}, the second term on the rhs is also $o\left(N^{-1}\right)$. Since $\max_i\left\|\hat{b}_{2,i}\right\| = o_P(1)$ and each element $x_{2,it}$ can plays the same role as u_{it}, the third term on the rhs is also $o\left(N^{-1}\right)$. Then (i) follows.

(ii) Note that $\frac{1}{T^2}x_{1,i}'M_{F_1^0}\hat{u}_i^* = \frac{1}{T^2}x_{1,i}'\hat{u}_i^* - \frac{1}{T^2}x_{1,i}'P_{F_1^0}\hat{u}_i^*$. The first term is studied in (i). For the second term, we have

$$\max_i\left\|\frac{1}{T^2}x_{1,i}'P_{F_1^0}\ \hat{u}_i^*\right\|^2 \le d_T^2\left[\mu_{\min}\left(\frac{d_T}{T^2}F_1^{0\prime}F_1^0\right)\right]^{-1}\max_i\frac{\|x_{1,i}\|^2}{d_T T^2}\frac{\left\|F_1^{0\prime}\hat{u}_i^*\right\|^2}{T^4},$$

where $\liminf_{T\to\infty}\mu_{\min}\left(\frac{d_T}{T^2}F_1^{0\prime}F_1^0\right) \ge \rho_{\min}/2$ a.s. and $\limsup_{T\to\infty}\frac{\|x_{1,i}\|^2}{d_T T^2} \le (1+c)\rho_{\max}$ a.s. by Lemma A.2(i)-(ii). It follows that for some $c>0$

$$P\left(\max_i\left\|\frac{1}{T^2}x_{1,i}'P_{F_1^0}\hat{u}_i^*\right\| > cd_T\psi_{NT}\right) \le P\left(\max_i\frac{1}{T^2}\left\|F_1^{0\prime}\ \hat{u}_i^*\right\| > c\psi_{NT}\right) = o(N^{-1}),$$

where the equality holds by analogous arguments as used in the proof of (i). Consequently we have $P\left(\left\|\frac{1}{T^2}x_{1,i}'\ M_{F_1^0}\hat{u}_i^*\right\| > cd_T\psi_{NT}\right) = o(N^{-1})$. ■

Proof of Lemma A.8. (i) Note that $R_{1i} = \frac{-1}{T^2}x_{1,i}'(M_{\hat{F}_1}-M_{F_1^0})\hat{u}_i^*$, where recall that $\hat{u}_i^* = u_i + F_2^0\lambda_{2i}^0 - x_{2,i}\hat{b}_{2,i}$. By (B.2) and the proof of Lemma A.6(i), it suffices to study the probability bounds for $\max_i\|D_{li}\|$ where $l=1,2,3,4$. Let $\bar{\eta}_{1NT} = \frac{1}{T}\left\|\hat{F}_1 - F_1^0H_1\right\|$. Note that $\|D_{1i}\| \le (d_T^{1/2}\bar{\eta}_{1NT}^2)\frac{1}{d_T^{1/2}T}\|x_{1,i}\|\frac{1}{T}\|\hat{u}_i^*\|$. By Lemma S1.2 (iii) in SSP(2016b), and the fact that $\max_i\left\|\lambda_{2i}^0\right\| \le \bar{c}_\lambda$ by Assumption 3.2(i) and that $\max_i\left\|\hat{b}_{2,i}\right\| = o_P(1)$, we can show that $P(\max_i\frac{1}{T}\|\hat{u}_i^*\| \ge cT^{-1/2}(\log T)^3) = o(N^{-1})$ for any $c>0$. By Lemma A.2(i), $\frac{1}{d_T^{1/2}T}\|x_{1,i}\| = O_{a.s.}(1)$. It follows that

$$P\left(\|D_{1i}\| \ge cT^{-1/2}(\log T)^3 d_T^{1/2}\bar{\eta}_{1NT}^2\right) = o\left(N^{-1}\right).$$

For D_{2i}, we have $\|D_{2i}\| \leq d_T^{1/2}\bar{\eta}_{1NT}\|H_1\|\frac{\|x_{1,i}\|}{d_T^{1/2}T}\frac{1}{T^2}\left\|F_1^{0\prime}\hat{u}_i^*\right\|$. By Lemma A.2(i) in HJS, we can show that $P(\max_i \frac{1}{T^2}\left\|F_1^{0\prime}\hat{u}_i^*\right\| > c\psi_{NT}) = o\left(N^{-1}\right)$ for any $c > 0$. It follows that $P(\|D_{2i}\| \geq cd_T^{1/2}\bar{\eta}_{1NT}\psi_{NT}) = o\left(N^{-1}\right)$. Noting $\|D_{3i}\| \leq d_T^{1/2}\bar{\eta}_{1NT}\left\|H_1\right\|\frac{\left\|F_1^0\right\|}{T}\frac{\|x_{1,i}\|}{d_T^{1/2}T}\frac{1}{T}\left\|\hat{u}_i^*\right\|$, we have

$$P\left(\|D_{3i}\| \geq cT^{-1/2}(\log T)^3 d_T^{1/2}\bar{\eta}_{1NT}\right) = o\left(N^{-1}\right).$$

Next, $\|D_{4i}\| \leq d_T^{1/2}\left\|I_{r_1} - (\frac{1}{T^2}H_1'F_1^{0\prime}F_1^0H_1)^{-1}\right\|\left\|H_1\right\|^2\frac{\left\|F_1^0\right\|}{T}\frac{\|x_{1,i}\|}{d_T^{1/2}T}\frac{\left\|F_1^{0\prime}\hat{u}_i^*\right\|}{T^2}$. Using the fact that $\frac{1}{d_T^{1/2}T}\left\|x_{1,i}\right\| = o_{a.s.}(1)$, $\frac{1}{T^2}H_1'F_1^{0\prime}F_1^0H_1 - I_{r_1} = O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})$, and $P(\max_i \frac{1}{T^2}\left\|F_1^{0\prime}\hat{u}_i^*\right\| > c\psi_{NT}) = o(N^{-1})$, we can show that

$$P\left(\|D_{4i}\| \geq cd_T^{1/2}(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})\psi_{NT}\right) = o\left(N^{-1}\right).$$

Noting that $\eta_{1NT} = O_P(d_T^{1/2}T^{-1})$ by Theorem 3.2(i) and $\bar{\eta}_{1NT} = O_P(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})$, we have

$$\begin{aligned} & d_T^{1/2}\bar{\eta}_{1NT}[T^{-1/2}(\log T)^3\bar{\eta}_{1NT} + \psi_{NT} + T^{-1/2}(\log T)^3] + \left(\eta_{1NT} + T^{-1/2}C_{NT}^{-1}\right)\psi_{NT} \\ = \;& O_P\left[d_T^{1/2}(\psi_{NT} + T^{-1/2}(\log T)^3)(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})\right]. \end{aligned}$$

Then we have $P\left(\max_i \|R_{1i}\| > cd_T^{1/2}(\psi_{NT} + T^{-1/2}(\log T)^3)(\eta_{1NT} + T^{-1/2}C_{NT}^{-1})\right) = o(N^{-1})$.

(ii) By the proof of Lemma A.6(ii), we have $\frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 = \frac{-1}{T^2}x_{1,i}'M_{\hat{F}_1}(\hat{F}_1H_1^{-1} - F_1^0)\lambda_{1i}^0 = \frac{-1}{T^2}x_{1,i}'M_{\hat{F}_1}\left(\hat{F}_1H_1^{-1} - (I_1 + ... + I_{11})\,G_1\right)\lambda_{1i}^0 \equiv L_{1i} + ... + L_{11i}$. As in the proof of Lemma A.6(ii), we have

$$R_{2i} = \frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 - \frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_j\hat{b}_{1,j}a_{ji} + \frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}u_ja_{ji} = \bar{L}_{1i} + ... + \bar{L}_{11i},$$

where $\bar{L}_{li} = L_{li}$ for $l = 1, 3, 4, 5, 6, 8, ..., 11$, $\bar{L}_{2i} = \frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_{1,j}\hat{b}_{1,j}\lambda_{2j}^{0\prime}\frac{F_2^{0\prime}\hat{F}_1}{T^2}G_1\lambda_{1i}^0$, $\bar{L}_{7i} = \bar{L}_{7,1i} + \bar{L}_{7,1i}$, $\bar{L}_{7,1i} = \frac{-1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_{2,j}\hat{b}_{2,j}a_{ji}$ and $\bar{L}_{7,2i} = \frac{-1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}u_j\lambda_{2j}^{0\prime}\frac{F_2^{0\prime}\hat{F}_1}{T^2}G_1\lambda_{1i}^0$. It suffices to study $\bar{L}_{li}$ for $l = 1, ..., 11$. For $\bar{L}_{1i}$, we have $\|\bar{L}_{1i}\| \leq d_T^{3/2}\|G_1\|\left\|\lambda_{1i}^0\right\|\frac{1}{T}\left\|\hat{F}_1\right\|\max_{1\leq j\leq N}\frac{\|x_{1,j}\|^3}{d_T^{3/2}T^3}$ $\times\frac{1}{N}\sum_{j=1}^N\left\|\hat{b}_{1,j}\right\|^2$. Noting that $\max_i \frac{1}{d_T^{3/2}T^3}\left\|x_{1,i}\right\|^3 = O_{a.s.}(1)$ by Lemma A.2(i), $\frac{1}{T}\left\|\hat{F}_1\right\| = \sqrt{r_1}$, and $\max_i \|\lambda_{1i}^0\| \leq \bar{c}_\lambda$ by Assumption 3.2(i) and the Bernstein inequality, it is easy to show that $P(\max_i \left\|\bar{L}_{1i}\right\| > cd_T^{3/2}\eta_{1NT}^2) = o(N^{-1})$. Similarly, we can show that

$$\begin{aligned} &P(\max_i \left\|\bar{L}_{2i}\right\| > cd_TT^{-1}\eta_{1NT}) = o(N^{-1}),\ P(\max_i \left\|\bar{L}_{3i}\right\| > cd_T(T^{-1}\eta_{1NT} + T^{-1/2}d_T^{1/2}\eta_{1NT}^2)) = o(N^{-1}), \\ &P(\max_i \left\|\bar{L}_{4i}\right\| > cd_T(T^{-1}\eta_{1NT} + d_T^{1/2}\eta_{1NT}^2)) = o(N^{-1}),\ P(\max_i \left\|\bar{L}_{5i}\right\| > cd_TT^{-1}\eta_{1NT}) = o(N^{-1}), \\ &P(\max_i \left\|\bar{L}_{6i}\right\| > cd_T(T^{-1}\eta_{1NT} + d_T^{-1/2}T^{-2})) = o(N^{-1}), \\ &P(\max_i \left\|\bar{L}_{li}\right\| > cd_TT^{-2}) = o(N^{-1}) \text{ for } l = 7, 8, 9, \end{aligned}$$

and $P(\max_i \|\bar{L}_{li}\| > cd_T^{1/2} N^{-1/2} T^{-1}) = o(N^{-1})$ for $l = 10, 11$. Consequently, we have $P(\max_i \|R_{2i}\| > cd_T^{1/2} \varsigma_{2NT}) = o(N^{-1})$.

(iii) Following the proof of Lemma A.6(iii), we have

$$\|R_{3i}\| \leq \frac{1}{T^{1/2}} \frac{\|x_{1,i}\|}{T} \left\|\lambda_{1i}^0\right\| \left\|P_{F_1^0} - P_{\hat{F}_1}\right\| \left\|(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}\right\| \left\{\frac{1}{N}\sum_{j=1}^N \lambda_{1j}^0\right\}^{1/2} \left\{\frac{1}{NT}\sum_{j=1}^N \|u_j\|^2\right\}^{1/2}.$$

Then $P\left(\max_i \|R_{3i}\| > cd_T^{1/2}\varsigma_{1NT}\right) = o(N^{-1})$.

(iv) Write $R_{4i} = \frac{1}{T^2} x_{1,i}' M_{F_1^0} \hat{u}_i^* - \frac{1}{NT^2}\sum_{j=1}^N a_{ij} x_{1,i}' M_{F_1^0}\hat{u}_j^* = \frac{1}{T^2} x_{1,i}' M_{F_1^0}\hat{u}_i^* - \frac{1}{NT^2}\sum_{j=1}^N a_{ij} x_{1,i}'\hat{u}_j^* + \frac{1}{NT^2} \times \sum_{j=1}^N a_{ij} x_{1,i}' P_{F_1^0}\hat{u}_j^*$. By Lemma A.7(ii), $P\left(\max_i \left\|\frac{1}{T^2} x_{1,i}' M_{F_1^0}\hat{u}_i^*\right\| > cd_T\psi_{NT}\right) = o(N^{-1})$. For $\frac{1}{NT^2}\sum_{j=1}^N a_{ij} x_{1,i}'\hat{u}_j^*$, we have

$$\left\|\frac{1}{NT^2}\sum_{j=1}^N a_{ij} x_{1,i}'\hat{u}_j^*\right\| \leq \left\|\lambda_{1i}^0\right\| \left\|(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}\right\| \frac{1}{N}\sum_{j=1}^N \left\|\lambda_{1j}^0\right\| \max_{i,j} \frac{1}{T^2}\left\|x_{1,i}'\hat{u}_j^*\right\|.$$

Following the analysis in (i), we can show that $P(\max_{i,j} \frac{1}{T^2}||x_{1,i}'\hat{u}_j^*|| > c\psi_{NT}) = o(N^{-1})$. So

$$P\left(\max_i \left\|\frac{1}{NT^2}\sum_{j=1}^N a_{ij}\ x_{1,i}'\hat{u}_j^*\right\| > c\psi_{NT}\right) = o(N^{-1}).$$

Similarly, we can show that $P(\max_i ||\frac{1}{NT^2}\sum_{j=1}^N a_{ij} x_{1,i}' P_{F_1^0}\hat{u}_j^*|| > c\psi_{NT}) = o(N^{-1})$. Consequently we have $P\left(\|R_{4i}\| > cd_T\psi_{NT}\right) = o(N^{-1})$.

(v) By the proof of Theorem 3.2(i), we have $(\hat{Q}_1 - \hat{Q}_2)\hat{\boldsymbol{b}}_1 = R$. Let $S_i = (0_{p_1\times p_1}, \ldots, 0_{p_1\times p_1}, I_{p_1}, 0_{p_1\times p_1}, \ldots, 0_{p_1\times p_1})$ be a $p_1 \times Np_1$ selection matrix such that $S_i\hat{\boldsymbol{b}}_1 = \hat{b}_{1,i}$. It follows that

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

where the second equality follows from the fact that $\operatorname{tr}(A_1A_2A_3A_4) = \operatorname{vec}(A_1)'(A_2 \otimes A_3')\operatorname{vec}(A_4')$ for conformable matrices A_1, A_2, A_3, and A_4. By Assumption 3.2(v), we have that $P(\mu_{\min}(\hat{Q}_1 - \hat{Q}_2) \geq c\rho_{\min}) = 1 - o(N^{-1})$. By the proof of Theorem 3.2, we have $R_i = R_{1i} + R_{2i} - R_{3i} + R_{4i} - R_{5i}$. By Lemma A.6(ii) and Lemma A.7(i)-(ii), we directly obtain that $P\left(\max_i \|R_{1i} + R_{2i} - R_{3i} + R_{4i}\| > c\psi_{NT}\right) =$

$o(N^{-1})$. For $\|R_{5i}\|$, we have

$$\|R_{5i}\| \leq \frac{\lambda}{2}\sum_{j=1}^{K}\prod_{l=1,l\neq j}^{K}\left\|\hat{\beta}_{1,i}-\hat{\alpha}_l\right\| \leq \frac{\lambda}{2}\sum_{j=1}^{K}\left|\prod_{l=1,l\neq j}^{K}\left\|\hat{\beta}_{1,i}-\hat{\alpha}_l\right\| - \prod_{l=1,l\neq j}^{K}\|\beta_{1,i}^0-\hat{\alpha}_l\|\right| + \frac{\lambda}{2}\sum_{j=1}^{K}\prod_{l=1,l\neq j}^{K}\|\beta_{1,i}^0-\hat{\alpha}_l\|$$

$$\leq C_{KNT}(\alpha)K\lambda\left\|\hat{b}_{1,i}\right\| + \frac{\lambda}{2}C_{KNT}(\alpha)K + \frac{\lambda}{2}\sum_{j=1}^{K}\prod_{l=1,l\neq j}^{K}\left\|\beta_{1,i}^0-\hat{\alpha}_l\right\|,$$

where we use the fact $\left|\prod_{k=1}^K\|\hat{\beta}_{1,i}-\alpha_k\| - \prod_{k=1}^K\left\|\beta_{1,i}^0-\alpha_k\right\|\right| \leq C_{KNT}(\alpha)(1+2\|\hat{b}_{1,i}\|)$ in the proof of Theorem 3.2(i). Noting that $\frac{\lambda}{2}C_{KNT}(\alpha)K + \frac{\lambda}{2}\sum_{j=1}^K\prod_{l=1,l\neq j}^K\left\|\beta_{1,i}^0-\hat{\alpha}_l\right\| = O(\lambda)$, it follows that for sufficiently large N,

$$\|\hat{b}_{1,i}\|^2 \leq \left[\mu_{\min}\left(\hat{Q}_1-\hat{Q}_2\right)\right]^{-2}\|R_i\|^2$$
$$\leq \left[\mu_{\min}\left(\hat{Q}_1-\hat{Q}_2\right)\right]^{-2}\left(2\left\|R_{1i}+R_{2i}-R_{3i}+R_{4i}\right\|^2 + 4c\lambda^2 + 4C_{KNT}(\alpha)^2K^2\lambda^2\left\|\hat{b}_{1,i}\right\|^2\right).$$

That is, $\|\hat{b}_{1,i}\|^2 \leq \frac{\left[\mu_{\min}(\hat{Q}_1-\hat{Q}_2)\right]^{-2}(2\|R_i+R_{5i}\|^2+4\lambda^2)}{(1-4C_{KNT}(\alpha)^2K^2\lambda^2)}$. Combining the above results, we have

$$P\left(\max_i\|\hat{b}_{1,i}\|^2 \geq c\left(\psi_{NT}^2+\lambda^2(\log T)^{\epsilon}\right)\right)$$
$$\leq P\left(\max_i\|\hat{b}_{1,i}\|^2 \geq c\left(\psi_{NT}^2+\lambda^2(\log T)^{\epsilon}\right), \mu_{\min}\left(\hat{Q}_1-\hat{Q}_2\right) \leq c\rho_{\min}\right)$$
$$+P\left(\mu_{\min}\left(\hat{Q}_1-\hat{Q}_2\right) > c\rho_{\min}\right)$$
$$\leq P\left(2\left\|R_{1i}+R_{2i}-R_{3i}+R_{4i}\right\|^2 > c\psi_{NT}^2\rho_{\min}^2\right) + P\left(C\lambda^2 \geq c\lambda^2(\log T)^{\epsilon}\rho_{\min}^2\right) + o(N^{-1})$$
$$= o(N^{-1}) + 0 + o(N^{-1}) = o(N^{-1}).$$

(vi) The proof closely follows that of (i)-(v) and thus omitted.

(vii) By the definition of R_{2i}, $\frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 = R_{2i} + \frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_{1,j}\hat{b}_{1,j}a_{ji} - \frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}u_ja_{ji}$. We have studied R_{2i} in (ii) and it remains to analyze the last two terms. Noting that

$$\frac{1}{NT^2}\left\|\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_{1,j}\hat{b}_{1,j}a_{ji}\right\| \leq d_T\frac{\max_i\|x_{1,i}\|^2}{d_TT^2}\left\|\lambda_{1i}^0\right\|\left\{\frac{1}{N}\sum_{j=1}^N\left\|\hat{b}_{1,j}\right\|^2\right\}^{1/2},$$

we can show that $P\left(\max_i\frac{1}{NT^2}\left\|\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}x_{1,j}\hat{b}_{1,j}a_{ji}\right\| \geq d_T\eta_{1NT}\right) = o\left(N^{-1}\right)$. Noting that

$$\frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{\hat{F}_1}u_ja_{ji} = \frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'M_{F_1^0}u_ja_{ji} + \frac{1}{NT^2}\sum_{j=1}^N x_{1,i}'(M_{\hat{F}_1}-M_{F_1^0})u_ja_{ji}.$$

By (B.1), $\frac{-1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 = \frac{1}{T^2}x_{1,i}'(P_{\hat{F}_1}-P_{F_1^0H_1})F_1^0\lambda_{1i}^0 = \frac{1}{T^2}x_{1,i}'(\hat{p}_1+\hat{p}_2+\hat{p}_3+\hat{p}_4)F_1^0\lambda_{1i}^0 \equiv E_{1i} +$

$E_{2i}+E_{3i}+E_{4i}$. Note that

$$\begin{aligned}
\|E_{1i}\| &\leq \bar{\eta}_{1NT}^2 \frac{\|x_{1,i}\|}{T}\frac{\left\|F_1^0\right\|}{T}\left\|\lambda_{1i}^0\right\|, \\
\|E_{2i}\| &\leq \bar{\eta}_{1NT} \frac{\|x_{1,i}\|}{T}\frac{\left\|H_1'F_1^{0\prime}F_1^0\right\|}{T^2}\left\|\lambda_{1i}^0\right\|, \\
\|E_{3i}\| &\leq \frac{1}{T^2}\left\|(\hat{F}_1-F_1^0H_1)'F_1^0\right\|\frac{\|x_{1,i}\|}{T}\frac{\left\|F_1^0H_1\right\|}{T}\left\|\lambda_{1i}^0\right\|, \text{ and} \\
\|E_{4i}\| &\leq \left\|I_{r_1}-\left(\frac{1}{T^2}H_1'F_1^{0\prime}F_1^0H_1\right)^{-1}\right\|\frac{\|x_{1,i}\|}{T}\frac{\left\|F_1^0H_1\right\|}{T}\frac{\left\|H_1'F_1^{0\prime}F_1^0\right\|}{T^2}\left\|\lambda_{1i}^0\right\|,
\end{aligned}$$

where $\bar{\eta}_{1NT}=\frac{1}{T}\left\|\hat{F}_1-F_1^0H_1\right\|=O_P(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})$ by Theorem 3.1(iii), $\frac{1}{T^2}\left\|(\hat{F}_1-F_1^0H_1)'F_1^0\right\|$ $=O_P(\eta_{1NT}+T^{-1}\delta_{NT}^{-1}+T^{-5/4}C_{NT}^{-1/2})$ by Lemma A.5(iii), and $\left\|I_{r_1}-(\frac{1}{T^2}H_1'F_1^{0\prime}F_1^0H_1)^{-1}\right\|=O_P(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})$ by the proof of Lemma A.5(i). Using the fact that $\max_i \frac{1}{d_T T^2}\|x_i\|^2=O_{a.s.}(1)$ and $\max_i\|\lambda_{1i}^0\|\leq\bar{c}_\lambda$, we can use the uniform bound for each of the above four terms to obtain

$$P\left(\max_i \frac{1}{T^2}\left\|x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right\|\geq d_T^{1/2}(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})\right)=o\left(N^{-1}\right).$$

Then (vii) follows.

(viii) By the proof of Theorem 3.2, we have $\hat{b}_{2,i}=\left(\frac{1}{T}x_{2,i}'M_{\hat{F}_1}x_{2,i}\right)^{-1}\frac{1}{T}x_{2,i}'M_{\hat{F}_1}\left(u_i+F_2^0\lambda_{2i}^0\right)$ $+\frac{1}{T}x_{2,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0+\frac{1}{T}x_{2,i}'M_{\hat{F}_1}x_{1,i}\hat{b}_{1,i}$. Note that $\frac{1}{T}x_{2,i}'M_{\hat{F}_1}\left(u_i+F_2^0\lambda_{2i}^0\right)=\frac{1}{T}x_{2,i}'M_{F_1^0}\left(u_i+F_2^0\lambda_{2i}^0\right)+\frac{1}{T}x_{2,i}'(P_{F_1^0}-P_{\hat{F}_1})\left(u_i+F_2^0\lambda_{2i}^0\right)$ By Lemma S1.2 (iii) in SSP(2016b), we can show that

$$\begin{aligned}
P(\max_i \frac{1}{T}\|x_{2,i}'M_{F_1^0}\left(u_i+F_2^0\lambda_{2i}^0\right)\| &\geq cT^{-1/2}(\log T)^3)=o(N^{-1}) \\
P(\max_i \frac{1}{T}\|x_{2,i}'(P_{F_1^0}-P_{\hat{F}_1})\left(u_i+F_2^0\lambda_{2i}^0\right)\| &\geq cd_T^{1/2}T^{-1/2}(\log T)^3)=o(N^{-1})
\end{aligned}$$

for any $c>0$. The proof follows closely that of (vii). We have

$$P\left(\max_i \frac{1}{T}\|x_{2,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\|\geq cT^{1/2}(\log T)^3\left(\eta_{1NT}+T^{-1/2}C_{NT}^{-1}\right)\right)=o(N^{-1})$$

Note that $\frac{1}{T}x_{2,i}'M_{\hat{F}_1}x_{1,i}\hat{b}_{1,i}=\frac{1}{T}x_{2,i}'M_{F_1^0}x_{1,i}\hat{b}_{1,i}+\frac{1}{T}x_{2,i}'(P_{F_1^0}-P_{\hat{F}_1})x_{1,i}\hat{b}_{1,i}$,

$$\begin{aligned}
\frac{1}{T}\left\|x_{2,i}'M_{F_1^0}x_{1,i}\hat{b}_{1,i}\right\| &\leq \frac{\left\|x_{2,i}'F_1^0\right\|}{T}\frac{\left\|F_1^0\right\|}{T}\frac{\|x_{1,i}\|}{T}\left\|\hat{b}_{1,i}\right\|, \text{ and} \\
\frac{1}{T}\left\|x_{2,i}'(P_{F_1^0}-P_{\hat{F}_1})x_{1,i}\hat{b}_{1,i}\right\| &\leq \sqrt{T}\frac{\|x_{2,i}\|}{\sqrt{T}}\frac{\|x_{1,i}\|}{T}\left\|P_{F_1^0}-P_{\hat{F}_1}\right\|\left\|\hat{b}_{1,i}\right\|,
\end{aligned}$$

where $P\left(\max_i \frac{1}{T}\|x_{2,i}'F_1^0\|\geq cT\psi_{NT}\right)=o(N^{-1})$ and $P\left(\max_i ||\hat{b}_{1,i}||\geq c\left(\psi_{NT}+\lambda(\log T)^{\epsilon/2}\right)\right)=o(N^{-1})$. It follows that $P\left(\max_i \frac{1}{T}\|x_{2,i}'M_{F_1^0}x_{1,i}\hat{b}_{1,i}\|\geq cd_T^{1/2}T\psi_{NT}c\left(\psi_{NT}+\lambda(\log T)^{\epsilon/2}\right)\right)=o(N^{-1})$ and $P\{\max_i$

$\frac{1}{T}||x_{2,i}'(P_{F_1^0}-P_{\hat{F}_1})x_{1,i}\hat{b}_{1,i}|| \geq c d_T^{1/2}T^{1/2}(\log T)^3(\eta_{1NT}+T^{-1/2}C_{NT}^{-1})\left(\psi_{NT}+\lambda(\log T)^{\epsilon/2}\right)\} = o(N^{-1})$. Then (viii) follows. ■

Proof of Lemma A.9. (i) By Lemma A.6(ii), we have

$$\frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 = \frac{-1}{T^2}x_{1,i}'M_{\hat{F}_1}(\hat{F}_1H_1^{-1}-F_1^0)\lambda_{1i}^0 = \frac{-1}{T^2}x_{1,i}'M_{\hat{F}_1}\left[I_1+...+I_{11}\right]G_1\lambda_{1i}^0 = L_{1i}+...+L_{11i}.$$

Let $\bar{L}_{li}, l=1,...,11$, be as defined in the proof of Lemma A.6(ii). Then, following the proof of Lemma A.6(ii), we can readily show that $\frac{1}{N_k}\sum_{i\in\hat{G}_k}(\bar{L}_{1i}+\bar{L}_{2i}+...+\bar{L}_{7,2i})+\frac{1}{N_k}\sum_{i\in\hat{G}_k}(\bar{L}_{8,i}+\bar{L}_{9,i}) = O_P(d_T\eta_{1NT}^2+T^{-1}d_T^{1/2}\eta_{1NT}+T^{-2}) = o_P\left(N^{-1/2}T^{-1}\right)$. For $\bar{L}_{7,1i}$, we have

$$\begin{aligned}\frac{1}{N_k}\sum_{i\in\hat{G}_k}\bar{L}_{7,1i} =& \frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{NT^2}\sum_{j=1}^{N}x_{1,i}'M_{F_1^0}x_{2,j}\hat{b}_{2,j}a_{ij}+\frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{NT^2}\sum_{j=1}^{N}x_{1,i}'\left(P_{F_1^0}-P_{\hat{F}_1}\right)x_{2,j}\hat{b}_{2,j}a_{ij}\\ \equiv& I+II.\end{aligned}$$

Note that $\|I\| \leq \frac{1}{\sqrt{NT}}\max_{1\leq i,j\leq N}\|a_{ij}\|\left\{\frac{1}{N}\sum_{j=1}^{N}\left\|\hat{b}_{2,j}\right\|^2\right\}^{1/2}\left\{\frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{NT^2}\sum_{j=1}^{N}\left\|x_{1,i}'M_{F_1^0}x_{2,j}\right\|^2\right\}^{1/2}$ $= o_P\left(N^{-1/2}T^{-1}\right)$ by Lemma A.8(viii) and the fact that $\max_{1\leq i,j\leq N}\|a_{ij}\| = O_P(1)$ and that $\frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{NT^2}\sum_{j=1}^{N}\left\|x_{1,i}'M_{F_1^0}x_{2,j}\right\|^2 = O_P(1)$. For II, we can decompose $P_{F_1^0}-P_{\hat{F}_1}$ as in (B.1) and use similar arguments as used in the proof of Lemma A.9(iii) below to show that $\|II\| = o_P\left(N^{-1/2}T^{-1}\right)$. Then $\frac{1}{N_k}\sum_{i\in\hat{G}_k}\bar{L}_{7,1i} = o_P\left(N^{-1/2}T^{-1}\right)$. It follows that

$$\begin{aligned}\frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 =& \frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{NT^2}\sum_{j=1}^{N}x_{1,i}'M_{\hat{F}_1}x_{1,j}\hat{b}_{1,j}a_{ij}-\frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{NT^2}\sum_{j=1}^{N}x_{1,i}'M_{\hat{F}_1}u_ja_{ij}\\ &+\frac{1}{N_k}\sum_{i\in\hat{G}_k}L_{10i}+\frac{1}{N_k}\sum_{i\in\hat{G}_k}L_{11i}+o_P\left(N^{-1/2}T^{-1}\right).\end{aligned}$$

Next, we can show that $\frac{1}{N_k}\sum_{i\in\hat{G}_k}L_{10i} = \frac{-1}{N_kT^2}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}(F_1^0-\hat{F}_1H_1^{-1})\frac{\Lambda_1^{0\prime}\Lambda_2^0}{N}\frac{F_2^{0\prime}\hat{F}_1}{T^2}G_1\lambda_{1i}^0 = o_P\left(N^{-1/2}T^{-1}\right)$ by using the fact that $\frac{1}{T}\left\|F_1^0-\hat{F}_1H_1^{-1}\right\| = O_P(d_T^{1/2}T^{-1}+T^{-1/2}C_{NT}^{-1})$, $\frac{1}{N}\Lambda_1^{0\prime}\Lambda_2^0 = O_P\left(N^{-1/2}\right)$, and $\frac{1}{T^2}F_2^{0\prime}\hat{F}_1 = \frac{1}{T^2}F_2^{0\prime}F_1^0H_1+\frac{1}{T^2}F_2^{0\prime}(\hat{F}_1-F_1^0H_1) = O_P(T^{-1}+T^{-1/2}\eta_{1NT}+T^{-1}C_{NT}^{-1})$. In addition, $\frac{1}{N_k}\sum_{i\in\hat{G}_k}L_{11i} = \frac{-1}{N_kT^2}\sum_{i\in\hat{G}_k}x_{1,i}'M_{\hat{F}_1}F_2^0\frac{\Lambda_2^{0\prime}\Lambda_1^0}{N}\frac{F_1^{0\prime}\hat{F}_1}{T^2}G_1\lambda_{1i}^0 = -\frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{NT^2}\sum_{j=1}^{N}x_{1,i}'M_{\hat{F}_1}$ $\times F_2^0\lambda_{2j}^0a_{ij}$. It follows that

$$\begin{aligned}\frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{T^2}x_{1,i}'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 =& \frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{NT^2}\sum_{j=1}^{N}x_{1,i}'M_{\hat{F}_1}x_{1,j}'\hat{b}_{1,j}a_{ij}-\frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{NT^2}\sum_{j=1}^{N}x_{1,i}'M_{\hat{F}_1}u_ja_{ij}\\ &-\frac{1}{N_k}\sum_{i\in\hat{G}_k}\frac{1}{NT^2}\sum_{j=1}^{N}x_{1,i}'M_{\hat{F}_1}F_2^0\lambda_{2j}^0a_{ij}+o_P(N^{-1/2}T^{-1}).\end{aligned}$$

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

$$\begin{aligned}\frac{1}{N_k T^2} \sum_{i \in \hat{G}_k} x_{1,i}' M_{\hat{F}_1} x_{1,i} =& \frac{1}{N_k T^2} \sum_{i \in G_k^0} x_{1,i}' M_{\hat{F}_1} x_{1,i} + \frac{1}{N_k T^2} \sum_{i \in \hat{G}_k \setminus G_k^0} x_{1,i}' M_{\hat{F}_1} x_{1,i} - \frac{1}{N_k T^2} \sum_{i \in G_k^0 \setminus \hat{G}_k} x_{1,i}' M_{\hat{F}_1} x_{1,i} \\ =& \frac{1}{N_k T^2} \sum_{i \in G_k^0} x_{1,i}' M_{\hat{F}_1} x_{1,i} + \mathcal{C}_{1NK} - \mathcal{C}_{2NK}, \text{ say.}\end{aligned}$$

By Theorem 3.3, we have $P(\|\mathcal{C}_{1NK}\| \geq c(N^{-1/2}T^{-1})) \leq P(\hat{F}_{k,NT}) \to 0$ and $P(\| \mathcal{C}_{2NK}\| \geq c(N^{-1/2}T^{-1})) \leq P(\hat{E}_{k,NT}) \to 0$. It follows that

$$\frac{1}{N_k T^2} \sum_{i \in \hat{G}_k} x_{1,i}' M_{\hat{F}_1} x_{1,i} = \frac{1}{N_k T^2} \sum_{i \in G_k^0} x_{1,i}' M_{\hat{F}_1} x_{1,i} + o_P(N^{-1/2}T^{-1}).$$

Next, $\frac{1}{N_k T^2} \sum_{i \in G_k^0} x_{1,i}' M_{\hat{F}_1} x_{1,i} = \frac{1}{N_k T^2} \sum_{i \in G_k^0} x_{1,i}' M_{F_1^0} x_{1,i} + \frac{1}{N_k T^2} \sum_{i \in G_k^0} x_{1,i}' (P_{F_1^0} - P_{\hat{F}_1}) x_{1,i}$, where

$$\left\| \frac{1}{N_k T^2} \sum_{i \in G_k^0} x_{1,i}' (P_{F_1^0} - P_{\hat{F}_1}) x_{1,i} \right\| \leq \left\| P_{F_1^0} - P_{\hat{F}_1} \right\| \frac{1}{N_k} \sum_{i \in G_k^0} \frac{\|x_{1,i}\|^2}{T^2} = O_P(\eta_{1NT} + T^{-1/2} C_{NT}^{-1}).$$

by Lemma A.5(i). Thus $\frac{1}{N_k T^2} \sum_{i \in \hat{G}_k} x_{1,i}' M_{\hat{F}_1} x_{1,i} = \frac{1}{N_k T^2} \sum_{i \in G_k^0} x_{1,i}' M_{F_1^0} x_{1,i} + o_P(1)$.

(iii) Using the same arguments as those in the proof of (ii), we can readily show that

$$\begin{aligned}&\frac{1}{\sqrt{N_k} T} \sum_{i \in \hat{G}_k} x_{1,i}' M_{\hat{F}_1} \left(\left(u_i + F_2^0 \lambda_{2i}^0\right) - \frac{1}{N} \sum_{j=1}^N \left(u_j + F_2^0 \lambda_{2j}^0\right) a_{ij} \right) \\ =& U_{k,NT} + \frac{1}{\sqrt{N_k} T} \sum_{i \in G_k^0} x_{1,i}' \left(M_{\hat{F}_1} - M_{F_1^0} \right) \left(\left(u_i + F_2^0 \lambda_{2i}^0\right) - \frac{1}{N} \sum_{j=1}^N \left(u_j + F_2^0 \lambda_{2j}^0\right) a_{ij} \right) + o_P(1).\end{aligned}$$

We first consider $\frac{1}{\sqrt{N_k} T} \sum_{i \in G_k^0} x_{1,i}' (M_{\hat{F}_1} - M_{F_1^0}) \left(u_i + F_2^0 \lambda_{2i}^0\right)$. By (B.1), $\frac{1}{\sqrt{N_k} T} \sum_{i \in G_k^0} x_{1,i}' (M_{\hat{F}_1} - M_{F_1^0}) \left(u_i + F_2^0 \lambda_{2i}^0\right) = \sum_{l=1}^4 \frac{1}{\sqrt{N_k} T} \sum_{i \in G_k^0} x_{1,i}' \hat{p}_l \left(u_i + F_2^0 \lambda_{2i}^0\right) \equiv \sum_{l=1}^4 D_l$. Noting that $\frac{1}{T} \left\| \hat{F}_1 - F_1^0 H_1 \right\| = O_P(\eta_{1NT} + T^{-1/2} C_{NT}^{-1}) = O_P(d_T^{1/2} T^{-1} + (NT)^{-1/2})$, we can show that

$$\begin{aligned}\|D_1\| &\leq \left\{ \frac{1}{T} \left\| \hat{F}_1 - F_1^0 H_1 \right\|^2 \right\} \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left\| \frac{1}{\sqrt{N_k} T} \sum_{i \in G_k^0} x_{1,is} \left(u_{it} + \lambda_{2i}^{0\prime} F_{2t}^0\right) \right\|^2 \right)^{1/2} \\ &= O_P(d_T T^{-1} + N^{-1}) O_P(1) = o_P(1),\end{aligned}$$

and

$$
\begin{aligned}
\|D_2\| &= \frac{1}{\sqrt{N_k}T}\left\|\sum_{i\in G_k^0} x'_{1,i}\frac{1}{T^2}(\hat{F}_1 - F_1^0H_1)H_1'F_1^{0\prime}\left(u_i + F_2^0\lambda_{2i}^0\right)\right\| \\
&\leq \frac{\sqrt{N_k}}{T}\left\|\hat{F}_1 - F_1^0H_1\right\|\|H_1\|\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\|x_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\left\|F_1^{0\prime}\left(u_i + F_2^0\lambda_{2i}^0\right)\right\|^2\right\}^{1/2} \\
&= \sqrt{N_k}O_P(d_T^{1/2}T^{-1} + (NT)^{-1/2})O_P(1) = o_P(1).
\end{aligned}
$$

Similarly, we have $\|D_4\| = o_P(1)$. For D_3, we can apply analogous arguments as used in the proof of Lemma 5(v) to show that

$$
\begin{aligned}
\|D_2\| &= \frac{1}{\sqrt{N_k}T}\left\|\sum_{i\in G_k^0} x'_{1,i}\frac{1}{T^2}F_1^0H_1(\hat{F}_1 - F_1^0H_1)'\left(u_i + F_2^0\lambda_{2i}^0\right)\right\| \\
&\leq \frac{\sqrt{N_k}}{T}\|F_1^0H_1\|\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\|x_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\left\|(\hat{F}_1 - F_1^0H_1)'\left(u_i + F_2^0\lambda_{2i}^0\right)\right\|^2\right\}^{1/2} \\
&= \sqrt{N_k}O_P(1)\,O_P(Td_T\eta_{1NT}^2 + \delta_{NT}^{-2}) = o_P(1).
\end{aligned}
$$

It follows that $\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0} x'_{1,i}(M_{\hat{F}_1} - M_{F_1^0})\left(u_i + F_2^0\lambda_{2i}^0\right) = o_P(1)$. Analogously, we can show that $\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0} x'_{1,i}(M_{\hat{F}_1} - M_{F_1^0})\frac{1}{N}\sum_{j=1}^N\left(u_j + F_2^0\lambda_{2j}^0\right)a_{ij} = o_P(1)$. Then (iii) follows.

(iv) As in (ii), we can readily show that

$$
\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j\in\hat{G}_{k'}} x'_{1,i}M_{\hat{F}_1}x_{1,j}a_{ij} = \frac{1}{N_kT^2}\sum_{i\in G_k^0}\frac{1}{N}\sum_{j\in G_{k'}^0} x'_{1,i}M_{\hat{F}_1}x_{1,j}a_{ij} + o_P(N_k^{-1/2}T^{-1}).
$$

Using (B.1), we obtain the following decomposition

$$
\frac{-1}{N_kT^2}\sum_{i\in G_k^0}\frac{1}{N}\sum_{j\in G_{k'}^0} x'_{1,i}\left(M_{\hat{F}_1} - M_{F_1^0}\right)x_{1,j}a_{ij} = \sum_{l=1}^4\frac{1}{N_kT^2}\sum_{i\in G_k^0}\frac{1}{N}\sum_{j\in G_{k'}^0} x'_{1,i}\hat{p}_lx_{1,j}a_{ij} \equiv \sum_{l=1}^4\tilde{D}_l,\ \text{say.}
$$

Using arguments analogous to those in the proof of part (iii) we can readily show that $\tilde{D}_l = o_P(1)$ for $l = 1, 2, 3, 4$. Then $\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j\in\hat{G}_{k'}} x'_{1,i}M_{\hat{F}_1}x_{1,j}a_{ij} = \frac{1}{N_kT^2}\sum_{i\in G_k^0}\frac{1}{N}\sum_{j\in G_{k'}^0} x'_{1,i}M_{F_1^0}x_{1,j}a_{ij} + o_P(1)$.

(v) As in (ii), we can readily show that $\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x'_{1,i}M_{\hat{F}_1}x_{2,i}\hat{b}_{2,i} = \frac{1}{N_kT^2}\sum_{i\in G_k^0} x'_{1,i}M_{\hat{F}_1}x_{2,i}\hat{b}_{2,i} + o_P(N_k^{-1/2}T^{-1})$ Note that $\frac{1}{N_kT^2}\sum_{i\in G_k^0} x'_{1,i}M_{\hat{F}_1}x_{2,i}\hat{b}_{2,i} = \frac{1}{N_kT^2}\sum_{i\in G_k^0} x'_{1,i}M_{F_1^0}x_{2,i}\hat{b}_{2,i} + \frac{1}{N_kT^2}\sum_{i\in G_k^0} x'_{1,i}$ $(M_{\hat{F}_1} - M_{F_1^0})x_{2,i}\hat{b}_{2,i}$. The proof is close to (i) and (iii) and thus omitted here.■

Proof of Lemma A.10 (i) Note that $Q_{NT} = Q_{1NT} - Q_{2NT}$, where

$$
\begin{aligned}
Q_{1NT} &= \operatorname{diag}\left(\frac{1}{N_1T^2}\sum_{i\in G_1^0} x_{1,i}' M_{F_1^0} x_{1,i}, \ldots, \frac{1}{N_KT^2}\sum_{i\in G_K^0} x_{1,i}' M_{F_1^0} x_{1,i}\right), \\
Q_{2NT} &= \begin{pmatrix} Q_{2NT,11} & \cdots & Q_{2NT,1K} \\ \vdots & \ddots & \vdots \\ Q_{2NT,K1} & \cdots & Q_{2NT,KK} \end{pmatrix}, \text{ and } Q_{2NT,kl} = \frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j\in G_l^0} a_{ij} x_{1,i}' M_{F_1^0} x_{1,j}.
\end{aligned}
$$

It is sufficient to prove (i) by showing that $Q_{1NT} \xrightarrow{d} Q_1$ and $Q_{2NT} \xrightarrow{d} Q_2$ as $(N,T) \to \infty$ by the Cramér-Wold device, where

$$
\begin{aligned}
Q_1 =& \operatorname{diag}\left(\lim_{N\to\infty}\frac{1}{N_1}\sum_{i\in G_1^0}\mathbb{E}_{\mathcal{C}}\left(\int \tilde{B}_{2i}\tilde{B}_{2i}'\right), \ldots, \lim_{N\to\infty}\frac{1}{N_K}\sum_{i\in G_K^0}\mathbb{E}_{\mathcal{C}}\left(\int \tilde{B}_{2i}\tilde{B}_{2i}'\right)\right), \\
Q_2 =& \begin{pmatrix} Q_{2,11} & \cdots & Q_{2,1K} \\ \vdots & \ddots & \vdots \\ Q_{2,K1} & \cdots & Q_{2,KK} \end{pmatrix},
\end{aligned}
$$

$Q_{2,kl} = \lim_{N\to\infty}\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j\in G_l^0} a_{ij}\mathbb{E}_{\mathcal{C}}\left(\int \tilde{B}_{2i}\tilde{B}_{2j}\right)$, $\mathcal{C} = \sigma(F^0, \Lambda^0)$, and $\mathbb{E}_{\mathcal{C}}(\cdot)$ denotes expectation conditional on $\mathcal{C}$.

We first show $Q_{1NT} \xrightarrow{d} Q_1$ as $(N,T) \to \infty$. The kth block diagonal element of Q_{1NT} is given by

$$
\begin{aligned}
&\frac{1}{N_kT^2}\sum_{i\in G_k^0} x_{1,i}' M_{F_1^0} x_{1,i} \\
&= \frac{1}{N_k}\sum_{i\in G_k^0}\frac{1}{T^2}\sum_{t=1}^T x_{1,it}x_{1,it}' - \frac{1}{N_k}\sum_{i\in G_k^0}\left(\frac{1}{T^2}\sum_{t=1}^T x_{1,it}f_{1t}^{0\prime}\right)\left(\frac{1}{T^2}\sum_{t=1}^T f_{1t}^0 f_{1t}^{0\prime}\right)^{-1}\left(\frac{1}{T^2}\sum_{t=1}^T f_{1t}^0 x_{1,it}'\right) \\
&\equiv Q_{1kNT,1} - Q_{2kNT,2}, \text{ say.}
\end{aligned}
$$

We first establish the sequential limit. Let $(N,T)_{\text{seq}} \to \infty$ denote the sequential limit by passing $T \to \infty$ first and $N \to \infty$ later. Let $\tilde{x}_{1,i} = M_{F_1^0} x_{1,i}$. Denote the tth column of $\tilde{x}_{1,i}$ as $\tilde{x}_{1,it}$. Then as $T \to \infty$,

$$
\frac{1}{T^{1/2}}\tilde{x}_{1,it} = \frac{1}{T^{1/2}}\tilde{x}_{1,it} - \tilde{x}_{1,i}' F_1^0\left(F_1^{0\prime}F_1^0\right)^{-1}\frac{1}{T^{1/2}}f_{1t}^0 \Rightarrow B_{2i} - \int B_{2i}B_3'\left(\int B_3B_3'\right)^{-1} B_3 \equiv \tilde{B}_{2i},
$$

and by the continuous mapping theorem (CMT) $\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{1,i} = \frac{1}{T^2}x_{1,i}'x_{1,i} = \frac{1}{T}\sum_{t=1}^T\frac{1}{T^{1/2}}\tilde{x}_{1,it}\frac{1}{T^{1/2}}\tilde{x}_{1,it}'$ $\Rightarrow \int \tilde{B}_{2i}\tilde{B}_{2i}$. By the conditional law of large numbers with independent observations (conditional on $\mathcal{C}$), we have that as $N \to \infty$

$$
\frac{1}{N_k}\sum_{i\in G_k^0}\int \tilde{B}_{2i}\tilde{B}_{2i} \xrightarrow{p} \lim_{N\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0}\mathbb{E}_{\mathcal{C}}\left[\int \tilde{B}_{2i}\tilde{B}_{2i}\right].
$$

It follows that $\frac{1}{T^2} x_{1,i}' M_{F_1^0} x_{1,i} \xrightarrow{d} \lim_{N\to\infty} \frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E}_{\mathcal{C}} \left[\int \tilde{B}_{2i} \tilde{B}_{2i}\right]$ as $(N,T)_{\text{seq}} \to \infty$. To show the above limit is also the joint distributional limit, we need to verify condition (3.9) in Phillips and Moon (1999, hereafter PM). We do so by verifying the conditions in Theorem 1 of PM (1999) to obtain that as $(N,T) \to \infty$,

$$Q_{1kNT,1} \xrightarrow{d} \lim_{N\to\infty} \frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E}_{\mathcal{C}} (\int B_{2i} B_{2i}') \text{ and}$$

$$Q_{1kNT,2} \xrightarrow{d} \lim_{N\to\infty} \frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E}_{\mathcal{C}} \left(\int B_{2i} B_3' \left(\int B_3 B_3' \right)^{-1} \int B_3 B_{2i}' \right).$$

This implies that $\frac{1}{N_k T^2} \sum_{i\in G_k^0} x_{1,i}' M_{F_1^0} x_{1,i} \xrightarrow{d} \lim_{N\to\infty} \frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E}_{\mathcal{C}} \left(\int \tilde{B}_{2i} \tilde{B}_{2i}'\right)$ as $(N,T) \to \infty$. We focus on the study of $Q_{1kNT,1}$ as $Q_{1kNT,2}$ can be analogously studied.

It is easy to see that $\lim_{N\to\infty} \frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E}_{\mathcal{C}} (\int B_{2i} B_{2i}')$ is the sequential limit of $Q_{1kNT,1}$. We are left to verify the four conditions in Theorem 1 of PM (1999) that ensure their equation (3.9) holds. Let $\mathcal{X}_{i,T} \equiv \frac{1}{T^2} \sum_{t=1}^{T} x_{1,it} x_{1,it}'$ and $\mathcal{X}_i \equiv \int B_{2i} B_{2i}'$. Recall that M denotes a generic large constant. Our conditions ensure that $\sup_i \sup_T \mathbb{E} \|\mathcal{X}_{i,T}\|^2 \le M$. It follows that $\frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E} \|\mathcal{X}_{i,T}\| \le M$ and $\frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E} \|\mathcal{X}_{i,T}\| \mathbf{1}\{ \|\mathcal{X}_{i,T}\| > N\epsilon\} = 0$ for any $\epsilon > 0$, verifying conditions (i) and (iii) in PM (1999)'s Theorem 1. In addition, $\|\mathcal{X}_{i,T}\|^{1+\varsigma} \Rightarrow \|\mathcal{X}_{i,T}\|^{1+\varsigma}$ for all $\varsigma \in [0,1]$ by the continuos mapping theorem. This, in conjunction with the uniform integrability of $\{\|\mathcal{X}_{i,T}\|^{1+\varsigma}\}$ in T for all i and all $\varsigma \in [0,1)$ (implied by $\sup_i \sup_T \mathbb{E} \|\mathcal{X}_{i,T}\|^2 \le M$), and the Fatou lemma, implies that $\mathbb{E}(\mathcal{X}_{i,T}) \to \mathbb{E}(\mathcal{X}_i)$ and $\mathbb{E} \|\mathcal{X}_{i,T}\|^{1+\varsigma} \to \mathbb{E} \|\mathcal{X}_i\|^{1+\varsigma}$ for all $\varsigma \in [0,1)$ as $T \to \infty$. Then $\frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E} \|\mathcal{X}_i\|^{1+\varsigma} \le M < \infty$ for some $\varsigma > 0$ (see, e.g., Lemma 12 in PM (1999)), which implies that $\frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E} [\|\mathcal{X}_i\| \mathbf{1} \{\|\mathcal{X}_i\| > N\epsilon\}] = 0$, verifying condition (iv) in PM's Theorem 1. To verify condition (ii) in PM's Theorem 1, we apply the Skorohod representation theorem to construct $\left\{\mathcal{X}_{i,T}^*\right\}$ and $\{\mathcal{X}_i^*\}$ in some probability space such that $\mathcal{X}_{i,T}^* \overset{d}{=} \mathcal{X}_{i,T}$, $\mathcal{X}_i^* \overset{d}{=} \mathcal{X}_i$ for all i, and $\mathcal{X}_{i,T}^* \xrightarrow{a.s.} \mathcal{X}_i^*$, where $\overset{d}{=}$ and $\xrightarrow{a.s.}$ denote equality in distribution and almost sure convergence, respectively. Let $D_{i,T} = \mathcal{X}_{i,T}^* - \mathcal{X}_i^*$. Then $\{D_{i,T}\}$ are uniformly integrable in T for all i and $D_{i,T} \xrightarrow{a.s.} 0$. By the uniform integrability of $\{D_{i,T}\}$, for any $\epsilon > 0$ there exists $\delta = \delta(\epsilon)$ such that $\sup_i \sup_T \mathbb{E} [\|D_{i,T}\| \mathbf{1} \{\|D_{i,T}\| \ge \delta\}] \le \epsilon$. By the almost sure convergence of $D_{i,T}$ to zero and the dominated convergence theorem, $\lim_{T\to\infty} \sup_i \mathbb{E} [\|D_{i,T}\| \mathbf{1} \{\|D_{i,T}\| < \delta\}] = 0$. In addition, notice that

$$\begin{aligned}
\frac{1}{N_k} \sum_{i\in G_k^0} \|\mathbb{E}(\mathcal{X}_{i,T}) - \mathbb{E}(\mathcal{X}_i)\| &= \frac{1}{N_k} \sum_{i\in G_k^0} \left\|\mathbb{E}\left(\mathcal{X}_{i,T}^*\right) - \mathbb{E}(\mathcal{X}_i^*)\right\| \\
&\le \frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E} \left\|\mathcal{X}_{i,T}^* - \mathcal{X}_i^*\right\| = \frac{1}{N_k} \sum_{i\in G_k^0} \mathbb{E} \|D_{i,T}\|.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\limsup_{(N,T)\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\mathbb{E}\left(\mathcal{X}_{i,T}\right)-\mathbb{E}(\mathcal{X}_i)\right\| \\
&\le \limsup_{(N,T)\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0}\left[\mathbb{E}\left\|D_{i,T}\mathbf{1}\left\{\|D_{i,T}\|<\delta\right\}\right\|+\mathbb{E}\left\|D_{i,T}\mathbf{1}\left\{\|D_{i,T}\|\ge\delta\right\}\right\|\right]\le 0+\epsilon=\epsilon.
\end{aligned}$$

Since ϵ is arbitrary, we conclude that $\limsup_{(N,T)\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\mathbb{E}\left(\mathcal{X}_{i,T}\right)-\mathbb{E}(\mathcal{X}_i)\right\|=0$, which verifies condition (ii) in PM (1999)'s Theorem 1.

To show $Q_{2NT}\xrightarrow{d}Q_2$ as $(N,T)\to\infty$, we also establish the sequential limit first. Note that the (k,l)th block element of Q_{2NT} is given by $Q_{2NT,kl}=\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j\in G_l^0}a_{ij}x_{1,i}'M_{F_1^0}x_{1,j}$. As $T\to\infty$, $\frac{1}{T^{1/2}}\tilde{x}_{1,it}\Rightarrow B_{2,i}-\int B_{2,i}B_3'\left(\int B_3B_3'\right)^{-1}B_3\equiv\tilde{B}_{2i}$, and

$$\frac{1}{T^2}x_{1,i}'M_{F_1^0}x_{1,j}=\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{1,j}=\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{T^{1/2}}\tilde{x}_{1,it}\right)\left(\frac{1}{T^{1/2}}\tilde{x}_{1,jt}\right)'\Rightarrow\int\tilde{B}_{2i}\tilde{B}_{2j}\text{ by the CMT.}$$

By the conditional law of large numbers for second order U-statistics with independent observations (conditional on $\mathcal{C}$),

$$\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j\in G_l^0}a_{ij}\int\tilde{B}_{2i}\tilde{B}_{2j}\xrightarrow{p}\lim_{N\to\infty}\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j\in G_l^0}a_{ij}\mathbb{E}_{\mathcal{C}}\left(\int\tilde{B}_{2i}\tilde{B}_{2j}\right)\text{ as }N\to\infty.$$

It follows that

$$\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j\in G_l^0}a_{ij}x_{1,i}'M_{F_1^0}x_{1,j}\xrightarrow{d}\lim_{N\to\infty}\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j\in G_l^0}a_{ij}\mathbb{E}_{\mathcal{C}}\left(\int\tilde{B}_{2i}\tilde{B}_{2j}\right)\text{ as }(N,T)_{\text{seq}}\to\infty.$$

Let $\mathcal{X}_{ij,T}=\frac{1}{T^2}a_{ij}x_{1,i}'M_{F_1^0}x_{1,j}$ and $\mathcal{X}_{ij}=a_{ij}\int\tilde{B}_{2i}\tilde{B}_{2j}$. To obtain the joint limit, we can follow the proof of Theorem 1 in PM (1999) and find that it is sufficient to verify

(i1) $\limsup_{(N,T)\to\infty}\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j\in G_l^0}\mathbb{E}\left\|\mathcal{X}_{ij,T}\right\|<\infty$,

(i2) $\limsup_{(N,T)\to\infty}\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j\in G_l^0}\left\|\mathbb{E}(\mathcal{X}_{ij,T})-\mathbb{E}(\mathcal{X}_{ij})\right\|<\infty$,

(i3) $\limsup_{(N,T)\to\infty}\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j\in G_l^0}\mathbb{E}\left[\left\|\mathcal{X}_{ij,T}\right\|\mathbf{1}\left\{\left\|\mathcal{X}_{ij,T}\right\|>N\epsilon\right\}\right]=0\ \forall\ \epsilon>0$, and

(i4) $\limsup_{(N,T)\to\infty}\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j\in G_l^0}\mathbb{E}\left[\left\|\mathcal{X}_{ij}\right\|\mathbf{1}\left\{\left\|\mathcal{X}_{ij}\right\|>N\epsilon\right\}\right]=0\ \forall\ \epsilon>0$.

Note that $\{\mathcal{X}_{ij,T}\}$ is uniformly integrable in T for all i and j. We can follow step (1a) and verify the above conditions analogously. As a result,

$$\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j\in G_l^0}a_{ij}x_{1,i}'M_{F_1^0}x_{1,j}\xrightarrow{d}\lim_{N\to\infty}\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j\in G_l^0}a_{ij}\mathbb{E}_{\mathcal{C}}\left(\int\tilde{B}_{2i}\tilde{B}_{2j}\right)\text{ as }(N,T)\to\infty.$$

(ii) First, we observe that

$$U_{kNT}=U_{kNT}^{u}+U_{kNT}^{f_2},\tag{B.3}$$

where $U_{kNT}^{u}=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}x_{1,i}'M_{F_1^0}(u_i-\frac{1}{N}\sum_{j=1}^{N}u_ja_{ij})$ and $U_{kNT}^{f_2}=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}x_{1,i}'M_{F_1^0}(F_2^0\lambda_{2i}^0-$

$\frac{1}{N}\sum_{j=1}^{N} F_2^0\lambda_{2j}^0 a_{ij})$. We study U_{kNT}^{u} and $U_{kNT}^{f_2}$ in turn.

For U_{kNT}^{u}, we make the decomposition

$$\begin{aligned}
U_{kNT}^{u} &= \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}x_{1,i}'M_{F_1^0}u_i - \frac{1}{\sqrt{N_k}}\sum_{i=1}^{N}\frac{1}{N}\sum_{j\in G_k^0}a_{ij}\frac{1}{T}x_{1,j}'M_{F_1^0}u_i \\
&= \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}[x_{1,i}' - \mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)]M_{F_1^0}u_i \\
&+ \frac{1}{\sqrt{N_k}}\sum_{i=1}^{N}\left\{\frac{1}{T}\mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)\mathbf{1}\left\{i\in G_k^0\right\} - \frac{1}{N}\sum_{j\in G_k^0}a_{ij}\frac{1}{T}\mathbb{E}_{\mathcal{C}}(x_{1,j}')\right\}M_{F_1^0}u_i \\
&- \frac{1}{\sqrt{N_k}}\sum_{i=1}^{N}\frac{1}{N}\sum_{j\in G_k^0}a_{ij}\frac{1}{T}[x_{1,j} - \mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)]'M_{F_1^0}u_i \\
&\equiv U_{1kNT}^{u} + U_{2kNT}^{u} - U_{3kNT}^{u},\ \text{say.}
\end{aligned}$$

We will show that U_{1kNT}^{u} contributes to both the asymptotic bias and variance, U_{2kNT}^{u} contributes to the asymptotic variance, and U_{3kNT}^{u} is asymptotically negligible. We study these three terms in turn.

For U_{1kNT}^{u}, we make further decomposition:

$$U_{1kNT}^{u} = \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}[x_{1,i} - \mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)]'u_i - \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}[x_{1,i} - \mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)]'P_{F_1^0}u_i \equiv U_{1kNT,1}^{u} - U_{1kNT,2}^{u}.$$

Let $x_{1,it}^{\dagger} = x_{1,it} - \mathbb{E}_{\mathcal{C}}\left(x_{1,it}\right)$. Let $\phi_i^{\varepsilon\dagger}(L) = (\phi_i^{\varepsilon u}(L), \phi_i^{\varepsilon\varepsilon}(L), \phi_i^{\varepsilon x_2}(L))$, $\phi_i^{u\dagger}(L) = (\phi_i^{uu}(L), \phi_i^{u\varepsilon}(L), \phi_i^{ux_2}(L))$, $\phi_i^{\varepsilon, f_1 f_2} = (\phi_i^{\varepsilon f_1}(L), \phi_i^{\varepsilon f_2}(L))$, $v_{it}^{ux} = \left(v_{it}^{u}, v_{it}^{\varepsilon\prime}, v_{it}^{x_2\prime}\right)'$, and $v_t^{f_1 f_2} = (v_t^{f_1\prime}, v_t^{f_1\prime})'$. Noting that $\varepsilon_{it} = w_{it}^{\varepsilon} = \phi_i^{\varepsilon u}(L)v_{it}^{u} + \phi_i^{\varepsilon\varepsilon}(L)v_{it}^{\varepsilon} + \phi_i^{\varepsilon x_2}(L)v_{it}^{x_2} + \phi_i^{\varepsilon f_1}(L)v_t^{f_1} + \phi_i^{\varepsilon f_2}(L)v_t^{f_2}$ and by the independence of $\{v_{it}^{ux}\}$ and $\{v_s^{f_1 f_2}\}$, we have

$$\begin{aligned}
u_{it} =& \phi_i^{uu}(L)v_{it}^{u} + \phi_i^{u\varepsilon}(L)v_{it}^{\varepsilon} + \phi_i^{ux_2}(L)v_{it}^{x_2} = \phi_i^{u\dagger}(L)v_{it}^{ux} = S^u\phi_i^{\dagger}(L)v_{it}^{ux}, \\
\mathbb{E}_{\mathcal{C}}\left(\varepsilon_{it}\right) =& \phi_i^{\varepsilon f_1}(L)v_t^{f_1} + \phi_i^{\varepsilon f_2}(L)v_t^{f_2} = \phi_i^{\varepsilon, f_1 f_2}(L)v_t^{f_1 f_2}, \\
\varepsilon_{it} - \mathbb{E}_{\mathcal{C}}\left(\varepsilon_{it}\right) =& \phi_i^{\varepsilon u}(L)v_{it}^{u} + \phi_i^{\varepsilon\varepsilon}(L)v_{it}^{\varepsilon} + \phi_i^{\varepsilon x_2}(L)v_{it}^{x_2} = \phi_i^{\varepsilon\dagger}(L)v_{it}^{ux} = S^{\varepsilon}\phi_i^{\dagger}(L)v_{it}^{ux},
\end{aligned}$$

where

$$\phi_i^{\dagger}(L) = \begin{pmatrix} \phi_i^{u\dagger}(L) \\ \phi_i^{\varepsilon\dagger}(L) \end{pmatrix} = \begin{pmatrix} \phi_i^{uu}(L) & \phi_i^{u\varepsilon}(L) & \phi_i^{ux_2}(L) \\ \phi_i^{\varepsilon u}(L) & \phi_i^{\varepsilon\varepsilon}(L) & \phi_i^{\varepsilon x_2}(L) \end{pmatrix}_{(1+p_1)\times(1+p)},$$

$S^u = (1, 0_{1\times p})$, and $S^{\varepsilon} = (0_{p_1\times 1}, \iota_{p_1\times p})$. Let $V_{it}^{ux} = (V_{it}^{u}, V_{it}^{\varepsilon\prime}, V_{it}^{x_2\prime})' = \left(\sum_{s=1}^{t} v_{is}^{u}, \sum_{s=1}^{t} v_{is}^{\varepsilon\prime}, \sum_{s=1}^{t} v_{is}^{x_2\prime}\right)'$ and $w_{it}^{ux} = \left(w_{it}^{u}, w_{it}^{\varepsilon\prime}, w_{it}^{x_2\prime}\right)'$. Then by the panel BN decomposition,

$$w_{it}^{ux} = \phi_i^{\dagger}(1)v_{it}^{ux} + \tilde{w}_{i,t-1}^{ux} - \tilde{w}_{it}^{ux} \text{ and } \sum_{s=1}^{t} w_{is}^{ux} = \phi_i^{\dagger}(1)V_{it}^{ux} + \tilde{w}_{i0}^{ux} - \tilde{w}_{it}^{ux}, \tag{B.4}$$

where $\tilde{w}_{it}^{ux}=\sum_{j=0}^{\infty}\tilde{\phi}_{ij}^{\dagger}v_{i,t-j}^{ux}$ and $\tilde{\phi}_{ij}^{\dagger}=\sum_{s=j+1}^{\infty}\phi_{i,s}^{\dagger}$. Let $B_{1kNT,1}^{u}=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}S^{\varepsilon}\sum_{r=0}^{\infty}\sum_{l=0}^{\infty}\phi_{i,l+r}^{\dagger}\phi_{i,l}^{\dagger\prime}S^{w\prime}$. It follows that

$$
\begin{aligned}
&U_{1kNT,1}^{u}-B_{1kNT,1}^{u}\\
&=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^{T}x_{1,it}^{\dagger}u_{it}-B_{1kNT,1}^{u}\\
&=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}S^{\varepsilon}\left(\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{r=1}^{t}w_{ir}^{ux}w_{it}^{ux\prime}-\sum_{r=0}^{\infty}\sum_{l=0}^{\infty}\phi_{i,l+r}^{\dagger}\phi_{i,l}^{\dagger\prime}\right]\right)S^{w\prime}\\
&=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}S^{\varepsilon}\left\{\phi_i^{\dagger}(1)\frac{1}{T}\sum_{t=1}^{T}\left(V_{it}^{ux}v_{it}^{ux\prime}-I_{1+p}\right)\phi_i^{\dagger}(1)'+\frac{1}{T}\sum_{t=1}^{T-1}\left(w_{i,t+1}^{ux}\tilde{w}_{it}^{ux\prime}-\sum_{s=0}^{\infty}\phi_{i,s+1}^{\dagger}\tilde{\phi}_{i,s}^{\dagger}\right)\right.\\
&-\frac{1}{T}\sum_{s=0}^{\infty}\phi_{i,s+1}^{\dagger}\tilde{\phi}_{i,s}^{\dagger}-\frac{1}{T}\sum_{t=1}^{T}\left(\tilde{w}_{it}^{ux}v_{it}^{ux\prime}-\tilde{\phi}_{i0}^{\dagger}\right)\phi_i^{\dagger}(1)'+\frac{1}{T}\sum_{t=1}^{T}\tilde{w}_{i0}^{ux}v_{it}^{ux\prime}\phi_i^{\dagger}(1)'\\
&\left.-\frac{1}{T}\sum_{t=1}^{T}w_{it}^{ux}\tilde{w}_{iT}^{ux\prime}+\frac{1}{T}w_{i1}^{ux}\tilde{w}_{i0}^{ux\prime}\right\}S^{w\prime}\\
=&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}S^{\varepsilon}\left\{Q_{iT}+R_{1iT,1}+R_{1iT,2}+R_{1iT,3}+R_{1iT,4}+R_{1iT,5}+R_{1iT,6}\right\}S^{w\prime}.
\end{aligned}
$$

By Lemma A.7 in HJS, $\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}S^{\varepsilon}R_{1iT,l}S^{w\prime}=o_P(1)$ for $l=1,2,...,6$. It follows that

$$
U_{1kNT,1}^{u}-B_{1kNT,1}^{u}=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T}\sum_{t=1}^{T}\left(V_{it}^{ux}v_{is}^{ux\prime}-I_{1+p_1}\right)\phi_i^{\dagger}(1)'S^{w\prime}+o_P(1).
$$

Recall that $\varkappa_{ts}=f_{1t}^{0\prime}(F_1^{0\prime}F_1^0)^{-1}f_{1s}^0$ and $\bar{\varkappa}_{ts}=\mathbf{1}\{t=s\}-\varkappa_{ts}$. Let $B_{1kNT,2}^{u}=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}S^{\varepsilon}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\varkappa_{ts}\mathbf{1}\{s\le t\}\sum_{r=0}^{\infty}\sum_{l=0}^{\infty}\phi_{i,l+r}^{\dagger}\phi_{i,l}^{\dagger\prime}S^{w\prime}$. Then, using the BN decomposition in (B.4) and following the proof of Lemma A.7 in HJS, we can show that

$$
\begin{aligned}
U_{1kNT,2}^{u}-B_{1kNT,2}^{u}&=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\varkappa_{ts}x_{1,it}^{\dagger}u_{is}-B_{1kNT,2}^{u}\\
&=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}S^{\varepsilon}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\varkappa_{ts}\left[\sum_{r=1}^{t}w_{ir}^{ux}w_{is}^{ux\prime}-\mathbf{1}\{s\le t\}\sum_{r=0}^{\infty}\sum_{l=0}^{\infty}\phi_{i,l+r}^{\dagger}\phi_{i,l}^{\dagger\prime}\right]\right)S^{w\prime}\\
&=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\varkappa_{ts}[V_{it}^{ux}v_{is}^{ux\prime}-\mathbf{1}\{s\le t\}I_{1+p}]\phi_i^{\dagger}(1)'S^{w\prime}+o_P(1),
\end{aligned}
$$

where we use the fact that $\mathbb{E}_{\mathcal{C}}\left(V_{it}^{ux}v_{is}^{ux\prime}\right)=\mathbb{E}\left(V_{it}^{ux}v_{is}^{ux\prime}\right)=I_{1+p}$ if $s\le t$ and 0 if $s>t$ by the independence of v_{it}^{ux} over t. Then

$$
U_{1kNT}^{u}-B_{1kNT}^{u}=V_{1kNT}+o_P(1),
$$

where

$$
\begin{aligned}
V_{1kNT} =& \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} S^{\varepsilon} \phi_i^{\dagger}(1) \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left\{ \bar{\varkappa}_{ts} \left(V_{it}^{ux} v_{is}^{ux\prime} \right) - \left[\mathbf{1}\{t=s\} - \varkappa_{ts} \mathbf{1}\{s \le t\} \right] I_{1+p} \right\} \phi_i^{\dagger}(1)' S^{w\prime}, \\
B_{1kNT} =& \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[\mathbf{1}\{t=s\} - \varkappa_{ts} \mathbf{1}\{s \le t\} \right] \right) \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \Delta_{21,i},
\end{aligned}
$$

since $\Delta_{21,i} = S^{\varepsilon} \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r}^{\dagger} \phi_{i,l}^{\dagger\prime} S^{w\prime} = S_2 \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r} \phi_{i,l}' S_1'$ by construction.

For U_{2kNT}^{u}, we make further decomposition

$$
U_{2kNT}^{u} = \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \frac{1}{T} \mathbb{E}_{\mathcal{C}} \left(x_{1,i}' \right) M_{F_1^0} u_i - \frac{1}{\sqrt{N_k}} \sum_{i=1}^{N} \frac{1}{N} \sum_{j \in G_k^0} a_{ij} \frac{1}{T} \mathbb{E}_{\mathcal{C}} (x_{1,j}') M_{F_1^0} u_i \equiv U_{2kNT,1}^{u} - U_{2kNT,2}^{u}, \text{ say.}
$$

Apparently, $\mathbb{E}_{\mathcal{C}} \left(U_{2kNT,1}^{u} \right) = \mathbb{E}_{\mathcal{C}} \left(U_{2kNT,2}^{u} \right) = 0$, $\mathrm{Var}\left(U_{2kNT,1}^{u} | \mathcal{C} \right) = O_P(1)$ and $\mathrm{Var}\left(U_{2kNT,1}^{u} | \mathcal{C} \right) = O_P(1)$. We now show that $U_{2kNT,1}^{u}$ and $U_{2kNT,2}^{u}$ are asymptotically independent of V_{1kNT} conditional on $\mathcal{C}$. Note that

$$
\begin{aligned}
V_{1kNT} =& \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} S^{\varepsilon} \phi_i^{\dagger}(1) \frac{1}{T} \sum_{t=1}^{T} \left(V_{it}^{ux} v_{it}^{ux\prime} - I_{1+p} \right) \phi_i^{\dagger}(1)' S^{w\prime} \\
& - \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} S^{\varepsilon} \phi_i^{\dagger}(1) \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \varkappa_{ts} [V_{it}^{ux} v_{is}^{ux\prime} - \mathbf{1}\{s \le t\} I_{1+p}] \phi_i^{\dagger}(1)' S^{w\prime} \equiv V_{1kNT,1} - V_{1kNT,2}.
\end{aligned}
$$

Let c_1 and c_2 be arbitrary nonrandom $p \times 1$ vectors such that $\|c_1\| = \|c_2\| = 1$. Note that

$$
\begin{aligned}
& \mathrm{Cov} \left(c_1' V_{1kNT,1}, c_2' U_{2kNT,1}^{u} | \mathcal{C} \right) \\
&= \mathbb{E}_{\mathcal{C}} \left[\frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} c_1' S^{\varepsilon} \phi_i^{\dagger}(1) \frac{1}{T} \sum_{t=1}^{T} \left(V_{it}^{ux} v_{it}^{ux\prime} - I_{1+p} \right) \phi_i^{\dagger}(1)' S^{w\prime} \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \frac{1}{T} u_i' M_{F_1^0} \mathbb{E}_{\mathcal{C}} \left(x_{1,i} \right) c_2 \right] \\
&= \frac{1}{N_k} \sum_{i \in G_k^0} c_1' S^{\varepsilon} \phi_i^{\dagger}(1) \frac{1}{T^2} \sum_{t=1}^{T} \mathbb{E}_{\mathcal{C}} \left\{ \left(V_{it}^{ux} v_{it}^{ux\prime} - I_{1+p} \right) \phi_i^{\dagger}(1)' S^{w\prime} u_i' \right\} M_{F_1^0} \mathbb{E}_{\mathcal{C}} \left(x_{1,i} \right) c_2 \\
&= \frac{1}{N_k} \sum_{i \in G_k^0} c_1' S^{\varepsilon} \phi_i^{\dagger}(1) \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}_{\mathcal{C}} \left\{ \left(V_{it}^{ux} v_{it}^{ux\prime} - I_{1+p} \right) \phi_i^{\dagger}(1)' S^{w\prime} u_{is}' \right\} \sum_{r=1}^{T} \bar{\varkappa}_{sr} \mathbb{E}_{\mathcal{C}} \left(x_{1,ir} \right) c_2.
\end{aligned}
$$

Using the BN decomposition, we can readily show that

$$\frac{1}{N_k}\sum_{i\in G_k^0} c_1' S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_{\mathcal{C}}\left\{\left(V_{it}^{ux}v_{it}^{ux\prime}-I_{1+p}\right)\phi_i^{\dagger}(1)'\ S^{u\prime}u_{is}'\right\}\sum_{r=1}^{T}\bar{\varkappa}_{sr}\mathbb{E}_{\mathcal{C}}\left(x_{1,ir}\right)c_2$$

$$=\frac{1}{N_k}\sum_{i\in G_k^0} c_1' S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_{\mathcal{C}}\left\{\left(V_{it}^{ux}v_{it}^{ux\prime}-I_{1+p}\right)\phi_i^{\dagger}(1)'\ S^{u\prime}S^{u}\phi_i^{\dagger}(1)v_{is}^{ux}\right\}\ \sum_{r=1}^{T}\bar{\varkappa}_{sr}\mathbb{E}_{\mathcal{C}}\left(x_{1,ir}\right)c_2+o_P\left(1\right)$$

$$=\frac{1}{N_k}\sum_{i\in G_k^0} c_1' S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[V_{it}^{ux}v_{it}^{ux\prime}\phi_i^{\dagger}(1)'\ S^{u\prime}S^{u}\phi_i^{\dagger}(1)v_{is}^{ux}\right]\ \sum_{r=1}^{T}\bar{\varkappa}_{sr}\mathbb{E}_{\mathcal{C}}\left(x_{1,ir}\right)c_2+o_P\left(1\right)$$

$$=\frac{1}{N_k}\sum_{i\in G_k^0} c_1' S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}\left[V_{it}^{ux}v_{it}^{ux\prime}\phi_i^{\dagger}(1)'\ S^{u\prime}S^{u}\phi_i^{\dagger}(1)v_{it}^{ux}\right]\sum_{r=1}^{T}\bar{\varkappa}_{sr}\mathbb{E}_{\mathcal{C}}\left(x_{1,ir}\right)c_2+o_P\left(1\right)=o_P\left(1\right),$$

where the first inequality follows by the BN decomposition, the second equality follows by the fact that $E\left(v_{is}^{u\varepsilon}\right)=0$, the third inequality follows from $E[V_{it}^{ux}v_{it}^{ux\prime}\phi_i^{\dagger}(1)'\ S^{u}\phi_i^{\dagger}(1)v_{is}^{ux}]=0$ for $t\neq s$ and the last equality follows by straightforward moment calculations. Similarly, we have

$$\mathrm{Cov}\left(c_1'V_{1kNT,2},c_2'U_{2kNT,1}^{u}|\mathcal{C}\right)$$

$$=\mathbb{E}_{\mathcal{C}}\left[\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}c_1'S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\varkappa_{ts}[V_{it}^{ux}v_{is}^{ux\prime}-\mathbf{1}\left\{s\leq t\right\}I_{1+p}]\phi_i^{\dagger}(1)'S^{u\prime}\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}u_i'M_{F_1^0}\mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)c_2\right]$$

$$=\frac{1}{N_k}\sum_{i\in G_k^0}c_1'S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\varkappa_{ts}\mathbb{E}_{\mathcal{C}}\left\{\left[V_{it}^{ux}v_{is}^{ux\prime}-\mathbf{1}\left\{s\leq t\right\}I_{1+p}\right]\phi_i^{\dagger}(1)'S^{u\prime}u_i'\right\}M_{F_1^0}\mathbb{E}_{\mathcal{C}}\left(x_i\right)c_2$$

$$=\frac{1}{N_k}\sum_{i\in G_k^0}c_1'S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T^2}\sum_{t,s,r}\varkappa_{ts}\mathbb{E}_{\mathcal{C}}\left\{\left[V_{it}^{ux}v_{is}^{ux\prime}-\mathbf{1}\left\{s\leq t\right\}I_{1+p}\right]\phi_i^{\dagger}(1)'S^{u\prime}u_{ir}\right\}\sum_{l=1}^{T}\bar{\varkappa}_{rl}\mathbb{E}_{\mathcal{C}}\left(x_{1,il}\right)c_2$$

$$=\frac{1}{N_k}\sum_{i\in G_k^0}c_1'S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T^2}\sum_{t,s,r}\varkappa_{ts}\mathbb{E}_{\mathcal{C}}\left\{\left[V_{it}^{ux}v_{is}^{ux\prime}-\mathbf{1}\left\{s\leq t\right\}I_{1+p}\right]\phi_i^{\dagger}(1)'S^{u\prime}S^{u}\phi_i^{\dagger}(1)v_{ir}^{ux}\right\}\sum_{l=1}^{T}\bar{\varkappa}_{rl}\mathbb{E}_{\mathcal{C}}\left(x_{1,il}\right)c_2$$

$$+o_P\left(1\right)$$

$$=\frac{1}{N_k}\sum_{i\in G_k^0}c_1'S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T^2}\sum_{t,s,r}\varkappa_{ts}\mathbb{E}\left[V_{it}^{ux}v_{is}^{ux\prime}\phi_i^{\dagger}(1)'S^{u\prime}S^{u}\phi_i^{\dagger}(1)v_{ir}^{ux}\right]\sum_{l=1}^{T}\bar{\varkappa}_{rl}\mathbb{E}_{\mathcal{C}}\left(x_{1,il}\right)c_2+o_P\left(1\right)$$

$$=\frac{1}{N_k}\sum_{i\in G_k^0}c_1'S^{\varepsilon}\phi_i^{\dagger}(1)\frac{1}{T^2}\sum_{t=1}^{T}\varkappa_{ts}\mathbb{E}\left[V_{it}^{ux}v_{it}^{ux\prime}\phi_i^{\dagger}(1)'S^{u\prime}S^{u}\phi_i^{\dagger}(1)v_{it}^{ux}\right]\sum_{l=1}^{T}\bar{\varkappa}_{rl}\mathbb{E}_{\mathcal{C}}\left(x_{1,il}\right)c_2+o_P\left(1\right)=o_P\left(1\right),$$

where $\sum_{t,s,r}=\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}$. It follows that $\mathrm{Cov}\Big(c_1'V_{1kNT},c_2'U_{2kNT,1}^{u}|\mathcal{C}\Big)=o_P\left(1\right)$. Analogously, we can show that $\mathrm{Cov}\Big(c_1'V_{1kNT},c_2'U_{2kNT,2}^{u}|\mathcal{C}\Big)=o_P\left(1\right)$. Then $\mathrm{Cov}(c_1'V_{1kNT},c_2'U_{2kNT}^{u}|\mathcal{C})=o_P\left(1\right)$.

For U_{3kNT}^{u}, we can readily show that $U_{3kNT}^{u}=\bar{U}_{3kNT}^{u}+o_P\left(1\right)$ where $\bar{U}_{3kNT}^{u}=\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N}\frac{1}{N}\sum_{j\in G_k^0,j\neq i}\frac{1}{T}[x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)]'M_{F_1^0}a_{ij}u_i$. By the independence of (x_i,u_i) across i conditional on $\mathcal{C}$, we have

$E_{\mathcal{C}}\left(\bar{U}_{3kNT}^{u}\right)=0$ and $\operatorname{tr}\{\operatorname{Var}\left(\bar{U}_{3kNT}^{u}|\mathcal{C}\right)\}=\frac{1}{N^2N_kT^2}\sum_{i=1}^{N}\sum_{j\in G_k^0,j\neq i}\sum_{i_1=1}^{N}\sum_{j_1\in G_k^0,j_1\neq i_1}a_{ij}a_{i_1j_1}\,\mathbb{E}_{\mathcal{C}}\{u_{i_1}'$ $M_{F_1^0}[x_{1,j_1}-\mathbb{E}\left(x_{1,j_1}|\mathcal{C}\right)][x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)]'M_{F_1^0}u_i\}=O_P\left(N^{-1}\right)$, where we use the fact that $\mathbb{E}_{\mathcal{C}}\{u_{i_1}'M_{F_1^0}[x_{1,j_1}$ $-\mathbb{E}_{\mathcal{C}}\left(x_{1,j_1}\right)][x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)]'M_{F_1^0}u_i\}$ is nonzero if and only if $\#\{i,j,i_1,j_1\}=2$ or 1. It follows that $\bar{U}_{3kNT}^{u}=O_P\left(N^{-1/2}\right)$ and $U_{3kNT}^{u}=o_P\left(1\right).$

In sum, we have

$$U_{kNT}^{u}-B_{1kNT}=V_{1kNT}+V_{2kNT}+o_P\left(1\right), \tag{B.5}$$

where $V_{2kNT}=U_{2kNT}^{u}=\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N}\{\frac{1}{T}\mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)\mathbf{1}\left\{i\in G_k^0\right\}-\frac{1}{N}\sum_{j\in G_k^0}a_{ij}\frac{1}{T}\mathbb{E}_{\mathcal{C}}(x_{1,j}')\}M_{F_1^0}u_i$ and we have shown that V_{1kNT} and V_{2kNT} are asymptotically independent conditional on $\mathcal{C}$.

Now, we study $U_{kNT}^{f_2}$. We make the decomposition

$$\begin{aligned}
U_{kNT}^{f_2}=&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}x_{1,i}'M_{F_1^0}F_2^0\lambda_{2i}^0-\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N}\frac{1}{N}\sum_{j\in G_k^0}a_{ij}\frac{1}{T}x_{1,j}'M_{F_1^0}F_2^0\lambda_{2i}^0\\
&=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\left[x_{1,i}-\mathbb{E}_{\mathcal{C}}\left(x_i\right)\right]'M_{F_1^0}F_2^0\lambda_{2i}^0\\
&+\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N}\frac{1}{T}\left[\mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)\mathbf{1}\left\{i\in G_k^0\right\}-\frac{1}{N}\sum_{j\in G_k^0}a_{ij}\frac{1}{T}\mathbb{E}_{\mathcal{C}}(x_{1,j}')\right]M_{F_1^0}F_2^0\lambda_{2i}^0\\
&-\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N}\frac{1}{N}\sum_{j\in G_k^0}a_{ij}\frac{1}{T}\left[x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)\right]'M_{F_1^0}F_2^0\lambda_{2i}^0\equiv U_{1kNT}^{f_2}+U_{2kNT}^{f_2}-U_{3kNT}^{f_2}.
\end{aligned}$$

We show that $U_{1kNT}^{f_2}$ and $U_{2kNT}^{f_2}$ contribute to the asymptotic variance and bias, respectively, and $U_{3kNT}^{f_2}$ is asymptotically negligible. For $U_{1kNT}^{f_2}$, we have $\mathbb{E}[U_{1kNT}^{f_2}|\mathcal{C}]=0$, and $\operatorname{tr}\left[\operatorname{Var}(U_{1kNT}^{f_2}|\mathcal{C})\right]=\frac{1}{N_k}\sum_{i\in G_k^0}\frac{1}{T^2}\lambda_{2i}^{0\prime}f_2^{0\prime}M_{F_1^0}\mathbb{E}_{\mathcal{C}}\left\{\left[x_{1,i}-\mathbb{E}_{\mathcal{C}}\left(x_i\right)\right]\left[x_{1,i}-\mathbb{E}_{\mathcal{C}}\left(x_i\right)\right]'\right\}M_{F_1^0}F_2^0\lambda_{2i}^0=O_P\left(1\right)$. For $U_{2kNT}^{f_2}$, we have $U_{2kNT}^{f_2}=B_{2kNT}$. For $U_{3kNT}^{f_2}$, we have $\mathbb{E}_{\mathcal{C}}[U_{3kNT}^{f_2}]=0$, and

$$\begin{aligned}
&\operatorname{tr}\left[\operatorname{Var}(U_{3kNT}^{f_2}|\mathcal{C})\right]=\frac{1}{N_kN^2T^2}\sum_{i,l=1}^{N}\sum_{j,m\in G_k^0}a_{ij}a_{lm}\lambda_{2l}^{0\prime}F_2^{0\prime}M_{F_1^0}\mathbb{E}_{\mathcal{C}}\left\{[x_{1,m}-\mathbb{E}_{\mathcal{C}}(x_{1,m})][x_{1,j}-\mathbb{E}_{\mathcal{C}}(x_{1,j})]'\right\}M_{F_1^0}F_2^0\lambda_{2i}^0\\
&=\frac{1}{N_kN^2T^2}\sum_{i,l=1}^{N}\sum_{j\in G_k^0}a_{ij}a_{lj}\lambda_{2l}^{0\prime}F_2^{0\prime}M_{F_1^0}\mathbb{E}_{\mathcal{C}}\left\{\left[x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)\right]\left[x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)\right]'\right\}M_{F_1^0}F_2^0\lambda_{2i}^0\\
&=\frac{1}{N_kT^2}\sum_{j\in G_k^0}\lambda_{1j}^{0\prime}(\frac{\lambda_1^{0\prime}\lambda_1^0}{N})^{-1}\frac{\lambda_1^{0\prime}\lambda_2^0}{N}F_2^{0\prime}M_{F_1^0}\mathbb{E}_{\mathcal{C}}\left\{\left[x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)\right]\left[x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)\right]'\right\}M_{F_1^0}F_2^0\frac{\lambda_2^{0\prime}\lambda_1^0}{N}(\frac{\lambda_1^{0\prime}\lambda_1^0}{N})^{-1}\lambda_{1j}^0\\
&\leq\frac{1}{N}\left\|\frac{\lambda_1^{0\prime}\lambda_2^0}{N^{1/2}}\right\|^2\left\|(\frac{\lambda_1^{0\prime}\lambda_1^0}{N})^{-1}\right\|^2\frac{1}{N_kT^2}\sum_{j\in G_k^0}\left\|\lambda_{1j}^0\right\|^2\left\|F_2^{0\prime}M_{F_1^0}\mathbb{E}_{\mathcal{C}}\left\{[x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)][x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)]'\right\}M_{F_1^0}F_2^0\right\|\\
&=O_P\left(N^{-1}\right),
\end{aligned}$$

where the second equality follows from the independence of $\{x_{1,i}\}$ across i conditional on $\mathcal{C}$, the third equality follows the fact that $a_{ij}=\lambda_{1i}^{0\prime}\left(\lambda_1^{0\prime}\lambda_1^0/N\right)^{-1}\lambda_{1j}^0=a_{ji}$, and the last equality follows from the fact that $\frac{1}{N_kT^2}\sum_{j\in G_k^0}\left\|\lambda_{1j}^0\right\|^2\left\|F_2^{0\prime}M_{F_1^0}\mathbb{E}_{\mathcal{C}}\left\{\left[x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_j\right)\right]\left[x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)\right]'\right\}M_{F_1^0}F_2^0\right\| = O_P\left(1\right)$ by straightforward moment calculations. It follows that $U_{3kNT}^{f_2}=O_P\left(N^{-1/2}\right)=o_P\left(1\right)$. Then

$$U_{kNT}^{f_2}-B_{2kNT}=V_{3kNT}+o_P\left(1\right), \tag{B.6}$$

where $V_{3kNT}=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\left[x_{1,j}-\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)\right]'M_{F_1^0}F_2^0\lambda_{2i}^0.$

Combining (B.3), (B.5) and (B.6), we have $U_{kNT}-B_{1kNT}-B_{2kNT}=V_{kNT}+o_P\left(1\right)$, where $V_{kNT}=V_{1kNT}+V_{2kNT}+V_{3kNT}$. This completes the proof of (ii).

(iii) We have shown asymptotic independence between V_{1kNT} and V_{2kNT} conditional on $\mathcal{C}$. By the same token, we can show that V_{3kNT} is asymptotically independent of V_{1kNT} conditional on $\mathcal{C}$. Note that $V_{2kNT}=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\mathbb{E}\left(x_{1,i}'|\mathcal{C}\right)M_{F_1^0}u_i-\frac{1}{\sqrt{N_k}N}\sum_{i=1}^N\sum_{j\in G_k^0}a_{ij}\frac{1}{T}\mathbb{E}(x_{1,j}'|\mathcal{C})M_{F_1^0}u_i\equiv V_{2kNT,1}-V_{2kNT,2}$. We have

$$\begin{aligned}
&\operatorname{Cov}\left(c_1'V_{2kNT,1},c_2'V_{3kNT}|\mathcal{C}\right)\\
=&\frac{1}{N_kT^2}\sum_{i\in G_k^0}c_1'\mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)M_{F_1^0}u_i\mathbb{E}_{\mathcal{C}}\left\{\lambda_{2i}^{0\prime}F_2^{0\prime}M_{F_1^0}\left[x_{1,i}-\mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)\right]\right\}c_2\\
=&\frac{1}{N_kT^2}\sum_{i\in G_k^0}\operatorname{tr}\left\{c_2c_1'\mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)M_{F_1^0}\mathbb{E}_{\mathcal{C}}\left\{u_i\lambda_{2i}^{0\prime}F_2^{0\prime}M_{F_1^0}\left[x_{1,i}-\mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)\right]\right\}\right\}\\
=&\frac{1}{N_kT^2}\sum_{i\in G_k^0}\operatorname{tr}\left\{\left[\operatorname{vec}(c_2c_1'\mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)M_{F_1^0})\right]'\mathbb{E}_{\mathcal{C}}\left\{u_i\otimes\left[x_{1,i}-\mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)\right]'\right\}\operatorname{vec}(M_{F_1^0}F_2^0\lambda_{2i}^0)\right\},
\end{aligned}$$

which is $O_P\left(1\right)$ but not $o_P\left(1\right)$ in general unless $\operatorname{Cov}(u_{it},x_{1,it}|\mathcal{C})=0$ or $\mathbb{E}_{\mathcal{C}}\left(x_{1,it}\right)=0$, which we do not assume. Similarly,

$$\operatorname{Cov}\left(c_1'V_{2kNT,2},c_2'V_{3kNT}|\mathcal{C}\right)=\frac{1}{NN_kT^2}\sum_{i\in G_k^0}\sum_{j\in G_k^0}a_{ij}c_1'\mathbb{E}_{\mathcal{C}}(x_{1,j}')M_{F_1^0}\mathbb{E}_{\mathcal{C}}\left\{u_ic_2'\left[x_{1,i}-\mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)\right]'\right\}M_{F_1^0}F_2^0\lambda_{2i}^0,$$

which is $O_P\left(1\right)$ but not $o_P\left(1\right)$ in general. It follows that

$$\begin{aligned}
\operatorname{Var}\left(V_{kNT}|\mathcal{C}\right) &= \left\{\sum_{l=1}^3\operatorname{Var}\left(V_{lkNT}|\mathcal{C}\right)+\operatorname{Cov}\left(V_{2kNT},V_{3kNT}|\mathcal{C}\right)+\operatorname{Cov}\left(V_{2kNT},V_{3kNT}|\mathcal{C}\right)'\right\}+o_P\left(1\right)\\
&\equiv \Omega_{NT,kk}+o_P\left(1\right).
\end{aligned}$$

For any $k \neq l$, we have

$$
\begin{aligned}
\mathrm{Cov}\left(V_{kNT}, V_{lNT}|\mathcal{C}\right) =& \mathrm{Cov}\left(V_{2kNT}, V_{2lNT}|\mathcal{C}\right) \\
=& \frac{1}{N_k T^2}\mathrm{Cov}\left(\sum_{i=1}^{N}\frac{1}{N}\sum_{j\in G_k^0} a_{ij}\mathbb{E}_{\mathcal{C}}(x_{1,j}')M_{F_1^0}u_i, \sum_{i_1=1}^{N}\frac{1}{N}\sum_{j_1\in G_l^0} a_{i_1 j_1}\mathbb{E}_{\mathcal{C}}(x_{1,j_1}')M_{F_1^0}u_{i_1}|\mathcal{C}\right) \\
=& \frac{1}{N_k N^2 T^2}\sum_{i=1}^{N}\sum_{j\in G_k^0}\sum_{i_1=1}^{N}\sum_{j_1\in G_l^0} a_{ij}a_{i_1 j_1}\mathbb{E}_{\mathcal{C}}(x_{1,j}')M_{F_1^0}\mathbb{E}_{\mathcal{C}}(u_i u_{i_1}')M_{F_1^0}\mathbb{E}_{\mathcal{C}}(x_{1,j_1}') \\
=& \frac{1}{N_k N^2 T^2}\sum_{i=1}^{N}\sum_{j\in G_k^0}\sum_{j_1\in G_l^0} a_{ij}a_{ij_1}\mathbb{E}_{\mathcal{C}}(x_{1,j}')M_{F_1^0}\mathbb{E}(u_i u_i')M_{F_1^0}\mathbb{E}_{\mathcal{C}}(x_{1,j_1}') \equiv \Omega_{NT,kl},
\end{aligned}
$$

which is not vanishing unless $\mathbb{E}_{\mathcal{C}}(x_{1,it}) = 0$. Let Ω_{NT} denote that $Kp_1 \times Kp_1$ matrix with typical blocks $\Omega_{NT,kl}$ $(p_1 \times p_1)$ for $k, l = 1, ..., K$. Note that $V_{kNT} = \sum_{i=1}^{N} Z_{k,iNT}$, where

$$
\begin{aligned}
Z_{k,iNT} =& \frac{1}{\sqrt{N_k}T}S^{\varepsilon}\phi_i^{\dagger}(1)\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\bar{\varkappa}_{ts}V_{it}^{ux}v_{is}^{ux\prime} - \left[\mathbf{1}\left\{t=s\right\} - \varkappa_{ts}\mathbf{1}\left\{s\leq t\right\}\right]I_{1+p}\right\}\phi_i^{\dagger}(1)'S^{w\prime}\mathbf{1}\left\{i\in G_k^0\right\} \\
&+ \frac{1}{\sqrt{N_k}T}\left\{\mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)\mathbf{1}\left\{i\in G_k^0\right\} - \frac{1}{N}\sum_{j\in G_k^0} a_{ij}\mathbb{E}_{\mathcal{C}}(x_{1,j}')\right\}M_{F_1^0}u_i \\
&+ \frac{1}{\sqrt{N_k}T}\left[x_{1,i} - \mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)\right]' M_{F_1^0}F_2^0\lambda_{2i}^0\mathbf{1}\left\{i\in G_k^0\right\} \\
\equiv& Z_{k,iNT}(1) + Z_{k,iNT}(2) + Z_{k,iNT}(3).
\end{aligned}
$$

Let $V_{NT} = (V_{1NT}', ..., V_{KNT}')'$ and $Z_{iNT} = (Z_{1,iNT}', ..., Z_{K,iNT}')'$. Note that Z_{iNT} are independent across i conditional on $\mathcal{C}$. Let ω be a nonrandom $Kp_1 \times 1$ vector such that $\|\omega\| = 1$. By the Cramér-Wold device and the martingale CLT (e.g., Pollard (1984, p.171), we can show the asymptotic normality of V_{NT} by showing that

$$
\mathcal{Z}_1 \equiv \sum_{i=1}^{N}\mathbb{E}\left[\left|\omega' Z_{iNT}\right|^4|\mathcal{C}_{i-1,NT}\right] = o_P(1) \text{ and } \mathcal{Z}_2 \equiv \sum_{i=1}^{N}\left|\omega' Z_{iNT}\right|^2 - \omega'\Omega_{NT}\omega = o_P(1) \tag{B.7}
$$

where $\mathcal{C}_{i,NT} = \sigma\left(\mathcal{C}, \underline{x}_{1,i}, \underline{u}_i\right)$, the sigma-field generated from $\mathcal{C}$, $\underline{x}_{1,i} = (x_{1,1}, ..., x_{1,i})$ and $\underline{u}_i = (u_1, ..., u_i)$, and $\otimes_{NT} = \mathrm{Var}(V_{NT}|\mathcal{C})$ by the previous calculation and the independence of Z_{iNT} across i given $\mathcal{C}$.

We show the first claim (B.7) by the conditional Markov inequality. Let ω_k be a nonrandom $p_1 \times 1$ vector with $\|\omega_k\| \leq 1$. We can show that

$$
\mathbb{E}_{\mathcal{C}}\left(\mathcal{Z}_1\right) = \sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left[\left|\omega' Z_{iNT}\right|^4|\right] = o_P(1)
$$

by showing that $\sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left[\left|\omega_k' Z_{k,iNT}(l)\right|^4|\right] = o_P(1)$ for $k = 1, ..., K$ and $l = 1, 2, 3$. We only show that $\sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left[\left|\omega_k' Z_{k,iNT}(3)\right|^4|\right] = o_P(1)$ for each $k \in \{1, ..., K\}$ since we can show $\sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left[\left|\omega_k' Z_{k,iNT}(l)\right|^4|\right] =$

$o_P(1)$ for $l=1,2$ analogously. By the BN decomposition in (B.4), $x_{it}^{\dagger} \equiv x_{it} - \mathbb{E}_{\mathcal{C}}(x_{it}) = \sum_{s=1}^{t} w_{is}^{ux} = \phi_i^{\dagger}(1) V_{it}^{ux} + \tilde{w}_{i0}^{ux} - \tilde{w}_{it}^{ux}$. Noting that

$$Z_{k,iNT}(3) = \frac{1}{\sqrt{N_k}T}\left[x_{1,i} - \mathbb{E}_{\mathcal{C}}(x_{1,i})\right]' M_{F_1^0} F_2^0 \lambda_{2i}^0 \mathbf{1}\left\{i \in G_k^0\right\} = \frac{1}{\sqrt{N_k}T}\sum_{t=1}^{T}\sum_{s=1}^{T} x_{it}^{\dagger}\overline{\chi}_{ts} f_{2s}^{0\prime}\lambda_{2i}^0 \mathbf{1}\left\{i \in G_k^0\right\},$$

we have

$$\begin{aligned}
&\sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left[\left|\omega_k' Z_{k,iNT}(3)\right|^4 |\right] \\
=&\frac{1}{N_k^2 T^4}\sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left[\omega_k' Z_{k,iNT}(3) Z_{k,iNT}(3)'\omega_k\omega_k' Z_{k,iNT}(3) Z_{k,iNT}(3)'\omega_k\right] \\
=&\frac{1}{N_k^2 T^4}\sum_{i\in G_k^0}\sum_{t_1,\ldots,t_4}\sum_{s_1,\ldots,s_4}\overline{\chi}_{t_1s_1}\overline{\chi}_{t_2s_2}\overline{\chi}_{t_3s_3}\overline{\chi}_{t_4s_4} f_{2s_1}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2s_2}^0 f_{2s_3}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2s_4}^0\mathbb{E}_{\mathcal{C}}\left[\omega_k' x_{it_1}^{\dagger} x_{it_2}^{\dagger\prime}\omega_k\ \omega_k' x_{it_3}^{\dagger} x_{it_4}^{\dagger\prime}\omega_k\right] \\
=&\frac{1}{N_k^2 T^4}\sum_{i\in G_k^0}\sum_{t_1,\ldots,t_4}\sum_{s_1,\ldots,s_4}\overline{\chi}_{t_1s_1}\overline{\chi}_{t_2s_2}\overline{\chi}_{t_3s_3}\overline{\chi}_{t_4s_4} f_{2s_1}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2s_2}^0 f_{2s_3}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2s_4}^0\mathbb{E}_{\mathcal{C}}\mathrm{tr}\left[\omega_k\omega_k' x_{it_1}^{\dagger} x_{it_2}^{\dagger\prime}\omega_k\ \omega_k' x_{it_3}^{\dagger} x_{it_4}^{\dagger\prime}\right] \\
=&\frac{1}{N_k^2 T^4}\sum_{i\in G_k^0}\sum_{t_1,\ldots,t_4}\sum_{s_1,\ldots,s_4}\overline{\chi}_{t_1s_1}\overline{\chi}_{t_2s_2}\overline{\chi}_{t_3s_3}\overline{\chi}_{t_4s_4} f_{2s_1}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2s_2}^0 f_{2s_3}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2s_4}^0 \\
&\times\left[\mathrm{vec}\left(\omega_k\omega_k'\right)\right]'\mathbb{E}_{\mathcal{C}}\left[\left(x_{it_1}^{\dagger} x_{it_2}^{\dagger\prime}\right)\otimes\left(x_{it_3}^{\dagger} x_{it_4}^{\dagger\prime}\right)\right]\mathrm{vec}\left(\omega_k\omega_k'\right).
\end{aligned}$$

Using the BN decomposition for $x_{it}^{\dagger}$, we can show that the last term has dominant term given by

$$\begin{aligned}
&\frac{1}{N_k^2 T^4}\sum_{i\in G_k^0}\sum_{1\le t_1,s_1\le T}\sum_{1\le t_2,s_2\le T}\sum_{1\le t_3,s_3\le T}\sum_{1\le t_4,s_4\le T}\overline{\chi}_{t_1s_1}\overline{\chi}_{t_2s_2}\overline{\chi}_{t_3s_3}\overline{\chi}_{t_4s_4} f_{2s_1}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2s_2}^0 f_{2s_3}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2s_4}^0 \\
&\times\left[\mathrm{vec}\left(\omega_k\omega_k'\right)\right]'\mathbb{E}\left[\left(\phi_i^{\dagger}(1) V_{it_1}^{ux} V_{it_2}^{ux\prime}\phi_i^{\dagger}(1)'\right)\otimes\left(\phi_i^{\dagger}(1) V_{it_3}^{ux} V_{it_4}^{ux\prime}\phi_i^{\dagger}(1)'\right)\right]\mathrm{vec}\left(\omega_k\omega_k'\right) \\
\le&\frac{M}{N_k^2 T^4}\sum_{i\in G_k^0}\sum_{t_1,t_2,t_3,t_4} f_{2t_1}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2t_2}^0 f_{2t_3}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2t_4}^0\left[\mathrm{vec}\left(\omega_k\omega_k'\right)\right]' \\
&\times\mathbb{E}\left[\left(\phi_i^{\dagger}(1) V_{it_1}^{ux} V_{it_2}^{ux\prime}\phi_i^{\dagger}(1)'\right)\otimes\left(\phi_i^{\dagger}(1) V_{it_3}^{ux} V_{it_4}^{ux\prime}\phi_i^{\dagger}(1)'\right)\right]\mathrm{vec}\left(\omega_k\omega_k'\right) \\
&+\frac{M}{N_k^2 T^4}\sum_{i\in G_k^0}\sum_{t_1,s_1}\sum_{t_2,s_2}\sum_{t_3,s_3}\sum_{t_4,s_4}\chi_{t_1s_1}\chi_{t_2s_2}\chi_{t_3s_3}\chi_{t_4s_4} f_{2s_1}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2s_2}^0 f_{2s_3}^{0\prime}\lambda_{2i}^0\lambda_{2i}^{0\prime} f_{2s_4}^0 \\
&\times\left[\mathrm{vec}\left(\omega_k\omega_k'\right)\right]'\mathbb{E}\left[\left(\phi_i^{\dagger}(1) V_{it_1}^{ux} V_{it_2}^{ux\prime}\phi_i^{\dagger}(1)'\right)\otimes\left(\phi_i^{\dagger}(1) V_{it_3}^{ux} V_{it_4}^{ux\prime}\phi_i^{\dagger}(1)'\right)\right]\mathrm{vec}\left(\omega_k\omega_k'\right),
\end{aligned}$$

where M is a generic constant that can vary across lines and the inequality follows from the Chebyshev inequality. One can readily show that the first term is $O_P\left(N^{-1}\right)$. For the second term, noting that

$\varkappa_{ts} = f_{1t}^{0\prime}\left(F_1^{0\prime}F_1^0\right)^{-1} f_{1s}^0$, it is bounded from above by

$$
\begin{aligned}
& M\left\|\left(\frac{F_1^{0\prime}F_1^0}{T^2}\right)^{-1}\right\|^4 \left\|\frac{1}{T}\sum_{s=1}^{T} f_{1s}^0 f_{2s}^{0\prime}\right\|^4 \frac{1}{N_k^2T^8}\sum_{i\in G_k^0}\sum_{t_1,t_2,t_3,t_4}\|f_{1t_1}^0\|\,\|f_{1t_2}^0\|\,\|f_{1t_3}^0\|\,\|f_{1t_4}^0\|\,\|\lambda_{2i}^0\|^4 \\
& \times\left\|\mathbb{E}\left[\left(\phi_i^\dagger(1)V_{it_1}^{ux}V_{it_2}^{ux\prime}\phi_i^\dagger(1)'\right)\otimes\left(\phi_i^\dagger(1)V_{it_3}^{ux}V_{it_4}^{ux\prime}\phi_i^\dagger(1)'\right)\right]\right\| \\
=& O_P(1)\,O_P(1)\,O_P\left(N^{-1}\right) = O_P\left(N^{-1}\right),
\end{aligned}
$$

where the last line follows because by Jensen's inequality and the independence between $\left\{f_{1t}^0\right\}$ and $\left\{\lambda_{2i}^0\right\}$,

$$
\begin{aligned}
& \frac{1}{N_k^2T^8}\sum_{i\in G_k^0}\sum_{t_1,\dots,t_4}\mathbb{E}\{\|f_{1t_1}^0\|\,\|f_{1t_2}^0\|\,\|f_{1t_3}^0\|\,\|f_{1t_4}^0\|\,\|\lambda_{2i}^0\|^4\,\|\,\mathbb{E}[\{\phi_i^\dagger(1)V_{it_1}^{ux}V_{it_2}^{ux\prime}\phi_i^\dagger(1)'\}\otimes\{\phi_i^\dagger(1)V_{it_3}^{ux}V_{it_4}^{ux\prime}\phi_i^\dagger(1)'\}]\|\|\} \\
\leq & \frac{M}{N_k^2T^8}\sum_{i\in G_k^0}\mathbb{E}\left[\|\lambda_{2i}^0\|^4\right] \\
& \times\sum_{t_1,t_2,t_3,t_4}\left\{\mathbb{E}\left\|f_{1t_1}^0\right\|^4\,\mathbb{E}\left\|f_{1t_2}^0\right\|^4\,\mathbb{E}\left\|f_{1t_3}^0\right\|^4\mathbb{E}\left\|f_{1t_4}^0\right\|^4\,\mathbb{E}\left\|V_{it_1}^{u\varepsilon}\right\|^4\,\mathbb{E}\left\|V_{it_2}^{u\varepsilon}\right\|^4\mathbb{E}\left\|V_{it_3}^{u\varepsilon}\right\|^4\,\mathbb{E}\left\|V_{it_4}^{u\varepsilon}\right\|^4\right\}^{1/4} \\
\leq & \frac{M}{N_k^2}\sum_{i\in G_k^0}\mathbb{E}\left[\|\lambda_{2i}^0\|^4\right]\left[\frac{1}{T^2}\sum_{t=1}^{T}\left\{\mathbb{E}[\|f_{1t}^0\|^4]\,\mathbb{E}[\|V_{it}^{u\varepsilon}\|^4]\right\}^{1/4}\right]^4 \\
\leq & \frac{M}{N_k^2}\sum_{i\in G_k^0}\mathbb{E}\left[\|\lambda_{2i}^0\|^4\right]\left[\frac{1}{T^2}\sum_{t=1}^{T}t\right]^4 = O\left(N^{-1}\right)
\end{aligned}
$$

and the last inequality follows from the fact that $\mathbb{E}[\left\|f_{1t}^0\right\|^4] \leq Mt^2$ and $\mathbb{E}[\|V_{it}^{u\varepsilon}\|^4] \leq Mt^2$. Consequently, we have shown that $\sum_{i=1}^N \mathbb{E}_{\mathcal{C}}\left[|\omega_k' Z_{k,iNT}(3)|^4\right] = o_P(1)$ for each $k = 1, ..., K$. Analogously, we can show that $\sum_{i=1}^N \mathbb{E}_{\mathcal{C}}\left[|\omega_k' Z_{k,iNT}(l)|^4\right] = o_P(1)$ for each $k = 1, ..., K$ and $l = 1, 2$. As a result, the first claim in (B.7) follows.

To show the second claim in (B.7), we first observe that $\mathbb{E}_{\mathcal{C}}(\mathcal{Z}_2) \equiv \omega'\sum_{i=1}^N \mathbb{E}_{\mathcal{C}}(Z_{iNT}Z_{iNT}')\,\omega - \omega'\otimes_{NT}\omega = 0$. The claim follows if we can show that $\mathrm{Var}(\mathcal{Z}_2|\mathcal{C}) \equiv \mathrm{Var}\left(\sum_{i=1}^N |\omega' Z_{iNT}|^2\,|\mathcal{C}\right) = o_P(1)$. By the independence of Z_{iNT} over i conditional on $\mathcal{C}$, we have $\mathrm{Var}(\mathcal{Z}_2|\mathcal{C}) = \sum_{i=1}^N \mathrm{Var}\left(|\omega' Z_{iNT}|^2\,|\mathcal{C}\right) \leq \sum_{i=1}^N \mathbb{E}_{\mathcal{C}}\left(|\omega' Z_{iNT}|^4\right) = o_P(1)$, where the last equality follows from in first claim in (B.7). Consequently, we have $V_{NT} \xrightarrow{d} N(0, \Omega_0)$ conditional on $\mathcal{C}$, where $\Omega_0 = \lim_{(N,T)\to\infty}\Omega_{NT}$. ■

Proof of Lemma A.11. (i) We first study $\|\hat{\lambda}_{1i} - H_1^{-1}\lambda_{1i}^0\|$. Noting that $\hat{\lambda}_{1i} = (\hat{F}_1'\hat{F}_1)^{-1}\hat{F}_1' e_i = \frac{1}{T^2}\hat{F}_1' e_i$ with $e_i \equiv y_i - x_i\hat{\beta}_i = \hat{F}_1 H_1^{-1}\lambda_{1i}^0 + (F_1^0 - \hat{F}_1 H_1^{-1})\lambda_{1i}^0 + F_2^0\lambda_{2i}^0 + u_i - x_{1,i}\hat{b}_{1,i} - x_{2,i}\hat{b}_{2,i}$, we have

$$
\left\|\hat{\lambda}_{1i} - H_1^{-1}\lambda_{1i}^0\right\| \leq \left\|\frac{1}{T^2}\hat{F}_1'(F_1^0 - \hat{F}_1 H_1^{-1})\lambda_{1i}^0\right\| + \left\|\frac{1}{T^2}\hat{F}_1'(u_i + F_2^0\lambda_{2i}^0 - x_{2,i}\hat{b}_{2,i})\right\| + \left\|\frac{1}{T^2}\hat{F}_1' x_{1,i}\hat{b}_{1,i}\right\| \equiv \sum_{l=1}^{3} E_{li}.
$$

By Lemma A.5(iii), we have $E_{1i} \leq \frac{1}{T^2}\left\|\hat{F}_1'(F_1^0 - \hat{F}_1 H_1^{-1})\right\| \|\lambda_{1i}^0\| = O_P(\eta_{1NT} + T^{-1}\delta_{NT}^{-1} + T^{-5/4}C_{NT}^{-1/2})$. For E_{2i}, we have $E_{2i} \leq \frac{1}{T^2}\left\|H_1' F_1^{0\prime}(u_i + F_2^0\lambda_{2i}^0 - x_{2,i}\hat{b}_{2,i})\right\| + \frac{1}{T^2}\left\|(\hat{F}_1 - F_1^0 H_1)'(u_i + F_2^0\lambda_{2i}^0 - x_{2,i}\hat{b}_{2,i})\right\| = O_P(T^{-1})$. In addition, we can show that $E_{3i} \leq \frac{1}{T^2}\left\|\hat{F}_1' x_{1,i}\hat{b}_{1,i}\right\| \leq \frac{1}{T^2}\left\|H_1' F_1^{0\prime} x_{1,i}\hat{b}_{1,i}\right\| + \frac{1}{T^2}\left\|(\hat{F}_1 - F_1^0 H_1)' x_{1,i}\hat{b}_{1,i}\right\|$ $= O_P(\eta_{1NT})$. It follows that $\|\hat{\lambda}_{1i} - H_1^{-1}\lambda_{1i}^0\| = O_P(\eta_{1NT} + T^{-1})$. Let $\Delta\hat{C}_{1,i} = \hat{F}_1\hat{\lambda}_{1i} - F_1^0\lambda_{1i}^0$. Then by the triangle inequality

$$\begin{aligned}
\|\Delta C_{1,i}\| &= \left\|\hat{F}_1\hat{\lambda}_{1i} - F_1^0\lambda_{1i}^0\right\| \\
&\leq \left\|(\hat{F}_1 H_1^{-1} - F_1^0)\lambda_{1i}^0\right\| + \left\|F_1^0 H_1(\hat{\lambda}_{1i} - H_1^{-1}\lambda_{1i}^0)\right\| + \left\|(\hat{F}_1 - F_1^0 H_1)(\hat{\lambda}_{1i} - H_1^{-1}\lambda_{1i}^0)\right\| \equiv \sum_{l=1}^{3}\|c_{li}\|.
\end{aligned}$$

It is easy to show that

$$\begin{aligned}
\frac{\|c_{1i}\|}{\sqrt{T}} &\leq \frac{1}{\sqrt{T}}\left\|\hat{F}_1 H_1^{-1} - F_1^0\right\| \|\lambda_{1i}^0\| = O_P(\sqrt{T}\eta_{1NT} + C_{NT}^{-1}), \\
\frac{\|c_{2i}\|}{\sqrt{T}} &\leq \frac{\|F_1^0 H_1\|}{T}\sqrt{T}\left\|\hat{\lambda}_{1i} - H_1^{-1}\lambda_{1i}^0\right\| = O_P(\sqrt{T}\eta_{1NT} + T^{-1/2}), \\
\frac{\|c_{3i}\|}{\sqrt{T}} &\leq \frac{1}{\sqrt{T}}\left\|\hat{F}_1 - F_1^0 H_1\right\| \left\|\hat{\lambda}_{1i} - H_1^{-1}\lambda_{1i}^0\right\| = O_P(\sqrt{T}\eta_{1NT}^2 + C_{NT}^{-1}\eta_{1NT}).
\end{aligned}$$

Consequently, we have $\frac{1}{\sqrt{T}}\|\hat{F}_1\hat{\lambda}_{1i} - F_1^0\lambda_{1i}^0\| = O_P(\sqrt{T}\eta_{1NT} + C_{NT}^{-1})$.

(ii) Given the fast convergence rate of $\hat{F}_1$ and $\hat{\alpha}_k$ and the established convergence rate of $\hat{\beta}_{2,i} = O_P(\sqrt{d_T}T^{-1/2})$ and $\hat{\lambda}_{1i}$ in part (i), the result follows from standard factor analysis. We also assume the stationary regressors are uncorrelated with the stationary common factors and factor loadings. Here, we only sketch the proof. Recall that $\hat{F}_2$ satisfies the following equation: $[\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}(y_i - x_{1,i}\hat{\alpha}_k - x_{2,i}\hat{\beta}_{2,i} - \hat{F}_1\hat{\lambda}_{1i})(y_i - x_{1,i}\hat{\alpha}_k - x_{2,i}\hat{\beta}_{2,i} - \hat{F}_1\hat{\lambda}_{1i})']\hat{F}_2 = \hat{F}_2 V_{2,NT}$. Note that $y_i - x_{1,i}\hat{\alpha}_k - x_{2,i}\hat{\beta}_{2,i} - \hat{F}_1\hat{\lambda}_{1i} = F_2^0\lambda_{2i}^0 + u_i - x_{1,i}(\hat{\alpha}_k - \alpha_k^0) - x_{2,i}\hat{b}_{2,i} + \Delta\hat{C}_{1,i} = F_2^0\lambda_{2i}^0 + \tilde{u}_i - x_{1,i}\hat{a}_k$, where $\tilde{u}_i = u_i - \Delta\hat{C}_{1,i} - x_{2,i}\hat{b}_{2,i}$ and $\hat{a}_k = \hat{\alpha}_k - \alpha_k^0$. Then by the proof of Theorem 3.1(iii), we have the following decomposition

$$\begin{aligned}
\hat{F}_2 V_{2,NT} =& \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k} x_{1,i}\hat{a}_k\hat{a}_k' x_{1,i}'\hat{F}_2 - \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k} x_{1,i}\hat{a}_k\lambda_{2i}^{0\prime}F_2^{0\prime}\hat{F}_2 - \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k} x_{1,i}\hat{a}_k\tilde{u}_i'\hat{F}_2 \\
& - \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k} F_2^0\lambda_{2i}^0\hat{a}_k' x_{1,i}'\hat{F}_2 - \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}\tilde{u}_i\hat{a}_k' x_{1,i}'\hat{F}_2 + \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k} F_2^0\lambda_{2i}^0\tilde{u}_i'\hat{F}_2 \\
& + \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}\tilde{u}_i\lambda_{2i}^{0\prime}F_2^{0\prime}\hat{F}_2 + \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}\tilde{u}_i\tilde{u}_i'\hat{F}_2 + \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k} F_2^0\lambda_{2i}^0\lambda_{2i}^{0\prime}F_2^{0\prime}\hat{F}_2 \\
& \equiv \tilde{L}_1 + ... + \tilde{L}_8 + \frac{1}{NT}\sum_{i=1}^{N} F_2^0\lambda_{2i}^0\lambda_{2i}^{0\prime}F_2^{0\prime}\hat{F}_2.
\end{aligned}$$

It follows that $\frac{1}{\sqrt{T}}\left\|\hat{F}_2 - F_2^0 H_2\right\| = \frac{1}{\sqrt{T}}\left(\left\|\tilde{L}_1\right\| + ... + \left\|\tilde{L}_8\right\|\right) V_{2,NT}^{-1}$, where $H_2 = (\frac{1}{N}\Lambda_2^{0\prime}\Lambda_2^0)(\frac{1}{T}F_2^{0\prime}\hat{F}_2)V_{2,NT}^{-1}$

$= O_P(1)$. One can readily show that

$$\begin{aligned}\frac{1}{\sqrt{T}}\|\tilde{L}_1\| \leq & T \max_i \frac{\|x_{1,i}\|^2}{T^2}\frac{\|\hat{F}_2\|}{\sqrt{T}}\frac{1}{N}\sum_{k=1}^{K}\|\hat{a}_k\|^2 = O_P(d_T T N^{-1} K \varrho_{NT}^2),\\ \frac{1}{\sqrt{T}}\|\tilde{L}_2\| \leq & \sqrt{\frac{TK}{N}} \max_i \frac{\|x_{1,i}\|}{T}\left\|\frac{F_2^{0\prime}\hat{F}_2}{T}\right\|\left\{\frac{1}{N}\sum_{i=1}^{N}\|\lambda_{2i}\|^2\right\}^{1/2}\left\{\frac{1}{K}\sum_{k=1}^{K}\|\hat{a}_k\|^2\right\}^{1/2}\\ & = O_P(T^{1/2}N^{-1/2}d_T^{1/2}\varrho_{NT}),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{\sqrt{T}}\|\tilde{L}_3\| & \leq \sqrt{\frac{TK}{N}} \max_i \frac{\|x_{1,i}\|}{T}\frac{\|\hat{F}_2\|}{\sqrt{T}}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\|\tilde{u}_i\|^2}{T}\right\}^{1/2}\left\{\frac{1}{K}\sum_{k=1}^{K}\|\hat{a}_k\|^2\right\}^{1/2}\\ & = O_P(T^{1/2}N^{-1/2}\sqrt{d_T}\varrho_{NT}).\end{aligned}$$

Analogously, we have $\frac{1}{\sqrt{T}}\|\tilde{L}_4\| = O_P(T^{1/2}N^{-1/2}\varrho_{NT})$ and $\frac{1}{\sqrt{T}}\|\tilde{L}_5\| = O_P(T^{1/2}N^{-1/2}\sqrt{d_T}\varrho_{NT})$. Since the remaining terms do not involve $\hat{\alpha}_k$, we can directly obtain from Bai and Ng (2002) that $\frac{1}{\sqrt{T}}\|\tilde{L}_{li}\| = O_P(C_{NT}^{-1})$ for $l = 6, 7, 8$. By the fact that $\hat{\alpha}_k$ are the group-specific C-Lasso estimators, we have $\varrho_{NT}^2 = \frac{1}{K}\sum_{k=1}^{K}\left\|\hat{\alpha}_k - \alpha_k^0\right\|^2 = O_P(N^{-1}T^{-2})$ before bias correction and $\varrho_{NT} = O_P(N^{-1/2}T^{-1})$ after bias correction. Thus, we have $T^{1/2}N^{-1/2}\sqrt{d_T}\varrho_{NT} = o(C_{NT}^{-1})$ and $\sqrt{d_T}TN^{-1}\varrho_{NT}^2 = o(C_{NT}^{-1})$. Consequently, $\frac{1}{\sqrt{T}}\left\|\hat{F}_2 - F_2^0 H_2\right\| = O_P\left(C_{NT}^{-1}\right)$.

(iii) Noting that $\hat{\lambda}_{2i} - H_2^{-1}\lambda_{2i}^0 = \frac{1}{T}\hat{F}_2'[(F_2^0 - \hat{F}_2 H_2^{-1})\lambda_{2i}^0 + \tilde{u}_i - x_{1,i}\hat{a}_k]$, we have

$$\begin{aligned}\frac{1}{\sqrt{N_k}}\sum_{i\in\hat{G}_k}\left(\hat{\lambda}_{2i} - H_2^{-1}\lambda_{2i}^0\right) & \leq \frac{1}{\sqrt{N_k}}\sum_{i\in\hat{G}_k}\frac{1}{T}\hat{F}_2'(F_2^0 - \hat{F}_2 H_2^{-1})\lambda_{2i}^0 + \frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}\hat{F}_2'\tilde{u}_i - \frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}\hat{F}_2' x_{1,i}\hat{a}_k\\ & \equiv \tilde{E}_1 + \tilde{E}_2 - \tilde{E}_3, \text{ say.}\end{aligned}$$

For $\tilde{E}_1$, we have $\left\|\tilde{E}_1\right\| = \frac{1}{T}\left\|\hat{F}_2'(F_2^0 - \hat{F}_2 H_2^{-1})\right\|\frac{1}{\sqrt{N_k}}\left\|\sum_{i\in\hat{G}_k}\lambda_{2i}^0\right\| = O_P(C_{NT}^{-2})O_P(\sqrt{N_k}) = o_P(1)$, where we use the fact that $T^{-1}\hat{F}_2'(F_2^0 - \hat{F}_2 H_2^{-1}) = O_P(C_{NT}^{-2})$. For $\tilde{E}_2$, we make the decomposition: $\tilde{E}_2 = \frac{1}{T\sqrt{N_k}}\sum_{i\in\hat{G}_k}H_2' F_2^{0\prime}\tilde{u}_i + \frac{1}{T\sqrt{N_k}}\sum_{i\in\hat{G}_k}(\hat{F}_2 - F_2^0 H_2)'\tilde{u}_i \equiv \tilde{E}_{2,1} + \tilde{E}_{2,2}$. Using $\tilde{u}_i = u_i - \Delta\hat{C}_{1,i} - x_{2,i}\hat{b}_{2,i}$, we can readily show that

$$\begin{aligned}\tilde{E}_{2,1} & = H_2'\frac{1}{T\sqrt{N_k}}\sum_{i\in\hat{G}_k}F_2^{0\prime}u_i - H_2'\frac{1}{T\sqrt{N_k}}\sum_{i\in\hat{G}_k}F_2^{0\prime}\Delta\hat{C}_{1,i} - H_2'\frac{1}{T\sqrt{N_k}}\sum_{i\in\hat{G}_k}F_2^{0\prime}x_{2,i}\hat{b}_{2,i}\\ & = O_P(T^{-1/2}) + o_P(1) + O_P(T^{-1/2}) = o_P(1).\end{aligned}$$

Using the decomposition of $\hat{F}_2 - F_2^0 H_2$ in (ii), we can readily show that $\tilde{E}_{2,2} = o_P(1)$. Then $\tilde{E}_2 =$

$o_P(1)$. For $\tilde{E}_3$, we have

$$\left\|\tilde{E}_3\right\| \leq \left\{\left\|\frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k} H_2'F_2^{0\prime}x_{1,i}\right\| + \left\|\frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}(\hat{F}_2 - F_2^0H_2)'x_{1,i}\right\|\right\}\|\hat{a}_k\|$$
$$= \left\{O_P(\sqrt{N_k}) + o_P(\sqrt{N_k})\right\}O_P(\varrho_{NT}) = o_P(1).$$

It follows that $\frac{1}{\sqrt{N_k}}\sum_{i\in\hat{G}_k}(\hat{\lambda}_{2i} - H_2^{-1}\lambda_{2i}^0) = o_P(1)$.

(iv) Noting that

$$\left\|\hat{\lambda}_{2i} - H_2^{-1}\lambda_{2i}^0\right\| \leq \left\|\frac{1}{T}\hat{F}_2'\left(F_2^0 - \hat{F}_2H_2^{-1}\right)\lambda_{2i}^0\right\| + \left\|\frac{1}{T}\hat{F}_2'\tilde{u}_i\right\| + \left\|\frac{1}{T}\hat{F}_2'x_{1,i}\hat{a}_k\right\|$$
$$\leq \left\|\frac{1}{T}\hat{F}_2'\left(F_2^0 - \hat{F}_2H_2^{-1}\right)\right\|\|\lambda_{2i}^0\| + T^{-1/2}\left\|\frac{1}{\sqrt{T}}\hat{F}_2'\tilde{u}_i\right\| + \left\|\frac{1}{T}\hat{F}_2'x_{1,i}\right\|\|\hat{a}_k\|$$
$$= O_P(C_{NT}^{-2}) + O_P(T^{-1/2}) + O_P(\varrho_{NT}) = O_P(C_{NT}^{-1}),$$

we have

$$\frac{1}{\sqrt{T}}\left\|\hat{F}_2'\hat{\lambda}_{2i} - F_2^0\lambda_{2i}^0\right\| \leq \frac{1}{\sqrt{T}}\left\|\left(\hat{F}_2H_2^{-1} - F_2^0\right)\lambda_{2i}^0\right\| + \frac{1}{\sqrt{T}}\left\|F_2^0H_2\left(\hat{\lambda}_{2i} - H_2^{-1}\lambda_{2i}^0\right)\right\|$$
$$+ \frac{1}{\sqrt{T}}\left\|\left(\hat{F}_2 - F_2^0H_2\right)\left(\hat{\lambda}_{2i} - H_2^{-1}\lambda_{2i}^0\right)\right\|$$
$$= O_P(C_{NT}^{-1})O_P(1) + O_P(1)O_P(C_{NT}^{-1}) + O_P(C_{NT}^{-1})O_P(C_{NT}^{-1}) = O_P(C_{NT}^{-1}).$$

(v) Note that $\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{21,i} - \Delta_{21,i}) = \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}[\hat{\Delta}_{21,i} - \mathbb{E}_{\mathcal{C}}(\hat{\Delta}_{21,i})] + \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}[\mathbb{E}_{\mathcal{C}}(\hat{\Delta}_{21,i}) - \Delta_{21,i}] \equiv d_{1NT} + d_{2NT}$. Following the proof of Theorem 9 of Phillips and Moon (1999), we can show that

$$\mathbb{E}_{\mathcal{C}}\|d_{1NT}\|^2 = \frac{1}{N_k}\sum_{i\in G_k^0}\mathbb{E}_{\mathcal{C}}\left\|\hat{\Delta}_{21,i} - \mathbb{E}_{\mathcal{C}}(\hat{\Delta}_{21,i})\right\|^2 + o_P(1) = O_P(J/T) + o_P(1) = o_P(1),$$

and $\mathbb{E}_{\mathcal{C}}\left(d_{1NT}^2\right) \leq \sum_{i\in G_k^0}\left\|\mathbb{E}_{\mathcal{C}}(\hat{\Delta}_{21,i}) - \Delta_{21,i}\right\|^2 = O_P\left(N/J^{2q}\right) = o_P(1)$. It follows that $\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{21,i} - \Delta_{21,i}) = o_P(1)$.

(vi) We first obtain the rough probability bound. By Jensen's inequality and Lemma A.5(iii), we have

$$\left|\frac{\sqrt{N_k}}{T}\sum_{t=1}^T\sum_{s=1}^T(\hat{\varkappa}_{ts} - \varkappa_{ts})\mathbf{1}\{s\leq t\}\right| \leq \sqrt{N_k}\left\{\sum_{t=1}^T\sum_{s=1}^T(\hat{\varkappa}_{ts} - \varkappa_{ts})^2\right\}^2$$
$$= \sqrt{N_k}\|P_{\hat{F}_1} - P_{F_1^0}\| = O_P(\{\sqrt{d_T}\eta_{1NT}N_k + N_kT^{-1}\delta_{NT}^{-1}\}^{1/2}).$$

Note that $\eta_{1NT} = T^{-1}$ before bias correction, the last term may not be $o_P(1)$ under our conditions on (N,T). To obtain the tight probability bound, we make the decomposition:

$$
\begin{aligned}
&\frac{\sqrt{N_k}}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{\varkappa}_{ts}-\varkappa_{ts}\right)\mathbf{1}\left\{s\le t\right\}\\
=&\frac{\sqrt{N_k}}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\frac{1}{T^2}\hat{f}_{1t}'\hat{f}_{1s}-f_{1t}^{0\prime}\left(F_1^{0\prime}F_1^0\right)^{-1}f_{1s}^0\right]\mathbf{1}\left\{s\le t\right\}\\
=&\frac{\sqrt{N_k}}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\frac{1}{T^2}f_{1t}^{0\prime}\left(H_1H_1'-\left(\frac{1}{T^2}F_1^{0\prime}F_1^0\right)^{-1}\right)f_{1s}^0+\frac{1}{T^2}\left(\hat{f}_{1t}-H_1'f_{1t}^0\right)'H_1'f_{1s}^0\right.\\
&\left.+\frac{1}{T^2}f_{1t}^{0\prime}H_1\left(\hat{f}_{1s}-H_1'f_{1s}^0\right)+\frac{1}{T^2}\left(\hat{f}_{1t}-H_1'f_{1t}^0\right)'\left(\hat{f}_{1s}-H_1'f_{1s}^0\right)\right\}\mathbf{1}\left\{s\le t\right\}\\
\equiv&d_{3NT,1}+d_{3NT,2}+d_{3NT,3}+d_{3NT,4},\text{ say.}
\end{aligned}
$$

Following the proof of Lemma A.5(i), we can readily show that $d_{3NT,l}=o_P(1)$ for $l=1,2,3,4$. Then $\frac{\sqrt{N_k}}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{\varkappa}_{ts}-\varkappa_{ts}\right)\mathbf{1}\left\{s\le t\right\}=o_P\left(1\right).$

(vii) We first make the following decomposition

$$
\begin{aligned}
&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i}-\Delta_{24,i}\bar{\lambda}_{2i}^0)\\
=&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i}-\Delta_{24,i}H_2H_2^{-1}\bar{\lambda}_{2i}^0)\\
=&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{24,i}-\Delta_{24,i}H_2)(\hat{\bar{\lambda}}_{2i}-H_2^{-1}\bar{\lambda}_{2i}^0)\\
&+\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{24,i}-\Delta_{24,i}H_2)H_2^{-1}\bar{\lambda}_{2i}^0+\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\Delta_{24,i}H_2(\hat{\bar{\lambda}}_{2i}-H_2^{-1}\bar{\lambda}_{2i}^0)\equiv I_{11}+I_{12}+I_{13}.
\end{aligned}
$$

Following the proof of Theorem 9 of Phillips and Moon (1999), we can show that $\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\hat{\Delta}_{24,i}-\Delta_{24,i}H_2\right\|^2$ $=O_P(\frac{J}{T}+\frac{1}{J^{2q}})$ and $I_{12}=o_P\left(1\right)$. By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\hat{\bar{\lambda}}_{2i}-H_2^{-1}\bar{\lambda}_{2i}^0\right\|^2\\
\le&\frac{2}{N_k}\sum_{i\in G_k^0}\left\|\hat{\lambda}_{2i}-H_2^{-1}\lambda_{2i}^0\right\|^2+\frac{2}{N_k}\sum_{i\in G_k^0}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\hat{\lambda}_{2j}\hat{a}_{ij}-H_2^{-1}\lambda_{2j}^0a_{ij}\right)\right\|^2\\
\le&\frac{2}{N_k}\sum_{i\in G_k^0}\left\|\hat{\lambda}_{2i}-H_2^{-1}\lambda_{2i}^0\right\|^2+\frac{4}{N_k}\sum_{i\in G_k^0}\left\|\frac{1}{N}\sum_{j=1}^{N}\hat{\lambda}_{2j}\left(\hat{a}_{ij}-a_{ij}\right)\right\|^2+\frac{4}{N_k}\sum_{i\in G_k^0}\left\|\frac{1}{N}\sum_{j=1}^{N}(\hat{\lambda}_{2j}-H_2^{-1}\lambda_{2j}^0)a_{ij}\right\|^2.
\end{aligned}
$$

Following the analysis in (iii), we can readily show that

$$
\begin{aligned}
&\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\hat{\lambda}_{2i}-H_2^{-1}\lambda_{2i}^0\right\|^2\\
\leq&\frac{3}{N_k}\sum_{i\in G_k^0}\left\|\frac{1}{T}\hat{F}_2'(F_2^0-\hat{f}_2H_2^{-1})\lambda_{2i}^0\right\|^2+\frac{3}{N_k}\sum_{i\in G_k^0}\left\|\frac{1}{T}\hat{F}_2'\tilde{u}_i\right\|^2+\frac{3}{N_k}\sum_{i\in G_k^0}\left\|\frac{1}{T}\hat{F}_2'x_i\hat{a}_k\right\|^2\\
=&O_P(C_{NT}^{-2})+O_P(T^{-1})+O_P(\varrho_{NT}^2)=O_P(C_{NT}^{-2}).
\end{aligned}
$$

Similarly, by the Cauchy-Schwarz inequality, $\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\frac{1}{N}\sum_{j=1}^N(\hat{\lambda}_{2j}-H_2^{-1}\lambda_{2j}^0)a_{ij}\right\|^2\leq\frac{1}{N}\sum_{j=1}^N\left\|\hat{\lambda}_{2j}-H_2^{-1}\lambda_{2j}^0\right\|^2\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j=1}^N a_{ij}^2=O_P(C_{NT}^{-2})$. In addition, we can show that $\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j=1}^N\|\hat{a}_{ij}-a_{ij}\|^2=O_P(C_{NT}^{-2})$, which implies that $\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\frac{1}{N}\sum_{j=1}^N\hat{\lambda}_{2j}\left(\hat{a}_{ij}-a_{ij}\right)\right\|^2\leq\frac{1}{N}\sum_{j=1}^N\left\|\hat{\lambda}_{2j}\right\|^2\times\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j=1}^N\|\hat{a}_{ij}-a_{ij}\|^2=O_P(C_{NT}^{-2})$. Consequently, we have $\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\hat{\bar{\lambda}}_{2i}-H_2^{-1}\bar{\lambda}_{2i}^0\right\|^2=O_P(C_{NT}^{-2})$. This result, in conjunction with the result in (iv) and the Cauchy-Schwarz inequality, implies that $\|I_{11}\|\leq\sqrt{N_k}\left\{\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\hat{\Delta}_{24,i}-\Delta_{24,i}H_2\right\|^2\right\}^{1/2}\left\{\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\hat{\bar{\lambda}}_{2i}-H_2^{-1}\bar{\lambda}_{2i}^0\right\|^2\right\}^{1/2}=\sqrt{N_k}O_P\left(\frac{J^{1/2}}{T^{1/2}}+\frac{1}{J^q}\right)O_P\left(C_{NT}^{-1}\right)=o_P\left(1\right)$. By arguments like those used in the proof of (ii), we can show that $I_{13}=o_P\left(1\right)$. It follows that $\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}(\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i}-\Delta_{24,i}\bar{\lambda}_{2i}^0)=o_P\left(1\right)$.

(viii) We make the decomposition

$$
\begin{aligned}
&\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\sum_{t=1}^T\sum_{s=1}^T\left[\hat{\varkappa}_{ts}\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i}-\varkappa_{ts}\Delta_{24,i}\bar{\lambda}_{2i}^0\right]\mathbf{1}\left\{s\leq t\right\}\\
=&\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T(\hat{\varkappa}_{ts}-\varkappa_{ts})\mathbf{1}\left\{s\leq t\right\}\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left(\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i}-\Delta_{24,i}\bar{\lambda}_{2i}^0\right)\\
&+\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T(\hat{\varkappa}_{ts}-\varkappa_{ts})\mathbf{1}\left\{s\leq t\right\}\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\Delta_{24,i}\bar{\lambda}_{2i}^0\\
&+\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\varkappa_{ts}\mathbf{1}\left\{s\leq t\right\}\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left(\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i}-\Delta_{24,i}\bar{\lambda}_{2i}^0\right)\equiv I_{21}+I_{22}+I_{23}.
\end{aligned}
$$

Note that $I_{21}=o_P\left(1\right)$ by (v) and (vi), $I_{22}=o_P\left(1\right)$ by (v) and the fact that $\frac{1}{N_k}\sum_{i\in G_k^0}\Delta_{24,i}\bar{\lambda}_{2i}^0=O_P\left(1\right)$, and $I_{23}=o_P\left(1\right)$ by (vi) and the fact that $\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\varkappa_{ts}\mathbf{1}\left\{s\leq t\right\}=O_P\left(1\right)$. It follows that $\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\sum_{t=1}^T\sum_{s=1}^T[\hat{\varkappa}_{ts}\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i}-\varkappa_{ts}\Delta_{24,i}\bar{\lambda}_{2i}^0]\mathbf{1}\left\{s\leq t\right\}=o_P\left(1\right)$.

(ix) We define the following $T\times p_1$ matrices $\tilde{x}_{1,i}=x_{1,i}-\mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)$, $\mathfrak{X}_{k,i}=M_{F_1^0}\tilde{x}_{1,i}\mathbf{1}\left\{i\in G_k^0\right\}+M_{F_1^0}[\mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)\times\mathbf{1}\left\{i\in G_k^0\right\}-\frac{1}{N}\sum_{j\in G_k^0}a_{ij}\mathbb{E}_{\mathcal{C}}\left(x_{1,j}\right)]$, and $\mathbf{X}_{k,i}=M_{F_1^0}x_{1,i}\mathbf{1}\left\{i\in G_k^0\right\}-\frac{1}{N}\sum_{j\in G_k^0}a_{ij}M_{F_1^0}x_{1,j}$. Let $\mathfrak{X}_{k,it}'$ and $\mathbf{X}_{k,it}'$ denote the tth row of $\mathfrak{X}_{k,i}$ and $\mathbf{X}_{k,i}$, respectively, which is a $p_1\times 1$ vector. Let

$\mathfrak{X}_{it} = (\mathfrak{X}_{1,it}', ..., \mathfrak{X}_{K,it}')'$ and $\mathbf{X}_{it} = \left(\mathbf{X}_{1,it}', ..., \mathbf{X}_{K,it}'\right)'$. Recall that

$$\begin{aligned} U_{k,NT} &= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0} x_{1,i}' M_{F_1^0}\left(\left(u_i + F_2^0\lambda_{2i}^0\right) - \frac{1}{N}\sum_{j=1}^{N} a_{ij}(u_j + F_2^0\lambda_{2j}^0)\right) \\ &= \frac{1}{\sqrt{N_k}T}\sum_{i=1}^{N}\left(M_{F_1^0}x_{1,i}\mathbf{1}\left\{i\in G_k^0\right\} - \frac{1}{N}\sum_{j\in G_k^0} a_{ij}M_{F_1^0}x_{1,j}\right)'\left(u_i + F_2^0\lambda_{2i}^0\right) \\ &= \frac{1}{\sqrt{N_k}T}\sum_{i=1}^{N}\mathbf{X}_{k,i}'\left(u_i + F_2^0\lambda_{2i}^0\right) = \sum_{i=1}^{N} U_{k,iNT}, \end{aligned}$$

where $U_{k,iNT} = \frac{1}{\sqrt{N_k}T}\sum_{t=1}^{T}\mathbf{X}_{k,it}\left(u_{it} + \lambda_{2i}^{0\prime}f_{2t}^0\right)$. Let $U_{iNT} = \left(U_{1,iNT}', ..., U_{K,iNT}'\right)'$. Then

$$U_{iNT} = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\frac{1}{\sqrt{N_1}}\mathbf{X}_{1,it} \\ \vdots \\ \frac{1}{\sqrt{N_K}}\mathbf{X}_{K,it}\end{pmatrix}\left(u_{it} + \lambda_{2i}^{0\prime}f_{2t}^0\right) = \frac{\sqrt{D_{NK}}}{\sqrt{N}T}\sum_{t=1}^{T}\mathbf{X}_{it}\left(u_{it} + \lambda_{2i}^{0\prime}f_{2t}^0\right),$$

where D_{NK} =diag$(\frac{N}{N_1}, ..., \frac{N}{N_K}) \otimes I_{p_1}$, which is a $Kp_1 \times Kp_1$ diagonal matrix. Now we collect all asymptotic non-negligible components in $U_{k,NT}$ and define them as $\mathcal{Z}_{k,NT}$ as follows

$$\begin{aligned} \mathcal{Z}_{k,NT} &= U_{1k,NT}^{u} + U_{2k,NT}^{u} + U_{1k,NT}^{f_2} + U_{2k,NT}^{f_2} \\ &= \frac{1}{\sqrt{N_k}T}\sum_{i=1}^{N}[x_{1,i} - \mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)]'\mathbf{1}\left\{i\in G_k^0\right\} M_{F_1^0}u_i \\ &\quad + \frac{1}{\sqrt{N_k}T}\sum_{i=1}^{N}\left\{\mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)\mathbf{1}\left\{i\in G_k^0\right\} - \frac{1}{N}\sum_{j\in G_k^0} a_{ij}\mathbb{E}_{\mathcal{C}}(x_{1,j}')\right\} M_{F_1^0}u_i \\ &\quad + \frac{1}{\sqrt{N_k}T}\sum_{i=1}^{N}[x_{1,i} - \mathbb{E}_{\mathcal{C}}\left(x_{1,i}\right)]'\mathbf{1}\left\{i\in G_k^0\right\} M_{F_1^0}F_2^0\lambda_{2i}^0 \\ &\quad + \frac{1}{\sqrt{N_k}T}\sum_{i=1}^{N}\left[\mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)\mathbf{1}\left\{i\in G_k^0\right\} - \frac{1}{N}\sum_{j\in G_k^0} a_{ij}\frac{1}{T}\mathbb{E}_{\mathcal{C}}(x_{1,j}')\right] M_{F_1^0}F_2^0\lambda_{2i}^0 \\ &= \frac{1}{\sqrt{N_k}T}\sum_{i=1}^{N}\left[\tilde{x}_{1,i}'\mathbf{1}\left\{i\in G_k^0\right\} + \left(\mathbb{E}_{\mathcal{C}}\left(x_{1,i}'\right)\mathbf{1}\left\{i\in G_k^0\right\} - \frac{1}{N}\sum_{j\in G_k^0} a_{ij}\mathbb{E}_{\mathcal{C}}(x_{1,j}')\right)\right] M_{F_1^0}\left(u_i + F_2^0\lambda_{2i}^0\right) \\ &= \sum_{i=1}^{N}\frac{1}{\sqrt{N_k}T}\mathfrak{X}_{k,i}'\left(u_i + f_2^0\lambda_{2i}^0\right) = \sum_{i=1}^{N}\mathcal{Z}_{k,iNT}. \end{aligned}$$

where $\mathcal{Z}_{k,iNT} = \frac{1}{\sqrt{N_k}T}\sum_{t=1}^{T}\mathfrak{X}_{k,it}\left(u_{it} + \lambda_{2i}^{0\prime}f_{2t}^0\right)$. Similarly, letting $\mathcal{Z}_{iNT} = \left(\mathcal{Z}_{1,iNT}', ..., \ \mathcal{Z}_{K,iNT}'\right)'$, we have

$$\mathcal{Z}_{iNT}=\frac{\sqrt{D_{NK}}}{\sqrt{N}T}\sum_{t=1}^{T}\mathfrak{X}_{it}\left(u_{it}+\lambda_{2i}^{0\prime}f_{2t}^{0}\right).$$

By construction, we note that $U_{k,NT}=\mathcal{Z}_{k,NT}+o_P(1)$ and $U_{i,NT}=\mathcal{Z}_{i,NT}+o_P(N^{-1})$. Recall that $V_{kNT}=\sum_{i=1}^{N}Z_{k,iNT}$ and $U_{kNT}=V_{kNT}+B_{1k,NT}+B_{2k,NT}+o_P(1)$. Then we have

$$\mathcal{Z}_{k,iNT}=Z_{k,iNT}+B_{k,iNT}+o_P(N^{-1}),$$

where $B_{k,iNT}=B_{k,iNT,1}+B_{k,iNT,2}$, $B_{k,iNT,1}=\frac{1}{\sqrt{N_k}T}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\bar{\varkappa}_{ts}\right)\Delta_{21,i}\mathbf{1}\left\{i\in G_k^0\right\}$, and $B_{k,iNT,2}=\frac{1}{\sqrt{N_k}T}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\bar{\varkappa}_{ts}\right)\Delta_{24,i}\bar{\lambda}_{2i}\mathbf{1}\left\{i\in G_k^0\right\}$. Define $B_{iNT}=\left(B_{1,iNT}^{\prime},...,B_{K,iNT}^{\prime}\right)^{\prime}$. Note that Z_{iNT} are independent across i conditional on $\mathcal{C}$. Similarly, we have that $\mathcal{Z}_{iNT}$ are independent across i conditional on $\mathcal{C}$. Then we have $\Omega_{NT}=\text{Var}(V_{NT}|\mathcal{C})=\sum_{i=1}^{N}\text{Var}(Z_{iNT}|\ \mathcal{C})=\sum_{i=1}^{N}\text{Var}(\mathcal{Z}_{iNT}|\mathcal{C})+o_P(1)$, where $\sum_{i=1}^{N}\text{Var}(\mathcal{Z}_{iNT}|\mathcal{C})=\sum_{i=1}^{N}[\mathbb{E}_{\mathcal{C}}\left(\mathcal{Z}_{iNT}\mathcal{Z}_{iNT}^{\prime}\right)-\mathbb{E}_{\mathcal{C}}\left(\mathcal{Z}_{iNT}\right)\mathbb{E}_{\mathcal{C}}\left(\mathcal{Z}_{iNT}\right)^{\prime}]$. By construction, we have $\mathbb{E}_{\mathcal{C}}\left(\mathcal{Z}_{iNT}\right)=\mathbb{E}_{\mathcal{C}}\left(Z_{iNT}+B_{iNT}\right)+o_P(N^{-1})=B_{iNT}+o_P(N^{-1})$ and $\sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left(\mathcal{Z}_{iNT}\right)\mathbb{E}_{\mathcal{C}}\left(\mathcal{Z}_{iNT}\right)^{\prime}=\sum_{i=1}^{N}B_{iNT}B_{iNT}^{\prime}+o_P(1)$. Note that conditional on $\mathcal{C}$ the expression $\mathcal{Z}_{iNT}\mathcal{Z}_{iNT}^{\prime}-\mathbb{E}_{\mathcal{C}}\left(\mathcal{Z}_{iNT}\mathcal{Z}_{iNT}^{\prime}\right)$ is mean zero, and it is also independent across i. This together with the bounded moments implies that $\text{Var}(\sum_{i=1}^{N}\left(\mathcal{Z}_{iNT}\mathcal{Z}_{iNT}^{\prime}-\ \mathbb{E}_{\mathcal{C}}\left(\mathcal{Z}_{iNT}\mathcal{Z}_{iNT}^{\prime}\right)\right)|\mathcal{C})=o_P(1)$. Thus, we have

$$\begin{aligned}\sum_{i=1}^{N}\mathbb{E}_{\mathcal{C}}\left(\mathcal{Z}_{iNT}\mathcal{Z}_{iNT}^{\prime}\right) &= \sum_{i=1}^{N}\mathcal{Z}_{iNT}\mathcal{Z}_{iNT}^{\prime}+o_P(1)\\ &= \sum_{i=1}^{N}\frac{D_{NK}}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathfrak{X}_{it}\left(u_{it}+\lambda_{2i}^{0\prime}f_{2t}^{0}\right)\left(u_{is}+\lambda_{2i}^{0\prime}f_{2s}^{0}\right)\mathfrak{X}_{is}^{\prime}+o_P(1)\\ &= \frac{D_{NK}}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathfrak{X}_{it}\mathfrak{X}_{is}^{\prime}\left(u_{it}+\lambda_{2i}^{0\prime}f_{2t}^{0}\right)\left(u_{is}+\lambda_{2i}^{0\prime}f_{2s}^{0}\right)+o_P(1).\end{aligned}$$

By construction, we have $U_{i,NT}=\mathcal{Z}_{iNT}+o_P(N^{-1})$. Then we have $\sum_{i=1}^{N}\mathcal{Z}_{iNT}\mathcal{Z}_{iNT}^{\prime}=\sum_{i=1}^{N}U_{i,NT}U_{i,NT}^{\prime}+o_P(1)=\frac{D_{NK}}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{X}_{it}\mathbf{X}_{is}^{\prime}\left(u_{it}+\lambda_{2i}^{0\prime}f_{2t}^{0}\right)\left(u_{is}+\lambda_{2i}^{0\prime}f_{2s}^{0}\right)+o_P(1)$. It follows that

$$\Omega_{NT}=\frac{D_{NK}}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{X}_{it}\mathbf{X}_{is}^{\prime}\left(u_{it}+\lambda_{2i}^{0\prime}f_{2t}^{0}\right)\left(u_{is}+\lambda_{2i}^{0\prime}f_{2s}^{0}\right)-\sum_{i=1}^{N}B_{iNT}B_{iNT}^{\prime}.$$

Recall that $\hat{\Omega}_{NT}=\frac{\hat{D}_{NK}}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{X}}_{it}\hat{\mathbf{X}}_{is}^{\prime}\hat{u}_{it}^{*}\hat{u}_{is}^{*}-\sum_{i=1}^{N}\hat{B}_{iNT}\hat{B}_{iNT}^{\prime}$, where $\hat{\mathbf{X}}_{it}$, $\hat{u}_{it}^{*}$, and

$\hat{B}_{iNT}$ are as defined in Section 3.4. We decompose $\hat{\Omega}_{NT}-\Omega_{NT}$ as follows,

$$\begin{aligned}
\hat{\Omega}_{NT}-\Omega_{NT} &= \frac{\hat{D}_{NK}}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{X}}_{it}\hat{\mathbf{X}}_{is}'\left(\hat{u}_{it}^{*}\hat{u}_{is}^{*}-\left(u_{it}+\lambda_{2i}^{0\prime}f_{2t}^{0}\right)\left(u_{is}+\lambda_{2i}^{0\prime}f_{2s}^{0}\right)\right) \\
&\quad+\left(\hat{D}_{NK}-D_{NK}\right)\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{X}}_{it}\hat{\mathbf{X}}_{is}'\left(u_{it}+\lambda_{2i}^{0\prime}f_{2t}^{0}\right)\left(u_{is}+\lambda_{2i}^{0\prime}f_{2s}^{0}\right) \\
&\quad+\frac{D_{NK}}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{\mathbf{X}}_{it}\hat{\mathbf{X}}_{is}'-\mathbf{X}_{it}\mathbf{X}_{is}'\right)\left(u_{it}+\lambda_{2i}^{0\prime}f_{2t}^{0}\right)\left(u_{is}+\lambda_{2i}^{0\prime}f_{2s}^{0}\right) \\
&\quad-\left(\sum_{i=1}^{N}\hat{B}_{iNT}\hat{B}_{iNT}'-\sum_{i=1}^{N}B_{iNT}B_{iNT}'\right) \\
&\equiv \Omega_{NT,1}+\Omega_{NT,2}+\Omega_{NT,3}+\Omega_{NT,4}.
\end{aligned}$$

It suffices to prove $\|\Omega_{NT,l}\|_{sp}=o_P(1)$ for $l=1,2,3,4$. Let c_{Kp_1} be an arbitrary $Kp_1\times 1$ nonrandom vector with $\|c_{Kp_1}\|=1$. Note $\hat{u}_{it}^{*}=y_{it}-\hat{\alpha}_k'x_{1,it}-\hat{\beta}_{2,i}'x_{2,it}-\hat{\lambda}_{1i}'\hat{f}_{1t}$. By the triangle inequality,

$$\begin{aligned}
\left|c_{Kp_1}'\Omega_{NT,1}c_{Kp_1}\right| &= \left|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}c_{Kp_1}'\hat{D}_{NK}\hat{\mathbf{X}}_{it}\hat{\mathbf{X}}_{is}'c_{Kp_1}\left(\hat{u}_{it}^{*}\hat{u}_{is}^{*}-u_{it}^{*}u_{is}^{*}\right)\right| \\
&\le \left|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}c_{Kp_1}'\hat{D}_{NK}\hat{\mathbf{X}}_{it}\hat{\mathbf{X}}_{is}'c_{Kp_1}\hat{u}_{it}^{*}\left(\hat{u}_{is}^{*}-u_{is}^{*}\right)\right| \\
&\quad+\left|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}c_{Kp_1}'\hat{D}_{NK}\hat{\mathbf{X}}_{it}\hat{\mathbf{X}}_{is}'c_{Kp_1}\left(\hat{u}_{it}^{*}-u_{it}^{*}\right)u_{is}\right| \\
&\le \left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}c_{Kp_1}'\hat{D}_{NK}\hat{\mathbf{X}}_{it}\hat{u}_{it}^{*}\right)\left(\frac{1}{T}\sum_{s=1}^{T}\left(\hat{u}_{is}^{*}-u_{is}^{*}\right)\hat{\mathbf{X}}_{is}'c_{Kp_1}\right)\right| \\
&\quad+\left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}c_{Kp_1}'\hat{D}_{NK}\hat{\mathbf{X}}_{it}\left(\hat{u}_{it}^{*}-u_{it}^{*}\right)\right)\left(\frac{1}{T}\sum_{s=1}^{T}u_{is}^{*}\hat{\mathbf{X}}_{is}'c_{Kp_1}\right)\right| \\
&\equiv A_{1,1}+A_{1,2}.
\end{aligned}$$

Note that $A_{1,1}\le\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{D}_{NK}\hat{\mathbf{X}}_{it}\hat{u}_{it}^{*}\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\hat{u}_{is}^{*}-u_{is}^{*}\right)\hat{\mathbf{X}}_{is}'\right\|^2\right)^{1/2}=O_P(1)o_P(1)=o_P(1)$ by Lemmas A.3(ii) and A.11(i), where

$$\begin{aligned}
\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\hat{u}_{is}^{*}-u_{is}^{*}\right)\hat{\mathbf{X}}_{is}'\right\| &\le \left\|\left(\hat{\alpha}_k-\alpha_k^0\right)'\frac{1}{T}\sum_{s=1}^{T}x_{1,is}\hat{\mathbf{X}}_{is}'\right\|+\left\|\hat{b}_{2i}'\frac{1}{T}\sum_{s=1}^{T}x_{2,is}\hat{\mathbf{X}}_{is}'\right\|+\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\hat{\lambda}_{1i}'\hat{f}_{1s}-\lambda_{1i}^{0\prime}f_{1s}^{0}\right)\hat{\mathbf{X}}_{is}'\right\| \\
&= O_P(N^{-1/2})+O_P(\sqrt{d_T}T^{-1/2})+O_P(\sqrt{T}\eta_{1NT}+C_{NT}^{-1})=o_P(1).
\end{aligned}$$

Similarly, we can show that $A_{1,2}=o_P(1)$. It follows that $\|\Omega_{NT,1}\|_{sp}=o_P(1)$.

To prove that $\|\Omega_{NT,2}\|_{sp}=o_P(1)$, we need to show $\left\|\hat{D}_{NK}-D_{NK}\right\|_{sp}=o_P(1)$. By Theorem 3.3, it directly implies that $P\left(\hat{N}_k=N_k\right)\to 1$. Then it follows that $\|\Omega_{NT,2}\|_{sp}=o_P(1)$.

To prove that $\|\Omega_{NT,3}\|_{sp} = o_P(1)$, we observe that

$$\begin{aligned}
\left|c'_{Kp_1}\Omega_{NT,3}c_{Kp_1}\right| &= \left|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} c'_{Kp_1}D_{NK}\left(\hat{\mathbf{X}}_{it}\hat{\mathbf{X}}'_{is}-\mathbf{X}_{it}\mathbf{X}'_{is}\right)c_{Kp_1}u^*_{it}u^*_{is}\right| \\
&\leq \left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}c'_{Kp_1}D_{NK}\hat{\mathbf{X}}_{it}u^*_{it}\right)\left(\frac{1}{T}\sum_{s=1}^{T}u^*_{is}\left(\hat{\mathbf{X}}'_{is}-\mathbf{X}'_{is}\right)c_{Kp_1}\right)\right| \\
&\quad + \left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}c'_{Kp_1}D_{NK}\left(\hat{\mathbf{X}}_{it}-\ \mathbf{X_{it}}\right)u^*_{it}\right)\left(\frac{1}{T}\sum_{s=1}^{T}u^*_{is}\mathbf{X}'_{is}c_{Kp_1}\right)\right| \equiv A_{2,1}+A_{2,2}.
\end{aligned}$$

Note that $A_{2,1} \leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}D_{NK}\hat{\mathbf{X}}_{it}u^*_{it}\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}u^*_{is}(\hat{\mathbf{X}}_{is}-\mathbf{X}_{is})'\right\|^2\right)^{1/2}$ and $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}D_{NK}\hat{\mathbf{X}}_{it}u^*_{it}\right\|^2 = O_P(1)$. It remains to show $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}u^*_{is}(\ \hat{\mathbf{X}}_{is}-\mathbf{X}_{is})'\right\|^2 = o_P(1)$ by using that $\frac{1}{T}\sum_{s=1}^{T}u^*_{is}\left(\hat{\mathbf{X}}_{is}-\mathbf{X}_{is}\right)' = \frac{1}{T}u^{*\prime}_i\begin{pmatrix}\hat{\mathbf{X}}_{1,i}-\mathbf{X}_{1,i}\\ \vdots \\ \hat{\mathbf{X}}_{K,i}-\mathbf{X}_{K,i}\end{pmatrix}$, where $\frac{1}{T}u^{*\prime}_i(\hat{\mathbf{X}}_{k,i}-\mathbf{X}_{k,i}) = \frac{1}{T}u^{*\prime}_i[M_{\hat{F}_1}x_{1,i}\mathbf{1}\{i\in\hat{G}_k\}-M_{F^0_1}x_{1,i}\mathbf{1}\left\{i\in G^0_k\right\}]-\frac{1}{T}u^{*\prime}_i[\frac{1}{N}\sum_{j\in\hat{G}_k}\hat{a}_{ij}M_{\hat{F}_1}x_{1,j}-\frac{1}{N}\sum_{j\in G^0_k}a_{ij}M_{F^0_1}x_{1,j}]$. By similar arguments to those in the proof of Lemma A.9, we can show that $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}u^*_{is}(\hat{\mathbf{X}}_{is}-\mathbf{X}_{is})'\right\|^2 = o_P(1)$. Then $A_{2,1} = o_P(1)$. Similarly, we can show that $A_{2,2} = o_P(1)$. It follows that $\|\Omega_{NT,3}\|_{sp} = o_P(1)$.

By the proof of Theorem 3.5, we already show that $\hat{B}_{k,NT}-B_{kNT} = o_P(1)$. It follows that $\hat{B}_{k,iNT}-B_{k,iNT} = o_P(N^{-1})$. Since $\Omega_{NT,4} = \sum_{i=1}^{N}\hat{B}_{iNT}(\hat{B}_{iNT}-B_{iNT})+\sum_{i=1}^{N}(\hat{B}_{iNT}-B_{iNT})B'_{iNT}$, we have $\|\Omega_{NT,4}\|_{sp} = o_P(1)$. It follows that $\hat{\Omega}_{NT}-\Omega_{NT} = o_P(1)$.

The proof that $\hat{\Omega}^+_{NT}-\Omega^+_{NT} = o_P(1)$ is analogous and thus omitted. ■

Proof of Lemma A.12. Let $\hat{e}_i = y_i - x_{1,i}\hat{\beta}_{1,i} - x_{2,i}\hat{\beta}_{2,i} = F^0_1\lambda^0_{1i} + \hat{u}^*_i - x_{1,i}\hat{b}_{1,i}$, where $\hat{u}^*_i = u_i + F^{0\prime}_2\lambda^0_{2i} - x_{2,i}\hat{b}_{2,i}$ with a typical element denoted as $\hat{u}^*_{it}$. Then $V_1(r_1,\hat{F}^{r_1}_1) = \frac{1}{NT}\sum_{i=1}^{N}\hat{e}'_iM_{\hat{F}^{r_1}_1}\hat{e}_i$ and $V_1(r_1,F^0_1H^{r_1}_1) = \frac{1}{NT}\sum_{i=1}^{N}\hat{e}'_iM_{F^0_1H_1}\hat{e}_i$. Noting that $\hat{u}^*_{it}$ behaves like a zero mean $I\left(0\right)$ process, $\hat{\beta}_{1,i}$ and $\hat{\beta}_{2,i}$ are T^{-1}- and $T^{-1/2}$-consistent, respectively, when $r_1 \geq r^0_1$, the proof follows from obvious modifications to Lemmas C.2-C.4 in Bai (2004). ■

Proof of Lemma A.13. Note that we determine the number of unobserved stationary factors based on the resultant residuals

$$\hat{r}_{it} = y_{it} - \hat{\beta}'_{1,i}x_{1,it} - \hat{\beta}'_{2,i}x_{2,it} - \hat{\lambda}'_{1i}\hat{f}_{1t} = \lambda^{0\prime}_{2i}f^0_{2t} + u_{it} + v_{it},$$

where $v_{it} = -(\hat{b}'_{1,i}x_{1,it} + \hat{b}'_{2,i}x_{2,it} + \hat{\lambda}'_{1i}\hat{f}_{1t} - \lambda^{0\prime}_{1i}f^0_{1t})$ signifies the parameter estimation error from early stages. Given the preliminary consistency of $\hat{\beta}_{1,i}$, $\hat{\beta}_{2,i}$ and $\hat{\lambda}'_{1i}\hat{f}_{1t}$, and the fact that f^0_{2t} is stationary, the proof of the lemma follows from that of Lemma A.10 in Su and Ju (2018) and is omitted. ■

Proof of Lemma A.14. Here we consider the case where the model contains both stationary and nonstationary common factors as analyses of the other cases are similar to but simpler than this case. Let $\hat{F}(K,\lambda) = (\hat{f}_1(K,\lambda),...,\hat{f}_T(K,\lambda))'$. Noting that $\hat{e}_i\left(K\right) \equiv y_i - x_{1,i}\hat{\alpha}^{cup\prime}_{\hat{G}_k(K,\lambda)} - x_{2,i}\hat{\beta}_{2,i} -$

$\hat{F}(K,\lambda)\hat{\lambda}_i(K,\lambda) = u_i - x_{1,i}[\hat{\alpha}^{cup}_{\hat{G}_k(K,\lambda)} - \beta^0_{1,i}] - x_{2,i}[\hat{\beta}_{2,i} - \beta^0_{2,i}] + [F^0\lambda^0_i - \hat{F}(K,\lambda)\hat{\lambda}_i(K,\lambda)]$, we make the following decomposition on $\hat{\sigma}^2_{\hat{G}(K_0,\lambda)} = \frac{1}{NT}\sum_{i=1}^N \hat{e}_i(K_0)'\hat{e}_i(K):$

$$\begin{aligned}
\hat{\sigma}^2_{\hat{G}(K_0,\lambda)} &= \frac{1}{NT}\sum_{i=1}^N u_i'u_i + \frac{1}{NT}\sum_{i=1}^N \left(F^0\lambda^0_i - \hat{F}(K_0,\lambda)\hat{\lambda}_i(K_0,\lambda)\right)'\left(F^0\lambda^0_i - \hat{F}(K_0,\lambda)\hat{\lambda}_i(K_0,\lambda)\right) \\
&\quad + \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k(K_0,\lambda)} \left(\hat{\alpha}^{cup}_{\hat{G}_k(K_0,\lambda)} - \beta^0_{1,i}\right)' x_{1,i}'x_{1,i}\left(\hat{\alpha}^{cup}_{\hat{G}_k(K_0,\lambda)} - \beta^0_{1,i}\right) \\
&\quad + \frac{1}{NT}\sum_{i=1}^N \left(\hat{\beta}_{2,i} - \beta^0_{2,i}\right)' x_{2,i}'x_{2,i}\left(\hat{\beta}_{2,i} - \beta^0_{2,i}\right) \\
&\quad + \frac{2}{NT}\sum_{i=1}^N \left(F^0\lambda^0_i - \hat{F}(K_0,\lambda)\hat{\lambda}_i(K_0,\lambda)\right)' u_i + \frac{2}{NT}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k(K_0,\lambda)} \left(\hat{\alpha}^{cup}_{\hat{G}_k(K_0,\lambda)} - \beta^0_{1,i}\right)' x_{1,i}'u_i \\
&\quad + \frac{2}{NT}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k(K_0,\lambda)} \left(\hat{\alpha}^{cup}_{\hat{G}_k(K_0,\lambda)} - \beta^0_{1,i}\right)' x_{1,i}'\left(F^0\lambda_i - \hat{F}(K_0,\lambda)\hat{\lambda}_i(K_0,\lambda)\right) \\
&\quad + \frac{2}{NT}\sum_{i=1}^N \left(\hat{\beta}_{2,i} - \beta^0_{2,i}\right)' x_{2,i}'u_i + \frac{2}{NT}\sum_{i=1}^N \left(\hat{\beta}_{2,i} - \beta^0_{2,i}\right)' x_{2,i}'\left(F^0\lambda_i - \hat{F}(K_0,\lambda)\hat{\lambda}_i(K_0,\lambda)\right) \\
&\quad + \frac{2}{NT}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k(K_0,\lambda)} \left(\hat{\alpha}^{cup}_{\hat{G}_k(K_0,\lambda)} - \beta^0_{1,i}\right)' x_{1,i}'x_{2,i}\left(\hat{\beta}_{2,i} - \beta^0_{2,i}\right) \\
&\equiv \frac{1}{NT}\sum_{i=1}^N u_i'u_i + \sum_{l=1}^9 R_{1,lNT}.
\end{aligned}$$

It is easy to show that

$$|R_{1,1NT}| = \frac{1}{NT}\sum_{i=1}^N \left\|\hat{F}\hat{\lambda}_i - F^0\lambda^0_i\right\|^2 \le \frac{2}{NT}\sum_{i=1}^N \left\|\hat{F}_1\hat{\lambda}_{1i} - F^0_1\lambda^0_{1i}\right\|^2 + \frac{2}{NT}\sum_{i=1}^N \left\|\hat{F}_2\hat{\lambda}_{2i} - F^0_2\lambda^0_{2i}\right\|^2 = O_P(C_{NT}^{-2}),$$

by using arguments as in the proofs of Lemmas A.5 and A.11. Similarly,

$$|R_{1,2NT}| \le T\max_{1\le k\le K_0}\left\|\hat{\alpha}^{cup}_{\hat{G}_k(K_0,\lambda)} - \beta^0_k\right\|^2 \frac{1}{NT^2}\sum_{i=1}^N \left\|x_{1,i}'x_{1,i}\right\| = TO_P\left(N^{-1}T^{-2}\right) = O_P\left(N^{-1}T^{-1}\right).$$

by Theorem 3.5. Similarly, $|R_{1,3NT}| \le \max_i \frac{1}{T}\|x_{2,i}\|^2 \frac{1}{NT}\sum_{i=1}^N \left\|\hat{\beta}_{2,i}(K_0,\lambda) - \beta^0_{2,i}\right\|^2 = O_P\left(T^{-1}\right)$. By the Cauchy-Schwarz inequality, $|R_{1,6NT}| \le 2\{|R_{1,1NT}||R_{1,2NT}|\}^{1/2} = o_P(C_{NT}^{-2})$. In addition, we can show that $R_{1,lNT} = o_P(C_{NT}^{-2})$ for $l = 4,5,7,8,$ and 9. It follows that $\hat{\sigma}^2_{\hat{G}(K_0,\lambda)} = \frac{1}{NT}\sum_{i=1}^N u_i'u_i + O_P(C_{NT}^{-2})$.

When $K > K_0$, we use $\mathbf{1}\{i\in\hat{G}_k(K,\lambda)\} = \mathbf{1}\{i\in G^0_k\} + \mathbf{1}\{i\in\hat{G}_k(K,\lambda)\backslash G^0_k\} - \mathbf{1}\{i\in G^0_k\backslash\hat{G}_k(K,\lambda)\}$ to obtain $\hat{\sigma}^2_{G(K,\lambda)} = \frac{1}{NT}\sum_{k=1}^K\sum_{i\in\hat{G}_k(K,\lambda)}\sum_{t=1}^T \left[y_{it} - \hat{\alpha}^{cup\prime}_{\hat{G}_k(K,\lambda)}x_{1,it} - \hat{\beta}'_{2,i}x_{2,it} - \hat{\lambda}_i(K,\lambda)'\hat{f}_t(K,\lambda)\right]^2 =$

$R_{2,1NT}(K) + R_{2,2NT}(K) - R_{2,3NT}(K) + R_{2,4NT}(K)$, where

$$R_{2,1NT}(K) = \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in G_k^0}\sum_{t=1}^{T}\left[y_{it} - \hat{\alpha}_{\hat{G}_k(K,\lambda)}^{cup\prime} x_{1,it} - \hat{\beta}_{2,i}^{\prime} x_{2,it} - \hat{\lambda}_i(K,\lambda)^{\prime}\hat{f}_t(K,\lambda)\right]^2,$$

$$R_{2,2NT}(K) = \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in \hat{G}_k(K,\lambda)\backslash G_k^0}\sum_{t=1}^{T}\left[y_{it} - \hat{\alpha}_{\hat{G}_k(K,\lambda)}^{cup\prime} x_{1,it} - \hat{\beta}_{2,i}^{\prime} x_{2,it} - \hat{\lambda}_i(K,\lambda)^{\prime}\hat{f}_t(K,\lambda)\right]^2,$$

$$R_{2,3NT}(K) = \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in G_k^0\backslash \hat{G}_k(K,\lambda)}\sum_{t=1}^{T}\left[y_{it} - \hat{\alpha}_{\hat{G}_k(K,\lambda)}^{cup\prime} x_{1,it} - \hat{\beta}_{2,i}^{\prime} x_{2,it} - \hat{\lambda}_i(K,\lambda)^{\prime}\hat{f}_t(K,\lambda)\right]^2, \text{ and}$$

$$R_{2,4NT}(K) = \frac{1}{NT}\sum_{k=K_0+1}^{K}\sum_{i\in \hat{G}_k(K,\lambda)}\sum_{t=1}^{T}\left[y_{it} - \hat{\alpha}_{\hat{G}_k(K,\lambda)}^{cup\prime} x_{1,it} - \hat{\beta}_{2,i}^{\prime} x_{2,it} - \hat{\lambda}_i(K,\lambda)^{\prime}\hat{f}_t(K,\lambda)\right]^2.$$

Following the proof of Lemma A.11 in Su and Ju (2018), we can show that, after some relabeling the indices for the group-specific parameters,

$$\hat{\alpha}_{\hat{G}_k(K,\lambda)}^{cup} - \alpha_k^0 = O_P(N^{-1/2}T^{-1}) \text{ for } k = 1, ..., K_0,$$
$$\sum_{i\in G_k^0} P\left(\hat{E}_{kNT,i}\right) = o(1) \text{ and } \sum_{i\in G_k^0} P(\hat{F}_{kNT,i}) = o(1) \text{ for } k = 1, ..., K.$$

Then $\sum_{i=1}^{N} P\left(i \in \hat{G}_k(K,\lambda) \text{ for } k = 0, K_0+1, ..., K\right) = o(1)$, which ensures that $R_{2,lNT}(K) = o_P((NT)^{-1})$ for all $l = 2,3,4$. Given the consistency of $\hat{\alpha}_{\hat{G}_k(K,\lambda)}^{cup}$ for $k = 1, ..., K_0$, we can establish the consistency of $\hat{\lambda}_i(K,\lambda)$ and $\hat{f}_t(K,\lambda)$ as in the case where $K = K_0$. With these results, we can show that $R_{2,1NT}(K) = \frac{1}{NT}\sum_{i=1}^{N} u_i^{\prime}u_i + O_P(C_{NT}^{-2})$. The probability order for the remainder term in $R_{2,1NT}(K)$ can be improved in some cases: (1) When there are no unobserved common factor, no stationary regressors and endogeneity in $x_{1,it}$, we can show that $R_{2,1NT}(K) = \frac{1}{NT}\sum_{i=1}^{N} u_i^{\prime}u_i + O_P((NT)^{-1})$ by using the fact that $\hat{\alpha}_{\hat{G}_k(K,\lambda)}^{cup} - \alpha_k^0 = O_P(N^{-1/2}T^{-1})$ for $k = 1, ..., K_0$ when $K \geq K_0$; (2) when there is stationary regressor $x_{2,it}$ but no unobserved factor in the model, we can show that $R_{2,1NT}(K) = \frac{1}{NT}\sum_{i=1}^{N} u_i^{\prime}u_i + O_P(T^{-1})$; (3) when there exists common nonstationary factor but no common stationary factor or stationary regressor $x_{2,it}$, $R_{2,1NT}(K) = \frac{1}{NT}\sum_{i=1}^{N} u_i^{\prime}u_i + O_P(N^{-1} + T^{-2}) = \frac{1}{NT}\sum_{i=1}^{N} u_i^{\prime}u_i + O_P(N^{-1})$. So the results in Lemma A.14 follows. ■

C Discussion on the Identification of $\beta_{1,i}^0$

In this appendix, we formally discuss the identification issue regarding the key parameter vector of interest, namely, $\beta_{1,i}^0$. Recall that $\boldsymbol{\beta}_1^0 = (\beta_{1,1}^0, ..., \beta_{1,N}^0)$. The major difficulty lies in the fact that the dimensions of $\text{vec}(\boldsymbol{\beta}_1^0)$ and $\text{vec}(F_1^0)$ all increase to infinity as $(N,T) \to \infty$ so that the usual identification arguments (uniform convergence along with identification uniqueness) do not apply. In fact, for the factor matrix F_1^0, we are not able to identify the matrix itself but instead $P_{F_1^0}$, which indicates the space spanned by the columns of F_1^0. Despite these difficulties, we argue here that the identification of the $\beta_{1,i}^0$'s is buried in the proof of Theorem 3.1 in the paper.

To proceed, recall that $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2)$, $\mathbf{b}_l = (b_{l,1}, ..., b_{l,N})^{\prime}$ and $b_{l,1} = \beta_{l,i} - \beta_{l,i}^0$ for $l = 1,2$ and

$i = 1, ..., N$. As in Bai (2009, p.1264) and Su and Ju (2018, Proof of Theorem 3.1), it is easy to argue that the objective function cannot achieve its minimum for very large value of $\frac{1}{N}\|\mathbf{b}\|^2$ so that there is no loss of generality to restrict our attention on the case where $\frac{1}{N}\|\mathbf{b}\|^2 \leq M$ and M is a large positive constant that does not grow with N or T. Recall that $Q_{NT}^{\lambda,K}(\boldsymbol{\beta}, \boldsymbol{\alpha}, F_1) = Q_{NT}(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, F_1) + \frac{\lambda}{N}\sum_{i=1}^N \prod_{k=1}^K \|\beta_{1,i} - \alpha_k\|$ and $S_{NT}(\boldsymbol{\beta}_1, F_1) = \frac{1}{N}\sum_{i=1}^N S_{i,NT}(\beta_{1,i}, F_1)$, where

$$\begin{aligned} Q_{NT}(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, F_1) &= \frac{1}{NT^2}\sum_{i=1}^N \left(y_i - x_{1,i}\beta_{1,i} - x_{2,i}\beta_{2,i}\right)' M_{F_1}\left(y_i - x_{1,i}\beta_{1,i} - x_{2,i}\beta_{2,i}\right) \text{ and} \\ S_{i,NT}(\beta_{1,i}, F_1) &= \frac{1}{T^2}(x_{1,i}b_{1,i} - F_1^0\lambda_{1i}^0)' M_{F_1}(x_{1,i}b_{1,i} - F_1^0\lambda_{1i}^0). \end{aligned}$$

Apparently, if we have **homogenous** panels as in Bai (2009), then we can write $b_{l,i} = \beta_{l,i} - \beta_{l,i}^0 = b_l$ (and similarly $\beta_{l,i} = \beta_l$ and $\beta_{l,i}^0 = \beta_l^0$) to obtain

$$\begin{aligned} S_{NT}(\boldsymbol{\beta}_1, F_1) &= b_1'\frac{1}{NT^2}\sum_{i=1}^N x_{1,i}' M_{F_1} x_{1,i} b_1 + \operatorname{tr}\left\{\frac{F_1^0 M_{F_1} F_1^0}{T^2}\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right\} - 2b_1'\frac{1}{NT^2}\sum_{i=1}^N x_{1,i}' M_{F_1} F_1^0 \lambda_{1i}^0 \\ &= b_1' A b_1 + \eta' B \eta - 2b_1' C' \eta, \end{aligned} \tag{C.1}$$

where $A = \frac{1}{NT^2}\sum_{i=1}^N x_{1,i}' M_{F_1} x_{1,i}$, $B = (\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N} \otimes I_T)$, $C = \frac{1}{NT}\sum_{i=1}^N \lambda_{1i}^0 \otimes M_{F_1} x_{1,i}$, and $\eta = \text{vec}\left(M_{F_1}F_1^0\right)/T$. Note that we suppress the dependence of A, C and η on F_1. Completing the squares, we have

$$S_{NT}(\boldsymbol{\beta}_1, F_1) = b_1' D\left(F_1\right) b_1 + \theta' B \theta, \tag{C.2}$$

where $D\left(F_1\right) = A - C'B^{-1}C$, and $\theta = \eta - B^{-1}Cb_1$. Then, under the key identification condition

$$\inf_{F_1 \in \mathcal{F}_1} \mu_{\min}(D\left(F_1\right)) \geq \underline{c} \text{ for some constant } \underline{c} > 0 \tag{C.3}$$

where $\mathcal{F}_1 = \{F_1 \in \mathbb{R}^{T \times r} : \frac{1}{T^2}F_1'F_1 = I_{r_1}\}$, we can follow Bai (2009) to first establish the consistency of the estimator of the finite dimensional parameter β_l^0 and then establish the consistency of the estimator of $P_{F_1^0}$. As Bai (2009) remarks, the identification condition in (C.3) rules out common regressors and time-invariant regressors. He discusses how to relax the condition in (C.3) to

$$\mu_{\min}(D\left(F_1^0\right)) \geq \underline{c} \text{ for some constant } \underline{c} > 0, \tag{C.4}$$

such that both time-invariant and common regressors can be allowed in the regression provided that they do not form collinearity with the common factors or factor loadings. The discussion essentially hinges on the analysis of the expression of $S_{NT}(\boldsymbol{\beta}_1, F_1)$ in (C.2). As one can imagine, similar relaxations would hold for our nonstationary panels if the slope coefficients were indeed homogeneous.

Below we first outline the major challenges in the formal establishment of the identification conditions and then explain how we establish the consistency result in Theorem 3.1 with the implicit use of the identification conditions. Note that even in the stationary homogenous panel, Bai (2009) only considers the latter directly.

By the proof of Theorem 3.1, we have

$$Q_{NT}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, F_1) - Q_{NT}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, F_1^0) = S_{NT}(\boldsymbol{\beta}_1, F_1) + O_P((T/d_T)^{-1/2}),$$

where $O_P((T/d_T)^{-1/2})$ holds uniformly in $(\mathbf{b}, F_1) \in \{\mathbf{b} \in \mathbb{R}^{(p_1+p_2)\times N}, F_1 \in \mathbb{R}^{T\times r_1} : \frac{1}{T^2}F_1'F_1 = I_{r_1}$ and $\frac{1}{N}\|\mathbf{b}\|^2 \le M\}$. Since we restrict our attention to the case where $\{\beta_{1,i}^0\}$ form into some finite K groups, they are be regarded as uniformly bounded. As a result, we can restrict the parameter space for $\beta_{1,i}$ and α_k to be bounded so that $\frac{1}{N}\sum_{i=1}^N \prod_{k=1}^K \|\beta_{1,i} - \alpha_k\| = O(1)$ uniformly in $(\boldsymbol{\beta}_1\ \boldsymbol{\alpha})$. Then

$$
\begin{aligned}
& Q_{NT}^{K,\lambda}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}, F_1) - Q_{NT}^{K,\lambda}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, \boldsymbol{\alpha}^0, F_1^0) \\
= \; & \frac{1}{N}\sum_{i=1}^N [Q_{NT,i}(\beta_{1,i}, \beta_{2,i}, F_1) - Q_{NT,i}(\beta_{1,i}^0, \beta_{2,i}^0, F_1^0)] + \frac{\lambda}{N}\sum_{i=1}^N \prod_{k=1}^K \|\beta_{1,i} - \alpha_k\| \\
= \; & S_{NT}(\boldsymbol{\beta}_1, F_1) + O_P((T/d_T)^{-1/2})
\end{aligned}
$$

where we also apply the fact that $\lambda = o\left(T^{-1/2}\right)$ under Assumption 3.3(iv) to obtain the last equality.

Apparently, $S_{i,NT}(\beta_{1,i}, \beta_{2,i}, F_1) \ge 0$ and it attains its unique minimum value 0 at $(\beta_{1,i}, F_1) = (\beta_{1,i}^0, F_1^0)$. Similarly, $S_{NT}(\boldsymbol{\beta}_1, F_1)$ attains its unique minimum value 0 at $(\boldsymbol{\beta}_1, F_1) = (\boldsymbol{\beta}_1^0, F_1^0)$. We show that $(\boldsymbol{\beta}_1^0, F_1^0)$ is the unique point at which $S_{NT}(\boldsymbol{\beta}_1, F_1)$ achieves its minimum, where uniqueness with respect to F_1^0 is up to a rotation as in the stationary case. This is because $M_{F_1^0 H_1} = M_{F_1^0}$ and $S_{NT}(\boldsymbol{\beta}_1^0, F_1^0 H_1) = 0$ for any nonsingular matrix H_1. For ease of discussion, we assume that $F_1^0 \in \mathcal{F}_1$ (otherwise, we can always focus on its rotational version such that $F_1^0 H_1 \in \mathcal{F}_1$). Let

$$
(\boldsymbol{\beta}_1^*, F_1^*) = \underset{\boldsymbol{\beta}_1, F_1 \in \mathcal{F}_1}{\arg\min}\, S_{NT}(\boldsymbol{\beta}_1, F_1).
$$

We need to show that $(\boldsymbol{\beta}_1^*, F_1^*) = (\boldsymbol{\beta}_1^0, F_1^0)$. We consider three cases: (1) $F_1^* = F_1^0$, (2) $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_1^0$, and (3) $F_1^* \ne F_1^0$ and $\boldsymbol{\beta}_1^* \ne \boldsymbol{\beta}_1^0$.

In Case (1), we argue that if $F_1^* = F_1^0$, then we must have $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_1^0$. In the case of $F_1^* = F_1^0$, we have

$$
\begin{aligned}
0 \; &= \; S_{NT}(\boldsymbol{\beta}_1^*, F_1^*) = S_{NT}(\boldsymbol{\beta}_1^*, F_1^0) = \frac{1}{N}\sum_{i=1}^N \frac{1}{T^2}(x_{1,i}b_{1,i}^* - F_1^0\lambda_{1i}^0)' M_{F_1^0}(x_{1,i}b_{1,i}^* - F_1^0\lambda_{1i}^0) \\
&= \; \frac{1}{N}\sum_{i=1}^N b_{1,i}^{*\prime}\left(\frac{1}{T^2}x_{1,i}' M_{F_1^0} x_{1,i}\right) b_{1,i}^*.
\end{aligned}
$$

Consequently, we must have $b_{1,i}^{*\prime}\left(\frac{1}{T^2}x_{1,i}' M_{F_1^0} x_{1,i}\right) b_{1,i}^* = 0$ for each i. The identification condition in Assumption 3.2(iv) is more sufficient to ensure $\frac{1}{T^2}x_{1,i}' M_{F_1^0} x_{1,i}$ to be uniformly asymptotically positive definite. As a result, we must have $b_{1,i}^* = \beta_{1,i}^* - \beta_{1,i}^0 = 0$ for all i. That is, $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_1^0$.

In Case (2), we argue that if $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_1^0$, then we must have $F_1^* = F_1^0$. In the case of $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_1^0$, we have

$$
0 = S_{NT}(\boldsymbol{\beta}_1^*, F_1^*) = S_{NT}(\boldsymbol{\beta}_1^0, F_1^*) = \frac{1}{N}\sum_{i=1}^N \frac{1}{T^2}\lambda_{1i}^{0\prime} F_1^{0\prime} M_{F_1^*} F_1^0 \lambda_{1i}^0 = \operatorname{tr}\left(\frac{1}{T^2}F_1^{0\prime} M_{F_1^*} F_1^0 \frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0\right).
$$

Assumption 3.2(i) ensures that $\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0$ is asymptotically positive definite. It follows that $F_1^{0\prime} M_{F_1^*} F_1^0 = 0$ or equivalently $M_{F_1^*} F_1^0 = 0$. Then we must have $F_1^* = F_1^0$.

Now, we consider Case (3). Suppose that $F_1^* \neq F_1^0$ and $\boldsymbol{\beta}_1^* \neq \boldsymbol{\beta}_1^0$. Let $\chi_i = x_{1,i} b_{1,i}^* - F_1^0 \lambda_{1i}^0$. Then

$$0 = S_{NT}(\boldsymbol{\beta}_1^*, F_1^*) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \chi_i' M_{F_1^*} \chi_i. \tag{C.5}$$

Observe that $M_{F_1^*} \chi_i$ denotes the residual vector in the least squares projection of χ_i onto F_1^* :

$$\chi_i = F_1^* \pi_{1i} + \hat{v}_i \ \forall i$$

where $\pi_{1i} = (F_1^{*\prime} F_1^*)^{-1} F_1^{*\prime} \chi_i = F_1^{*\prime} \chi_i / T^2$ and $\hat{v}_i = M_{F_1^*} \chi_i$. (C.5) implies that $\hat{v}_i = 0 \ \forall i$ so that $x_{1,i} b_{1,i}^* - F_1^0 \lambda_{1i}^0 = F_1^* \pi_i \ \forall i$, or equivalently

$$x_{1,i} b_{1,i}^* = (F_1^0, F_1^*) \begin{pmatrix} \lambda_{1i}^0 \\ \pi_{1i} \end{pmatrix} \ \forall i.$$

But by the identification condition in Assumption 3.2(iv), the above system of equations can hold only if $b_{1,i}^* = 0$ and $F_1^* \pi_{1i} = -F_1^0 \lambda_{1i}^0 \ \forall i$ (implying that $F_1^* = F_1^0$ and $\pi_{1i} = -\lambda_{1i}^0$). Thus a contradiction arises and we cannot have $F_1^* \neq F_1^0$ and $\boldsymbol{\beta}_1^* \neq \boldsymbol{\beta}_1^0$.

D The PPC-based Estimation Procedure

In this appendix, we provide more details on the practical implementation of the PPC-based estimation procedure. It consists of five steps.

1. **Obtain the initial estimates.** By setting $r_1 = r_{\max}$, we obtain the initial estimates $\tilde{\beta}_{1,i}$, $\tilde{\beta}_{2,i}$ and $\tilde{F}_1$ from the following set of nonlinear equations:

$$\tilde{\beta}_i = \left(\tilde{\beta}_{1,i}', \tilde{\beta}_{2,i}'\right)' = \left(x_i' M_{\tilde{F}_1} x_i\right)^{-1} x_i' M_{\tilde{F}_1} y_i,$$

$$\tilde{F}_1 \tilde{V}_{1,NT} = \left[\frac{1}{NT^2} \sum_{i=1}^{N} (y_i - x_i \tilde{\beta}_i)(y_i - x_i \tilde{\beta}_i)\right] \tilde{F}_1,$$

where $M_{\tilde{F}_1} = I_T - \frac{1}{T^2} \tilde{F}_1 \tilde{F}_1'$, $\frac{1}{T^2} \tilde{F}_1' \tilde{F}_1 = I_{r_1}$, and $\tilde{V}_{1,NT}$ is a diagonal matrix.

2. **Determine the number of common factors.** We separately determine the number of nonstationary factors and stationary factors.

   (a) Determine the number of nonstationary common factors by choosing r_1 to minimize the following information criterion (IC)

$$IC_1(r_1) = \log V_1(r_1, \hat{F}_1^{r_1}) + r_1 g_1(N, T),$$

   where $V_1(r_1, \hat{F}_1^{r_1}) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \hat{\beta}_{1,i}' x_{1,it} - \hat{\beta}_{2,i}' x_{2,it} - \hat{\lambda}_{1i}^{r_1 \prime} \hat{f}_{1t}^{r_1})^2$, $g_1(N, T) = \alpha_T g_2(N, T)$, and $\alpha_T = \frac{T}{4 \log \log T}$.

   (b) Determine the number of stationary common factors by choosing r_2 to minimize the following IC

$$IC_2(r) = \log V_2(r_2, \hat{F}_2^{r_2}) + r_2 g_2(N, T),$$

where $V_2(r_2, \hat{F}_2^{r_2}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{r}_{it} - \hat{\lambda}_{2.i}^{r_2\prime}\hat{f}_{2t}^{r_2})^2$, $\hat{r}_{it} = y_{it} - \hat{\beta}_{1,i}^{\prime}x_{1,it} - \hat{\beta}_{2,i}^{\prime}x_{2,it} - \hat{\lambda}_{1i}^{\prime}\hat{f}_{1t}$ and $g_2(N,T) = \frac{N+T}{NT}\log\left(C_{NT}^2\right)$ or $g_2(N,T) = \frac{N+T}{NT}\log\left(\frac{N+T}{NT}\right)$ as in Bai and Ng (2002).

3. **Determining the number of groups.** Let $\Lambda \equiv \left\{\lambda = c_j T^{-3/4},\ c_j = c_0\gamma^j \text{ for } j = 0, ..., J\right\}$ for some $c_0 > 0$ and $\gamma > 1$. Given any $K \in \{1, 2, ..., K_{\max}\}$ and $\lambda \in \Lambda$, compute $\mathrm{IC}_3(\hat{K}(\lambda), \lambda)$, where $\hat{K}(\lambda) = \arg\min_{1\leq K\leq K_{\max}}\mathrm{IC}_3\left(K,\lambda\right)$. Choose $\hat{\lambda} \in \Lambda$ such that $\mathrm{IC}_3(\hat{K}\left(\lambda\right), \lambda)$ is minimized. Estimate the number of group by $\hat{K} = \min_{\lambda\in\Lambda}\hat{K}(\lambda)$ as recommended by Su, Shi, and Phillips (2016a). We find in simulations $c_0 = 0.05$, $\gamma = 2$, and $J = 3$ work fairly well for all DGPs under our investigation. If $\hat{K} = 1$, stop here and estimate a homogeneous nonstationary panel as usual. Otherwise, move to the next step.

4. **PPC-based estimation.**

   (a) Given $\lambda = \lambda(N,T) =$ and $\hat{K} > 1$, $\hat{r}_1$ and $\hat{r}_2$, solve the following PPC criterion function to obtain estimates of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$:

   $$Q_{NT}^{\lambda,K}(\boldsymbol{\beta}, \boldsymbol{\alpha}, F_1) = Q_{NT}(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, F_1) + \frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\left\|\beta_{1,i} - \alpha_k\right\|,$$

   where $Q_{NT}(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, F_1) = \frac{1}{NT^2}\sum_{i=1}^{N}\left(y_i - x_{1,i}\beta_{1,i} - x_{2,i}\beta_{2,i}\right)^{\prime} M_{F_1}\left(y_i - x_{1,i}\beta_{1,i} - x_{2,i}\beta_{2,i}\right)$, and $\lambda = \lambda(N,T)$ is a tuning parameter.

   (b) Given C-Lasso estimates $(\hat{\alpha}_k, \hat{\beta}_1, \hat{\beta}_2)$, solve the following eigen-decomposition equation to obtain estimates of F_1

   $$\hat{F}_1 V_{1,NT} = \left[\frac{1}{NT^2}\sum_{i=1}^{N}(y_i - x_{1,i}\hat{\beta}_{1,i} - x_{2,i}\hat{\beta}_{2,i})(y_i - x_{1,i}\hat{\beta}_{1,i} - x_{2,i}\hat{\beta}_{2,i})^{\prime}\right]\hat{F}_1,$$

   where $\frac{1}{T^2}\hat{F}_1^{\prime}\hat{F}_1 = I_{r_1}$ and $V_{1,NT}$ is a diagonal matrix.

   (c) Given the estimates $\hat{\beta}_{1,i}, \hat{\alpha}_k$, and $\hat{F}_1$, we obtain the cointegration residuals $\hat{r}_{it} = y_{it} - \hat{\beta}_{1,i}^{\prime}x_{1,it} - \hat{\lambda}_{1i}^{\prime}\hat{f}_{1t}$. The LS estimator of $(\hat{\beta}_{2,i}, \hat{F}_2)$ is the solution to the following set of nonlinear equations:

   $$\hat{\beta}_{2,i} = \left(x_{2,i}^{\prime}M_{\hat{F}_2}x_{2,i}\right)^{-1}x_{2,i}^{\prime}M_{\hat{F}_2}\hat{r}_i,$$
   $$\hat{F}_2\tilde{V}_{2,NT} = \left[\frac{1}{NT}\sum_{i=1}^{N}(\hat{r}_i - x_{2,i}\hat{\beta}_{2,i})(\hat{r}_i - x_{2,i}\hat{\beta}_{2,i})^{\prime}\right]\hat{F}_2,$$

   where $\frac{1}{T}\hat{F}_2^{\prime}\hat{F}_2 = I_{r_2}$ and $V_{2,NT}$ is a diagonal matrix.

   (d) Iterate above steps until convergence and obtain jointly $(\hat{\beta}_{1,i}, \hat{\beta}_{2,i}, \hat{\alpha}_k, \hat{F}_1, \hat{F}_2)$. Obtain the C-Lasso estimates $\{\hat{\alpha}_k\}$ for the group-specific parameters and $\{\hat{G}_k, k = 1, ..., \hat{K}\}$ for the estimated group membership.

5. **Post-Lasso estimator with bias correction.**

   (a) Given the estimated groups, $\{\hat{G}_k, k = 1, ..., \hat{K}\}$, we obtain the continuous updated estimators $\hat{\alpha}^{cup}_{\hat{G}_k}$, $\hat{F}_1$ and $\hat{F}_2$ by iteratively solving the following equations:

$$\hat{\alpha}^{fm}_{\hat{G}_k} = \left(\sum_{i\in\hat{G}_k} x'_{1,i} M_{\hat{F}_1} x_{1,i}\right)^{-1} \left\{\sum_{i\in\hat{G}_k} x'_{1,i} M_{\hat{F}_1} \hat{y}_i^{+} - T\sqrt{N_k}\left(\hat{B}^{+}_{kNT,1} + \hat{B}_{kNT,2}\right)\right\},$$

$$\hat{F}_1 V_{1,NT} = \left[\frac{1}{NT^2}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}(\hat{y}_i - x_{1,i}\hat{\alpha}^{fm}_{\hat{G}_k} - x_{2,i}\hat{\beta}_{2,i})(\hat{y}_i - x_{1,i}\hat{\alpha}^{fm}_{\hat{G}_k} - x_{2,i}\hat{\beta}_{2,i})'\right]\hat{F}_1,$$

$$\hat{F}_2 V_{2,NT} = \left[\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}(\hat{y}_i - x_{1.i}\hat{\alpha}^{fm}_{\hat{G}_k} - x_{2,i}\hat{\beta}_{2,i} - \hat{F}_1\hat{\lambda}_{1i})(\hat{y}_i - x_{1,i}\hat{\alpha}^{fm}_{\hat{G}_k} - x_{2,i}\hat{\beta}_{2,i} - \hat{F}_1\hat{\lambda}_{1i})'\right]\hat{F}_2,$$

where $\hat{B}_{NT,l} = (\hat{B}'_{1NT,l}, ..., \hat{B}'_{KNT,l})'$ for $l = 1, 2$, $\hat{B}_{kNT,1} = \frac{1}{\sqrt{\hat{N}_k T}}\sum_{i\in\hat{G}_k}(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\bar{\varkappa}}_{ts})\hat{\Delta}_{21,i}$, $\hat{B}_{kNT,2} = \frac{1}{\sqrt{\hat{N}_k T}}\sum_{i\in\hat{G}_k}(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\bar{\varkappa}}_{ts})\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i}$, $\hat{B}^{+}_{kNT,1} = \frac{1}{\sqrt{\hat{N}_k T}}\sum_{i\in\hat{G}_k}(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\bar{\varkappa}}_{ts})\hat{\Delta}^{+}_{21,i}$, $\hat{\bar{\varkappa}}_{ts} = \mathbf{1}\{t = s\} - \hat{\varkappa}_{ts}$, $\hat{\varkappa}_{ts} = \hat{f}'_{1t}(\hat{F}'_1\hat{F}_1)^{-1}\hat{f}_{1s} = \hat{f}'_{1t}\hat{f}_{1s}/T^2$, $\hat{\bar{\lambda}}_{2i} = \hat{\lambda}_{2i} - \frac{1}{N}\sum_{j=1}^{N}\hat{\lambda}_{2j}\hat{a}_{ij}$, and $\hat{a}_{ij} = \hat{\lambda}'_{1i}(\frac{1}{N}\hat{\Lambda}'_1\hat{\Lambda}_1)^{-1}\hat{\lambda}_{1j}$. Note that $\hat{F}_1$, $V_{1,NT}$, $\hat{F}_2$, $V_{2,NT}$, and $\{\hat{\lambda}_{1i}, \hat{\lambda}_{2i}\}$ are also updated continuously in the procedure to obtain $\hat{\alpha}^{cup}_{\hat{G}_k}$.

(b) Estimate Ω_{NT} and Ω^{+}_{NT} consistently by

$$\hat{\Omega}_{NT} = \frac{\hat{D}_{NK}}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{X}}_{it}\hat{\mathbf{X}}'_{is}\hat{u}^{*}_{it}\hat{u}^{*}_{is} - \sum_{i=1}^{N}\hat{B}_{iNT}\hat{B}'_{iNT},$$

$$\hat{\Omega}^{+}_{NT} = \frac{\hat{D}_{NK}}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{X}}_{it}\hat{\mathbf{X}}'_{is}\hat{u}^{*+}_{it}\hat{u}^{*+}_{is} - \sum_{i=1}^{N}\hat{B}^{+}_{iNT}\hat{B}^{+\prime}_{iNT},$$

where $\hat{\mathbf{X}}_{it} = (\hat{\mathbf{X}}'_{1,it}, ..., \hat{\mathbf{X}}'_{K,it})'$, $\hat{\mathbf{X}}'_{k,it}$ is the tth row of $\hat{\mathbf{X}}_{k,i}$, $\hat{\mathbf{X}}_{k,i} = M_{\hat{F}_1}x_{1,i}\mathbf{1}\{i \in \hat{G}_k\} - \frac{1}{N}\sum_{j\in\hat{G}_k}\hat{a}_{ij}M_{\hat{F}_1}x_{1,j}$, $\hat{D}_{NK} = \mathrm{diag}\,(\frac{N}{\hat{N}_1}, ..., \frac{N}{\hat{N}_K})\otimes I_p$, $\hat{B}_{iNT} = (\hat{B}'_{1,iNT}, ..., \hat{B}'_{K,iNT})'$, $\hat{B}_{k,iNT} = \hat{B}_{k,iNT,1} + \hat{B}_{k,iNT,2}$, $\hat{B}_{k,iNT,1} = \frac{1}{\sqrt{\hat{N}_k T}}(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\bar{\varkappa}}_{ts})\hat{\Delta}_{21,i}\mathbf{1}\{i \in \hat{G}_k\}$, $\hat{B}_{k,iNT,2} = \frac{1}{\sqrt{\hat{N}_k T}}$ $(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\bar{\varkappa}}_{ts})\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i}\mathbf{1}\{i \in \hat{G}_k\}$, $\hat{u}^{*}_{it} = y_{it} - \hat{\alpha}^{fm\prime}_{k}x_{it} - \hat{\beta}'_{2,i}x_{2,it} - \hat{\lambda}'_{1i}\hat{f}_{1t}$ for $i \in \hat{G}_k$, $\hat{B}^{+}_{iNT} = (\hat{B}^{+\prime}_{1,iNT}, ..., \hat{B}^{+\prime}_{K,iNT})'$, $\hat{B}^{+}_{k,iNT} = \hat{B}^{+}_{k,iNT,1} + \hat{B}_{k,iNT,2}$, $\hat{B}^{+}_{k,iNT,1} = \frac{1}{\sqrt{\hat{N}_k T}}(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\bar{\varkappa}}_{ts})$ $\hat{\Delta}^{+}_{21,i}\mathbf{1}\{i \in \hat{G}_k\}$, and $\hat{u}^{*+}_{it} = \hat{y}^{+}_{it} - \hat{\alpha}^{fm\prime}_{k}x_{1,it} - \hat{\beta}'_{2,i}x_{2,it} - \hat{\lambda}'_{1i}\hat{f}_{1t}$ for $i \in \hat{G}_k$.

E Some Additional Simulation Results

In this appendix, we report some additional simulation results for DGPs 1-6. In addition, we follow the editor's suggestion and consider two additional DGPs, namely DGPs 7-8, to closely mimic the empirical application.

E.1 Additional simulation results for DGPs 1-6

First, we consider the performance of our classification and estimation procedure for DGPs 1-6 when $N = 200$ and $T = 40$. Here N and T differ to a larger extent than their values in Tables 3-5 in the paper. The results are reported in Table A.1. Comparing the results in Table A.1 with those in Tables 3-5 suggests that our post-Lasso estimates (bias corrected or fully modified) and Cup-Lasso estimates perform qualitatively similarly to those in Tables 3-5.

Now we consider two DGPs that mimic the data in the empirical applications where the sample sizes, $(N,T)=(24,34)$, are relatively small. We now consider $(N,T)=c\cdot(24,34)$ for $c=1,\,2,\,3$. By increasing the value of c from 1 to 3, we should be able to observe the improved performance of our estimators. We generate the data as follows:

$$\begin{cases} y_{it}=\beta_{1,i}x_{1,it}+\beta_{2,i}x_{2,it}+\beta_{3,i}x_{3,it}+c_1\lambda_{1i}'f_{1t}+u_{it} \\ x_{it}=x_{it-1}+\varepsilon_{it} \\ f_{1t}=f_{1,t-1}+\nu_t \end{cases}, \tag{E.1}$$

where $i=1,...,N$, $t=1,...,T$, the dimension of f_{1t} is $r_1=1$, and $x_{it}=(x_{1,it},x_{2,it})'$ in DGP 7 and $x_{it}=(x_{1,it},x_{2,it},x_{3,it})'$ in DGP 8 below.

DGP 7 *(Mimicking Model (5.1) in Table 7)* The observations are drawn from three groups with $N_1:N_2:N_3=7:7:10$ such that $N=\sum_{j=1}^{3}N_j=24c$ and $T=34c$ for $c=1,2,3$. Let $\beta_{3,i}=0$ and $c_1=0.5$ in (E.1). The factor loadings λ_{1i} are i.i.d. $\lambda_{1i}\sim N(0.1,1)$ and $\mu_\lambda=0.1$. Let $\beta_i=(\beta_{1,i},\beta_{2,i})'$. The long-run slope coefficients β_i exhibit the group structure in (2.3) for $K=3$ and the true values for the group-specific parameters are

$$(\alpha_1^0,\alpha_2^0,\alpha_3^0)=\left(\begin{pmatrix}0.289\\-0.147\end{pmatrix},\begin{pmatrix}0.101\\0.120\end{pmatrix},\begin{pmatrix}0.058\\0.086\end{pmatrix}\right)$$

which are as estimated for Model (5.1) from the real empirical data in our applications. The errors $w_{it}=(u_{it},\varepsilon_{it}',\Delta f_{1t}')'$ are generated from the linear process $w_{it}=\sum_{j=0}^{\infty}\phi_{ij}v_{i,t-j}$, where

$$\phi_{ij}=L(j)\Omega^{1/2},\ L(j)=j^{-3.5},\ \Omega=\begin{pmatrix}0.25&\Omega_{12}&\Omega_{13}\\\Omega_{21}&\Omega_{22}&\Omega_{23}\\0&0_{1\times2}&\Omega_{33}\end{pmatrix},\ v_{it}=(v_{it}^{u\varepsilon\prime},v_t^{f_1\prime},v_t^{f_2\prime})',\ v_{it}^{u\varepsilon}\sim\text{i.i.d. }N(0,I_3),$$

and $v_t^{f_1}\sim$ i.i.d. $N(0,1)$. Let $\Omega_{12}=\Omega_{21}'=\Omega_{23}'=\begin{pmatrix}0.2,0.2\end{pmatrix}$, $\Omega_{13}=0.2$, $\Omega_{22}=\begin{pmatrix}1&0.2\\0.2&1\end{pmatrix}$, and $\Omega_{33}=1$.

DGP 8 *(Mimicking Model (5.2) in Table 7)* The observations are drawn from three groups with $N_1:N_2:N_3=7:7:10$ such that $N=\sum_{j=1}^{3}N_j=24c$ and $T=34c$ for $c=1,2,3$. Let $c_1=0.5$ in (E.1). The factor loadings λ_{1i} are i.i.d. $\lambda_{1i}\sim N(0.1,1)$. Let $\beta_i=(\beta_{1,i},\beta_{2,i},\beta_{3,i})'$. The long-run slope coefficients β_i exhibit the group structure in (2.3) for $K=3$ and the true values for the group-specific parameters are

$$(\alpha_1^0,\alpha_2^0,\alpha_3^0)=\left(\begin{pmatrix}0.464\\-0.413\\1.405\end{pmatrix},\begin{pmatrix}0.055\\0.022\\0.550\end{pmatrix},\begin{pmatrix}-0.104\\0.219\\0.567\end{pmatrix}\right)$$

which are as estimated for Model (5.2) from the real empirical data in our applications. The errors $w_{it}=(u_{it},\varepsilon_{it}',\Delta f_{1t}')'$ are generated from the linear process $w_{it}=\sum_{j=0}^{\infty}\phi_{ij}v_{i,t-j}$, where

$$\phi_{ij} = L(j)\Omega^{1/2},\ L(j) = j^{-3.5},\ \Omega = \begin{pmatrix} 0.25 & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ 0 & 0_{1\times 3} & \Omega_{33} \end{pmatrix},\ v_{it} = (v_{it}^{u\varepsilon\prime}, v_t^{f_1\prime}, v_t^{f_2\prime})',\ v_{it}^{u\varepsilon} \sim \text{i.i.d. } N(0, I_4),$$

and $v_t^{f_1} \sim$ i.i.d. $N(0,1)$. Let $\Omega_{12} = \Omega_{21}' = \Omega_{23}' = \Big(0.2, 0.2, 0.2\Big)$, $\Omega_{13} = 0.2$, $\Omega_{22} = \begin{pmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{pmatrix}$, and $\Omega_{33} = 1$.

Table A.2 reports the simulation results for DGPs 7-8. We summarize the main findings from Table A.2. First, the classification result is not as good as those in Tables 1-5 when $(N,T) = (24, 34)$. This is as expected as on average we have only 8 individuals in each group and the large sample theory cannot work very well in such as case. But as both N and T increase, we observe that the classification results improve quickly. Second, the Cup-Lasso estimator generally performs better than the two post-Lasso estimators and thus it is recommended for empirical applications. In particular, as both N and T increases, the performance of all estimators improve and the coverage of the Cup-Lasso estimator gets closer to the oracle one.

REFERENCES

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

Bai, J., 2004. Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics* 122, 137-183.

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

Bernstein, D. S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory.* Princeton University Press, Princeton.

Donsker, M.D., Varadhan, S.R.S., 1977. On laws of the iterated logarithm for local times. *Communications on Pure and Applied Mathematics* 30, 707-753.

Huang, W., Jin, S., Su, L., 2020. Identifying latent group patterns in cointegrated panels. *Econometric Theory*, forthcoming.

Lai, T.L., Wei, C.Z., 1982. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics*, 154-166.

Park, J. Y., Phillips, P. C., 1988. Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4, 468-497.

Phillips, P.C., 1995. Fully modified least squares and vector autoregression. *Econometrica* 63, 1023-1078.

Phillips, P.C., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057-1111.

Phillips, P.C., Solo, V., 1992. Asymptotics for linear processes. *The Annals of Statistics* 20, 971-1001.

Su, L., Ju, G., 2018. Identifying latent group patterns in panel data models with interactive fixed effects. *Journal of Econometrics* 206, 554-573.

Su, L., Shi, Z., Phillips, P. C., 2016a. Identifying latent structures in panel data. *Econometrica* 84, 2215-2264.

Su, L., Shi, Z., Phillips, P. C., 2016b. Supplement to "Identifying latent structures in panel data." *Econometrica Supplemental Material* 84, http://dx.doi.org/10.3982/ECTA12560.

Table A.1 Classification and point estimation of α_1 in DGPs 1-6

		c_λ	0.1				0.2			
N	T		% Correct classification	RMSE	Bias	% Coverage	% Correct classification	RMSE	Bias	% Coverage
DGP 1										
200	40	C-Lasso	99.99	0.0039	0.0001	93.20	99.98	0.0038	0.0001	94.92
		post-Lasso^{bc}	99.99	0.0039	0.0000	94.34	99.98	0.0039	0.0000	94.34
		post-Lasso^{fm}	99.99	0.0039	0.0000	94.34	99.98	0.0039	0.0000	94.34
		Cup-Lasso	99.99	0.0039	0.0000	94.34	99.98	0.0039	0.0000	94.34
		Oracle	-	0.0039	0.0000	95.10	-	0.0039	0.0000	95.10
DGP 2										
200	40	C-Lasso	99.98	0.0065	0.0052	62.66	99.97	0.0060	0.0048	67.82
		post-Lasso^{bc}	99.98	0.0038	0.0003	93.32	99.97	0.0038	0.0003	93.32
		post-Lasso^{fm}	99.98	0.0038	0.0003	94.34	99.97	0.0037	0.0003	94.16
		Cup-Lasso	99.98	0.0038	0.0003	94.34	99.97	0.0037	0.0003	94.16
		Oracle	-	0.0037	0.0003	94.26	-	0.0037	0.0003	94.26
DGP 3										
200	40	C-Lasso	98.73	0.0251	0.0144	40.40	98.58	0.0272	0.0133	42.60
		post-Lasso^{bc}	98.73	0.0234	0.0023	87.84	98.58	0.0233	0.0024	87.80
		post-Lasso^{fm}	98.73	0.0234	0.0025	87.22	98.58	0.0234	0.0026	87.36
		Cup-Lasso	100.00	0.0057	0.0023	88.88	99.98	0.0057	0.0024	88.64
		Oracle	-	0.0057	0.0023	88.88	-	0.0057	0.0023	88.88
DGP 4										
200	40	C-Lasso	98.82	0.0230	0.0124	51.02	98.67	0.0245	0.0114	52.76
		post-Lasso^{bc}	98.82	0.0193	0.0019	89.88	98.67	0.0190	0.0018	90.16
		post-Lasso^{fm}	98.82	0.0193	0.0022	89.96	98.67	0.0190	0.0021	89.92
		Cup-Lasso	99.97	0.0072	0.0020	91.06	99.92	0.0071	0.0020	91.02
		Oracle	-	0.0071	0.0020	91.14	-	0.0071	0.0020	91.14
DGP 5										
200	40	C-Lasso	98.69	0.0239	0.0143	40.52	98.59	0.0256	0.0135	42.30
		post-Lasso^{bc}	98.69	0.0222	0.0025	88.72	98.59	0.0215	0.0024	88.86
		post-Lasso^{fm}	98.69	0.0223	0.0027	88.44	98.59	0.0216	0.0026	88.46
		Cup-Lasso	100.00	0.0057	0.0026	89.24	99.98	0.0057	0.0026	89.16
		Oracle	-	0.0057	0.0026	90.84	-	0.0057	0.0026	90.84
DGP 6										
200	40	C-Lasso	99.93	0.0212	0.0191	26.56	99.92	0.0206	0.0183	28.88
		post-Lasso^{bc}	99.93	0.0140	-0.0102	70.66	99.92	0.0140	-0.0103	70.50
		post-Lasso^{fm}	99.93	0.0080	0.0009	93.10	99.92	0.0079	0.0009	93.02
		Cup-Lasso	100.00	0.0075	0.0008	93.16	99.99	0.0075	0.0008	93.24
		Oracle	-	0.0075	0.0008	92.62	-	0.0075	0.0008	92.62

Table A.2 Classification and point estimation of α_1 in DGPs 7-8

		c_λ	0.1				0.2			
N	T		% Correct classification	RMSE	Bias	% Coverage	% Correct classification	RMSE	Bias	% Coverage
DGP 7										
24	34	C-Lasso	77.78	0.0951	0.0156	79.17	76.52	0.0929	0.0145	76.67
		post-Lasso^{bc}	77.78	0.0350	-0.0049	79.57	76.52	0.0427	-0.0026	78.69
		post-Lasso^{fm}	77.78	0.0353	-0.0056	80.74	76.52	0.0430	-0.0032	79.31
		Cup-Lasso	82.78	0.0248	-0.0001	78.23	81.81	0.0254	0.0006	77.13
		Oracle	-	0.0151	0.0009	88.52	-	0.0151	0.0009	88.52
48	68	C-Lasso	86.97	0.0939	0.0072	72.18	87.16	0.0755	0.0069	71.33
		post-Lasso^{bc}	86.97	0.0192	-0.0043	82.72	87.16	0.0201	-0.0033	83.15
		post-Lasso^{fm}	86.97	0.0192	-0.0042	83.45	87.16	0.0202	-0.0032	83.72
		Cup-Lasso	92.94	0.0062	-0.0011	87.37	93.55	0.0059	-0.0007	88.04
		Oracle	-	0.0048	0.0005	92.14	-	0.0048	0.0005	92.14
72	102	C-Lasso	91.57	0.0714	0.0009	67.77	92.02	0.0521	0.0017	67.36
		post-Lasso^{bc}	91.57	0.0147	-0.0034	85.72	92.02	0.0144	-0.0032	86.05
		post-Lasso^{fm}	91.57	0.0147	-0.0033	86.62	92.02	0.0144	-0.0030	86.85
		Cup-Lasso	96.67	0.0030	-0.0005	91.06	97.19	0.0028	-0.0003	91.78
		Oracle	-	0.0025	0.0003	92.42	-	0.0025	0.0003	92.42
DGP 8										
24	34	C-Lasso	82.63	0.0903	0.0136	78.22	79.82	0.0912	0.0164	76.65
		post-Lasso^{bc}	82.63	0.0657	-0.0177	83.58	79.82	0.0694	-0.0137	83.85
		post-Lasso^{fm}	82.63	0.0672	-0.0176	83.25	79.82	0.0713	-0.0131	82.25
		Cup-Lasso	96.39	0.0275	0.0036	83.27	92.13	0.0355	0.0047	80.90
		Oracle	-	0.0241	0.0017	82.72	-	0.0241	0.0017	82.72
48	68	C-Lasso	89.24	0.0553	-0.0013	65.03	86.77	0.0567	0.0027	66.22
		post-Lasso^{bc}	89.24	0.0531	-0.0156	85.82	86.77	0.0515	-0.0135	83.98
		post-Lasso^{fm}	89.24	0.0532	-0.0155	86.52	86.77	0.0515	-0.0134	84.55
		Cup-Lasso	99.56	0.0109	0.0001	90.23	98.12	0.0099	0.0003	89.37
		Oracle	-	0.0056	0.0002	90.47	-	0.0056	0.0002	90.47
72	102	C-Lasso	91.95	0.0379	0.0000	58.53	90.77	0.0393	0.0017	58.80
		post-Lasso^{bc}	91.95	0.0424	-0.0119	88.12	90.77	0.0408	-0.0112	87.90
		post-Lasso^{fm}	91.95	0.0421	-0.0116	87.72	90.77	0.0407	-0.0109	88.15
		Cup-Lasso	99.97	0.0027	0.0003	92.33	99.71	0.0028	0.0003	92.25
		Oracle	-	0.0027	0.0003	92.33	-	0.0027	0.0003	92.33