Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Economics | School of Economics

1-2019

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

Wenxin HUANG

Peter C. B. PHILLIPS

Liangjun SU

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

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

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

[b] Yale University, University of Auckland

University of Southampton & Singapore Management University

[c] School of Economics, Singapore Management University

January 23, 2019

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 the good finite sample performance of the proposed method. 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 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 the funding support provided by the Lee Kong Chian Fund for Excellence. Phillips acknowledges research support from the Kelly Fund at the University of Auckland. 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 seriously distort the reliability of classical methods. The goal of this paper is to study efficient estimation 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 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. 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 in the world. Similarly, both stock market shocks and macro-economic news affect security prices. But it is hard to decide 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 the *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), 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 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 in 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 our Lasso-type estimators. The presence of unobserved common factors complicates our asymptotic analysis in several ways. First, we establish the 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). In addition, our group-specific estimators enjoy the oracle property in the latent group literature, which essentially says that the three bias-corrected estimators are asymptotically equivalent to the corresponding infeasible ones that are obtained with the knowledge of exact individual's group identity.

Since we allow for both contemporaneous and serial correlation in the errors, nonstationary regressors, and unobserved common factors, we have the usual endogeneity bias in nonstationary panels, which originates 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 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. In addition, 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 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 the numbers of groups and common factors are known. For practical work we propose three information criteria to determine the number of groups, the total number of common factors, and the number of nonstationary common factors. 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 the 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 productivities 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 shocks. 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. 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 even when individual group membership is unknown. The penalization method borrows from the *C-Lasso* formulation in SSP, but is modified here by using the principal component method to simultaneously account for cross-section dependence. There are various papers that 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. (2018) 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 non-standard limit theory; see, e.g., Phillips and Moon (1999), Kao and Chiang (2001), and Pedroni (2004). In this context, we maintain the simplicity of asymptotic 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, Bai and Ng (2002, 2004), Phillips and Sul (2003), Pesaran (2006), Bai (2009), and Moon and Weider (2017). Compared with existing work, our approach accommodates both stationary and nonstationary common factors and provides 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 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 Appendix A. Further technical details can be found in the additional online supplement that is available at http://www.mysmu.edu/faculty/ljsu/Publications/HPS19_suppl.pdf.

NOTATION. We write integrals such as $\int_0^1 W(s)ds$ more 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 group structure.

2.1 A nonstationary panel with latent grouping and cross-section dependence

Suppose that (y_{it}, x_{it}) are generated as follows

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

where y_{it} is a scalar, x_{it} is a $p \times 1$ vector of nonstationary regressors of order one ($I(1)$ process) for all i, ε_{it} is assumed to have zero mean and finite long-run variance, and the β_i^0 are $p \times 1$ vectors of parameters that denote long-run cointegration relationships. 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, we assume that u_{it} is 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, following the group formulation in SSP, we assume that the cointegrating vectors β_i^0 are heterogeneous across different groups and homogeneous within a group:

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

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 3.6, we propose an information criterion to determine the number of groups.

If f_t contains only stationary common factors, we may still obtain consistent but typically biased estimators of the long-run relationships involving β_i by the penalized least squares (PLS) method proposed by Huang et al. (2018) without considering the cross-section dependence issue. The bias may arise from the contemporaneous and serial correlations between the innovation processes of the nonstationary regressors x_{it} and the unobserved stationary factors. In contrast, if f_t contains nonstationary factors, the PLS method does not in general yield consistent estimators of β_i due to the presence of spurious regression effects. This complication calls for new estimation methodology.

To proceed, let

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

The true values of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, Λ, Λ_l, F, and F_l are denoted $\boldsymbol{\alpha}^0$, $\boldsymbol{\beta}^0$, Λ^0, Λ_l^0, F^0, and F_l^0, respectively. We also use α_k^0, β_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 $\alpha_k, \beta_i, \lambda_i = (\lambda_{1i}', \lambda_{2i}')'$, 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 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 β_i and unobserved common factors f_t, and to identify the group structure of these long-run relationships.

Combining (2.1)-(2.2) yields

$$y_{it} = \beta_i^{0\prime} x_{it} + \lambda_i^{0\prime} f_t^0 + u_{it} = \beta_i^{0\prime} x_{it} + \lambda_{1i}^{0\prime} f_{1t}^0 + \lambda_{2i}^{0\prime} f_{2t}^0 + u_{it}, \tag{2.4}$$

or in vector form:

$$y_i = x_i\beta_i^0 + F^0\lambda_i^0 + u_i = x_i\beta_i^0 + F_1^0\lambda_{1i}^0 + F_2^0\lambda_{2i}^0 + u_i, \tag{2.5}$$

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

Ideally, one might attempt to estimate both the stationary and nonstationary common components along with the parameters of interest, β_i. But due to the fact that the stationary factors and nonstationary factors 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 still obtain consistent estimates of β_i by taking into account the nonstationary factor component and ignoring the stationary factor component. This motivates the following sequential approach to estimate the unknown parameters in the model. We first estimate the nonstationary factor component along with β_i and then estimate the stationary factor component from the resultant residuals. The stepwise procedure is as follows.

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

$$\mathrm{SSR}(\boldsymbol{\beta}, F_1, \Lambda_1) = \sum_{i=1}^{N}(y_i - x_i\beta_i - F_1\lambda_{1i})'(y_i - x_i\beta_i - F_1\lambda_{1i}) \tag{2.6}$$

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:

$$\tilde{\beta}_i = \left(x_i'M_{\tilde{F}_1}x_i\right)^{-1} x_i'M_{\tilde{F}_1}y_i, \tag{2.7}$$

$$\tilde{F}_1\tilde{V}_{1,NT} = \left[\frac{1}{NT^2}\sum_{i=1}^{N}(y_i - x_i\tilde{\beta}_i)(y_i - x_i\tilde{\beta}_i)'\right]\tilde{F}_1, \tag{2.8}$$

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 (2.8), 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 minimizing the following PPC criterion function to obtain estimates of $(\boldsymbol{\beta}, \boldsymbol{\alpha}, F_1)$:

$$Q_{NT}^{\lambda,K}(\boldsymbol{\beta}, \boldsymbol{\alpha}, F_1) = Q_{NT}(\boldsymbol{\beta}, F_1) + \frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\beta_i - \alpha_k\|, \tag{2.9}$$

where $Q_{NT}(\boldsymbol{\beta}, F_1) = \frac{1}{NT^2}\sum_{i=1}^{N}(y_i - x_i\beta_i)' M_{F_1}(y_i - x_i\beta_i)$, and $\lambda = \lambda(N,T)$ is a tuning parameter. Minimizing the PPC criterion function in (2.9) produces the C-Lasso estimators $(\hat{\beta}_i, \hat{\alpha}_k, \hat{F}_1)$ of (β_i, α_k, F_1) where $\hat{F}_1 = (\hat{f}_{11}, ..., \hat{f}_{1T})'$. Note that

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

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 (2.10), 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 $\hat{\lambda}_{1i}' = \frac{1}{T^2}(y_i - x_i\hat{\beta}_i)'\hat{F}_1$. Define the resulting estimated groups

$$\hat{G}_k = \{i \in \{1, 2, ..., N\} : \hat{\beta}_i = \hat{\alpha}_k\} \text{ for } k = 1, ..., K. \tag{2.11}$$

Step 3. Given the estimates $\hat{\beta}_i$, $\hat{\alpha}_k$, and $\hat{F}_1$, we obtain the estimator of the stationary factor F_2 by $\hat{F}_2$, which solves the following eigen-decomposition problem:

$$\hat{F}_2 V_{2,NT} = \left[\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}(y_i - x_i\hat{\alpha}_k - \hat{F}_1\hat{\lambda}_{1i})(y_i - x_i\hat{\alpha}_k - \hat{F}_1\hat{\lambda}_{1i})'\right]\hat{F}_2, \tag{2.12}$$

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 (2.12), arranged in decreasing order.

Let $\hat{\boldsymbol{\beta}} \equiv (\hat{\beta}_1, ..., \hat{\beta}_N)$ and $\hat{\boldsymbol{\alpha}} \equiv (\hat{\alpha}_1, ..., \hat{\alpha}_K)$. We will study the asymptotic properties of $\hat{\beta}_i$, $\hat{\alpha}_k$, 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.

3 Asymptotic Theory

3.1 Main assumptions

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

$$Q_2(F_1) = \begin{pmatrix} \frac{1}{NT^2} x_1' M_{F_1} x_1 a_{11} & \frac{1}{NT^2} x_1' M_{F_1} x_2 a_{12} & \cdots & \frac{1}{NT^2} x_1' M_{F_1} x_N a_{1N} \\ \frac{1}{NT^2} x_2' M_{F_1} x_1 a_{21} & \frac{1}{NT^2} x_2' M_{F_1} x_2 a_{22} & \cdots & \frac{1}{NT^2} x_2' M_{F_1} x_N a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{NT^2} x_N' M_{F_1} x_1 a_{N1} & \frac{1}{NT^2} x_N' M_{F_1} x_2 a_{N2} & \cdots & \frac{1}{NT^2} x_N' M_{F_1} x_N a_{NN} \end{pmatrix},$$

where F_1 satisfies $\frac{1}{T^2} F_1' F_1 = I_{r_1}$. Note that $Q_2(F_1)$ is an $Np \times Np$ 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 places. Let $w_{it} = (u_{it}, \varepsilon_{it}', \Delta f_{1t}^{0\prime}, f_{2t}^{0\prime},)'$. Let $\Omega_i = \sum_{j=-\infty}^{\infty} \mathbb{E}(w_{ij} w_{i0}')$, 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 $\Delta_i = \sum_{j=0}^{\infty} \mathbb{E}(w_{i0} w_{ij}') = \Gamma_i + \Sigma_i$ of $\{w_{it}\}$. Conformably with w_{it}, Ω_i and Δ_i can be partitioned as follows

$$\Omega_i = \Gamma_i' + \Gamma_i + \Sigma_i = \begin{pmatrix} \Omega_{11,i} & \Omega_{12,i} & \Omega_{13,i} & \Omega_{14,i} \\ \Omega_{21,i} & \Omega_{22,i} & \Omega_{23,i} & \Omega_{24,i} \\ \Omega_{31,i} & \Omega_{32,i} & \Omega_{33} & \Omega_{34} \\ \Omega_{41,i} & \Omega_{42,i} & \Omega_{43} & \Omega_{44} \end{pmatrix} \text{ and } \Delta_i = \begin{pmatrix} \Delta_{11,i} & \Delta_{12,i} & \Delta_{13,i} & \Delta_{14,i} \\ \Delta_{21,i} & \Delta_{22,i} & \Delta_{23,i} & \Delta_{24,i} \\ \Delta_{31,i} & \Delta_{32,i} & \Delta_{33} & \Delta_{34} \\ \Delta_{41,i} & \Delta_{42,i} & \Delta_{43} & \Delta_{44} \end{pmatrix}.$$

Let S_1, S_2, S_3, and S_4 denote, respectively, the $1 \times (1+p+r)$, $p \times (1+p+r)$, $r_1 \times (1+p+r)$, and $r_2 \times (1+p+r)$ selection matrices for which $S_1 w_{it} = u_{it}$, $S_2 w_{it} = \varepsilon_{it}$, $S_3 w_{it} = \Delta f_{1t}^0$, and $S_4 w_{it} = f_{2t}^0$. Let $S_{23} = (S_2', S_3')'$, a $(p + r_1) \times (1+p+r)$ selection matrix.

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 $v_{it} = (v_{it}^u, v_{it}^{\varepsilon\prime}, v_t^{f_1\prime}, v_t^{f_2\prime})'$ is a $(1+p+r_1+r_2) \times 1$ random vector that is i.i.d. over t with zero mean and variance matrix I_{1+p+r}; $\sup_{N\geq 1} \max_{1\leq i\leq N} \mathbb{E}(\|v_{it}\|^{2q+\epsilon}) < M$, where $q > 4$ and ϵ is an arbitrarily small positive constant; v_{it}^u, v_{it}^{ε}, $v_t^{f_1}$, and $v_t^{f_2}$ are mutually independent; and $(v_{it}^u, v_{it}^{\varepsilon\prime})'$ 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) $(u_{it}, \varepsilon_{it})$ are independent across i conditional on $\mathcal{C}$.

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

Following Phillips and Solo (1992; PS), 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:

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

Since nonstationary and stationary common factors do not depend on i, we have $\phi^{f_1 u}(L) = \phi^{f_1\varepsilon}(L) = \phi^{f_2 u}(L) = \phi^{f_2\varepsilon}(L) = 0$. Moreover, we assume that $\phi_i^{uf_2}(L) = 0$. This assumption indicates that there exists no serial correlation or contemporaneous correlation between the regression error u_{it} and the unobserved stationary common factors f_{2t}^0, and it ensures the consistency of our initial estimators. 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, PS) decomposition

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

where $\tilde{w}_{it} = \sum_{j=0}^{\infty} \tilde{\phi}_{ij} v_{i,t-j}$ and $\tilde{\phi}_{ij} = \sum_{s=j+1}^{\infty} \phi_{is}$. Assumption 3.1(ii) imposes a uniform 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 (1992). 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 $(x_{it}', \ f_{1t}^{0\prime})'$. Assumption 3.1(iii) allows $(u_{it}, \varepsilon_{it})$ to be cross-sectionally dependent but they become independent across i given $\mathcal{C}$. Assumption 3.1(iv) 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 $B_i = (B_{1i}, B_{2i}', B_3', B_4')'$ is a $(1 + p + r_1 + r_2) \times 1$ vector Brownian motion with covariance matrix Ω_i.

Assumption 3.2 *(i) As* $N \to \infty$, $\frac{1}{N}\Lambda^{0\prime}\Lambda^0 \xrightarrow{p} \Sigma_\lambda > 0$. $\sup_{N\geq 1} \max_{1\leq i\leq N} \mathbb{E}\|\lambda_i^0\|^{2q} \leq M$ *for some* $q \geq 4$ *and* $\Lambda_1^{0\prime}\Lambda_2^0 = O_P\left(N^{1/2}\right)$.

(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) There exists a constant $\rho_{\min} > 0$ *such that* $P\left(\min_{1\leq i\leq N}\inf_{F_1}\mu_{\min}\left(Q_1(F_1) - Q_2(F_1)\right) \geq c\rho_{\min}\right) = 1 - o(N^{-1})$*, where the* inf *is taken with respect to* F_1 *such that* $\frac{1}{T^2}F_1'F_1 = I_{r_1}$.

Assumption 3.2(i)-(iii) imposes the standard moment conditions in the factor literature; see, e.g., Bai and Ng (2002, 2004). The last condition in Assumption 3.2(i) indicates that the stationary factor loadings and the nonstationary factor loadings can only be weakly correlated, which will greatly facilitate the derivation. 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 error processes. Assumption 3.2(iv) assumes $Q_1(F_1) - Q_2(F_1)$ is positive definite in the limit across i when F_1 satisfies the restriction $\frac{1}{T^2}F_1'F_1 = I_{r_1}$. This assumption is the identification condition for β_i, which is related to ASSUMPTION A in Bai (2009, p.1241). Since F_1 is to be estimated, the identification condition for β_i is imposed on the set of F_1 satisfying the restriction $\frac{1}{T^2}F_1'F_1 = I_{r_1}$.

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* $(N,T) \to \infty$, $\lambda d_T \to 0$, $\lambda T N^{-1/q} d_T^{-2}/\left(\log T\right)^{1+\epsilon} \to \infty$*, and* $d_T^2 N^{1/q} T^{-1} \times \left(\log T\right)^{1+\epsilon} \to 0$.

Assumptions 3.3(i)-(ii) were used in SSP. 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), Su et al. (2017), and Su and Ju (2018). Assumptions 3.3(iii)-(iv) impose 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 $\alpha \in (0, \frac{q-1}{q})$.

3.2 Preliminary rates of convergence

Let $\hat{b}_i = \hat{\beta}_i - \beta_i^0$, $\delta_{NT} = \min(\sqrt{N}, T)$, $C_{NT} = \min(\sqrt{N}, \sqrt{T})$, $\eta_{NT}^2 = \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{b}_i\right\|^2$, 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 $\hat{\beta}_i$ and $\hat{F}_1$.

Theorem 3.1 *Suppose that Assumptions 3.1-3.2 hold. Then*

(i) $\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\beta}_i-\beta_i^0\right)'\frac{1}{T^2}x_i'M_{\hat{F}_1}x_i\left(\hat{\beta}_i-\beta_i^0\right)=o_P(1)$,

(ii) $\left\|P_{\hat{F}_1}-P_{F_1^0}\right\|=o_P(1)$,

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

(iv) $\frac{1}{T}\|\hat{F}_1-F_1^0H_1\|=O_P(\eta_{NT})+\frac{1}{\sqrt{T}}O_P(C_{NT}^{-1})$.

Theorem 3.1(i) establishes the weighted mean-square consistency of $\{\hat{\beta}_i\}$. Theorem 3.1(ii) shows that the spaces spanned by the columns of $\hat{F}_1$ and F_1^0 are asymptotically the same. Given the weighted mean-square consistency and Assumption 3.2(iv), we can further establish the non-weighted mean-square consistency of β_i in Theorem 3.1(iii). As expected, Theorem 3.1(iv) 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 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) $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i-\beta_i^0\|^2=O_P(d_TT^{-2})$,

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

(iii) $(\hat{\alpha}_{(1)},...,\hat{\alpha}_{(K)})-(\alpha_1^0,...,\alpha_K^0)=O_P(d_TT^{-1})$ *for some suitable permutation* $(\hat{\alpha}_{(1)},...,\hat{\alpha}_{(K)})$ *of* $(\hat{\alpha}_1,...,\hat{\alpha}_K)$,

(iv) $T^{-1}\|\hat{F}_1-F_1^0H_1\|^2=O_P(N^{-1}+d_T^2T^{-1})$.

Theorem 3.2(i)-(ii) establish the mean-square and point-wise convergence of the slope coefficients β_i. The usual super consistency of nonstationary estimators $\hat{\beta}_i$ 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 $\left(x_{it},f_{1t}^0\right)$. 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(iv). 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, we say that a classification method is individual 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 and 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 (2.11). In practice, we classify $i\in\hat{G}_k$ if $\hat{\beta}_i=\hat{\alpha}_k$ for some $k=1,...,K$, and $i\in\hat{G}_l$ for some $l=1,...,K$ if $||\hat{\beta}_i-\hat{\alpha}_l||=\min\{||\hat{\beta}_i-\hat{\alpha}_1||,...,||\hat{\beta}_i-\hat{\alpha}_K||\}$ and $\sum_{k=1}^K\mathbf{1}\{\hat{\beta}_i=\hat{\alpha}_k\}=0$. Since Theorem 3.3 ensures $\sum_{k=1}^K P(\hat{\beta}_i=\hat{\alpha}_k)\to 1$ 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 (2.11) 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

$$U_{kNT}=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_i'M_{F_1^0}\left(\left(u_i+F_2^0\lambda_{2i}^0\right)-\frac{1}{N}\sum_{j=1}^N\left(u_j+F_2^0\lambda_{2j}^0\right)a_{ij}\right),$$

$$B_{kNT,1}=\sum_{i=1}^N B_{k,iNT,1}=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\left(\sum_{t=1}^T\sum_{s=1}^T\mathbf{1}\left\{t=s\right\}-\varkappa_{ts}\mathbf{1}\left\{s\leq t\right\}\right)\Delta_{21,i},$$

$$B_{kNT,2}=\sum_{i=1}^N B_{k,iNT,2}=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\mathbb{E}_{\mathcal{C}}\left(x_i\right)'M_{F_1^0}F_2^0\left(\lambda_{2i}^0-\frac{1}{N}\sum_{j=1}^N\lambda_{2j}^0a_{ij}\right),$$

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

where $a_{ij}=\lambda_{1i}^{0\prime}(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}\lambda_{1j}^0$, $\varkappa_{ts}=f_{1t}^{0\prime}(F_1^{0\prime}F_1^0)^{-1}f_{1s}^0$, $\bar{\varkappa}_{ts}=\mathbf{1}\left\{t=s\right\}-\varkappa_{ts}$, $v_{is}^{u\varepsilon}=(v_{is}^u,v_{is}^{\varepsilon\prime})'$, $V_{it}^{u\varepsilon}=\sum_{s=1}^{t}v_{is}^{u\varepsilon}$, $\mathbb{E}_{\mathcal{C}}\left(\cdot\right)=\mathbb{E}\left(\cdot|\mathcal{C}\right)$, $\phi_i^{\dagger}\left(L\right)=\begin{pmatrix}\phi_i^{u\dagger}\left(L\right)\\ \phi_i^{\varepsilon\dagger}\left(L\right)\end{pmatrix}=\begin{pmatrix}\phi_i^{uu}\left(L\right) & \phi_i^{u\varepsilon}\left(L\right)\\ \phi_i^{\varepsilon u}\left(L\right) & \phi_i^{\varepsilon\varepsilon}\left(L\right)\end{pmatrix}$, $S^u=\left(1,0_{1\times p}\right)$, and $S^{\varepsilon}=\left(0_{p\times 1},I_p\right)$. Let $Q_{1NT}=\text{diag}\left(\frac{1}{N_1T^2}\sum_{i\in G_1^0}x_i'M_{F_1^0}x_i,\ldots,\frac{1}{N_KT^2}\sum_{i\in G_K^0}x_i'M_{F_1^0}x_i\right)$ and $Q_{2NT}=\begin{pmatrix}Q_{2NT,11} & \cdots & Q_{2NT,1K}\\ \vdots & \ddots & \vdots\\ Q_{2NT,K1} & \cdots & Q_{2NT,KK}\end{pmatrix}$, where $Q_{2NT,kl}=\frac{1}{N_kNT^2}\sum_{i\in G_k^0}\sum_{j\in G_l^0}x_i'M_{F_1^0}x_ja_{ij}$ for $k,l=1,...,K$. Let

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

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

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 (2.9). Then*

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

(ii) $\sqrt{N}T\,vec(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)-\sqrt{D_{NK}}Q_{NT}^{-1}B_{NT}\Rightarrow\mathcal{MN}(0,\ D_0Q_0^{-1}\Omega_0Q_0^{-1})$ *as* $(N,T)\to\infty$,

where $D_{NK}=diag\left(\frac{N}{N_1},...,\frac{N}{N_K}\right)\otimes I_p$, $D_0=diag\left(\frac{1}{\tau_1},...,\frac{1}{\tau_K}\right)\otimes I_p$, $\Omega_0=\lim_{(N,T)\to\infty}\Omega_{NT}$, *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 the $\hat{\alpha}_k$. 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_{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 our estimators $\hat{\boldsymbol{\alpha}}$ in Theorem 3.4(ii).

As we show 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 $\hat{w}_{it} = (\hat{u}_{it}, \Delta x_{it}', \Delta \hat{f}_{1t}', \hat{f}_{2t}')'$. 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 $S_j \hat{\Delta}_i S_l'$ for $j, l = 1, ..., 4$.

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 $\int_{-\infty}^{\infty} \omega(x)^2 dx \le \infty$, $\omega(0) = 1$, $\omega(x) = 0$ for $|x| \ge 1$, and $\lim_{|x|\to 1} \omega(x)/(1-|x|)^q = c > 0$ for some $q \in (0, \infty)$.*

(ii) As $(N, T) \to \infty$, $N/J^{2q} \to 0$ and $J/T \to 0$.

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

$$\hat{y}_{it}^{+} = y_{it} - \hat{\Omega}_{12,i}\hat{\Omega}_{22,i}^{-1}\Delta x_{it}. \tag{3.2}$$

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

$$\hat{\Delta}_{12,i}^{+} = \hat{\Delta}_{12,i} - \hat{\Omega}_{12,i}\hat{\Omega}_{22i}^{-1}\hat{\Delta}_{22,i}. \tag{3.3}$$

Note that (3.2) and (3.3) 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 $\hat{\alpha}_{\hat{G}_k}^{fm}$, and the continuous updated estimators of $\hat{F}_1$ and $\hat{F}_2$ by iteratively solving (3.5) to (3.7), such that

$$\text{vec}\left(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc}\right) = \text{vec}\left(\hat{\boldsymbol{\alpha}}\right) - \frac{1}{\sqrt{N}T}\sqrt{D_{NK}}\hat{Q}_{NT}^{-1}\left(\hat{B}_{NT,1} + \hat{B}_{NT,2}\right), \tag{3.4}$$

$$\hat{\alpha}_{\hat{G}_k}^{fm} = \left(\sum_{i\in\hat{G}_k} x_i' M_{\hat{F}_1} x_i\right)^{-1}\left\{\sum_{i\in\hat{G}_k} x_i' M_{\hat{F}_1}\hat{y}_i^{+} - T\sqrt{N_k}\left(\hat{B}_{kNT,1}^{+} + \hat{B}_{kNT,2}\right)\right\}, \tag{3.5}$$

$$\hat{F}_1 V_{1,NT} = \left[\frac{1}{NT^2}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}(\hat{y}_i - x_i\hat{\alpha}_{\hat{G}_k}^{fm})(\hat{y}_i - x_i\hat{\alpha}_{\hat{G}_k}^{fm})'\right]\hat{F}_1, \tag{3.6}$$

$$\hat{F}_2 V_{2,NT} = \left[\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}(\hat{y}_i - x_i\hat{\alpha}_{\hat{G}_k}^{fm} - \hat{F}_1\hat{\lambda}_{1i})(\hat{y}_i - x_i\hat{\alpha}_{\hat{G}_k}^{fm} - \hat{F}_1\hat{\lambda}_{1i})'\right]\hat{F}_2, \tag{3.7}$$

where $\hat{B}_{NT,l} = (\hat{B}_{1NT,l}', ..., \hat{B}_{KNT,l}')'$ for $l = 1,\ 2$, $\hat{B}_{kNT,1} = \frac{1}{\sqrt{\hat{N}_k}T}\sum_{i\in\hat{G}_k}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\bar{\varkappa}}_{ts}\right)\hat{\Delta}_{21,i}$, $\hat{B}_{kNT,2} = \frac{1}{\sqrt{\hat{N}_k}T}\sum_{i\in\hat{G}_k}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\bar{\varkappa}}_{ts}\right)\hat{\Delta}_{24,i}\hat{\bar{\lambda}}_{2i}$, $\hat{B}_{kNT,1}^{+} = \frac{1}{\sqrt{\hat{N}_k}T}\sum_{i\in\hat{G}_k}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\bar{\varkappa}}_{ts}\right)\hat{\Delta}_{21,i}^{+}$, $\hat{\bar{\varkappa}}_{ts} = \mathbf{1}\{t = s\} - \hat{\varkappa}_{ts}$, $\hat{\varkappa}_{ts} = \hat{f}_{1t}'(\hat{F}_1'\hat{F}_1)^{-1}\hat{f}_{1s} = \hat{f}_{1t}'\hat{f}_{1s}/T^2$, $\hat{\bar{\lambda}}_{2i} = \hat{\lambda}_{2i} - \frac{1}{N}\sum_{j=1}^{N}\hat{\lambda}_{2j}\hat{a}_{ij}$, and $\hat{a}_{ij} = \hat{\lambda}_{1i}'(\frac{1}{N}\hat{\Lambda}_1'\hat{\Lambda}_1)^{-1}\hat{\lambda}_{1j}$. 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 (2.9), and (3.5) to (3.7), 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 (3.4), (3.6)-(3.7); let $\hat{\boldsymbol{\alpha}}_{\hat{G}}^{fm}$ be obtained by iteratively solving (3.5)-(3.7); and let $\hat{\boldsymbol{\alpha}}_{\hat{G}}^{cup}$ be obtained by iteratively solving (2.9) and (3.5)-(3.7). 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 $\Omega_0^+ = \lim_{N,T\to\infty} \Omega_{NT}^+$, $\Omega_{NT}^+ = Var\left(V_{NT}^+|\mathcal{C}\right)$ *, and* V_{NT}^+ *is defined in the proof of Theorem 3.5.*

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 $\hat{\mathbf{X}}_{k,i}$, $\hat{\mathbf{X}}_{k,i} = M_{\hat{F}_1} x_i \mathbf{1}\{i \in \hat{G}_k\} - \frac{1}{N}\sum_{j\in\hat{G}_k}\hat{a}_{ij}M_{\hat{F}_1}x_j$, $\hat{D}_{NK} = \text{diag}(\frac{N}{\hat{N}_1}, ..., \frac{N}{\hat{N}_K}) \otimes I_p$, $\hat{B}_{iNT} = (\hat{B}_{1,iNT}', ..., \hat{B}_{K,iNT}')'$, $\hat{B}_{k,iNT} = \hat{B}_{k,iNT,1} + \hat{B}_{k,iNT,2}$, $\hat{B}_{k,iNT,1} = \frac{1}{\sqrt{\hat{N}_k}T}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\hat{\varkappa}}_{ts}\right)\hat{\Delta}_{21,i}\mathbf{1}\{i \in \hat{G}_k\}$, $\hat{B}_{k,iNT,2} = \frac{1}{\sqrt{\hat{N}_k}T}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\hat{\varkappa}}_{ts}\right)\hat{\Delta}_{24,i}\hat{\hat{\lambda}}_{2i}\mathbf{1}\{i \in \hat{G}_k\}$, $\hat{u}_{it}^{*} = y_{it} - \hat{\alpha}_k^{fm\prime}x_{it} - \hat{\lambda}_{1i}'\hat{f}_{1t}$ for $i \in \hat{G}_k$, $\hat{B}_{iNT}^{+} = (\hat{B}_{1,iNT}^{+\prime}, ..., \hat{B}_{K,iNT}^{+\prime})'$, $\hat{B}_{k,iNT}^{+} = \hat{B}_{k,iNT,1}^{+} + \hat{B}_{k,iNT,2}$, $\hat{B}_{k,iNT,1}^{+} = \frac{1}{\sqrt{\hat{N}_k}T}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\hat{\hat{\varkappa}}_{ts}\right)\hat{\Delta}_{21,i}^{+}\mathbf{1}\{i \in \hat{G}_k\}$, and $\hat{u}_{it}^{*+} = \hat{y}_{it}^{+} - \hat{\alpha}_k^{fm\prime}x_{it} - \hat{\lambda}_{1i}'\hat{f}_{1t}$ 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 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 now introduce two information criteria to determine the number of unobserved factors before the PPC estimation procedure. Let r_1 denote a generic number of nonstationary factors. and r a generic total number of nonstationary and stationary factors. We use r_1^0 and r^0 to denote their true values and assume that r^0 is bounded above by a finite integer $r_{\max}$.

Bai et al. (2009) find that it is not necessary to distinguish I(0) and I(1) factors when one tries to determine the total number of factors based on the first-differenced model. After first differencing, (2.4) takes the form:

$$\Delta y_{it} = \beta_i^{0\prime}\Delta x_{it} + \lambda_i^{0\prime}\Delta f_t^0 + \Delta u_{it},\ t = 2, ..., T, \tag{3.8}$$

where, e.g., $\Delta y_{it} = y_{it} - y_{i,t-1}$. Since the true dimension r^0 is unknown, we start with a model with

r unobservable common factors. We write the factors as f_t^r and factor loadings as λ_i^r, where the superscript r denotes the dimension of the underlying factors or factor loadings. Let $G^r \equiv \Delta F^r = (G_2^r, ..., G_T^r)'$ where $G_t^r \equiv \Delta f_t^r$. We consider the following minimization problem:

$$\left\{\hat{G}^r, \hat{\Lambda}^r\right\} = \arg\min_{\Lambda^r, G^r} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it} - \hat{\beta}_i'\Delta x_{it} - \lambda_i^{r\prime} G_t^r)^2,$$
$$\text{s.t. } G^{r\prime}G^r/T = I_r \text{ and } \Lambda^{r\prime}\Lambda^r \text{ is diagonal},$$

where $\hat{G}^r = (\hat{G}_2^r, ..., \hat{G}_T^r)'$, $\hat{\Lambda}^r = (\hat{\lambda}_1^r, ..., \hat{\lambda}_N^r)'$, and the $\hat{\beta}_i$ are obtained from the model with $r_1 = r_{\max}$ nonstationary factors. It is easy to show that the $\hat{\beta}_i$ are T-consistent, which suffices for our purpose. It is well known that given $\hat{G}^r$, we can solve $\hat{\Lambda}^r = \hat{\Lambda}^r(\hat{G}^r)$ from the least squares regression as a function of $\hat{G}^r$. Then we can define $V_1(r, \hat{G}^r) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it} - \hat{\beta}_i'\Delta x_{it} - \hat{\lambda}_i^{r\prime}\hat{G}_t^r)^2$. Following Bai and Ng (2002) we consider the information criterion

$$IC_1(r) = \log V_1(r, \hat{G}^r) + r g_1(N, T), \tag{3.9}$$

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

Assumption 3.5 *As* $(N,T) \to \infty$, $g_1(N,T) \to 0$ *and* $C_{NT}^2 g_1(N,T) \to \infty$, *where* $C_{NT} = \min(\sqrt{N}, \sqrt{T})$.

Assumption 3.5 is common in the literature. It requires that $g_1(N,T)$ 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 $IC_1(r)$ to estimate r^0 consistently.

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

Theorem 3.6 shows that the total number of factors r^0 can be determined consistently by minimizing $IC_1(r)$.

As discussed in Section 3.4, ignoring the unobserved stationary factors will not affect the consistency of the long-run estimators but it does generate a bias term that is asymptotically non-negligible. For this reason, it is important to distinguish between nonstationary and stationary factors. Fortunately, it is possible to estimate the number of unobserved nonstationary factors, r_1^0, consistently based on the level data. 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, based on Theorem 3.6.

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. Given the preliminary T-consistent estimators $\hat{\beta}_i$ based on $r_{\max}$ nonstationary factors, we consider the following minimization problem:

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

Given $\hat{F}_1^{r_1} = (\hat{f}_{11}^{r_1}, ..., \hat{f}_{1T}^{r_1})'$, we can solve for $\hat{\Lambda}^{r_1} = (\hat{\lambda}_{11}^{r_1}, ..., \hat{\lambda}_{1N}^{r_1})'$ as a function of $\hat{F}_1^{r_1}$ by least squares regression. We suppress the dependence of $\hat{\Lambda}^{r_1}$ on $\hat{F}_1^{r_1}$ and define $V_2(r_1, \hat{F}_1^{r_1}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{\beta}_i' x_{it} - \hat{\lambda}_i^{r_1\prime}\hat{f}_{1t}^{r_1})^2$. Then we consider the information criterion:

$$IC_2(r_1) = \log V_2(r_1, \hat{F}_1^{r_1}) + r_1 g_2(N,T), \tag{3.10}$$

where $g_2(N,T)$ is a penalty function. Let $\hat{r}_1 = \arg\min_{0\leq r_1 \leq r_{\max}} IC_2(r_1)$. We add the following condition.

Assumption 3.6 *As* $(N,T) \to \infty$, $g_2(N,T)\frac{\log\log(T)}{T} \to 0$ *and* $g_2(N,T) \to \infty$.

Apparently, the conditions on $g_2(N,T)$ differ from the conventional conditions for the penalty function used in information criteria in the stationary framework (e.g., $g_1(N,T)$ in Assumption 3.5). In particular, we now require that $g_2(N,T)$ diverge to infinity rather than converge to zero. The intuition for this requirement is that the mean squared residual, $V_2(r_1, \hat{F}_1^{r_1})$, does not have a finite probability limit when the number of nonstationary common factors is under-specified. We can show that $\frac{\log\log T}{T}V_2(r_1, \hat{F}_1^{r_1})$ converges in probability to a positive constant when $0 \leq r_1 < r_1^0$. By contrast, we have $V_2(r_1, \hat{F}_1^{r_1}) - V_2(r_1^0, \hat{F}_1^{r_1^0}) = O_P(1)$ when $r_1 > r_1^0$.

The following theorem shows that use of $IC_2(r_1)$ determines r_1^0 consistently.

Theorem 3.7 *If Assumptions 3.1-3.3 and 3.6 hold, then* $P(\hat{r}_1 = r_1^0) \to 1$ *as* $(N,T) \to \infty$.

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

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

where $\alpha_T = \frac{T}{5\log\log T}$. We first estimate the total number of unobserved factors by $\hat{r}$ based on the first-differenced model, and next estimate the number of unobserved nonstationary factors by $\hat{r}_1$ based on the level model. A consistent estimator of r_2^0 is then given by $\hat{r}_2 \equiv \hat{r} - \hat{r}_1$.

3.6 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 (2.9), we obtain estimates $\hat{\beta}_i(K,\lambda)$, $\hat{\alpha}_k(K,\lambda)$, $\hat{\lambda}_{1i}(K,\lambda)$, and $\hat{f}_{1t}(K,\lambda)$ of $\beta_i^0, \alpha_k^0, \lambda_i^0$, and f_{1t}^0, in which notation the dependence of the estimates $\hat{\beta}_i, \hat{\alpha}_k, \hat{\lambda}_{1i}$, and $\hat{f}_{1t}$ on (K,λ) explicit. Let $\hat{G}_k(K,\lambda) = \{i \in \{1,2,...,N\} : \hat{\beta}_i(K,\lambda) = \hat{\alpha}_k(K,\lambda)\}$ 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_k^0. Define

$$V_3(K) = \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k(K,\lambda)}\sum_{t=1}^{T}\left[y_{it} - \hat{\alpha}^{cup\prime}_{\hat{G}_k(K,\lambda)}x_{it} - \hat{\lambda}_{1i}(K,\lambda)'\hat{f}_{1t}(K,\lambda)\right]^2.$$

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

$$IC_3(K,\lambda) = \log V_3(K) + pKg_3(N,T), \tag{3.11}$$

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

Let $\mathcal{G}^{(K)} = (G_{K,1},...,G_{K,K})$ be any K-partition of the set of individual index $\{1,2,...,N\}$. Define $\hat{\sigma}^2_{\mathcal{G}^{(K)}} = \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in G_{K,k}}\sum_{t=1}^{T}[y_{it} - \hat{\alpha}^{cup\prime}_{G_{K,k}}x_{it} - \hat{\lambda}_{1i}(\mathcal{G}^{(K)})'\hat{f}_{1t}(\mathcal{G}^{(K)})]^2$, where $\{\hat{\alpha}^{cup}_{G_{K,k}}, \hat{\lambda}_{1i}(\mathcal{G}^{(K)}),$ $\hat{f}_{1t}(\mathcal{G}^{(K)})\}$ is analogously defined as $\{\hat{\alpha}^{cup}_{\hat{G}_k(K,\lambda)}, \hat{\lambda}_{1i}(K,\lambda), \hat{f}_{1t}(K,\lambda)\}$ with $\{\hat{G}_k(K,\lambda)\}$ being replaced by $\{G_{K,k}\}$. Let $\sigma_0^2 = \text{plim}_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{i\in G_k^0}\sum_{t=1}^{T}[y_{it} - \alpha_k^{0\prime}x_{it} - \lambda_{1i}^{0\prime}f_{1t}^0]^2$. Define

$$\nu_{NT} = \begin{cases} (NT)^{-1/2} \text{ when there is no unobserved common factor,} \\ \delta_{NT}^{-1} \text{ when there are only unobserved nonstationary common factors,} \\ C_{NT}^{-1} \text{ when there are unobserved nonstationary and stationary common factors.} \end{cases},$$

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

We add the following assumption.

Assumption 3.7 (i) *As* $(N,T)\to\infty$, $\min_{1\le K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\hat{\sigma}^2_{\mathcal{G}^{(K)}} \xrightarrow{p} \underline{\sigma}^2 > \sigma_0^2$.

(ii) As $(N,T)\to\infty$, $g_3(N,T)\to 0$ and $g_3(N,T)/\nu_{NT}^2\to\infty$.

Assumption 3.7(i) requires that all under-fitted models yield asymptotic mean square errors larger than σ_0^2, 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 Assumption 3.1-3.4 and 3.7 hold. Then* $P(\hat{K}(\lambda) = K_0) \to 1$ *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 $\inf_{1 \le K \le K_{\max}, K \ne K_0} IC_3(K,\lambda) > IC_3(K_0,\lambda)$ 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

The simulations reported in this section are designed to evaluate the finite sample performance of the C-Lasso selection, the bias-corrected post-Lasso, the fully-modified post-Lasso regression, and the Cup-Lasso estimators, as well as the performance of the information criteria for determining the numbers of groups and common factors.

4.1 Data generating processes

We consider four 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 $T = 40, 80$. In all cases, the number of replications is 500.

DGP1 (Contemporaneous correlation among the errors, nonstationary regressors, and unobserved stationary common factors). The observations (y_{it}, x'_{it}) are generated from the model

$$\begin{cases} y_{it} = \beta'_i x_{it} + c_2 \lambda'_{2i} f_{2t} + u_{it} \\ x_{it} = x_{it-1} + \varepsilon_{it} \end{cases}, \tag{4.1}$$

where $x_{it} = (x_{1it}, x_{2it})'$ is a 2×1 vector of nonstationary regressors, and f_{2t} is a 2×1 vector of stationary common factors. The idiosyncratic errors $w_{it} = (u_{it}, \varepsilon'_{it}, f'_{2t})' = \Omega^{1/2} v_{it}$, where $\Omega^{1/2} = \begin{pmatrix} 0.5 & 0.2 & 0.2 & 0 & 0 \\ 0.2 & 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 & 0.2 \\ 0 & 0 & 0 & 1 & 0.2 \\ 0 & 0 & 0 & 0.2 & 1 \end{pmatrix}$, $v_{it} = (v_{it}^{u\varepsilon\prime}, v_t^{f_2\prime})'$, $v_{it}^{u\varepsilon} \sim$ i.i.d. $N(0, I_3)$ for $i = 1, ..., N$, and $v_t^{f_2} \sim$ i.i.d. $N(0, I_2)$. The factor loadings λ_{2i} are i.i.d. $N((0.1, 0.1)', I_2)$ for $i = 1, ..., N$. We set

$c_2 = 0.5$ to control the relative contribution of the unobserved common factors. 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

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

DGP2 (Weak dependence among the errors, nonstationary regressors, and unobserved nonstationary common factors). The observations $(y_{it}, x_{it}', f_{1t}')$ are generated from the model

$$\begin{cases} y_{it} = \beta_i' x_{it} + \lambda_{1i}' f_{1t} + u_{it} \\ x_{it} = x_{i,t-1} + \varepsilon_{it} \\ f_{1t} = f_{1,t-1} + \nu_t \end{cases}, \tag{4.2}$$

where $x_{it} = (x_{1it}, x_{2it})'$ is a 2×1 vector of nonstationary regressors, and f_{1t} is a 2×1 vector of nonstationary common factors. The idiosyncratic errors $w_{it} = (u_{it}, \varepsilon_{it}', \Delta f_{1t}')'$ are generated from a linear process: $w_{it} = \sum_{j=0}^{\infty} \phi_{ij} v_{i,t-j}$, where $\phi_{ij} = j^{-3.5} \Omega^{1/2}$, and $\Omega^{1/2} = \begin{pmatrix} 0.5 & 0.2 & 0.2 & 0 & 0 \\ 0.2 & 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 & 0.2 \\ 0 & 0 & 0 & 1 & 0.2 \\ 0 & 0 & 0 & 0.2 & 1 \end{pmatrix}$, $v_{it} = (v_{it}^{u\varepsilon\prime}, v_t^{f_1\prime})'$, $v_{it}^{u\varepsilon} \sim$ i.i.d. $N(0, I_3)$ for $i = 1, ..., N$, and $v_t^{f_1} \sim$ i.i.d. $N(0, I_2)$. The factor loadings of nonstationary common factors are i.i.d. $\lambda_{1i} \sim N((0.1, 0.1)', I_2)$ for $i = 1, ..., N$. The true coefficients of β_i are the same as in DGP1.

DGP3 (Weak dependence among the errors, nonstationary regressors, and unobserved mixed common factors). The observations $(y_{it}, x_{it}', f_{1t}', f_{2t}')$ are generated from the following model

$$\begin{cases} y_{it} = \beta_i' x_{it} + c_1(\lambda_{1i}' f_{1t}) + c_2(\lambda_{2i}' f_{2t}) + u_{it} \\ x_{it} = x_{i,t-1} + \varepsilon_{it} \\ f_{1t} = f_{1,t-1} + \nu_t \end{cases}, \tag{4.3}$$

where $x_{it} = (x_{1it}, x_{2it})'$ is a 2×1 vector of nonstationary regressors, f_{1t} is a 2×1 vector of nonstationary common factors, and f_{2t} contains one stationary common factor. The idiosyncratic errors $w_{it} = (u_{it}, \varepsilon_{it}', \Delta f_{1t}', f_{2t}')'$ are generated from the linear process $w_{it} = \sum_{j=0}^{\infty} \phi_{ij} v_{i,t-j}$, where $\phi_{ij} = j^{-3.5} \Omega^{1/2}$,

$$\Omega^{1/2} = \begin{pmatrix} 0.5 & 0.2 & 0.2 & 0 & 0 & 0 \\ 0.2 & 1 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 & 0.2 & 0.2 \\ 0 & 0 & 0 & 1 & 0.2 & 0.2 \\ 0 & 0 & 0 & 0.2 & 1 & 0.2 \\ 0 & 0 & 0 & 0.2 & 0.2 & 1 \end{pmatrix}, \; v_{it} = (v_{it}^{u\varepsilon\prime}, v_t^{f_1\prime}, v_t^{f_2\prime})', \; v_{it}^{u\varepsilon} \sim \text{i.i.d. } N(0, I_3) \text{ for } i = 1, ..., N,$$

and $(v_t^{f_1\prime}, v_t^{f_2\prime})' \sim$ i.i.d. $N(0, I_3)$. Let $c_1 = 1$ and $c_2 = 0.5$. The factor loadings $\lambda_i = (\lambda_{1i}', \lambda_{2i}')'$ are i.i.d. $\lambda_i \sim N((0.1, 0.1, 0.1)', I_3)$. The true coefficients of β_i are the same as in DGP1.

DGP4 (Weak dependence among the errors, nonstationary regressors, and unobserved mixed common factors). The settings of DGP4 are the same as those of DGP3, except that there is weak correlation among the factor loadings with $\lambda_i \sim$ i.i.d. $N((0.1, 0.1, 0.1)', \Omega_2)$, where $\Omega_2 = \begin{pmatrix} 1 & 0 & 2/\sqrt{N} \\ 0 & 1 & 2/\sqrt{N} \\ 2/\sqrt{N} & 2/\sqrt{N} & 1 \end{pmatrix}$.

4.2 Estimate the number of unobserved factors

We assess the performance of two information criteria proposed in Section 3.5 before determining the number of groups and running the PPC-based estimation procedure. We choose the BIC-type penalty function $g_1(N, T) = \frac{N+T}{NT} \log(\min(N, T))$ to determine the total number (r) of unobserved factors and $g_2(N, T) = \frac{T}{5 \log \log T} g_1(N, T)$ to determine the number (r_1) of unobserved nonstationary factors. Note that $r^0 = 2$, 2, 3, and 3 for DGPs 1-4, respectively, and $r_1^0 = 0$, 2, 2, and 2 for DGPs 1-4, respectively.

Table 1 displays the probability that a particular factor number from 0 to 5 is selected according to the information criteria proposed for the differenced and level data based on 500 replications. For the differenced data, the probabilities for selecting the total number of unobserved factors are higher than 99% in all DGPs when $N = 50$ and reach the unity when N increases to 100 in all cases under investigation. For the level data, the precision for selecting the number of nonstationary factors is not as good as that for selecting the total number of factors based on the differenced data, especially when $N = 100$ and $T = 40$ for DGP1 and when $N = 50$ and $T = 80$ for DGPs 2-4. But when N and T increase, the probabilities of selecting the true number of nonstationary factors approach 99% in all DGPs. In general, the simulation results show that the information criteria for the differenced data and level data work fairly well in finite samples.

Table 1: Frequency for selecting $r = 1, 2, ..., 5$ total factors and $r_1 = 0, 1, ..., 4$ nonstationary factors

			Differenced Data					Level Data				
	N	T	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	$r_1=0$	$r_1=1$	$r_1=2$	$r_1=3$	$r_1=4$
DGP1	50	40	0	1	0	0	0	0.992	0.008	0	0	0
	50	80	0	1	0	0	0	1	0	0	0	0
	100	40	0	1	0	0	0	0.950	0.050	0	0	0
	100	80	0	1	0	0	0	0.998	0.002	0	0	0
DGP2	50	40	0	1	0	0	0	0	0.010	0.990	0	0
	50	80	0	1	0	0	0	0.026	0.006	0.968	0	0
	100	40	0	1	0	0	0	0	0	1	0	0
	100	80	0	1	0	0	0	0.006	0	0.994	0	0
DGP3	50	40	0	0	1	0	0	0.006	0.068	0.922	0.004	0
	50	80	0	0	0.994	0.006	0	0.038	0.062	0.900	0	0
	100	40	0	0	1	0	0	0	0.004	0.934	0.062	0
	100	80	0	0	1	0	0	0.002	0.006	0.990	0.002	0
DGP4	50	40	0	0	1	0	0	0.006	0.058	0.932	0.004	0
	50	80	0	0	0.996	0.004	0	0.034	0.056	0.910	0	0
	100	40	0	0	1	0	0	0	0.002	0.960	0.038	0
	100	80	0	0	1	0	0	0.002	0.006	0.990	0.002	0

4.3 Determination of the number of groups

The results above show that the information criteria ($IC_1(r)$ and $IC_2(r_1)$) in Section 3.5 are useful in determining the number of nonstationary and stationary factors. We emphasize that these information criteria do not require 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 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 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 true number of groups is 3. The probabilities are higher than 98% in all cases and tend to the unity when T increases to 80. This indicates good finite sample performance of the criterion IC_3 in determining the number of groups.

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

	N	T	1	2	3	4	5	6
DGP1	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
DGP2	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
DGP3	50	40	0	0	0.986	0.014	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
DGP4	50	40	0	0	0.990	0.010	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

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. Tables 3 and 4 report classification and point estimation results from 500 replications for each DGP. As above, we set $\lambda = c_\lambda T^{-3/4}$ with $c_\lambda = 0.05, 0.1, 0.2, 0.4$ to check the sensitivity of classification and estimation performance. Due to space constraints, we only report results for $c_\lambda = 0.1$ and 0.2 in Tables 3-4 and for $\alpha_k = (\alpha_{1,k}, \alpha_{2,k})'$ we only report results for the estimation of the first coefficient $\alpha_{1,k}$ in each DGP.

Columns 4 and 8 in Tables 3-4 report the percentage of correct classification over the N cross sectional units, calculated as $\frac{1}{N}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k}\mathbf{1}\{\beta_i^0 = \alpha_k^0\}$, averaged over the 500 replications. Columns 5 to 7 and 9 to 11 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 $\frac{1}{N}\sum_{k=1}^{K} N_k \text{RMSE}(\hat{\alpha}_{1,k})$ with $\hat{\alpha}_{1,k}$ being the estimate of $\alpha_{1,k}$. We define the weighted average bias and 95% coverage probability analogously. The estimates of the long-run covariance matrix are obtained by using the Fejér kernel with bandwidth $J = 10$. Findings based on other kernels (the quadratic spectral kernel and Parzen kernel) and other choices of J are similar and are not reported. For comparison, we report 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. For comparison we also report estimation and inference results for the oracle estimates that are obtained by utilizing the true group structures $\{G_k^0\}$.

We summarize the general pattern of the findings on classification and estimation reported in Tables 3-4. First, the results with different c_λ's are similar, indicating some robustness in our algorithm to the choice of the tuning parameter λ. Second, for the classification results, the correct classification percentage approaches 100% when T increases. In particular and 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 iteration helps in finite samples to achieve better classification. Third, regarding parameter estimation Tables 3-4 show that the fully-modified procedure works slightly better than the direct bias-correction procedure. Therefore, we only provide the results for the Cup-Lasso estimates based on the fully-modified method. For DGP1, 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 and the Cup-Lasso estimates are found to perform as well as oracle estimation in terms of the reported RMSE, bias and coverage probability. For DGPs 2-4, the performance of the C-Lasso estimate is poorer due to the presence of unobserved nonstationary common factors. In addition, the Cup-Lasso estimates generally outperformed the two post-Lasso estimates due to the updated group classification results. 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 effect: 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

Table 3: Classification and point estimation of α_1 for DGP1 and DGP2

		c_λ	0.1				0.2			
N	T		% Correct classification	RMSE	Bias	% Coverage	% Correct classification	RMSE	Bias	%Coverage
DGP1										
50	40	C-Lasso	99.98	0.0098	0.0054	83.42	99.97	0.0092	0.0051	83.72
50	40	post-Lasso^{bc}	99.98	0.0081	0.0004	91.12	99.97	0.0080	0.0004	91.12
50	40	post-Lasso^{fm}	99.98	0.0080	0.0005	91.00	99.97	0.0079	0.0005	91.00
50	40	Cup-Lasso	99.98	0.0080	0.0005	91.00	99.97	0.0079	0.0005	91.00
50	40	Oracle	-	0.0079	0.0005	91.00	-	0.0079	0.0005	91.00
50	80	C-Lasso	100.00	0.0048	0.0026	84.12	100.00	0.0046	0.0025	84.22
50	80	post-Lasso^{bc}	100.00	0.0039	0.0001	91.32	100.00	0.0039	0.0001	91.32
50	80	post-Lasso^{fm}	100.00	0.0038	0.0002	92.04	100.00	0.0038	0.0002	92.04
50	80	Cup-Lasso	100.00	0.0038	0.0002	92.04	100.00	0.0038	0.0002	92.04
50	80	Oracle	-	0.0038	0.0002	92.04	-	0.0038	0.0002	92.04
100	40	C-Lasso	99.97	0.0075	0.0050	79.48	99.97	0.0071	0.0047	81.90
100	40	post-Lasso^{bc}	99.97	0.0056	0.0002	92.30	99.97	0.0055	0.0002	92.36
100	40	post-Lasso^{fm}	99.97	0.0055	0.0003	92.60	99.97	0.0055	0.0003	92.72
100	40	Cup-Lasso	99.97	0.0055	0.0003	92.60	99.97	0.0055	0.0003	92.72
100	40	Oracle	-	0.0054	0.0002	92.60	-	0.0054	0.0002	92.60
100	80	C-Lasso	100.00	0.0037	0.0024	80.04	100.00	0.0036	0.0023	80.90
100	80	post-Lasso^{bc}	100.00	0.0028	0.0000	92.24	100.00	0.0028	0.0000	92.24
100	80	post-Lasso^{fm}	100.00	0.0027	0.0001	92.60	100.00	0.0027	0.0001	92.60
100	80	Cup-Lasso	100.00	0.0027	0.0001	92.60	100.00	0.0027	0.0001	92.60
100	80	Oracle	-	0.0027	0.0001	92.60	-	0.0027	0.0001	92.60
DGP2										
50	40	C-Lasso	98.42	0.0420	0.0155	65.36	98.26	0.0443	0.0143	65.88
50	40	post-Lasso^{bc}	98.42	0.0305	0.0028	91.62	98.26	0.0311	0.0029	91.74
50	40	post-Lasso^{fm}	98.42	0.0305	0.0028	92.20	98.26	0.0311	0.0030	92.14
50	40	Cup-Lasso	100.00	0.0112	0.0021	90.28	99.98	0.0112	0.0021	90.28
50	40	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
50	80	post-Lasso^{bc}	99.34	0.0188	0.0009	91.34	99.31	0.0173	0.0014	91.74
50	80	post-Lasso^{fm}	99.34	0.0188	0.0014	91.28	99.31	0.0172	0.0018	91.62
50	80	Cup-Lasso	100.00	0.0050	0.0009	90.44	100.00	0.0050	0.0009	90.44
50	80	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
100	40	post-Lasso^{bc}	98.66	0.0225	0.0027	89.72	98.49	0.0222	0.0033	89.86
100	40	post-Lasso^{fm}	98.66	0.0226	0.0027	90.10	98.49	0.0223	0.0034	90.26
100	40	Cup-Lasso	100.00	0.0073	0.0025	89.78	99.98	0.0073	0.0025	89.78
100	40	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
100	80	post-Lasso^{bc}	99.41	0.0188	0.0009	92.72	99.38	0.0190	0.0009	92.84
100	80	post-Lasso^{fm}	99.41	0.0188	0.0014	93.08	99.38	0.0190	0.0013	93.20
100	80	Cup-Lasso	100.00	0.0035	0.0010	93.12	100.00	0.0035	0.0010	93.12
100	80	Oracle	-	0.0035	0.0010	93.12	-	0.0035	0.0010	93.12

Table 4: Classification and point estimation of α_1 for DGP3 and DGP4

		c_λ	0.1				0.2			
N	T		% Correct classification	RMSE	Bias	% Coverage	% Correct classification	RMSE	Bias	%Coverage
DGP3										
50	40	C-Lasso	97.93	0.0500	0.0148	72.98	97.73	0.0543	0.0141	71.40
50	40	post-Lasso^{bc}	97.93	0.0379	0.0009	91.20	97.73	0.0405	0.0014	90.86
50	40	post-Lasso^{fm}	97.93	0.0380	0.0012	91.30	97.73	0.0405	0.0017	90.86
50	40	Cup-Lasso	99.96	0.0141	0.0011	90.00	99.91	0.0140	0.0012	89.84
50	40	Oracle	-	0.0139	0.0011	89.98	-	0.0139	0.0011	89.98
50	80	C-Lasso	99.07	0.0446	0.0080	67.44	99.07	0.0451	0.0075	66.46
50	80	post-Lasso^{bc}	99.07	0.0300	0.0003	92.04	99.07	0.0300	0.0002	92.10
50	80	post-Lasso^{fm}	99.07	0.0299	0.0007	92.16	99.07	0.0299	0.0007	92.22
50	80	Cup-Lasso	100.00	0.0066	0.0005	91.46	100.00	0.0066	0.0005	91.46
50	80	Oracle	-	0.0066	0.0005	91.46	-	0.0066	0.0005	91.46
100	40	C-Lasso	98.33	0.0320	0.0150	62.52	98.18	0.0352	0.0140	62.44
100	40	post-Lasso^{bc}	98.33	0.0286	0.0024	91.56	98.18	0.0286	0.0024	91.78
100	40	post-Lasso^{fm}	98.33	0.0287	0.0026	90.88	98.18	0.0286	0.0026	90.96
100	40	Cup-Lasso	99.96	0.0096	0.0020	91.98	99.93	0.0097	0.0020	91.86
100	40	Oracle	-	0.0095	0.0020	91.90	-	0.0095	0.0020	91.90
100	80	C-Lasso	99.38	0.0201	0.0074	56.96	99.34	0.0218	0.0070	57.54
100	80	post-Lasso^{bc}	99.38	0.0165	0.0002	93.40	99.34	0.0169	0.0001	93.40
100	80	post-Lasso^{fm}	99.38	0.0164	0.0007	93.34	99.34	0.0169	0.0006	93.44
100	80	Cup-Lasso	100.00	0.0046	0.0004	93.82	100.00	0.0046	0.0004	93.82
100	80	Oracle	-	0.0046	0.0004	93.82	-	0.0046	0.0004	93.82
DGP4										
50	40	C-Lasso	98.22	0.0479	0.0145	70.70	98.07	0.0511	0.0133	71.44
50	40	post-Lasso^{bc}	98.22	0.0337	0.0022	91.64	98.07	0.0335	0.0020	91.48
50	40	post-Lasso^{fm}	98.22	0.0338	0.0024	91.44	98.07	0.0335	0.0022	91.18
50	40	Cup-Lasso	99.97	0.0137	0.0015	89.98	99.93	0.0137	0.0015	90.10
50	40	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
50	80	post-Lasso^{bc}	99.10	0.0310	0.0008	91.52	99.09	0.0313	0.0007	91.40
50	80	post-Lasso^{fm}	99.10	0.0310	0.0012	91.14	99.09	0.0313	0.0012	91.02
50	80	Cup-Lasso	100.00	0.0065	0.0007	90.58	100.00	0.0065	0.0007	90.58
50	80	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
100	40	post-Lasso^{bc}	98.44	0.0277	0.0024	91.16	98.28	0.0282	0.0021	90.92
100	40	post-Lasso^{fm}	98.44	0.0279	0.0026	90.94	98.28	0.0283	0.0023	90.72
100	40	Cup-Lasso	99.97	0.0095	0.0021	91.12	99.94	0.0096	0.0021	91.22
100	40	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
100	80	post-Lasso^{bc}	99.45	0.0167	0.0007	92.62	99.43	0.0165	0.0006	92.66
100	80	post-Lasso^{fm}	99.45	0.0167	0.0011	92.70	99.43	0.0165	0.0011	92.88
100	80	Cup-Lasso	100.00	0.0047	0.0006	93.00	100.00	0.0047	0.0006	93.00
100	80	Oracle	-	0.0047	0.0006	93.00	-	0.0047	0.0006	93.00

(2004), Coe et al. (2009), and Ertur and Musolesi (2017).

In this section we apply our model and methodology to reinvestigate 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 TFP as the Solow residual, 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 is final output, L is labor force, K is capital stock, 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.

5.2 Data

We use the same dataset used by Coe et al. (2009, CHH2009 hereafter). The 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 $\log(F_{it})$, $\log(s_{it}^d)$, $\log(s_{it}^f)$, 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. We refer the readers directly to CHH2009 for details on the definitions and constructions of these variables, and 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 3.5 to estimate the number of unobserved factors. Following the simulation design, we set $g_1(N,T) = \frac{N+T}{NT} \log(\min(N,T))$ and $g_2(N,T) = \frac{T}{5 \log \log T} g_1(N,T)$. Based on the results for the differenced and level data, we obtain the estimates $\hat{r} = 1$ and $\hat{r}_1 = 1$. 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 ensuing empirical analysis.

Table 5: Information criterion for the determination of the number of groups

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

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 5 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 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 CHH2009.

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 6. First, a comparison between the estimates in CHH2009 and those obtained by pooled FM-OLS suggests that the estimate of the coefficient of $\log(s^d)$ in CHH2009 is qualitatively 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 CHH2009, this difference explains why the standard errors (s.e.'s) of our estimates are much larger than those in CHH2009. 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

Table 6: 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.

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 on 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 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

Table 7: 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				

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.

5.3.4 Classification results

Table 7 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 global R&D stock. By contrast, the remaining countries in Group 1 account for only 1.5% of global R&D stock. 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 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 the conventional testing using t, Wald, and F statistics for inference. The use of these methods in the empirical application provides new results that help to explain the growth convergence puzzle through the heterogeneous behavior of R&D spillover effects.

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., 2018. Identifying latent group patterns in cointegrated panels. Working Paper, Singapore Management University.

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.

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.

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., 2017. Sieve estimation of time-varying panel data models with latent structures. *Journal of Business & Economic Statistics*, forthcoming.

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

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

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

A Proof of the Main Results in Section 3

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

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

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

(v) Without loss of generality, we set $x_{i0} = 0$ throughout the proof of the main results and supplementary Appendix.

To prove Theorem 3.1, we make use of the following four lemmas.

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

(i) $\frac{1}{T^2}x_i' M_{F_1^0} x_i \Rightarrow \int \tilde{B}_{2i}\tilde{B}_{2i}'$,

(ii) $\frac{1}{T}x_i' M_{F_1^0} u_i \Rightarrow \int (B_{2i} - \pi_i' B_3)\, dB_{1i} + (\Delta_{21,i} - \pi_i'\Delta_{31,i})$,

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

Lemma A.2 *Suppose that Assumptions 3.1-3.2 hold. Let $W_i = \left(x_i, F_1^0\right)$ and $d_T = \log\log T$, as in Assumption 3. Then for any fixed small constant $c \in (0, 1/2)$,*

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

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

(iii) $\limsup_{T\to\infty} \mu_{\max}\left(\frac{1}{d_T T^2}x_i' M_{F_1^0} x_i\right) \le (1+c)\rho_{\max}$ *a.s.*,

(iv) $\liminf_{T\to\infty} \mu_{\min}\left(\frac{d_T}{T^2}x_i' M_{F_1^0} x_i\right) \ge \rho_{\min}/2$ *a.s.*.

Lemma A.3 *Suppose that Assumptions 3.1-3.2 hold. Then*

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

(ii) $\frac{1}{N}\sum_{i=1}^N \left\|\frac{1}{T^2}x_i' M_{F_1^0} u_i^*\right\|^2 = O_P(d_T^2 T^{-2})$,

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

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

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

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

(i) $\sup_{F_1}\sup_{N^{-1}\|\mathbf{b}\|^2\le M}\left\|\frac{1}{NT^2}\sum_{i=1}^N b_i'x_i'M_{F_1}u_i^*\right\| = o_P(d_T^{-3})$,

(ii) $\sup_{F_1}\left\|\frac{1}{NT^2}\sum_{i=1}^N \lambda_{1i}^{0\prime}F_1^{0\prime}M_{F_1}u_i^*\right\| = o_P(d_T^{-3})$,

(iii) $\sup_{F_1}\left\|\frac{1}{NT^2}\sum_{i=1}^N u_i^{*\prime}P_{F_1}u_i^*\right\| = o_P(d_T^{-3})$,

where the sup *is taken with respect to* F_1 *such that* $\frac{F_1'F_1}{T^2} = I_{r_1}$ *and* u_i^* *is defined in Lemma A.3.*

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

$$\begin{aligned}Q_{i,NT}(\beta_i,F_1) - Q_{i,NT}(\beta_i^0,F_1^0) =& \frac{1}{T^2}(b_i'x_i'M_{F_1}x_ib_i + \lambda_{1i}^{0\prime}F_1^{0\prime}M_{F_1}F_1^0\lambda_{1i}^0 - 2b_i'x_i'M_{F_1}F_1^0\lambda_{1i}^0)\\ &+ \frac{1}{T^2}\left(2\lambda_{1i}^{0\prime}F_1^{0\prime}M_{F_1}u_i^* - 2b_i'x_i'M_{F_1}u_i^* - u_i^{*\prime}(P_{F_1} - P_{F_1^0})u_i^*\right), \quad \text{(A.1)}\end{aligned}$$

where $u_i^* = u_i + F_2^0\lambda_{2i}^0$. Let $S_{i,NT}(\beta_i,F_1) = \frac{1}{T^2}\left(b_i'x_i'M_{F_1}x_ib_i + \lambda_{1i}^{0\prime}F_1^{0\prime}M_{F_1}F_1^0\lambda_{1i}^0 - 2b_i'x_i'M_{F_1}F_1^0\lambda_{1i}^0\right)$. Then we have

$$\begin{aligned}&Q_{NT}(\boldsymbol{\beta},F_1) - Q_{NT}(\boldsymbol{\beta}^0,F_1^0)\\ =&\frac{1}{N}\sum_{i=1}^N S_{i,NT}(\beta_i,F_1) + \frac{1}{NT^2}\sum_{i=1}^N\left(2\lambda_{1i}^{0\prime}F_1^{0\prime}M_{F_1}u_i^* - 2b_i'x_i'M_{F_1}u_i^* - u_i^{*\prime}(P_{F_1} - P_{F_1^0})u_i^*\right)\\ =&\frac{1}{N}\sum_{i=1}^N S_{i,NT}(\beta_i,F_1) + o_P(d_T^{-3}), \quad \text{(A.2)}\end{aligned}$$

where the last three terms on the right side of (A.2) are $o_P(d_T^{-3})$ uniformly in $\{b_i\}$ and F_1 such that $\frac{F_1'F_1}{T^2} = I_{r_1}$ and $\frac{1}{N}\sum_{i=1}^N\|b_i\|^2 \le M$ by Lemma A.4(i)-(iii) and the fact that $\frac{1}{NT^2}\sum_{i=1}^N u_i^{*\prime}P_{F_1^0}u_i^* = o_P(d_T^{-3})$. Then we have

$$\begin{aligned}Q_{NT}^{K,\lambda}(\boldsymbol{\beta},\hat{\boldsymbol{\alpha}},F_1) - Q_{NT}^{K,\lambda}(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,F_1^0) =& \frac{1}{N}\sum_{i=1}^N[Q_{NT,i}(\beta_i,F_1) - Q_{NT,i}(\beta_i^0,F_1^0)] + \frac{\lambda}{N}\sum_{i=1}^N\prod_{k=1}^K\|\beta_i - \hat{\alpha}_k\|\\ \ge& S_{NT}(\boldsymbol{\beta},F_1) + o_P(d_T^{-3}). \quad \text{(A.3)}\end{aligned}$$

where $S_{NT}(\boldsymbol{\beta},F_1) = \frac{1}{N}\sum_{i=1}^N S_{i,NT}(\beta_i,F_1)$. Then by (A.2) and (A.3) and the fact that $Q_{NT}^{K,\lambda}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\hat{F}_1) - Q_{NT}^{K,\lambda}(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,F_1^0) \le 0$, we have

$$S_{NT}(\hat{\boldsymbol{\beta}},\hat{F}_1) = \frac{1}{NT^2}\sum_{i=1}^N\left[\hat{b}_i'x_i'M_{\hat{F}_1}x_i\hat{b}_i + \lambda_{1i}^{0\prime}F_1^{0\prime}M_{\hat{F}_1}F_1^0\lambda_{1i}^0 - 2\hat{b}_i'x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right] = o_P(d_T^{-3}). \quad \text{(A.4)}$$

Similarly, by (A.2), (A.3) and Lemma A.4(i)-(iii), we have

$$
\begin{aligned}
Q_{NT}^{K,\lambda}(\boldsymbol{\beta},\hat{\boldsymbol{\alpha}},\hat{F}_1)-Q_{NT}^{K,\lambda}(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\hat{F}_1)=&\frac{1}{N}\sum_{i=1}^{N}[Q_{NT,i}(\beta_i,\hat{F}_1)-Q_{NT,i}(\beta_i^0,\hat{F}_1)]+\frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\beta_i-\hat{\alpha}_k\| \\
\geq&\frac{1}{NT^2}\sum_{i=1}^{N}\left[b_i'x_i'M_{\hat{F}_1}x_ib_i-2b_i'x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right]+o_P(d_T^{-3}). \qquad (A.5)
\end{aligned}
$$

This, in conjunction with the fact that $Q_{NT}^{K,\lambda}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\hat{F}_1)-Q_{NT}^{K,\lambda}(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\hat{F}_1)\leq 0$, implies that

$$
\frac{1}{NT^2}\sum_{i=1}^{N}\left[\hat{b}_i'x_i'M_{\hat{F}_1}x_i\hat{b}_i-2\hat{b}_i'x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right]\leq o_P(d_T^{-3}). \qquad (A.6)
$$

Combining (A.4) and (A.6) yields that $o_P(d_T^{-3})=\frac{1}{NT^2}\sum_{i=1}^N\lambda_{1i}^{0\prime}F_1^{0\prime}M_{\hat{F}_1}F_1^0\lambda_{1i}^0=\mathrm{tr}[(\frac{1}{T^2}F_1^{0\prime}M_{\hat{F}_1}F_1^0)(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)]$ $\geq\mathrm{tr}(\frac{1}{T^2}F_1^{0\prime}M_{\hat{F}_1}F_1^0)\mu_{\min}(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)$. It follows that $\mathrm{tr}(\frac{1}{T^2}F_1^{0\prime}M_{\hat{F}_1}F_1^0)=o_P(d_T^{-3})$ 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

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

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 $\frac{1}{T^2}\hat{F}_1'P_{F_1^0}\hat{F}_1-I_{r_1}=o_P(d_T^{-3})$, which further implies that $\left\|P_{\hat{F}_1}-P_{F_1^0}\right\|^2=2\mathrm{tr}\left(I_{r_1}-\frac{1}{T^2}\hat{F}_1'P_{F_1^0}\hat{F}_1\right)=o_P(d_T^{-3})$. By the Cauchy-Schwarz inequality and (A.6),

$$
o_P(d_T^{-3})\geq\frac{1}{NT^2}\sum_{i=1}^{N}\hat{b}_i'x_i'M_{\hat{F}_1}x_i\hat{b}_i-2\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\hat{b}_i'x_i'M_{\hat{F}_1}x_i\hat{b}_i\right\}^{1/2}\left\{\frac{1}{NT^2}\lambda_{1i}^{0\prime}F_1^{0\prime}M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right\}^{1/2}. \quad (A.8)
$$

This result, in conjunction with (A.7), implies that $\frac{1}{NT^2}\sum_{i=1}^N\hat{b}_i'x_i'M_{\hat{F}_1}x_i\hat{b}_i=o_P(d_T^{-3})$. So we have shown parts (i) and (ii) in the theorem.

(iii) By the results in parts (i) and (ii) and Lemma A.2(i) and (iv), we have

$$
\begin{aligned}
o_P(d_T^{-3})&=\frac{1}{N}\sum_{i=1}^{N}\hat{b}_i'\left(\frac{1}{T^2}x_i'M_{\hat{F}_1}x_i\right)\hat{b}_i \\
&=\frac{1}{N}\sum_{i=1}^{N}\hat{b}_i'\left(\frac{1}{T^2}x_i'M_{F_1^0}x_i\right)\hat{b}_i+\frac{1}{N}\sum_{i=1}^{N}\hat{b}_i'\left(\frac{1}{T^2}x_i'(M_{\hat{F}_1}-M_{F_1^0})x_i\right)\hat{b}_i \\
&\geq\frac{1}{d_T}\min_{1\leq i\leq N}\mu_{\min}\left(\frac{d_T}{T^2}x_i'M_{F_1^0}x_i\right)\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_i\|^2-\max_{1\leq i\leq N}\frac{\|x_i\|^2}{T^2}\|P_{F_1^0}-P_{\hat{F}_1}\|\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_i\|^2 \\
&\geq\frac{1}{d_T}\left(\frac{1}{2}\rho_{\min}-o_P(d_T^{-1})\right)\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_i\|^2,
\end{aligned}
$$

where the second inequality follows from the fact that $\min_{1\le i\le N}\mu_{\min}\left(\frac{d_T}{T^2}x_i'M_{F_1^0}x_i\right)\ge\frac{1}{2}\rho_{\min}>0$ a.s. by Lemma A.2(iv), and $\max_{1\le i\le N}\frac{\|x_i\|^2}{T^2}\le\max_{1\le i\le N}d_T\mu_{\max}\left(\frac{x_i'x_i}{d_TT^2}\right)=O_P(d_T)$ by Lemma A.2(i). Then we have $\frac{1}{N}\sum_{i=1}^N\|\hat{b}_i\|^2=o_P(d_T^{-2})=o_P(1)$.

(iv) 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 and unobserved stationary common factors. We estimate $\hat{F}_1$ from equation (2.10) in Section 2.2 as follows

$$\left[\frac{1}{NT^2}\sum_{i=1}^N(y_i-x_i\hat{\beta}_i)(y_i-x_i\hat{\beta}_i)'\right]\hat{F}_1=\hat{F}_1V_{1,NT}. \tag{A.9}$$

Combining (A.9) and the fact that $y_i-x_i\hat{\beta}_i=-x_i\hat{b}_i+F^0\lambda_i^0+u_i=-x_i\hat{b}_i+F_1^0\lambda_{1i}^0+F_2^0\lambda_{2i}^0+u_i$, we have

$$\begin{aligned}\hat{F}_1V_{1,NT}=&\frac{1}{NT^2}\sum_{i=1}^N x_i\hat{b}_i\hat{b}_i'x_i'\hat{F}_1-\frac{1}{NT^2}\sum_{i=1}^N x_i\hat{b}_i\lambda_i^{0\prime}F^{0\prime}\hat{F}_1-\frac{1}{NT^2}\sum_{i=1}^N x_i\hat{b}_iu_i'\hat{F}_1\\&-\frac{1}{NT^2}\sum_{i=1}^N F^0\lambda_i^0\hat{b}_i'x_i'\hat{F}_1-\frac{1}{NT^2}\sum_{i=1}^N u_i\hat{b}_i'x_i'\hat{F}_1+\frac{1}{NT^2}\sum_{i=1}^N F^0\lambda_i^0u_i'\hat{F}_1\\&+\frac{1}{NT^2}\sum_{i=1}^N u_i\lambda_i^{0\prime}F^{0\prime}\hat{F}_1+\frac{1}{NT^2}\sum_{i=1}^N u_iu_i'\hat{F}_1+\frac{1}{NT^2}\sum_{i=1}^N F_2^0\lambda_{2i}^0\lambda_{2i}^{0\prime}F_2^{0\prime}\hat{F}_1\\&+\frac{1}{NT^2}\sum_{i=1}^N F_1^0\lambda_{1i}^0\lambda_{2i}^{0\prime}F_2^{0\prime}\hat{F}_1+\frac{1}{NT^2}\sum_{i=1}^N F_2^0\lambda_{2i}^0\lambda_{1i}^{0\prime}F_1^{0\prime}\hat{F}_1+\frac{1}{NT^2}\sum_{i=1}^N F_1^0\lambda_{1i}^0\lambda_{1i}^{0\prime}F_1^{0\prime}\hat{F}_1\\\equiv&I_1+...+I_{11}+\frac{1}{NT^2}\sum_{i=1}^N F_1^0\lambda_{1i}^0\lambda_{1i}^{0\prime}F_1^{0\prime}\hat{F}_1,\text{ say.}\end{aligned}$$

It follows that $\hat{F}_1V_{1,NT}-F_1^0(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1)=I_1+...+I_{11}$. 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 $\hat{F}_1H_1^{-1}-F_1^0=[I_1+...+I_{11}]\,(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1)^{-1}(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}$ and $\frac{1}{T}\left\|\hat{F}_1H^{-1}-F_1^0\right\|\le\frac{1}{T}(\|I_1\|+...+\|I_{11}\|)\left\|(\frac{1}{T^2}F_1^{0\prime}\hat{F}_1)^{-1}\right\|\times\left\|(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)^{-1}\right\|$. It remains to analyze $\|I_l\|$ for $l=1,2,...,11$. For I_1, we have that by the result in (iii),

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

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

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

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

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

where $\frac{1}{N}\sum_{i=1}^{N}\frac{\|u_i\|^2}{T} = O_P(1)$ by Assumption 3.1(i). Similarly, for I_4 and I_5,

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

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

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

where $u = (u_1, ..., u_N)'$ and we have used the fact that $\frac{1}{NT}\left\|\Lambda^{0\prime}u\right\|^2 = O_P(1)$ by Assumption 3.2(iii). Analogously, we can show that $\frac{1}{T}\|I_7\| = O_P(T^{-1/2}N^{-1/2})$. For I_8,

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

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 $\frac{1}{T}\|I_8\| = O_P(N^{-1/2}T^{-1} + T^{-3/2})$. 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 $\frac{1}{T}\left\|\hat{F}_1H_1^{-1} - F_1^0\right\| = O_P(\sqrt{d_T}\eta_{NT}) + \frac{1}{\sqrt{T}}O_P(C_{NT}^{-1})$. 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. Then*
(i) $\frac{1}{T}F_1^{0\prime}(\hat{F}_1 - F_1^0 H_1) = O_P(T\sqrt{d_T}\eta_{NT} + \delta_{NT}^{-1})$,
(ii) $\frac{1}{T}\hat{F}_1^{\prime}(\hat{F}_1 - F_1^0 H_1) = O_P(T\sqrt{d_T}\eta_{NT} + \delta_{NT}^{-1})$,
(iii) $\|P_{\hat{F}_1} - P_{F_1^0}\|^2 = O_P(\sqrt{d_T}\eta_{NT} + T^{-1}\delta_{NT}^{-1})$,
(iv) $\frac{1}{T}u_j^{*\prime}\left(\hat{F}_1 H_1^{-1} - F_1^0\right) = O_P(\sqrt{Td_T}\eta_{NT} + \delta_{NT}^{-1})$ for each $j = 1, ..., N$.

Lemma A.6 *Suppose that Assumptions 3.1-3.2 hold. Let* $R_{1i} = \frac{1}{T^2}x_i^{\prime}(P_{F_1^0} - P_{\hat{F}_1})u_i^*$, $R_{2i} = \frac{1}{T^2}x_i^{\prime}M_{\hat{F}_1}F_1^0\lambda_{1i}^0$ $-\frac{1}{NT^2}\sum_{j=1}^N x_i^{\prime}M_{\hat{F}_1}x_j a_{ij}\hat{b}_j + \frac{1}{NT^2}\sum_{j=1}^N a_{ij}x_i^{\prime}M_{\hat{F}_1}u_j$, $R_{3i} = \frac{1}{NT^2}\sum_{j=1}^N a_{ij}x_i^{\prime}(P_{F_1^0} - P_{\hat{F}_1})u_j$, *and* $R_{4i} = \frac{1}{T^2}x_i^{\prime}M_{F_1^0}u_i^* - \frac{1}{NT^2}\sum_{j=1}^N a_{ij}x_i^{\prime}M_{F_1^0}u_j$. *Then*
(i) $R_{1i} = O_P(\varsigma_{1NT})$ *for each* $i = 1, ..., N$, *and* $N^{-1}\sum_{i=1}^N \|R_{1i}\|^2 = O_P(\varsigma_{1NT}^2)$,
(ii) $R_{2i} = O_P(\varsigma_{2NT})$ *for each* $i = 1, ..., N$, *and* $N^{-1}\sum_{i=1}^N \|R_{2i}\|^2 = O_P(\varsigma_{2NT}^2)$,
(iii) $R_{3i} = O_P(\varsigma_{3NT})$ *for each* $i = 1, ..., N$, *and* $N^{-1}\sum_{i=1}^N \|R_{3i}\|^2 = O_P(\varsigma_{3NT}^2)$,
(iv) $R_{4i} = O_P(T^{-1})$ *for each* $i = 1, ..., N$, *and* $N^{-1}\sum_{i=1}^N \|R_{4i}\|^2 = O_P(T^{-2})$,
where $\varsigma_{1NT} = T^{-1/2}\sqrt{d_T}\eta_{NT} + d_T\eta_{NT}^2 + T^{-1}C_{NT}^{-1}$, $\varsigma_{2NT} = T^{-1}\sqrt{d_T}\eta_{NT} + d_T\eta_{NT}^2 + T^{-1}\delta_{NT}^{-1}$, *and* $\varsigma_{3NT} = T^{-1/2}d_T^{1/4}\eta_{NT}^{1/2} + T^{-1}\delta_{NT}^{-1/2}$.

Proof of Theorem 3.2. (i) Based on the sub-differential calculus, a necessary condition for $\hat{\beta}_i$, $\hat{\alpha}_k$, and $\hat{F}_1$ to minimize the objective function (2.9) is, for each $i = 1, ..., N$, that $0_{p\times 1}$ belongs to the sub-differential of $Q_{NT}^{\lambda,K}(\boldsymbol{\beta}, \boldsymbol{\alpha}, F_1)$ with respect to β_i (resp. α_k) evaluated at $\{\hat{\beta}_i\}$, $\{\hat{\alpha}_k\}$ and $\hat{F}_1$. That is, for each $i = 1, ..., N$ and $k = 1, ..., K$, we have

$$0_{p\times 1} = -\frac{2}{T^2}x_i^{\prime}M_{\hat{F}_1}(y_i - x_i\hat{\beta}_i) + \lambda\sum_{j=1}^K \hat{e}_{ij}\prod_{l=1,l\neq j}^K \|\hat{\beta}_i - \hat{\alpha}_l\|, \tag{A.10}$$

where $\hat{e}_{ij} = \frac{\hat{\beta}_i - \hat{\alpha}_j}{\|\hat{\beta}_i - \hat{\alpha}_j\|}$ if $\|\hat{\beta}_i - \hat{\alpha}_j\| \neq 0$ and $\|\hat{e}_{ij}\| \leq 1$ if $\|\hat{\beta}_i - \hat{\alpha}_j\| = 0$. Noting that $y_i = x_i\beta_i^0 + \hat{F}_1 H_1^{-1}\lambda_{1i}^0 + u_i^* + (F_1^0 - \hat{F}_1 H_1^{-1})\lambda_{1i}^0$, (A.10) implies that

$$\hat{Q}_{i,xx}\hat{b}_i = \frac{1}{T^2}x_i^{\prime}M_{\hat{F}_1}u_i^* + \frac{1}{T^2}x_i^{\prime}M_{\hat{F}_1}F_1^0\lambda_{1i}^0 - \frac{\lambda}{2}\sum_{j=1}^{K_0}\hat{e}_{ij}\prod_{l=1,l\neq j}^K \|\hat{\beta}_i - \hat{\alpha}_l\|, \tag{A.11}$$

which can be rewritten as

$$\hat{Q}_{i,xx}\hat{b}_i = -\frac{1}{NT^2}\sum_{j=1}^N x_i^{\prime}M_{\hat{F}_1}x_j a_{ij}\hat{b}_j + R_i, \tag{A.12}$$

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 $R_{5i} = \frac{\lambda}{2}\sum_{j=1}^K \hat{e}_{ij}\prod_{l=1,l\neq j}^K \|\hat{\beta}_i - \hat{\alpha}_l\|$. By Lemma A.6(i)-(iv), we have that $\sum_{l=1}^4 \frac{1}{N}\sum_{i=1}^N \|R_{li}\|^2 = O_P(T^{-1}d_T^{1/2}\eta_{NT} + d_T^2\eta_{NT}^4 + T^{-2}C_{NT}^{-2} + T^{-2}\delta_{NT}^{-1} + T^{-2}) = O_P(T^{-1}d_T^{1/2}\eta_{NT} + d_T^2\eta_{NT}^4 + T^{-2})$. 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 $\frac{1}{N}\sum_{i=1}^N \|R_i\|^2 = O_P(T^{-1}d_T^{1/2}\eta_{NT} + d_T^2\eta_{NT}^4 + T^{-2} + \lambda^2)$.

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

$$\hat{Q}_2 = \begin{pmatrix} \frac{1}{NT^2} x_1' M_{\hat{F}_1} x_1 a_{11} & \frac{1}{NT^2} x_1' M_{\hat{F}_1} x_2 a_{12} & \cdots & \frac{1}{NT^2} x_1' M_{\hat{F}_1} x_N a_{1N} \\ \frac{1}{NT^2} x_2' M_{\hat{F}_1} x_1 a_{21} & \frac{1}{NT^2} x_2' M_{\hat{F}_1} x_2 a_{22} & \cdots & \frac{1}{NT^2} x_2' M_{\hat{F}_1} x_N a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{NT^2} x_N' M_{\hat{F}_1} x_1 a_{N1} & \frac{1}{NT^2} x_N' M_{\hat{F}_1} x_2 a_{N2} & \cdots & \frac{1}{NT^2} x_N' M_{\hat{F}_1} x_N a_{NN} \end{pmatrix}.$$

Let $R = (R_1', ..., R_N')'$. Then (A.12) implies that $(\hat{Q}_1 - \hat{Q}_2)\hat{\mathbf{b}} = R$. It follows that

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

By Assumption 3.2(v), we have that $\mu_{\min}(\hat{Q}_1 - \hat{Q}_2) \geq \rho_{\min}/2 > 0$ w.p.a.1. Then $\frac{1}{N}\|\hat{\boldsymbol{b}}\|^2 \leq \frac{\rho_{\min}^2}{4N}\sum_{i=1}^N \|R_i\|^2 = O_P(T^{-1}d_T^{1/2}\eta_{NT} + d_T^2\eta_{NT}^4 + T^{-2} + \lambda^2) = O_P\left(d_T T^{-2} + \lambda^2\right)$. Consequently, $\frac{1}{N}\sum_{i=1}^N \|\hat{b}_i\|^2 = O_P\left(d_T T^{-2} + \lambda^2\right)$.

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

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

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

$$\begin{aligned} &T^2\left[Q_{NT}^{\lambda,K}(\boldsymbol{\beta} + d_T^{1/2}T^{-1}\boldsymbol{v}, \hat{\boldsymbol{\alpha}}, \hat{F}_1) - Q_{NT}^{\lambda,K}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, \hat{F}_1)\right] \\ &\geq \frac{d_T^{1/2}}{N}\sum_{i=1}^N\left(\frac{d_T^{1/2}}{T^2}v_i'x_i'M_{\hat{F}_1}x_i v_i - \frac{2}{T}v_i'x_i'M_{\hat{F}_1}(F_1^0 - \hat{F}_1H_1)\lambda_{1i}^0 - \frac{2}{T}v_i'x_i'M_{\hat{F}_1}u_i^*\right) \\ &= \frac{d_T}{N}\sum_{i=1}^N \frac{1}{T^2}v_i'x_i'M_{\hat{F}_1}x_i v_i \\ &\quad - \frac{2d_T^{1/2}}{N}\sum_{i=1}^N v_i'\left\{T \cdot R_{2i} + \frac{1}{T}x_i'M_{\hat{F}_1}u_i^* + \frac{1}{NT}\sum_{j=1}^N a_{ij}x_i'M_{\hat{F}_1}x_j\hat{b}_j - \frac{1}{NT}\sum_{j=1}^N a_{ij}x_i'M_{\hat{F}_1}u_j\right\} \\ &\equiv D_{1NT} - 2D_{2NT}, \end{aligned}$$

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

$D_{2NT,1} = \frac{T^2}{d_T N}\sum_{i=1}^N \|\bar{R}_{2i}\|^2$, $D_{2NT,2} = \frac{1}{d_T NT^2}\sum_{i=1}^N \|x_i' M_{\hat{F}_1} u_i^*\|^2$, $D_{2NT,3} = \frac{1}{d_T N^3 T^2}\sum_{i=1}^N \sum_{j=1}^N \|a_{ij} x_i' M_{\hat{F}_1} \hat{b}_j\|^2$, and $D_{2NT,4} = \frac{1}{d_T N^3 T^2}\sum_{i=1}^N \sum_{j=1}^N \|a_{ij} x_i' M_{\hat{F}_1} u_j\|^2$. By Lemmas A.6(i)-(ii) and A.5(iii), $D_{2NT,1} = \frac{T^2}{d_T} O_P(T^{-2} d_T \eta_{NT}^2 + d_T^2 \eta_{NT}^4 + T^{-2}\delta_{NT}^{-2}) = o_P(1)$, and $D_{2NT,2} \leq \frac{2T^2}{d_T N}\sum_{i=1}^N \left\| \frac{1}{T^2} x_i'(M_{\hat{F}_1} - M_{F_1^0}) u_i^* \right\|^2 + \frac{2}{d_T N}\sum_{i=1}^N \left\| \frac{1}{T} x_i' M_{F_1^0} u_i^* \right\|^2 = \frac{T^2}{d_T} O_P(T^{-1} d_T \eta_{NT}^2 + d_T^2 \eta_{NT}^4 + T^{-2} C_{NT}^{-2}) + \frac{1}{d_T} O_P(1) = o_P(1)$. Next,

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

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

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

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, $T^2[Q_{NT}^{\lambda,K}(\boldsymbol{\beta} + d_T^{1/2} T^{-1}\boldsymbol{v}, \hat{\boldsymbol{\alpha}}, \hat{F}_1) - Q_{NT}^{\lambda,K}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, \hat{F}_1)] > 0$ 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.11). For the first term, we have by Lemma A.6(i)

$$
\begin{aligned}
\left\| \frac{1}{T^2} x_i' M_{\hat{F}_1} u_i^* \right\| &\leq \left\| \frac{1}{T^2} x_i' M_{F_1^0} u_i^* \right\| + \left\| \frac{1}{T^2} x_i'(M_{\hat{F}_1} - M_{F_1^0}) u_i^* \right\| \\
&= O_P(T^{-1}) + O_P(T^{-1/2}\sqrt{d_T}\eta_{NT} + d_T\eta_{NT}^2 + T^{-1} C_{NT}^{-1}) = O_P(d_T T^{-1}). \qquad \text{(A.13)}
\end{aligned}
$$

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

$$\left\|\frac{1}{T^2}x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right\| \leq \|R_{2i}\| + \left\|\frac{1}{NT^2}\sum_{j=1}^N x_i'M_{\hat{F}_1}x_j\hat{b}_j a_{ij}\right\| + \left\|\frac{1}{NT^2}\sum_{j=1}^N x_i'M_{\hat{F}_1}u_j a_{ij}\right\|$$

$$= O_P(T^{-1}\sqrt{d_T}\eta_{NT} + d_T\eta_{NT}^2 + T^{-1}\delta_{NT}^{-1}) + O_P(\eta_{NT}) + O_P(d_T T^{-1}) = O_P(d_T T^{-1}). \tag{A.14}$$

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

(iii) Let $P_{NT}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{N}\sum_{i=1}^N \prod_{k=1}^K \|\beta_i - \alpha_k\|$ and $\hat{c}_{iNT}(\alpha) = \prod_{k=1}^{K-1}\|\hat{\beta}_i - \alpha_k\| + \prod_{k=1}^{K-2}\|\hat{\beta}_i - \alpha_k\| \times \|\beta_i^0 - \alpha_K\| + ... + \prod_{k=2}^K \|\beta_i^0 - \alpha_k\|$. By SSP, we have that as $(N,T) \to \infty$, $\left|\prod_{k=1}^K \|\hat{\beta}_i - \alpha_k\| - \prod_{k=1}^K \|\beta_i^0 - \alpha_k\|\right| \leq \hat{c}_{iNT}(\alpha)\|\hat{\beta}_i - \beta_i^0\|$, where $\hat{c}_{iNT}(\alpha) \leq C_{KNT}(\alpha)(1 + 2\|\hat{\beta}_i - \beta_i^0\|)$ and $C_{KNT}(\alpha) = \max_{1\leq i\leq N}\max_{1\leq s\leq k\leq K-1} \prod_{k=1}^s c_{ks}\|\beta_i^0 - \alpha_k\|^{K-1-s} = \max_{1\leq l\leq K}\max_{1\leq s\leq k\leq K_0-1}\prod_{k=1}^s c_{ks}\|\alpha_l^0 - \alpha_k\|^{K-1-s} = O(1)$ with c_{ks} being finite integers. It follows that as $(N,T) \to \infty$

$$|P_{NT}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}) - P_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\alpha})| \leq C_{KNT}(\alpha)\frac{1}{N}\sum_{i=1}^N \|\hat{b}_i\| + 2C_{KNT}(\alpha)\frac{1}{N}\sum_{i=1}^N \|\hat{b}_i\|^2$$

$$\leq C_{KNT}(\alpha)\left\{\frac{1}{N}\sum_{i=1}^N \|\hat{b}_i\|^2\right\}^{1/2} + O_P(d_T T^{-2}) = O_P(d_T^{1/2}T^{-1}). \tag{A.15}$$

By (A.15) and the fact that $P_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0) = 0$ and that $P_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - P_{NT}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}^0) \leq 0$. we have

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

$$= \frac{1}{N}\sum_{i=1}^N\prod_{k=1}^K \|\beta_i^0 - \hat{\alpha}_k\| + O_P(d_T^{1/2}T^{-1})$$

$$= \frac{N_1}{N}\prod_{k=1}^K \|\hat{\alpha}_k - \alpha_1^0\| + \frac{N_2}{N}\prod_{k=1}^K \|\hat{\alpha}_k - \alpha_2^0\| + ... + \frac{N_K}{N}\prod_{k=1}^K \|\hat{\alpha}_k - \alpha_K^0\| + O_P(d_T^{1/2}T^{-1}). \tag{A.16}$$

By Assumption 3.3(i), $N_k/N \to \tau_k \in (0,1)$ for each $k = 1, ...K$. So (A.16) 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(iv) and Theorem 3.2(i), we have $\frac{1}{T}\|\hat{F}_1 - F_1^0 H_1\|^2 = O_P(T d_T \eta_{NT}^2 + C_{NT}^{-2}) = O_P(d_T^2 T^{-1} + N^{-1})$. ■

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) $P\left(\max_{1\leq i\leq N}\left\|\frac{1}{T^2}x_i'u_i^*\right\| > c\psi_{NT}\right) = o(N^{-1})$,
(ii) $P\left(\max_{1\leq i\leq N}\left\|\frac{1}{T^2}x_i'M_{F_1^0}u_i^*\right\| > c d_T\psi_{NT}\right) = o(N^{-1})$.

Lemma A.8 *Suppose that Assumptions 3.1-3.3 hold. Then for any* $c>0$,

(i) $P\left(\max_{1\le i\le N}\|R_{1i}\|>c(d_T\eta_{NT}+T^{-1/2}d_T^{1/2}C_{NT}^{-1})\left(\psi_{NT}+T^{-1/2}(\log T)^3\right)\right)=o(N^{-1})$,

(ii) $P\left(\max_{1\le i\le N}\|R_{2i}\|>cd_T^{1/2}N^{(1/2q)}\varsigma_{2NT}\right)=o(N^{-1})$,

(iii) $P\left(\max_{1\le i\le N}\|R_{3i}\|>cd_T^{1/2}N^{(1/2q)}\varsigma_{3NT}\right)=o(N^{-1})$,

(iv) $P\left(\max_{1\le i\le N}\|R_{4i}\|>c(d_T+N^{(1/2q)})\psi_{NT}\right)=o(N^{-1})$,

(v) $P\left(\max_{1\le i\le N}\left\|\hat{\beta}_i-\beta_i^0\right\|>c\left(N^{(1/2q)}\psi_{NT}+\lambda(\log T)^{\epsilon/2}\right)\right)=o(N^{-1})$ *for any* $\epsilon>0$,

(vi) $P\left(\frac{1}{N}\sum_{i=1}^N\left\|\hat{\beta}_i-\beta_i^0\right\|^2>cd_T^2\psi_{NT}^2\right)=o(N^{-1})$ *for any* $\epsilon>0$,

(vii) $P\left(\max_{1\le i\le N}\left\|\frac{1}{T^2}x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0\right\|>cN^{1/2q}(d_T\eta_{NT}+T^{-1/2}d_T^{1/2}C_{NT}^{-1})\right)=o(N^{-1})$.

Proof of Theorem 3.3. (i) Fix $k\in\{1,...,K\}$. By the consistency of $\hat{\alpha}_k$ and $\hat{\beta}_i$, we have $\hat{\beta}_i-\hat{\alpha}_l\xrightarrow{p}\alpha_k^0-\alpha_l^0\neq 0$ for all $i\in G_k^0$ and $l\neq k$. Now, suppose that $\|\hat{\beta}_i-\hat{\alpha}_k\|\neq 0$ for some $i\in G_k^0$. Then the first order condition (with respect to β_i) for the minimization of the objective function (2.8) implies that

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

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

$$P\left(\max_{i\in G_k^0}\left\|\hat{\beta}_i-\beta_i^0\right\|>c\Psi_{NT}\right)=o(N^{-1})\text{ and }P\left(\frac{1}{N}\sum_{i=1}^N\left\|\hat{\beta}_i-\beta_i^0\right\|^2>cd_T^2\psi_{NT}^2\right)=o(N^{-1}).\quad\text{(A.17)}$$

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

$$P(\|\hat{\alpha}_k-\alpha_k^0\|>cd_T\psi_{NT})=o(N^{-1}),\text{ and }P(\max_{i\in G_k^0}\left|\hat{c}_{ki}-c_k^0\right|\geq c_k^0/2)=o(N^{-1}).\quad\text{(A.18)}$$

By (A.17)-(A.18) and the fact that $\max_{i\in G_k^0}\frac{1}{T^2}x_i'M_{\hat{F}_1}x_i\le cd_T\rho_{\max}$ 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) and A.8(i), (vii), we have $P\left(\max_{i\in G_k^0}\|\hat{A}_{1i}\|>cTb_T\psi_{NT}\right)=o(N^{-1})$, $P\left(\max_{i\in G_k^0}\|\hat{A}_{3i}\|>cN^{1/2q}(Td_T\eta_{NT}+T^{1/2}d_T^{1/2}C_{NT}^{-1})\right)=o(N^{-1})$, and $P\left(\max_{i\in G_k^0}\|\hat{A}_{2i}\|>c(Td_T\eta_{NT}+T^{1/2}d_T^{1/2}C_{NT}^{-1})\left(\psi_{NT}+T^{-1/2}(\log T)^3\right)\right)=o(N^{-1})$.

For $\hat{A}_{5i}$, we have

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

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

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

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

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

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

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

where the last inequality follows because $T\lambda \gg Tb_T\psi_{NT}$ by Assumption 3.3(iv). Consequently, we can conclude that w.p.a.1, $\hat{\beta}_i - \hat{\alpha}_k$ must be in a position where $\|\beta_i - \alpha_k\|$ is not differentiable with respect to β_i for any $i \in G_k^0$. That is, $P(\|\hat{\beta}_i - \hat{\alpha}_k\| = 0|i \in G_k^0) = 1 - o(N^{-1})$ 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 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)* $\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0=\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N x_i'M_{\hat{F}_1}x_ja_{ij}\hat{b}_j-\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N a_{ij}x_i'M_{\hat{F}_1}u_j$ $-\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N a_{ij}x_i'M_{\hat{F}_1}F_2^0\lambda_{2j}^0+o_P(N^{-1/2}T^{-1})$,

(ii) $\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_i'M_{\hat{F}_1}x_i=\frac{1}{N_kT^2}\sum_{i\in G_k^0}x_i'M_{F_1^0}x_i+o_P(1)$,

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

(iv) $\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j\in\hat{G}_l}x_i'M_{\hat{F}_1}x_ja_{ij}=\frac{1}{N_kT^2}\sum_{i\in G_k^0}\frac{1}{N}\sum_{j\in G_l^0}x_i'M_{F_1^0}x_ja_{ij}+o_P(1)$.

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

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

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

(iii) $V_{NT}\overset{d}{\to}N(0,\Omega_0)$ *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 $\{\hat{\beta}_i\}$ and $\{\hat{\alpha}_k\}$ to minimize the objective function in (2.9) is that for each $i=1,...,N$ (resp. $k=1,...,K$), the null vector $0_{p\times 1}$ belongs to the sub-differential of $Q_{NT}^{\lambda,K}(\boldsymbol{\beta},\boldsymbol{\alpha},\hat{F}_1)$ with respect to β_i (resp. α_k) evaluated at $\{\hat{\beta}_i\}$ and $\{\hat{\alpha}_k\}$. That is, for each $i=1,...,N$ and $k=1,...,K$, we have

$$0_{p\times 1}=-\frac{2}{NT^2}x_i'M_{\hat{F}_1}(y_i-x_i\hat{\beta}_i)+\frac{\lambda}{N}\sum_{j=1}^K\hat{e}_{ij}\prod_{l=1,l\neq j}^K\|\hat{\beta}_i-\hat{\alpha}_l\|, \tag{A.19}$$

$$0_{p\times 1}=\frac{\lambda}{N}\sum_{i=1}^N\hat{e}_{ik}\prod_{l=1,l\neq k}^K\|\hat{\beta}_i-\hat{\alpha}_l\|, \tag{A.20}$$

where $\hat{e}_{ij}=\frac{\hat{\beta}_i-\hat{\alpha}_j}{\|\hat{\beta}_i-\hat{\alpha}_j\|}$ if $\|\hat{\beta}_i-\hat{\alpha}_j\|\neq 0$ and $\|\hat{e}_{ij}\|\leq 1$ if $\|\hat{\beta}_i-\hat{\alpha}_j\|=0$. First, we observe that $\|\hat{\beta}_i-\hat{\alpha}_k\|=0$ for any $i\in\hat{G}_k$ by the definition of $\hat{G}_k$, implying that $\hat{\beta}_i-\hat{\alpha}_l\to\alpha_k^0-\alpha_l^0\neq 0$ 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 $\hat{e}_{ij}=\frac{\hat{\beta}_i-\hat{\alpha}_j}{\|\hat{\beta}_i-\hat{\alpha}_j\|}=\frac{\hat{\alpha}_k-\hat{\alpha}_j}{\|\hat{\alpha}_k-\hat{\alpha}_j\|}$ w.p.a.1 for any $i\in\hat{G}_k$ and $j\neq k$. This further implies that w.p.a.1 $\sum_{i\in\hat{G}_k}\sum_{j=1,j\neq k}^K\hat{e}_{ij}\prod_{l=1,l\neq j}^K\|\hat{\beta}_i-\hat{\alpha}_l\|=\sum_{i\in\hat{G}_k}\sum_{j=1,j\neq k}^K\frac{\hat{\alpha}_k-\hat{\alpha}_j}{\|\hat{\alpha}_k-\hat{\alpha}_j\|}\prod_{l=1,l\neq j}^K\|\hat{\alpha}_k-\hat{\alpha}_l\|=0_{p\times 1}$, and

$$\begin{aligned}0_{p\times 1}&=\sum_{i=1}^N\hat{e}_{ik}\prod_{l=1,l\neq k}^K\|\hat{\beta}_i-\hat{\alpha}_l\|\\&=\sum_{i\in\hat{G}_k}\hat{e}_{ik}\prod_{l=1,l\neq k}^K\|\hat{\alpha}_k-\hat{\alpha}_l\|+\sum_{i\in\hat{G}_0}\hat{e}_{ik}\prod_{l=1,l\neq k}^K\|\hat{\beta}_i-\hat{\alpha}_l\|+\sum_{j=1,j\neq k}^K\sum_{i\in\hat{G}_j}\hat{e}_{ik}\prod_{l=1,l\neq k}^K\|\hat{\alpha}_j-\hat{\alpha}_l\|\\&=\sum_{i\in\hat{G}_k}\hat{e}_{ik}\prod_{l=1,l\neq k}^K\|\hat{\alpha}_k-\hat{\alpha}_l\|+\sum_{i\in\hat{G}_0}\hat{e}_{ik}\prod_{l=1,l\neq k}^K\|\hat{\beta}_i-\hat{\alpha}_l\|.\end{aligned} \tag{A.21}$$

Then by (A.19)–(A.21) we have

$$\frac{2}{N_kT^2}\sum_{i\in\hat{G}_k}x_i'M_{\hat{F}_1}(y_i-x_i\hat{\alpha}_k)+\frac{\lambda}{N}\sum_{i\in\hat{G}_0}\hat{e}_{ik}\prod_{l=1,l\neq k}^K\|\hat{\beta}_i-\hat{\alpha}_l\|=0_{p\times 1}. \tag{A.22}$$

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 $y_i = x_i\alpha_k^0 + F_1^0\lambda_{1i}^0 + u_i^*$ when $i \in G_k^0$, we have

$$\begin{aligned}\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x_iM_{\hat{F}_1}y_i =& \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x_i'M_{\hat{F}_1}x_i\beta_i^0 + \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 + \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x_i'M_{\hat{F}_1}u_i^* \\ =& \frac{1}{N_kT^2}\sum_{i\in G_k^0} x_i'M_{\hat{F}_1}x_i\alpha_k^0 + \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k\setminus G_k^0} x_i'M_{\hat{F}_1}x_i\beta_i^0 - \frac{1}{N_kT^2}\sum_{i\in G_k^0\setminus\hat{G}_k} x_i'M_{\hat{F}_1}x_i\alpha_k^0 \\ &+ \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 + \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x_i'M_{\hat{F}_1}(u_i + F_2^0\lambda_{2i}^0). \qquad \text{(A.23)}\end{aligned}$$

Combining (A.22) and (A.23) yields

$$\begin{aligned}\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x_i'M_{\hat{F}_1}x_i(\hat{\alpha}_k - \alpha_k^0) =& \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 + \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x_i'M_{\hat{F}_1}\left(u_i + F_2^0\lambda_{2i}^0\right) \\ &+ \hat{C}_{1k} - \hat{C}_{2k} + \hat{C}_{3k}, \qquad \text{(A.24)}\end{aligned}$$

where $\hat{C}_{1k} = \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k\setminus G_k^0} x_i'M_{\hat{F}_1}x_i\beta_i^0$, $\hat{C}_{2k} = \frac{1}{N_kT^2}\sum_{i\in G_k^0\setminus\hat{G}_k} x_i'M_{\hat{F}_1}x_i\alpha_k^0$, and $\hat{C}_{3k} = \frac{\lambda}{2N_k}\sum_{i\in\hat{G}_0}\hat{e}_{ik} \times \prod_{l=1,l\neq k}^K \|\hat{\beta}_i - \hat{\alpha}_l\|$. 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(i), we have as $\frac{\sqrt{N}}{T} \to 0$

$$\begin{aligned}\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k} x_i'M_{\hat{F}_1}F_1^0\lambda_{1i}^0 =& \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N x_i'M_{\hat{F}_1}x_ja_{ij}\hat{b}_j - \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N a_{ij}x_i'M_{\hat{F}_1}u_j \\ &- \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N a_{ij}x_i'M_{\hat{F}_1}F_2^0\lambda_{2j}^0 + o_P(N^{-1/2}T^{-1}). \qquad \text{(A.25)}\end{aligned}$$

In addition,

$$\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N x_i'M_{\hat{F}_1}x_ja_{ij}\hat{b}_j = \frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{l=1}^K\sum_{j\in\hat{G}_l} x_i'M_{\hat{F}_1}x_ja_{ij}\left(\hat{\alpha}_l - \alpha_l^0\right) + o_P(N^{-1/2}T^{-1}) \qquad \text{(A.26)}$$

by Theorem 3.3. Let $\hat{Q}_{1NT} = \text{diag}\left(\frac{1}{N_1T^2}\sum_{i\in\hat{G}_1} x_i'M_{\hat{F}_1}x_i, \ldots, \frac{1}{N_KT^2}\sum_{i\in\hat{G}_K} x_i'M_{\hat{F}_1}x_i\right)$ and $\hat{Q}_{2NT}$ is a

$Kp \times Kp$ matrix with typical blocks $\frac{1}{NN_kT}\sum_{i\in\hat{G}_k}\sum_{j\in\hat{G}_l} a_{ij}x_i'M_{\hat{F}_1}x_j$ such that

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

Combining (A.24)–(A.26), 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

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

and $D_{NK}=\text{diag}(\frac{N}{N_1},...,\frac{N}{N_K})\otimes I_p$. 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,

$$\sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0) = Q_{NT}^{-1}\sqrt{D_{NK}}\left(V_{NT}+B_{NT}\right)+o_P(1), \tag{A.27}$$

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

(ii) By Lemma A.10 (i) and (iii), we have

$$Q_{NT}\xrightarrow{d} Q_0 \text{ and } V_{NT}\xrightarrow{d} N(0,\Omega_0) \text{ conditional } \mathcal{C}. \tag{A.28}$$

Combining (A.27)–(A.28) yields $\sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0) - \sqrt{D_{NK}}Q_{NT}^{-1}B_{NT}\xrightarrow{d} MN(0, D_0Q_0^{-1}\Omega_0Q_0^{-1})$. ■

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

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

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

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

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

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

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

It suffices to show that $\sqrt{NT}\text{vec}(\hat{\boldsymbol{\alpha}}^{bc}_{\hat{G}} - \boldsymbol{\alpha}^0) = \sqrt{D_{NK}}Q_{NT}^{-1}V_{NT} + o_P(1)$ 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

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

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

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

where $\bar{\lambda}_{2i}^0 = \lambda_{2i}^0 - \frac{1}{N}\sum_{j=1}^N\lambda_{2j}^0 a_{ij}$. Let $\phi^{f_2,f_1f_2} = (\phi^{f_2f_1}(L), \phi^{f_2f_2}(L))$, $\phi_i^{\varepsilon,f_1f_2} = (\phi_i^{\varepsilon f_1}(L), \phi_i^{\varepsilon f_2}(L)) = (\phi^{\varepsilon f_1}(L), \phi^{\varepsilon f_2}(L))$, and $v_t^{f_1f_2} = (v_t^{f_1\prime}, v_t^{f_1\prime})'$. Note that $\varepsilon_{it} = w_{it}^{\varepsilon} = \phi_i^{\varepsilon u}(L)v_{it}^u + \phi_i^{\varepsilon\varepsilon}(L)v_{it}^{\varepsilon} + \phi^{\varepsilon f_1}(L)v_t^{f_1} +$

$\phi^{\varepsilon f_2}(L)\, v_t^{f_2}$. By the BN decomposition and the independence of $\{v_{it}^{u\varepsilon}\}$ and $\{v_s^{f_1 f_2}\}$, we have

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

where $V_t^{f_1 f_2} = (V_t^{f_1 \prime}, V_t^{f_2 \prime})' = \left(\sum_{m=1}^t v_m^{f_1 \prime}, \sum_{m=1}^t v_m^{f_2 \prime} \right)'$, w_{it} and $\tilde{w}_{it}$ are defined in Assumption 3.1. Let $B_{kNT,21}^* = \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} S_2 \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r} \phi_{i,l}' S_4' \bar{\lambda}_{2i}^0$. It follows that

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

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 Huang et al. (2018), we can show that $\frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} S_2 R_{iT,l}^{f_2} S_4' \bar{\lambda}_{2i}^0 = o_P(1)$ for $l = 1, 2, ..., 6$ and $\frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \mathbb{E}(Q_{iT}^{f_2}) = 0$. It follows that $B_{kNT,21} = B_{kNT,21}^* + o_P(1) = \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \Delta_{24,i} \bar{\lambda}_{2i}^0 + o_P(1)$. Analogously, we have $B_{kNT,22} = B_{kNT,22}^* + o_P(1)$, where $B_{kNT,22}^* = \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \varkappa_{ts} \mathbf{1}\{s \leq$

$t\} \times S_2 \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r}\phi_{i,l}' S_4' \bar{\lambda}_{2i}^0$. Let $B_{kNT,2}^* = B_{kNT,21}^* - B_{kNT,22}^*$. Then

$$
\begin{aligned}
B_{kNT,2}^* =& \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\mathbf{1}\{s=t\} - \varkappa_{ts}\mathbf{1}\left\{s \le t\right\}\right) S_2 \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r}\phi_{i,l}' S_4' \bar{\lambda}_{2i}^0 \\
=& \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t} \bar{\varkappa}_{ts} \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \left(\phi_{l+r}^{\varepsilon f_1}\phi_l^{f_2 f_1} + \phi_{l+r}^{\varepsilon f_2}\phi_l^{f_2 f_2}\right) \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \bar{\lambda}_{2i}^0 \\
=& \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t} \bar{\varkappa}_{ts} \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \Delta_{24,i} \bar{\lambda}_{2i}^0.
\end{aligned}
$$

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 $\tilde{B}_{kNT,2} = \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t} \hat{\varkappa}_{ts} \hat{\Delta}_{24,i} \hat{\bar{\lambda}}_{2i}$. 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 $\tilde{B}_{kNT,2} - B_{kNT,2}^* = \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} (\hat{\Delta}_{24,i} \hat{\bar{\lambda}}_{2i} - \Delta_{24,i} \bar{\lambda}_{2i}^0) - \frac{1}{\sqrt{N_k} T} \sum_{i \in G_k^0} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{1}\left\{s \le t\right\} [\hat{\varkappa}_{ts} \hat{\Delta}_{24,i} \hat{\bar{\lambda}}_{2i} - \varkappa_{ts} \Delta_{24,i} \bar{\lambda}_{2i}^0] = o_P(1) - o_P(1) = o_P(1)$ by Lemma A.11(vii)-(viii). Consequently, $\hat{B}_{kNT,2} - B_{kNT,2} = o_P(1)$.

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

(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 $y_i^+ = x_i \alpha_k^0 + F_1^0 \lambda_{1i}^0 + F_2^0 \lambda_{2i}^0 + u_i^+$, by (A.24) and (A.25) and Theorem 3.3, we have

$$
\begin{aligned}
\frac{1}{N_k T^2} \sum_{i \in \hat{G}_k} x_i' M_{\hat{F}_1} x_i (\tilde{\alpha}_{G_k}^{fm} - \alpha_k^0) =& \frac{1}{N_k T^2} \sum_{i \in G_k^0} x_i' M_{\hat{F}_1} \left(u_i^+ + F_2^0 \lambda_{2i}^0\right) + \frac{1}{N_k T^2} \sum_{i \in \hat{G}_k} x_i' M_{\hat{F}_1} F_1^0 \lambda_{1i}^0 \\
& - \frac{1}{\sqrt{N_k} T} B_{kNT,1}^+ - \frac{1}{\sqrt{N_k} T} B_{kNT,2} + o_P(N^{-1/2} T^{-1}). \qquad \text{(A.29)}
\end{aligned}
$$

Combining (A.26), (A.29) and Lemma A.9(i) yields

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

By (A.26) and Lemma A.10 (i)-(iii), we have $\sqrt{N}T \text{vec}(\tilde{\boldsymbol{\alpha}}_{G}^{fm} - \boldsymbol{\alpha}^0) = (\hat{Q}_{1NT} - \hat{Q}_{2NT})^{-1} \sqrt{D_{NK}} [(U_{NT}^{u+} +$

$U_{NT}^{f_2}) - B_{NT,1}^{+} - B_{NT,2}] + o_P(1) = \sqrt{D_{NK}} Q_{NT}^{-1} V_{NT}^{+} + o_P(1)$, where

$$U_{k,NT}^{u+} = \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} x_i' M_{F_1^0} \left(u_i^{+} - \frac{1}{N} \sum_{j=1}^{N} a_{ij} u_j^{+} \right),$$

$$U_{k,NT}^{f_2} = \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} x_i' M_{F_1^0} \left(F_2^0 - \frac{1}{N} \sum_{j=1}^{N} a_{ij} F_2^0 \right),$$

$$V_{kNT,1}^{+} = \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} S^{\varepsilon} \phi_i^{\dagger}(1) \sum_{t=1}^{T} \sum_{s=1}^{T} \left\{ \bar{\varkappa}_{ts} \left(V_{it}^{u\varepsilon} v_{is}^{u\varepsilon,+\prime} \right) - \left[\mathbf{1}\{t=s\} - \varkappa_{ts} \mathbf{1}\{s \le t\} \right] I_{1+p} \right\} \phi_i^{\dagger}(1)' S^{w\prime},$$

$$V_{kNT,2}^{+} = \frac{1}{\sqrt{N_k}} \sum_{i=1}^{N} \left\{ \frac{1}{T} \mathbb{E}\left(x_i' | \mathcal{C} \right) \mathbf{1}\left\{ i \in G_k^0 \right\} - \frac{1}{N} \sum_{j \in G_k^0} a_{ij} \frac{1}{T} \mathbb{E}(x_j' | \mathcal{C}) \right\} M_{F_1^0} u_i^{+},$$

$$V_{kNT,3} = \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} [x_i - \mathbb{E}_{\mathcal{C}}(x_i)]' M_{F_1^0} F_2^0 \lambda_{2i}^0,$$

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 $w_{it}^{+} = (u_{it}^{+}, \Delta x_{it}', \Delta f_{1t}', f_{2t}',)'$ 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 $\Omega_{NT}^{+} = \text{Var}\left(V_{NT}^{+} | \mathcal{C} \right)$. Then we have

$$\sqrt{N}T \text{vec}(\tilde{\boldsymbol{\alpha}}_G^{fm} - \boldsymbol{\alpha}^0) \xrightarrow{d} \mathcal{MN}(0,\ D_0 Q_0^{-1} \Omega_0^{+} Q_0^{-1}).$$

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

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

where $\tilde{U}^{u+}_{kNT} = \frac{1}{\sqrt{N_k}T} \sum_{i \in G^0_k} x_i' M_{\hat{F}_1} \left(u^+_i - \frac{1}{N} \sum_{j=1}^N a_{ij} u^+_j \right)$ and $\tilde{U}^{u+}_{kNT} - U^{u+}_{kNT} = o_P(1)$ 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 3.5 hold. Then*

(i) For any $1 \leq r \leq r^0$, $V_1(r, \hat{G}^r) - V_1(r, G^0 H^r) = O_P(C^{-1}_{NT})$,

(ii) For each r *with* $0 \leq r < r^0$, *there exist a positive number* c_r *such that* $\text{plim inf}_{(N,T)\to\infty} [V_1(r, G^0 H^r) - V_1(r^0, G^0)] = c_r$,

(iii) For any fixed r, *with* $r^0 \leq r \leq r_{\max}$, $V_1(r, \hat{G}^r) - V_1(r^0, \hat{G}^{r^0}) = O_P(C^{-2}_{NT})$,

where $V_1(r, G^0 H^r)$ *is defined analogously to* $V_1(r, \hat{G}^r)$ *with* $\hat{G}^r$ *replaced by* $G^0 H^r$, $H^r = (N^{-1}\Lambda^{0\prime}\Lambda^0) \times (T^{-1} G^{0\prime} \hat{G}^r)$, *and* $G^0 = \Delta F^0$.

Lemma A.13 *Suppose that Assumptions 3.1-3.3 and 3.6 hold. Then*

(i) For any $1 \leq r_1 \leq r^0_1$, $V_2(r_1, \hat{F}^{r_1}_1) - V_2(r_1, F^0_1 H^{r_1}_1) = O_P(\sqrt{T})$,

(ii) For any $1 \leq r_1 < r^0_1$, $\text{plim inf}_{(N,T)\to\infty} d_T T^{-1} [V_2(r_1, F^0_1 H^{r_1}_1) - V_2(r_1, F^0_1)] = d_{r_1}$ *for some* $d_{r_1} > 0$,

(iii) For any $r^0_1 \leq r_1 \leq r_{\max}$, $V_2(r_1, \hat{f}^{r_1}) - V_2(r^0_1, \hat{f}^{r^0_1}) = O_P(1)$,

where $V_2(r_1, F^0_1 H^{r_1}_1)$ *is defined analogously to* $V_2(r_1, \hat{F}^{r_1}_1)$ *with* $\hat{F}^{r_1}_1$ *replaced by* $F^0_1 H^{r_1}_1$, *and* $H^{r_1}_1 = (N^{-1}\Lambda^{0\prime}\Lambda^0) \times (T^{-2} F^{0\prime} \hat{F}^{r_1})$.

Proof of Theorem 3.6. Noting that $IC_1(r) - IC_1(r^0) = V_1(r, \hat{G}^r) - V_1(r^0, \hat{G}^{r^0}) - (r^0 - r) g_1(N, T)$, it suffices to show that $P\left(V_1(r, \hat{G}^r) - V_1(r^0, \hat{G}^{r^0}) < (r^0 - r) g_1(N, T) \right) \to 0$ as $(N, T) \to \infty$ when $r \neq r^0$. We consider the under- and over-fitted models, respectively. When $0 \leq r < r^0$, we make the following decomposition:

$$V_1(r,\hat{G}^r)-V_1(r^0,\hat{G}^{r^0}) =[V_1(r,\hat{G}^r)-V(r,G^0H^r)]+[V_1(r,G^0H^r)-V_1(r^0,G^0H^{r^0})] + [V_1(r^0,G^0H^{r^0})-V_1(r^0,\hat{G}^{r^0})] \equiv DV_{1,1}+DV_{1,2}+DV_{1,3}.$$

$DV_{1,l}=O_P(C_{NT}^{-1})$ for $l=1,3$ by Lemma A.12(i). Noting that $V_1(r^0,G^0H^{r^0})=V_1(r^0,G^0)$, $\text{plim inf}_{(N,T)\to\infty} DV_{1,2}=c_r$ when $r<r^0$ by Lemma A.12(ii). It follows that $P(IC_1(r)<IC_1(r^0))\to 0$ as $g_1(N,T)\to 0$ as $(N,T)\to\infty$ under Assumption 3.5.

Now, we consider the case where $r^0<r\leq r_{\max}$. Note that $C_{NT}^2[V_1(r,\hat{G}^r)-V_1(r^0,\hat{G}^{r^0})]=O_P(1)$ and $C_{NT}^2(r-r^0)g_1(N,T)>C_{NT}^2g_1(N,T)\to\infty$ by Lemma A.12(iii) and Assumption 3.5, we have $P(IC_1(r)<IC_1(r^0))=P(V_1(r,\hat{G}^r)-V_1(r^0,\hat{G}^{r^0})<(r^0-r)g_1(N,T))\to 0$ as $(N,T)\to\infty$. ■

Proof of Theorem 3.7. Noting that $IC_2(r_1)-IC_2(r_1^0)=V_2(r_1,\hat{F}_1^{r_1})-V_2(r_1^0,\hat{F}_1^{r_1^0})-(r_1^0-r_1)g_2(N,T)$, it suffices to show that $P\left(V_2(r_1,\hat{F}_1^{r_1})-V_2(r_1^0,\hat{F}_1^{r_1^0})<(r_1^0-r_1)g_2(N,T)\right)\to 0$ as $(N,T)\to\infty$ when $r\neq r^0$. First, when $r_1<r_1^0$, we consider the decomposition

$$V_2(r_1,\hat{F}_1^{r_1})-V_2(r_1^0,\hat{F}_1^{r_1^0})=[V_2(r_1,\hat{F}_1^{r_1})-V_2(r_1,F_1^0H_1^{r_1})]+[V_2(r_1,F_1^0H_1^{r_1})-V_2(r_1^0,F_1^0H_1^{r_1^0})] + [V(r_1^0,F_1^0H_1^{r_1^0})-V(r_1^0,\hat{F}_1^{r_1^0})]\equiv DV_{2,1}+DV_{2,2}+DV_{2,3}.$$

By Lemma A.13, $DV_{2,1}=O_P(T^{1/2})$, $DV_{2,2}$ is of exact probability order $O_P(T/\log\log T)$, and $DV_{2,3}=O_P(1)$. It follows that

$$P(IC_2(r_1)<IC_2(r_1^0))=P\left(V_2(r_1,\hat{F}_1^{r_1})-V_2(r_1^0,\hat{F}_1^{r_1^0})<(r_1^0-r_1)g_2(N,T)\right)\to 0$$

as $g_2(N,T)(\log\log T)/T\to 0$ under Assumption 3.5.

Next, for $r_1>r_1^0$, we have $V(r_1,\hat{F}_1^{r_1})-V(r_1^0,\hat{F}_1^{r_1^0})=O_P(1)$ for $r_1>r_1^0$ by Lemma A.13(iii), and $(r_1-r_1^0)g_2(N,T)\to\infty$ by Assumption 3.5. This implies that $P(IC_2(r_1)-IC_2(r_1^0)<0)=P(V_2(r_1,\hat{F}_1^{r_1})-V_2(r_1^0,\hat{F}_1^{r_1^0})<(r_1^0-r_1)g_2(N,T))\to 0$ 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* $\max_{K_0\leq K\leq K_{\max}}|\hat{\sigma}^2_{G_{(K,\lambda)}}-\hat{\sigma}^2_{\hat{G}(K_0,\lambda)}|=O_P(\nu_{NT}^2)$, *where* $\hat{\sigma}^2_{G_{(K,\lambda)}}=\frac{1}{NT}\sum_{k=1}^K\sum_{i\in\hat{G}_k(K,\lambda)}\sum_{t=1}^T[y_{it}-\hat{\alpha}^{cup\prime}_{\hat{G}_k(K,\lambda)}x_{it}-\hat{\lambda}_{1i}(K,\lambda)'\hat{f}_{1t}(K,\lambda)]^2$ *and* ν_{NT} *is defined in Section 3.6.*

Proof of Theorem 3.8. First, we show that

$$IC_3(K_0,\lambda)=\ln[V_3(K_0)]+pK_0g_3(N,T) = \ln\frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k(K_0,\lambda)}\sum_{t=1}^T\left[y_{it}-\hat{\alpha}^{fm\prime}_{\hat{G}_k(K_0,\lambda)}x_{it}-\hat{\lambda}_{1i}(K_0,\lambda)'\hat{f}_{1t}(K_0,\lambda)\right]^2+o(1)\xrightarrow{p}\ln(\sigma_0^2).$$

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

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

By Assumption 3.6 and Slutsky's Lemma, we can demonstrate

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

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 is no unobserved common factor and no endogeneity in x_{it}, $\delta^2_{NT}[\hat{\sigma}^2_{\hat{G}(K,\lambda)} - \hat{\sigma}^2_{\hat{G}(K_0,\lambda)}] = O_P(1)$ when there are only unobserved nonstationary common factors and $C^2_{NT}[\hat{\sigma}^2_{\hat{G}(K,\lambda)} - \hat{\sigma}^2_{\hat{G}(K_0,\lambda)}] = O_P(1)$ when there are both nonstationary and stationary common factors. Then by Lemma 14,

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

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