Nonstationary Panel Model with Latent Group Structures and Cross-sectional Dependence*

Wenxin Huang[a], Peter C.B. Phillips[b], and Liangjun Su[a]
[a] School of Economics, Singapore Management University
[b] Yale University, University of Auckland
University of Southampton & Singapore Management University

November 20, 2017

Abstract

This article proposes a novel approach, based on Lasso, to handle unobserved heterogeneity in nonstationary panel model with cross-sectional dependence. We employ the penalized principal component (PPC, hereafter) estimation method to jointly estimate the group-specific long-run relations, unobserved common factors and identify individuals' membership. We obtain three types of estimators–C-Lasso, post-Lasso and Cup-Lasso estimators by iteratively performing the PPC-based method. In post-Lasso and Cup-Lasso estimators, we apply the fully modified procedure for bias-correction. Taken together, our estimators achieve the oracle property so that the group-specific coefficients can be estimated as well as if the individuals' membership were known. We establish the convergence rates and limiting distributions of the C-Lasso, post-Lasso and the Cup-Lasso estimators, which are normal and permit inference using standard test statistics. An empirical example is presented based on growth convergence puzzle through the channel of global technology diffusions. It empirically confirms the multiple steady states of growth convergence.

JEL Classification: C13; C33; C38; C51; F60; O32; O40.

Keywords: Nonstationary; Parameter heterogeneity; Latent group patterns; Penalized principal component; Cross-sectional dependence; Lasso; Growth convergence puzzle

---

*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. 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

Econometric methods for nonstationary panel model have been extensively used in applied economics. The asymptotic properties for it are well explored in classical settings, such as the assumptions of common long-run relations and cross-sectional independence among individuals. Even though those assumptions offer estimation efficiency and help us simplify the asymptotic theory, they are often hard to meet in real-world problems. In one case, researchers often face the issue of unobserved parameter heterogeneity that figures within the model, including the "convergence club" (Durlauf and Johnson (1995), Quah (1997)), "Lucas paradox" (Lucas (1990)), the relation between income and democracy (Acemoglu et al. (2008)), and "resources curse" (Van der Ploeg (2011)). In another case, globalization and international spillovers raise to a new challenge–cross-sectional dependence among individuals. In general, both the presence of unobserved heterogeneity and cross-sectional dependence can substantially complicate statistical inference for nonstationary panels. The classical settings, which ignore these two issues, lead to inconsistent estimation and imprecise inference. Essentially, the main contributions of our paper are not only to offer a flexible and comprehensive econometric model, which is closer to the real economic questions, but also seek to maintain certain degree of parsimony by imposing the latent group structures.

In this paper, we propose a novel econometric method that allow us to study the unobserved parameter heterogeneity and cross-sectional dependence simultaneously in nonstationary panel model, especially account for the problems facing in economic growth literature. Recently there is a growing list of theoretical papers accounting for the unobserved heterogeneity in large dimensional panel models by imposing latent group patterns, see Bonhomme and Manresa (2015), Su, Shi and Phillips (2016, SSP hereafter), Su et al. (2017). Our baseline model is obtained from SSP, where they employ the *Classifier-Lasso* (C-Lasso, hereafter) technique to study the unobserved grouped patterns in stationary panel data. The results of C-Lasso estimators identify group membership and estimate the group-specific slope coefficients simultaneously via shrinkage procedure. Huang et al. (2017) establish the asymptotic theory for the latent group patterns in cointegrated panel models. They do not allow for cross-sectional dependence. Under our grouped nonstationary panel model, cross-sectional dependence is characterized by unobserved common factors, which can be either stationary or nonstationary. Then we provide a purely data-driven method, based on penalized principal component (PPC, hereafter) method, to account for the unobserved parameter heterogeneity and cross-sectional dependence simultaneously in nonstationary panels. Our framework allows us to jointly estimate the group-specific long-run relations, unobserved common factors and identify individuals' membership. Thus, our results generalize the SSP model by allowing nonstationary variables and cross-sectional dependence.

By using PPC-based methods, the grouped nonstationary panel model provides a practical approach to maintaining the estimation efficiency gains from employing the panel data while allowing some degree of freedom on parameter heterogeneity. We show that PPC-based methods provide consistent estimators to the group-specific long-run relations and unobserved common factors even when individuals' membership were unknown. Moreover, the PPC-based methods simultaneously account for the issue of cross-sectional dependence in panels. It is commonly acknowledged that many economic variables exhibit common patterns across individuals due to global shocks, spatial effects or as a result of social interactions. When existing unobserved common patterns, the classical least square methods lead to several problems, such as biased inference, inconsistent estimators and spurious regression, see Baltagi and Pesaran (2007) and Bai et al.(2009). There are two main approaches to study cross-sectional dependence–the factor structure approach and the spatial approach, see Bai and Ng (2002, 2004) and Lee (2004). In this sense, our PPC-based method is also related to the factor structure model, see Bai and Ng (2002), Bai and Ng (2004), Pesaran (2006), and Moon and Weidner (2015). The multi-factor structure model assumes that cross-sectional dependence is characterized by the common factors, which can be estimated by either principal component method or cross-sectional mean method. In this paper, we employ the principal component method, proposed by Bai and Ng (2004), to estimate the unobserved common patterns.

The international R&D spillovers model of Coe and Helpman(1995) (CH, hereafter) motivates our empirical application. They estimate total factor productivity (TFP, hereafter) on domestic R&D capital stock and foreign R&D capital stock to study the two sources of technology changes–domestic innovations and catch-up effects. The innovation part is explained by the increasing function of TFP on domestic R&D capital stock. And the catch-up effects through the channel of international R&D spillovers favor the hypothesis of convergence across economies. Our main empirical interest is to explain the "growth convergence puzzle" through the technology spillovers model. There are two main problems in the econometric methodology of CH model. First, the important assumption underlying the CH model is that all countries obey a common linear specification. However, cross-countries productivity behavior typically reaches multiple steady states. And most theoretical growth models suggest the multiple regimes of convergence across economies, see Solow (1956) and Eaton and Kortum (2002), which imply the unobserved parameter heterogeneity within the economic growth model. It is a natural relaxation to allow the parameters vary across countries. Second, those economic variables, like TFP and R&D stocks, apparently share some common patterns, such as global technology trends, international financial crisis shocks, and oil price shocks. Obviously, the CH model fails to account for the unobserved parameter heterogeneity and common patterns due to the limitations in econometric methodologies. As a result, the international R&D spillovers model may be misspecified, which leads to biased estimates and incorrect inference. In general, our econometric model yields a direct solution for the unobserved heterogeneity and cross-sectional dependence, first, to allow the latent group structures in parameters of interests and, second, to estimate the unobserved common patterns directly from data. From those two features we simultaneously identify the multiple regimes of convergence across economies from the international R&D spillovers model and account for unobserved technology trends across countries.

In this paper, we first introduce a nonstationary panel model with latent group structures and cross-sectional dependence, where the slope coefficients are heterogeneous across groups and homogeneous within a group. Then we propose a penalized principal component-based (PPC-based, hereafter) method that jointly estimates the group-specific long-run relations, unobserved common factors and infers group membership. Further, we iteratively perform the PPC-based method and obtain three types of estimators–C-Lasso, post-Lasso and continuous-updated-Lasso (Cup-Lasso, hereafter) estimators. In asymptotic theory, we establish the preliminary rates of convergence for the group-specific long-run relations and unobserved common factors. Based on the preliminary rates of convergence, we establish the classification consistency, which indicates that all individuals are classified into correct group with a probability approaching one (w.p.a.1). Third, our long-run estimators have asymptotic biases since we allow for weakly dependent error processes and unobserved stationary common factors. The first source of biases is common acknowledged in nonstationary time series due to serial correlation and endogeneity issues. An additional bias comes from the unobserved stationary common factors. Therefore we employ the fully modified procedures, proposed by Phillips and Hansen (1990) for bias-correction. After bias-correction, our estimators achieve the $\sqrt{N}T$ consistency in homogeneous nonstationary panel model. Fourth, we establish the oracle properties of the C-Lasso estimators, post-Lasso and Cup-Lasso versions, which are asymptotically equivalent to the corresponding infeasible estimators, obtained by knowing the exact individuals' group membership. At last, we develop the limiting distributions of group-specific estimators, which help to make inference about our group-specific long-run relations. Three information criteria are introduced to estimate the number of unobserved common factors and the number of groups. We demonstrate that those information criteria can select the correct number of unobserved common factors and group w.p.a.1. Our simulation results show good finite sample performance for both estimation and classification.

Because our PPC-based estimation method allows us to account for the unobserved heterogeneity and cross-sectional dependence simultaneously, it is the best fitted method to study the heterogeneous behavior in growth convergence model. In empirical application, we report both pooled FMOLS estimates and group-specific Cup-Lasso estimates with cross-sectional dependence, comparing with Coe et al. (2009)'s (CH2009, hereafter) estimates. The pooled FMOLS estimates are quantitatively similar to CH2009 ones. It confirms the international R&D spillovers after controlling unobserved global trends. Then we notice that the group-specific Cup-Lasso estimates show heterogeneous behavior. It indicates multiple regimes of growth convergence. In addition, we document the group classification results. The countries are classified into three groups–"Convergence", "Divergence", and "Balance". The major sources of technology changes in "Convergence" group come from global technology diffusions. As a result, the catch-up effects through the channel of technology diffusion are the main forces towards convergence in income. On the contrary, countries in "Divergence" group show an opposite story. The technology changes rely mainly on their domestic R&D stock. They fail to benefit from international R&D spillovers. For the "Balance" group, they have balanced sources in technology growth both innovations and international spillovers. In general, we identify multiple regimes of growth convergence across economies through the channel of technology changes.

Our econometric theory also speak to recent literature trying to detect the unobserved heterogeneity by grouping or clustering. For example, Yuan and Lin (2006) consider the problem of selecting grouped variables for accurate prediction in regression. Qian and Su (2016) study the unobserved group structure on time dimension to detect the structure breaks. Bonhomme and Manresa (2015) allow time-varying grouped patterns of heterogeneity in stationary linear panel models. Their focus is to detect the latent group patterns in fixed effects. Sarafidis and Weber (2015) propose a partially heterogeneous model for the panel data with fixed T, where the cross-sectional individuals are grouped into clusters. One advantage of our approach is that we allow nonstationary variables and estimate the long-run cointegrating relations with latent group structures. Our approach also simultaneously handle the unobserved common patterns across individuals and the unobserved heterogeneity. Lastly, our paper is closely related to the long literature on economic growth, in particular the analysis of global technology diffusion. In this context, our results using grouped nonstationary panel models with cross-sectional dependence provide a purely data-driven method for the "convergence puzzle". It empirically identifies the multiple steady states in growth convergence.

This paper is structured as follows. Section 2 introduces nonstationary panel model with latent group structures and cross-sectional dependence and proposes a penalized principal component estimation. Section 3 explains the main assumptions and establishes the asymptotic properties of three types of Lasso estimators. Section 4 reports simulation results. Section 5 studies the heterogeneous behavior of growth convergence. Section 6 concludes. All proofs are relegated to the appendix.

NOTATION. Hereafter, we write the integral $\int_0^1 W(s)ds$ as $\int W$ and define $\Omega^{1/2}$ to be any matrix such that $\Omega = (\Omega^{1/2})(\Omega^{1/2})'$, and $BM(\Omega)$ to denote Brownian motion with the covariance matrix Ω. For any $m \times n$ real matrix A, we write the Frobenius norm $\|A\|$, the spectral norm $\|A\|_{sp}$, the transpose A'. The operator $\overset{d}{\to}$ denotes convergence in distribution, $\overset{p}{\to}$ convergence in probability, $\Rightarrow$ weak convergence, $a.s.$ almost surely, and $[x]$ the largest integer less than or equal to x. When A is symmetric, we use $\mu_{\max}(A)$ and $\mu_{\min}(A)$ to denote its largest and smallest eigenvalues, respectively. Let $M < \infty$ be a generic positive number, not depending on T or N. We also define the matrix that projects onto orthogonal space of A as $M_A = I_T - A(AA')^{-1}A'$. Let $0_{p\times 1}$ denote a $p \times 1$ vector of zeros and $\mathbf{1}\{\cdot\}$ the indicator function. We use "p.d." and "p.s.d." to abbreviate "positive definite" and "positive semidefinite", respectively. Unless indicated explicitly, we use $(N, T) \to \infty$ to stand for that N and T pass jointly to infinity.

2 Model and Estimation

In this section, we first introduce the nonstationary panel model with latent group structures and cross-sectional dependence. Then we propose a penalized principal component method to estimate the model.

2.1 Nonstationary panel model with latent group structures and cross-sectional dependence

The generating process of (y_{it}, x_{it}) is 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, e_{it} is the error term and assumed to be cross-sectionally dependent due to unobserved common factors, ε_{it} is assumed to have zero mean and finite long-run variance, and β_i^0 is a $p \times 1$ vector of unknown long-run cointegrating relations. We assume that the long-run cointegrating relations β_i are heterogeneous across groups and homogeneous within a group. And we denote the true values of β_i as β_i^0, to follow the latent group structures, such that

$$\beta_i^0 = \begin{cases} \alpha_1^0 & \text{if } i \in G_1^0 \\ \vdots & \vdots \\ \alpha_K^0 & \text{if } i \in G_K^0 \end{cases}, \tag{2.2}$$

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, we assume that the number of group K is known and fixed but each individual's group membership is unknown. A information criterion is proposed to determine the number of groups in Section 3.6.

Since e_{it} is assumed to be cross-sectionally dependent, it allows for unobserved common patterns across individual i. By Bai and Ng (2004), we impose a multi-factor structure on e_{it} to model cross-sectional dependence. That is,

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

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), λ_i 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 and assumed to cross-sectionally independent. We emphasize that cross-sectional dependence only comes from common factors f_t such that e_{it} and e_{jt} are correlated due to common factors f_t in the form of $E(e_{it}e_{jt}) = \lambda_i' E(f_t f_t')\lambda_j \neq 0$.

If f_t only contains stationary factors, in some cases we still can obtain consistent estimators of β_i by the penalized-least-squares based (hereafter, PLS-based) method, proposed by Huang et al. (2017) when ignoring cross-sectional dependence. However, if there are serial correlations between dependent variable x_{it} and unobserved common factors f_t, ignoring those factors f_t yields biased inference for β_i. Furthermore, the unobserved nonstationary factors will lead to inconsistency due to spurious regression. In general, we fail to obtain consistent and unbiased group-specific estimators by the PLS-based method when existing cross-sectional dependence in nonstationary panel models.

Now we impose the multi-factor error structure to the first equation of (2.1) as follows

$$y_{it} = \beta_i^{0\prime} x_{it} + \lambda_i^{0\prime} f_t^0 + u_{it}. \tag{2.3}$$

Our estimation procedures are performed on model (2.3) by penalized principal component method, proposed in Section 2.2. Let

$$\boldsymbol{\alpha} \equiv (\alpha_1, ..., \alpha_{K_0}), \quad \boldsymbol{\beta} \equiv (\beta_1, ..., \beta_N), \quad \Lambda = (\lambda_1, ..., \lambda_N)', \quad \text{and} \quad f = (f_1, ..., f_T)'.$$

The true values of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \Lambda$ and f are denoted as $\boldsymbol{\alpha}^0, \boldsymbol{\beta}^0, \Lambda^0$, and f^0 respectively. We also use $\alpha_k^0, \beta_i^0, \lambda_i^0$ and f_t^0 denote the true value of $\alpha_k, \beta_i, \lambda_i$ and f_t. Our interest is to infer each individual's group identity and obtain consistent estimators of both group-specific long-run relations α_k and unobserved common factors f_t.

2.2 Penalized principal component estimation

In this section, we propose an iterative PPC-based procedure to simultaneously estimate the long-run relations β_i, unobserved common factors f_t and identify group membership. Here, we rewrite the nonstationary panel model with latent group structures and multi-factor error structure (2.3) in vector form,

$$y_i = x_i\beta_i^0 + f^0\lambda_i^0 + u_i = x_i\beta_i^0 + f_1^0\lambda_{1i}^0 + f_2^0\lambda_{2i}^0 + u_i, \tag{2.4}$$

where $f^0 = (f_1^0, f_2^0)$, $\lambda_i^0 = (\lambda_{1i}^{0\prime}, \lambda_{2i}^{0\prime})'$, $y_i = (y_{i1}, ..., y_{iT})'$, x_i, f_1^0, f_2^0, and u_i are analogously defined. As we discussed in Section 2.1, we can still obtain consistent estimates of β_i when ignoring the unobserved stationary common factors. The principal component estimators of β_i and f_1^0 are obtained from the following least objective function

$$SSR(\beta_i, f_1, \Lambda_1) = \sum_{i=1}^{N}(y_i - x_i\beta_i - f_1\lambda_{1i})'(y_i - x_i\beta_i - f_1\lambda_{1i}), \tag{2.5}$$

subject to the constraint $\frac{f_1' f_1}{T^2} = I_{r_1}$ and $\Lambda_1'\Lambda_1$ being diagonal. Define the projection matrix $M_{f_1} = I_T - P_{f_1} = I_T - \frac{f_1 f_1'}{T^2}$. We can obtain the least squares estimator of β_i for each given f_1 is

$$\hat{\beta}_i = (x_i' M_{f_1} x_i)^{-1} x_i' M_{f_1} y_i.$$

Given β_i, the variable $e_i = y_i - x_i\beta_i = f\lambda_i + u_i$ has a pure factor structure. Let $e = (e_1, e_2, ..., e_N)$, a $T \times N$ matrix and $\Lambda_1 = (\lambda_{11}, ..., \lambda_{1N})'$ a $N \times r_1$ matrix. We can obtain the following least squares objective function for f_1

$$tr\left[(e - f_1\Lambda_1')(e - f_1\Lambda_1')'\right].$$

By Bai (2009), we can concentrating out Λ_1 by its least square estimator, such that $\Lambda_1 = e' f_1 (f_1' f_1)^{-1} = e' f_1 / T^2$. The objective function (2.5) becomes

$$tr\left(e' M_{f_1} e\right) = tr(e'e) - tr\left(f_1' e e' f_1 / T^2\right).$$

The final least squares estimator $(\hat{\beta}, \hat{f}_1)$ is the solution of the set of nonlinear equations,

$$\hat{\beta}_i = \left(x_i' M_{\hat{f}_1} x_i\right)^{-1}\left(x_i' M_{\hat{f}_1} y_i\right), \tag{2.6}$$

$$\hat{f}_1 V_{1,NT} = \left[\frac{1}{NT^2}\sum_{i=1}^{N}\left(y_i - x_i\hat{\beta}_i\right)\left(y_i - x_i\hat{\beta}_i\right)'\right]\hat{f}_1, \tag{2.7}$$

where $M_{\hat{f}_1} = I_T - \frac{1}{T^2}\hat{f}_1\hat{f}_1'$, $\frac{1}{T^2}\hat{f}_1'\hat{f}_1 = I_{r_1}$, and $V_{1,NT}$ is a diagonal matrix consisting of the r_1 largest eigenvalues of the matrix inside the brackets, arranged in decreasing order. Based on (2.6) and (2.7), we can further show that $\hat{\Lambda}_1'\hat{\Lambda}_1$ is a diagonal matrix with descending diagonal elements as follows,

$$\frac{1}{N}\hat{\Lambda}_1'\hat{\Lambda}_1 = T^{-2}\hat{f}_1'\left(\frac{1}{NT^2}\sum_{i=1}^{N}\left(y_i - x_i\hat{\beta}_i\right)\left(y_i - x_i\hat{\beta}_i\right)'\hat{f}_1\right) = \left(\frac{1}{T^2}\hat{f}_1'\hat{f}_1\right)V_{1,NT} = V_{1,NT}.$$

Given the initial estimates of β_i and f_1 obtained from (2.6) and (2.7), we propose the penalized principal component method to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, where $\boldsymbol{\beta}$ exhibits the latent group structures. The PPC criterion function is given by

$$Q_{NT}^{\lambda,K}(\boldsymbol{\beta},\boldsymbol{\alpha},f_1) = Q_{NT}(\boldsymbol{\beta},f_1) + \frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\beta_i - \alpha_k\| \tag{2.8}$$

where $Q_{NT}(\boldsymbol{\beta},f_1) = \frac{1}{NT^2}\sum_{i=1}^{N}\left(y_i - x_i\beta_i\right)' M_{f_1}\left(y_i - x_i\beta_i\right)$, $\lambda = \lambda(N,T)$ is the tuning parameter. Minimizing the PPC criterion function in (2.8) produces the *Classifier-Lasso* (C-Lasso, hereafter) estimators of β_i and α_k, respectively. Then we update the estimates of the nonstationary common factors f_1 as follows

$$\hat{f}_1 V_{1,NT} = \left[\frac{1}{NT^2}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}(y_i - x_i\hat{\alpha}_k)(y_i - x_i\hat{\alpha}_k)'\right]\hat{f}_1. \tag{2.9}$$

with the identification restrictions: $\frac{1}{T^2}\hat{f}_1'\hat{f}_1 = I_{r_1}$ and $\hat{\Lambda}_1'\hat{\Lambda}_1$ is a diagonal matrix with descending diagonal elements. Since we allow for both stationary and nonstationary common factors, we minimize the following equation to obtain consistent estimates of the stationary common factors f_2,

$$\hat{f}_2 V_{2,NT} = \left[\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k}(y_i - x_i\hat{\alpha}_k - \hat{f}_1\hat{\lambda}_{1i})(y_i - x_i\hat{\alpha}_k - \hat{f}_1\hat{\lambda}_{1i})'\right]\hat{f}_2. \tag{2.10}$$

with the identification restrictions: $\frac{1}{T}\hat{f}_2'\hat{f}_2 = I_{r_2}$ and $V_{2,NT}$ is a diagonal matrix with descending diagonal elements. After obtaining the estimates of f_2, we apply bias-correction in post-Lasso estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. The biases emerge from the unobserved stationary common factors, endogeneity, and serial correlation issues from the weakly dependent error terms.

Now we summarize the estimation procedures in PPC-based estimation method. We first obtain the prior estimates of $\hat{\beta}_i$ and $\hat{f}_1$ by solving equations (2.6) and (2.7). Second, we minimize the above PPC criterion function (2.8), which produces the C-Lasso estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$. Third, with C-Lasso estimates of α, we update the estimates of nonstationary common factor f_1 by (2.9) and estimate the stationary common factors f_2 by (2.10). Forth, we apply bias-correction by the fully modified method in the post-Lasso estimator of $\boldsymbol{\alpha}$, which is explained in Section 3.4. We iterate steps 2–4 until achieving convergence to obtain the Cup-Lasso estimators. Our estimators, which we will refer to as "C-Lasso", "post-Lasso", and "Cup-Lasso", are based on the optimal group on the cross-sectional individuals, according to the PPC criterion function. The triplet $(\hat{\beta},\hat{\alpha},\hat{f}_1)$ jointly minimizes the objective function (2.8). Let $\hat{\beta}_i$ and $\hat{\alpha}_k$ denote the i^{th} and k^{th} columns of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$, respectively, i.e., $\hat{\boldsymbol{\beta}} \equiv (\hat{\beta}_1,...,\hat{\beta}_N)$ and $\hat{\boldsymbol{\alpha}} \equiv (\hat{\alpha}_1,...,\hat{\alpha}_K)$. We will study the asymptotic properties of the C-Lasso, post-Lasso and Cup-Lasso estimators below.

3 Asymptotic Theory

3.1 Main assumptions

In this subsection, we introduce the main assumptions that are needed to study the asymptotic properties of our estimators $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\alpha}}$ and $\hat{f}_1$.

Let $Q_{ixx}(f_1) = \frac{1}{T^2} x_i' M_{f_1} x_i$, $Q_1(f_1) =$diag$(Q_{1,xx}, ..., Q_{N,xx})$, 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 $w_{it} = (u_{it}, \varepsilon_{it}', \Delta f_{1t}^{0\prime}, f_{2t}^{0\prime},)'$. Let M be a generic constant that can vary across lines.

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

(ii) $\max_{1 \leq i \leq N} \sum_{j=0}^{\infty} j^k \|\phi_{ij}\| < \infty$ and $|\phi_i(1)| \neq 0$ for some $k \geq 2$.

(iii) u_{it} and ε_{it} are cross-sectionally independent conditional on $\mathcal{C}$.

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

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

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

Since both 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 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 for our initial estimators. The finite $2q + \varepsilon$ moments for $q > 4$ ensure the validity of the law of large numbers (LLN) and functional central limit theory (FCLT) for the weakly dependent linear process $\{w_{it}\}$. We will frequently apply the Beveridge and Nelson (BN) 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 $\tilde{\phi}_{ij} = \sum_{s=j+1}^{\infty} \phi_s$. Assumption 3.1(ii) gives the summability conditions on the coefficients matrix ϕ_{ij}. By Lemma (BN) in Phillips and Solo (1992), we have $\sum_{j=1}^{\infty} j^k \|\phi_{ij}\|^k < \infty \rightarrow \sum_{j=0}^{\infty} \|\tilde{\phi}_{ij}\|^k < \infty$, which implies that $\tilde{w}_{it}$ has Wold decomposition and behaves like a stationary process. Specifically, we have $\sum_{j=0}^{\infty} \left\|\tilde{\phi}_{ij}\right\|^2 < \infty$ under $\sum_{j=1}^{\infty} j^{1/2}\|\phi_{ij}\| < \infty$. The suitable choice of k ensures that finite kth moment of $\tilde{w}_{it}$. In our case, we need strong conditions to ensure the uniform behavior across i. The second part of Assumption 3.1(ii) rules out potential cointegration relation among x_{it} and f_{1t}^0. Assumption 3.1(iii) emphasizes that the cross-sectional dependence only comes from the unobserved common factors. Assumption 3.1(iv) ensures that the factor loadings are independent of the generalization of the error processes both over t and across i.

Assumption 3.1 ensures the multivariate invariance principle for the partial sum process of w_{it}. That is,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} w_{it} \Rightarrow B_i(r) \equiv BM_i(\Omega_i) \text{ as } T \rightarrow \infty \text{ for all } i,$$

where $B_i = (B_{1i}, B_{2i}', B_3', B_4')'$ is a $(1+p+r_1+r_2) \times 1$ vector of Brownian motions with long-run covariance matrix Ω_i. We can also define the temporal variance $\Sigma_i = E(w_{i0}w_{i0}')$ and the one-sided long-run covariance matrix $\Delta_i = \sum_{j=0}^{\infty} E(w_{i0}w_{ij}') = \Gamma_i + \Sigma_i$ of $\{w_{it}\}$, where Ω_i has the following partition

$$\Omega_i = \sum_{j=-\infty}^{\infty} E(w_{ij}w_{i0}') = \Gamma_i' + \Gamma_i + \Sigma_i = \begin{pmatrix} \Omega_{11,i} & \Omega_{12,i} & \Omega_{13,i} & \Omega_{14,i} \\ \Omega_{21,i} & \Omega_{22,i} & \Omega_{23,i} & \Omega_{24,i} \\ \Omega_{31,i} & \Omega_{32,i} & \Omega_{33} & \Omega_{34} \\ \Omega_{41,i} & \Omega_{42,i} & \Omega_{43} & \Omega_{44} \end{pmatrix}.$$

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 such that $S_1 w_{it} = u_{it}, S_2 w_{it} = \varepsilon_{it}, S_3 w_{it} = \Delta f_{1t}^0$, and $S_4 w_{it} = f_{2t}^0$.

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

(ii) $E\|\Delta f_{1t}^0\|^{2q+\epsilon} \le M$ *and* $E\|f_{2t}^0\|^{2q+\epsilon} \le M$ *for some* $\epsilon > 0$, $q \ge 4$ *and for all* t. *As* $T \rightarrow \infty$, $\frac{1}{T^2}\sum_{t=1}^T f_{1t}^0 f_{1t}^{0\prime} \xrightarrow{d} \int B_3 B_3$ *and* $\frac{1}{T}\sum_{t=1}^T f_{2t}^0 f_{2t}^{0\prime} \xrightarrow{p} \Sigma_{44} > 0$, *where* B_3 *is a* r_1*-vector of Brownian motions with long-run covariance matrix* $\Omega_{33} > 0$.

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

(iv) There exists a constant $\rho_{\min} > 0$ *such that* $P\left(\min_{1\le i\le N} \inf_{f_1} \mu_{\min}\left(Q_1(f_1) - 2Q_2(f_1)\right) \ge c\rho_{\min}\right) = 1 - o(N^{-1})$, *where the* inf *is taken respect to* f_1 *such that* $\frac{1}{T^2}f_1'f_1 = I_{r_1}$.

Assumption 3.1(i)-(iii) imposes standard moment conditions in the factor literature; see, e.g., Bai and Ng (2002, 2004). The last condition in 3.1(i) indicates that the stationary factor loadings and the nonstationary factor loadings can be only weakly correlated, which will greatly facilitate the derivation. Assumption 3.1(iii) imposes conditions on the error processes $\{u_{it}\}$, which are adapted from Bai (2003) and allow for weak forms of cross-sectional and serial dependence in the error processes. Assumption 3.1(iv) assumes $Q_1(f_1) - 2Q_2(f_1)$ 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 $(N,T) \to \infty$, $N/T^2 \to c_1 \in [0,\infty)$, $T/N^2 \to c_2 \in [0,\infty)$.

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

Assumption 3.3(i)-(ii) are borrowed from 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 the group-specific parameters. Similar conditions are assumed in the panel literature with latent group patterns, see, e.g., Bonhomme and Manresa (2015), Ando and Bai (2016), and Su et al. (2017). Assumption 3.3(iii)-(iv) imposes conditions to control the relative rates at which N and T pass to infinity. Note that N can pass to infinity at a faster or slower rate than T. The involving of d_T is due to the law of iterated logarithm, such that $d_T = O(\log\log T)$. One can verify that the range of values for λ to 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 $H_1 = (\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0)(\frac{1}{T^2}f_1^{0\prime}\hat{f}_1) \times V_{1,NT}^{-1}$. The consistency of $\hat{\beta}_i$ and $\hat{f}_1$ is ensured by the following theorem.

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) $\frac{1}{T}\|\hat{f}_1 - f_1^0 H_1\| = O_P(\eta_{NT}) + \frac{1}{\sqrt{T}} O_P(C_{NT}^{-1})$.

Theorem 3.1(i) establishes the weighted mean square consistency of $\{\hat{\beta}_i\}$. 3.1(ii) shows that the space 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 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. Compared to Bai and Ng (2004) and Bai et al. (2009), our results allow for both 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 and the estimated factors as well.

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

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

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

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

(iv) $T^{-1}\|\hat{f}_1 - f_1^0 H_1\|^2 = O_P(N^{-1} + d_T^2 T^{-1})$.

Theorem 3.2(i)-(ii) establishes the mean-square and point-wise convergence of the slope coefficients β_i, respectively. The usual super consistency of nonstationary estimators $\hat{\beta}_i$ is preserved if $\lambda = O(T^{-1})$ despite the fact that we ignore the 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, $\alpha_1^0, ..., \alpha_{K_0}^0$, 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's. Let $\hat{G}_k = \{i \in \{1, 2, ..., N\} : \hat{\beta}_i = \hat{\alpha}_k\}$ for $k = 1, ..., K$. Let $\hat{G}_0$ denote the group of individuals in $\{1, 2, ..., N\}$ that are not classified into any of the K groups.

3.3 Classification consistency

In this subsection, we study the 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 $k = 1, ...K^0$. 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 the 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 the uniform classification consistency– all individuals within a certain group, say G_k^0, can be simultaneously 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.

3.4 The oracle properties, post-Lasso and Cup-Lasso estimators

In this subsection, we study the oracle properties of PPC-based estimators. To proceed, we add some notations. For $k=1,...,K$, we define

$$
\begin{aligned}
U_{kNT} &= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0} x_i M_{f_1^0}\left(\left(u_i + f_2^0\lambda_{2i}^0\right) - \frac{1}{N}\sum_{j=1}^{N}\left(u_j + f_2^0\lambda_{2j}^0\right)a_{ij}\right),\\
B_{kNT,1} &= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathbf{1}\left\{t=s\right\} - \varkappa_{ts}\mathbf{1}\left\{s\le t\right\}\right]\right)\Delta_{21,i},\\
B_{kNT,2} &= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0} E_{\mathcal{C}}\left(x_i\right)' M_{f_1^0}\left(\lambda_{2i}^0 - \frac{1}{N}\sum_{j=1}^{N}\lambda_{2j}^0 a_{ij}\right),\\
V_{kNT} &= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0} S^{\varepsilon}\phi_i^{\dagger}(1)\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\bar{\varkappa}_{ts}\left(V_{it}^{u\varepsilon}v_{is}^{u\varepsilon\prime}\right) - \left[\mathbf{1}\left\{t=s\right\} - \varkappa_{ts}\mathbf{1}\left\{s\le t\right\}\right]I_{1+p}\right\}\phi_i^{\dagger}(1)'S^{u\prime}\\
&\quad + \frac{1}{\sqrt{N_k}T}\sum_{i=1}^{N}\left\{E_{\mathcal{C}}\left(x_i'\right)\mathbf{1}\left\{i\in G_k^0\right\} - \frac{1}{N}\sum_{j\in G_k^0} a_{ij}E_{\mathcal{C}}(x_j')\right\} M_{f_1^0}u_i\\
&\quad + \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\left[x_i - E_{\mathcal{C}}\left(x_i\right)\right]' M_{f_1^0} f_2^0\lambda_{2i}^0,
\end{aligned}
$$

where $\varkappa_{ts} = f_{1t}^{0\prime}(f_1^{0\prime}f_1^0)^{-1}f_{1s}^0$, $\bar{\varkappa}_{ts} = \mathbf{1}\left\{t=s\right\} - \varkappa_{ts}$, $\mathcal{C} = \sigma\left(\Lambda^0, f^0\right)$, $E_{\mathcal{C}}\left(\cdot\right) = E_{\mathcal{C}}\left(\cdot|\mathcal{C}\right)$,

$$
\phi_i^{\dagger}\left(L\right) = \begin{pmatrix} \phi_i^{u\dagger}\left(L\right) \\ \phi_i^{\varepsilon\dagger}\left(L\right)\end{pmatrix} = \begin{pmatrix} \phi_i^{uu}\left(L\right) & \phi_i^{u\varepsilon}\left(L\right) \\ \phi_i^{\varepsilon u}\left(L\right) & \phi_i^{\varepsilon\varepsilon}\left(L\right)\end{pmatrix},\ S^u = \left(1, 0_{1\times p}\right),\ \text{and } S^{\varepsilon} = \left(0_{p\times 1}, I_p\right).
$$

Let

$$
Q_{1NT} = \operatorname{diag}\left(\frac{1}{N_1T^2}\sum_{i\in G_1^0} x_i' M_{f_1^0} x_i, \ldots, \frac{1}{N_KT^2}\sum_{i\in G_K^0} x_i' M_{f_1^0} x_i\right) \text{ and } Q_{2NT} = \begin{pmatrix} Q_{2NT,11} & \cdots & Q_{2NT,1K} \\ \vdots & \ddots & \vdots \\ Q_{2NT,K1} & \cdots & Q_{2NT,KK}\end{pmatrix},
$$

where $Q_{2NT,kl} = \frac{1}{N_kNT^2}\sum_{i\in G_k^0}\sum_{j\in G_l^0} x_i' M_{f_1^0} x_j a_{ij}$ for $k,l=1,...,K$. Let

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

where $Q_{1,k} = \lim_{N\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0} E_{\mathcal{C}}\left(\int \tilde{B}_{2i}\tilde{B}_{2i}'\right)$, $Q_{2,kl} = \lim_{N\to\infty}\frac{1}{NN_k}\sum_{i\in G_k^0}\sum_{j\in G_l^0} a_{ij}E_{\mathcal{C}}\left(\int \tilde{B}_{2,i}\tilde{B}_{2,j}\right)$, 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 $U_{NT} = \left(U_{1NT}',\ldots,U_{KNT}'\right)'$, $B_{NT} = \left(B_{1NT}',\ldots,B_{KNT}'\right)'$, $V_{NT} = \left(V_{1NT}',\ldots,V_{KNT}'\right)'$

and $B_{kNT} = B_{kNT,1} + B_{kNT,2}$. The following theorem reports the Bahadur-type representation and asymptotic distribution of $\text{vec}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0)$.

Theorem 3.4 *Suppose that assumptions 3.1-3.3 hold and $\sqrt{N} = o(T)$. Let $\hat{\alpha}_k$ be obtained by solving (2.8). Then*

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

(ii) $\sqrt{N}T vec(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) - \sqrt{D_N} Q_{NT}^{-1} B_{NT} \xrightarrow{d} MN(0, \lim_{N\to\infty} D_N Q_0^{-1} \Omega_0 Q_0^{-1})$ *as* $(N,T) \to \infty$,

where $D_N = diag\left(\frac{N}{N_1}, ..., \frac{N}{N_K}\right)$, $\Omega_0 = \lim_{(N,T)\to\infty} \Omega_{NT}$, *and* $\Omega_{NT} = Var(V_{NT}|\mathcal{C})$.

Theorem 3.4 indicates that V_{NT} and B_{NT} are associated with the asymptotic variance and bias of $\hat{\alpha}_k$'s, respectively. Note that $B_{kNT} = B_{kNT,1} + B_{kNT,2}$, which indicates the two sources of biases. The appearance of $B_{kNT,1}$ results from the correlation between (x_{it}, f_{1t}) and u_{it} and the serial correlation among the innovation process $\{w_{it}\}$. Apparently, the presence of the unobserved nonstationary factors f_{1t}^0 complicates the formula for $B_{kNT,1}$ through the term $\varkappa_{ts}(= f_{1t}^{0\prime}(f_1^{0\prime} f_1^0)^{-1} f_{1s}^0)$. The second source of asymptotic bias is due to the unobserved stationary factors f_{2t}^0 so that $B_{kNT,2} = 0$ if f_{2t}^0 is absent from the model. 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 bias term for panel cointegrating regression that is associated with the one-sided long-run covariance; see Phillips (1995) and Phillips and Moon (1999). Note that the ith element of V_{NT} is independent across i conditional on $\mathcal{C}$ and $E_{\mathcal{C}}(V_{NT}) = 0$. This makes it possible for us to derive a version of conditional CLT for V_{NT} and establish the limiting distribution of our estimators $\hat{\boldsymbol{\alpha}}$ in 3.4(ii).

As we show in the proof of Theorem 3.4, the asymptotic bias term B_{NT} is of $O(\sqrt{N_k})$, which implies the T-consistency of the C-Lasso estimators $\hat{\alpha}_k$. In order to obtain the $\sqrt{N}T$-convergence rate, we call upon various procedures to remove the asymptotic bias by constructing consistent estimates of B_{NT}.

3.4.1 The fully modified procedure

In this subsection, we first obtain the estimates of unobserved stationary factors f_{2t}^0 from (2.10). Then we employ the fully modified procedure of Phillips and Hansen (1990) and Phillips (1995) to make bias-corrections for endogeneity and serial correlation. Below we consider the three types of bias-corrected estimators: 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 Cup-Lasso estimator $\hat{\alpha}_{\hat{G}_k}^{cup}$.

Following Phillips and Hansen (1990), 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

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

where $\omega(\cdot)$ is a kernel function, H is the bandwidth parameter, and $\hat{\Gamma}_i(j) = \frac{1}{T}\sum_{t=1}^{T-j} \hat{w}_{it+j}\hat{w}_{it}'$ 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.

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

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

The endogeneity correction is achieved by modifying the variable y_{it} with the follow transformation

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

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

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

By Phillips (1995), (3.2) and (3.3) give correction for the endogeneity and serial correlation, respectively.

Therefore, we can obtain the bias-correction post-Lasso estimator $\hat{\alpha}_{\hat{G}}^{bc}$, fully modified post-Lasso estimator $\hat{\alpha}_{\hat{G}_k}^{fm}$, and the updated estimators of $\hat{f}_1$ and $\hat{f}_2$ by iteratively solving (3.4)-(3.7), such that

$$\operatorname{vec}\left(\hat{\alpha}_{\hat{G}}^{bc}\right) = \operatorname{vec}\left(\hat{\alpha}\right) - \frac{1}{\sqrt{N}T}\sqrt{D_N} Q_{NT}^{-1}\left(\hat{B}_{NT,1} + \hat{B}_{NT,2}\right), \tag{3.4}$$

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

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

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

where

$$\hat{B}_{kNT,1} = \frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathbf{1}\{t=s\} - \hat{\varkappa}_{ts}\mathbf{1}\{s\leq t\}\right]\right)\hat{\Delta}_{21,i},$$

$$\hat{B}_{kNT,1}^{+} = \frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathbf{1}\{t=s\} - \hat{\varkappa}_{ts}\mathbf{1}\{s\leq t\}\right]\right)\hat{\Delta}_{21,i}^{+},$$

$$B_{kNT,2} = \frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathbf{1}\{t=s\} - \hat{\varkappa}_{ts}\mathbf{1}\{s\leq t\}\right]\right)\hat{\Delta}_{24,i}\hat{\hat{\lambda}}_{2i},$$

$\hat{\varkappa}_{ts} = \hat{f}_{1t}'(\hat{f}_1'\hat{f}_1)^{-1}\hat{f}_{1s} = \hat{f}_{1t}'\hat{f}_{1s}/T^2$ and $\hat{\hat{\lambda}}_{2i} = \hat{\lambda}_{2i} - \frac{1}{N}\sum_{j=1}^{N}\hat{\lambda}_{2j}\hat{a}_{ij}$. We obtain the fully modified Cup-Lasso estimators $\hat{\alpha}^{cup}_{\hat{G}_k}$ by iteratively solving (2.8), and (3.5)-(3.7), where we update the group classification results in each iteration.

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

Theorem 3.5 *Suppose that assumptions 3.1-3.4 hold. Let $\hat{\boldsymbol{\alpha}}^{bc}_{\hat{G}}$ be obtained by iteratively solving (3.4), (3.6)-(3.7), $\hat{\boldsymbol{\alpha}}^{fm}_{\hat{G}}$ be obtained by iteratively solving (3.5)-(3.7) and $\hat{\boldsymbol{\alpha}}^{cup}_{\hat{G}}$ be obtained by iteratively solving (2.8) and (3.5)-(3.7). As $(N,T) \to \infty$ with $\sqrt{N} = o(T)$, we have*

(i) $\sqrt{N}T vec(\hat{\boldsymbol{\alpha}}^{bc}_{\hat{G}} - \boldsymbol{\alpha}^0) \xrightarrow{d} MN(0, \lim_{N\to\infty} D_N Q_0^{-1} \Omega_0 Q_0^{-1})$,

(ii) $\sqrt{N}T vec(\hat{\boldsymbol{\alpha}}^{fm}_{\hat{G}} - \boldsymbol{\alpha}^0) \xrightarrow{d} MN(0, \lim_{N\to\infty} D_N Q_0^{-1} \Omega_0^+ Q_0^{-1})$,

(iii) $\sqrt{N}T vec(\hat{\boldsymbol{\alpha}}^{cup}_{\hat{G}} - \boldsymbol{\alpha}^0) \xrightarrow{d} MN(0, \lim_{N\to\infty} D_N Q_0^{-1} \Omega_0^+ Q_0^{-1})$,

where $\Omega_0^+ = \lim_{N,T\to\infty} \Omega_{NT}^+$ *and* $\Omega_{NT}^+ = Var\big(V_{NT}^+ | \mathcal{C}\big)$.

All three types of estimators achieve $\sqrt{N}T$ consistency and have a mixture normal limit distribution. One can construct the asymptotic t-tests and Wald-tests as usual provided one can obtain consistent estimates of Q_0 and Ω_0^+. The procedure is standard given the estimated group structure.

3.5 Estimating the number of unobserved factors

In the previous subsections, we assume that the numbers of nonstationary and stationary factors, r_1 and r_2, are known. In this subsection, we propose 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. Let r denote a generic total number of nonstationary and stationary factors. We now use r_1^0 and r^0 to denote their true values, which are assumed to be 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 the first differencing, (2.3) 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 now write the factors and factor loadings respectively as f_t^r and λ_i^r, where the superscipt r highs the dimension of the underlying factors or factor loadings. Let $G^r \equiv \Delta f^r$ be a matrix of $(T-1) \times r$ unobserved differenced factors with a typical row given by $(G_t^r)' \equiv (\Delta f_t^r)'$. We consider the 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 $\hat{G}^r = (\hat{G}_2^{r\prime}, ..., \hat{G}_T^{r\prime})'$, $\hat{\Lambda}^r = (\hat{\lambda}_1^{r\prime}, ..., \hat{\lambda}_N^{r\prime})'$, and $\hat{\beta}_i$'s are obtained from the model with $r_1 = r_{\max}$ nonstationary factors. It is easy to show that $\hat{\beta}_i$'s are T-consistent, which suffices for our purpose. It is well known that given $\hat{G}^r$, we can solve $\hat{\Lambda}^r$ from the least squares regression as a function of $\hat{G}^r$. But we will suppress the dependence of $\hat{\Lambda}^r$ on $\hat{G}^r$ and 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 following 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 $\hat{r} = \arg\min_{0 \leq r \leq r_{\max}} IC_1(r)$. We add the following 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 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 consistently estimate r^0.

Theorem 3.6 *Suppose that 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 indicates that we can determine the total number of factors r^0 consistently by minimizing $IC_1(r)$.

As we have discussed in Section 3.4, ignoring the unobserved stationary factors will not affect the consistency of slope coefficient estimator, but 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 the consistent estimate of r_1^0, we can also obtain the consistent estimate of the number of unobserved stationary factors, r_2^0, based on Theorem **??**.

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 $\hat{\beta}_i$'s, we consider the following minimization problem

$$\left\{\hat{f}_1^{r_1}, \hat{\Lambda}^{r_1}\right\} = \arg\min_{\Lambda^{r_1}, f_t^{r_1}} \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T (y_{it} - \hat{\beta}_i' x_{it} - \lambda_i^{r_1\prime} f_{1t}^{r_1})^2,$$
$$\text{s.t. } f_1^{r_1\prime} f_1^{r_1}/T^2 = I_{r_1} \text{ and } \Lambda^{r_1\prime}\Lambda^{r_1} \text{is diagonal.}$$

Given $\hat{f}_1^{r_1} = (\hat{f}_{11}^{r_1\prime}, ..., \hat{f}_{1T}^{r_1\prime})'$, we can solve $\hat{\Lambda}^{r_1} = (\hat{\lambda}_{11}^{r_1\prime}, ..., \hat{\lambda}_{1N}^{r_1\prime})'$ as a function of $\hat{f}_1^{r_1}$ through the least squares regression. But we suppress the dependence of $\hat{\Lambda}^{r_1}$ on $\hat{f}_1^{r_1}$ and define

$$V_2(r_1, \hat{f}_1^{r_1}) = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T (y_{it} - \hat{\beta}_i' x_{it} - \hat{\lambda}_i^{r_1\prime} \hat{f}_{1t}^{r_1})^2.$$

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 $g_2(N,T)$ are quite different 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 as in Assumption 3.5. The intuition 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. In fact, 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 $0 \le r_1 < r_1^0$. On the other hand, 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 suggests that the use of $IC_2(r_1)$ helps to determine r_1^0 consistently.

Theorem 3.7 *Suppose that 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 below, 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 $\alpha_T = \frac{T}{4\log\log(T)}$. We first estimate the total number of unobserved factors by $\hat{r}$ based on the first-differenced model, and then estimate the number of unobserved nonstationary factors by $\hat{r}_1$ based on the level model. The 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

In this subsection, we propose a BIC-type information criterion to determine the number of groups, K. We assume that the true number of group, K_0, is bounded from above by a finite integer $K_{\max}$. We now consider the PPC criterion function

$$Q_{NT,\lambda}^K(\boldsymbol{\beta}, \boldsymbol{\alpha}, f_1) = Q_{NT}(\boldsymbol{\beta}, f_1) + \frac{\lambda}{N}\sum_{i=1}^N \prod_{k=1}^K \|\beta_i - \alpha_k\|,$$

where $1 \le K \le K_{\max}$. By minimizing the above criterion function, we obtain the 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, where we make by the $\hat{\beta}_i, \hat{\alpha}_k, \hat{\lambda}_{1i}$ 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 $\hat{K}(\lambda) = \arg\min_{1 \le K \le K_{\max}} GIC(K,\lambda)$.

Let $\mathcal{G}^{(K)} = (G_{K,1}, ..., G_{K,K})$ be any K-partition of the set of individual index $\{1, 2, ..., N\}$. Define $\hat{\sigma}^2_{\mathcal{G}^{(K)}} = \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_{K,k}}\sum_{t=1}^{T}[y_{it} - \hat{\alpha}^{cup\prime}_{\hat{G}_{K,k}} x_{it} - \hat{\lambda}_i(K,\lambda)'\hat{f}_t(K,\lambda)]^2$, where $\hat{\alpha}^{cup}_{\hat{G}_{K,k}}$ is analogously defined as $\hat{\alpha}^{cup}_{\hat{G}_k(K,\lambda)}$ with $\hat{G}_k(K,\lambda)$ being replaced by $G_{K,k}$. Let σ_0^2 =plim$_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{i\in G_k^0}\sum_{t=1}^{T}[y_{it} - \alpha_k^{0\prime}x_{it} - \lambda_i^{0\prime}f_t^0]^2$. Define

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

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

We add the following assumption.

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

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

Assumption 3.6(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 3.6(ii) imposes the typical conditions on the penalty function $g_3(N,T)$: it cannot shrink to zero either 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 3.6(ii), we have $\inf_{1\leq K\leq K_{\max},K\neq K_0} IC_3(K,\lambda) > IC_3(K_0,\lambda)$ as $(N,T)\to\infty$. Consequently, the minimizer of $IC_3(K,\lambda)$ with respect to K is equal to K_0 w.p.a.1 for a variety choices of λ.

4 Monte Carlo Simulations

In this section, we evaluate the finite sample performance of the C-Lasso, bias-corrected post-Lasso, fully-modified post-Lasso and Cup-Lasso estimators and that of the information criteria for determining the number of groups and the number of common factors.

4.1 Data generating processes

We consider four data generating processes (DGPs) that cover the cases of both stationary and nonstationary unobserved common factors. Throughout these DGPs, the observations in each DGP are drawn from three groups with $N_1 : N_2 : N_3 = 0.3 : 0.4 : 0.3$. There are four combinations of the sample sizes with $N = 50, 100$ and $T = 40, 80$.

DGP1 (Strictly exogenous nonstationary regressors and unobserved stationary common factors) The observations (y_{it}, x_{it}') are generated from the following model,

$$\begin{cases} y_{it} = \beta_i' x_{it} + c(\lambda_{2i}' f_{2t}) + u_{it} \\ x_{it} = x_{it-1} + \varepsilon_{it} \end{cases} \tag{4.1}$$

where $x_{it} = (x_{1it}, x_{2it})'$ is a 2×1 vector of nonstationary regressors, f_{2t} is a 2×1 vector of stationary common factors. Let $w_{it} = (u_{it}, \varepsilon_{it}', f_{2t}')' \sim$ i.i.d. $N(0, I_5)$. The factor loadings λ_{2i} are i.i.d. $N((1,1)', I_2)$ for $i = 1, ..., N$. We use c to control the importance of unobserved common factors and let $c = 0.5$. The long-run slope coefficients β_i exhibits the group-structure in (2.2) for $K = 3$ and the true values 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 (Weakly dependent nonstationary regressors and unobserved nonstationary common factors) The observations $(y_{it}, x_{it}', f_{1t}')$ are generated from the following model,

$$\begin{cases} y_{it} = \beta_i^{0\prime} x_{it} + c(\lambda_{1i}' f_{1,t}) + u_{it} \\ x_{it} = x_{it-1} + \varepsilon_{it} \\ f_{1t} = f_{1t-1} + \nu_t \end{cases} \tag{4.2}$$

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. The idiosyncratic errors $w_{it} = (u_{it}, \varepsilon_{it}', \Delta f_{1t}')'$ are generated from a linear process: $w_{it} = \sum_{j=0}^{\infty} \psi_{ij} v_{i,t-j}$, where v_{it} are i.i.d. $N(0, I_5), \psi_{ij} = j^{-3.5} * \Omega^{1/2}$, $\Omega^{1/2}$ is the symmetric square root of Ω, where $\Omega_{lm} = 0.2$ for $l \neq m$, $\Omega_{ll} = 1$ for $l = 2, 3, 4, 5$ and $\Omega_{11} = 0.25$. Let $c = 1$. The factor loadings of nonstationary common factors are i.i.d. $\lambda_{1i} \sim N((1,1)', I_2)$. The true coefficients of β_i are the same in DGP1.

DGP3 (Weakly dependent nonstationary regressors and mixed unobserved stationary and nonstationary common factors) The observations $(y_{it}, x_{it}', f_{1t}')$ are generated from the following model,

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

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 factors. The idiosyncratic errors $w_{it} = (u_{it}, \varepsilon_{it}', \Delta f_{1t}', f_{2t}')'$ are generated from a linear process: $w_{it} = \sum_{j=0}^{\infty} \psi_{ij} v_{i,t-j}$ where v_{it} are i.i.d. $N(0, I_6), \psi_{ij} = j^{-3.5} * \Omega^{1/2}$, $\Omega^{1/2}$ is the symmetric square root of Ω where $\Omega_{lm} = 0.2$ for $l \neq m$, $\Omega_{11} = 0.25$, and $\Omega_{ll} = 1$ for $l = 2, ..., 6$. Let $c_1 = 1$ and $c_2 = 0.5$. The factor loadings $\lambda_i = (\lambda_{1i}', \lambda_{2i}')'$ are i.i.d. $\lambda_{1i} \sim N((1,1,1)', I_3)$. The true coefficients of β_i are the same in DGP1.

DGP4 (Weakly dependent nonstationary regressors and mixed unobserved stationary and nonstationary

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	1.000	0	0	0	1.000	0	0	0	0
	50	80	0	1.000	0	0	0	1.000	0	0	0	0
	100	40	0	1.000	0	0	0	0.998	0.002	0	0	0
	100	80	0	1.000	0	0	0	1.000	0	0	0	0
DGP2	50	40	0	1.000	0	0	0	0	0.004	0.964	0.032	0
	50	80	0	1.000	0	0	0	0.004	0.016	0.976	0.004	0
	100	40	0	1.000	0	0	0	0	0.002	0.958	0.040	0
	100	80	0	1.000	0	0	0	0	0.002	0.976	0.022	0
DGP3	50	40	0	0	1.000	0	0	0.018	0.088	0.894	0	0
	50	80	0	0	1.000	0	0	0.006	0.026	0.968	0	0
	100	40	0	0	1.000	0	0	0	0.008	0.972	0.020	0
	100	80	0	0	1.000	0	0	0	0.012	0.988	0	0
DGP4	50	40	0	0	0.998	0.002	0	0.002	0.060	0.938	0	0
	50	80	0	0	1.000	0	0	0.004	0.016	0.980	0	0
	100	40	0	0	1.000	0	0	0	0.012	0.988	0	0
	100	80	0	0	1.000	0	0	0	0.008	0.992	0	0

common factors) The settings of DGP4 is the same with those of DGP3, except for allowing weakly correlation among factor loadings $\lambda_i \sim i.i.d.\ N((1,1,1)', I_3 * \Omega_2)$, where $\Omega_{2,lm} = 2/\sqrt{N}$ for $l \neq m$.

In all cases, the number of replications is 500.

4.2 Estimate number of unobserved factors

In this subsection, we assess the performance of two information criteria proposed in Section 3.5 before determining the number of group and PPC-based estimation procedure. We choice the BIC-type penalty function $g_1(N,T) = \frac{N+T}{NT}\log(\min(N,T))$ to determine the total number of unobserved factors and $g_2(N,T) = \frac{T}{4\log(\log(T))} \times g_1(N,T)$ to determine the number of unobserved nonstationary factors. Based on 500 replications for each DGP, Table 1 displays the probability that a particular factor size from 0 to 5 is selected according to the information criteria proposed for both differenced data and level data. In differenced data, when $T = 40$, the probabilities are more than 99% in all cases and tend to unit when $T = 80$ for selecting the total number of unobserved factors. The information criterion for level data performs as good as that in difference data when $T = 80$. When T=40, the probabilities are at least 90% in all cases. The simulation results show that our two information criteria in both differenced data and level one works fairly well.

4.3 Determine the number of groups

The results from previous subsection show that the information criteria are useful even though we have no information of latent group structures. This section focuses on the performance of the information criterion for determining the number of groups, where we assume that the number of unobserved factors

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

	N	T	1	2	3	4	5	6
DGP1	50	40	0	0	0.9860	0.0140	0	0
	50	80	0	0	0.9940	0.0060	0	0
	100	40	0	0	0.9700	0.0280	0	0.0020
	100	80	0	0	1.0000	0	0	0
DGP2	50	40	0	0	1.0000	0	0	0
	50	80	0	0	1.0000	0	0	0
	100	40	0	0	1.0000	0	0	0
	100	80	0	0	1.0000	0	0	0
DGP3	50	40	0	0	0.9760	0.0180	0.0060	0
	50	80	0	0	0.9980	0.0020	0	0
	100	40	0	0	0.9740	0.0240	0.0020	0
	100	80	0	0	1.0000	0	0	0
DGP4	50	40	0	0	0.9920	0.0060	0.0020	0
	50	80	0	0	1.0000	0	0	0
	100	40	0	0	0.9900	0.0100	0	0
	100	80	0	0	1.0000	0	0	0

are known. Here the penalty function $\rho(N,T) = \frac{1}{3} \times \log(\min(N,T))/\min(N,T)$, which satisfies the two restrictions proposed in Theorem 3.9. Due to space limitations, we report outcomes under the tuning parameter $\lambda = c_\lambda \times T^{-3/4}$, where $c_\lambda = 0.1$. Based on 500 replications for each DGP, Table 2 displays the probability that a particular group size from 1 to 6 is selected according to the information criterion. The true number of group is 3. When $N = 50$ the probabilities are more than 99% in all cases and tend to unit when T=80.

4.4 Classification and point estimation

In this subsection, we test the performance of classification and estimation when we have prior knowledge of the number of groups and that of unobserved factors. Table 3 and Table 4 report classification and point estimation results from 500 replications for each DGP. As shown in Table 3 and Table 4, we set the tuning parameter in the objective function (2.8) $\lambda = c_\lambda \times T^{-3/4}$ and choose a sequence of increasing constants of $c_\lambda = (0.025, 0.05, 0.1, 0.2)$[1] to test the sensitivity of classification and estimation performance. Here we only report the performance results for the first coefficient $\alpha_1 = \{\alpha_{1,k}\}_{k=1}^{K_0}$ in each model. In general, the outcomes are found robust over specified range of constants. Column 4 and 7 report the percentage of correct classification of the N units, calculated as $\frac{1}{N}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k} 1\{\beta_i^0 = \alpha_k^0\}$, averaged over the 500 replications. Column 5-6 and 8-9 summarize the estimation performance, such as root-mean-squared error (hereafter, RMSE), and bias. For simplicity we define weighted average RMSE and bias, as $\frac{1}{N}\sum_{k=1}^{K_0} N_k \mathrm{RMSE}(\hat{\alpha}_{1,k})$ with $\hat{\alpha}_{1,k}$ the same as bias. The estimate of the long-run covariance matrix is based on Fejer kernel with bandwidth set at 10. Results of other kernels (quadratic spectral kernel and Parzen kernel) are not reported, there are no essential differences for most cases. For comparison purpose,

[1]Due to space limitation, we only report the results when $c_\lambda = (0.1, 0.2)$. The rest results are available upon request.

we report the results of corresponding statistics of the C-Lasso, bias-corrected post-Lasso, fully-modified post-Lasso, Cup-Lasso, and oracle estimators. The oracle estimator utilizes the exact group identity G_k^0, which is infeasible in practice.

For classification results, the correct classification percentage approaches 100% when T increases. The results with different c_λ's are quite similar, indicating the robustness of our algorithm to the choice of tuning parameter. In particular, we iteratively minimize the PPC objective function to obtain the Cup-Lasso estimators. The correct classification percentage is higher than that of C-Lasso and post-Lasso estimators in all cases. For estimation performance, the RMSE, bias, and coverage of post-Lasso and Cup-Lasso estimators approach that of oracle ones in DGP1. Since we only introduce stationary factors and strictly exogenous nonstationary regressors, there is no asymptotic bias coming from the endogeneity and serial correlation. The RMSE and coverage of C-Lasso estimators are poor due to ignoring the unobserved stationary factors in PPC-based estimation procedure. In DGP2 and DGP3, the performance of C-Lasso estimator is poorer due to the additional sources of non-negligible bias from the endogeneity and serial correlation. And we show that the fully modified procedure work better compared to direct bias-correction procedure. The performance of Cup-Lasso estimators is better than that of post-Lasso ones due to updated group classification results. In general, the finite sample performance of the Cup-Lasso estimators is close to that of the oracle ones, which empirically confirms oracle efficiency of the Cup-Lasso estimators. In practice, we recommend Cup-Lasso estimators for estimation and inference.

5 Empirical Application: Growth Convergence Puzzle

Many researchers have explored the behavior of economic growth across multiple countries. The main question in this literature is whether economies exhibit convergence. Here we study the heterogeneous behavior of convergence through the channel of technology changes. The benchmark model is the international R&D spillovers model, proposed by Coe and Helpman (1995), where they regress the total factor productivity (TFP) on domestic R&D stock and foreign R&D stock. Their work suggests that the international R&D spillovers are some forces toward convergence through the channel of technology changes. There are two potential problems in their work. First, even though it is commonly accepted that there are multiple steady states for convergence across economies in theoretical growth model, we cannot empirically identify the heterogeneous behavior of convergence. In addition, they haven't account for the unobserved common patterns across countries. Since our PPC-based estimation method simultaneously accounts for the unobserved heterogeneity and cross-sectional dependence, it gives us a purely data-driven approach to study the heterogeneous behavior in economies' convergence. Furthermore, technology change is the main source of economies' growth. We specify the channel of convergence through technology changes by reestimating CH2009 dataset. Comparing to CH model, we allow for heterogeneous parameters and consider the unobserved common patterns across countries. Specifically, we impose the latent group structures on the long-run relations between technology changes, domestic R&D stock, and foreign R&D stock. These heterogeneous long-run relations explain the puzzle of economies convergence–some countries may fail to convergence in a long run.

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

		c_λ	0.1			0.2		
N	T		% Correct specification	RMSE	Bias	% Correct specification	RMSE	Bias
DGP1								
50	40	C-Lasso	99.68	0.0137	0.0049	99.70	0.0130	0.0047
50	40	post^{bc}-Lasso	99.68	0.0130	0.0003	99.70	0.0129	0.0002
50	40	post^{fm}-Lasso	99.68	0.0129	0.0004	99.70	0.0128	0.0003
50	40	Cup-Lasso	99.68	0.0126	-0.0002	99.70	0.0126	-0.0002
50	40	Oracle	-	0.0126	-0.0002	-	0.0126	-0.0002
50	80	C-Lasso	100	0.0081	0.0031	100	0.0077	0.0028
50	80	post^{bc}-Lasso	100	0.0070	0.0003	100	0.0070	0.0003
50	80	post^{fm}-Lasso	100	0.0069	0.0004	100	0.0069	0.0004
50	80	Cup-Lasso	100	0.0069	0.0004	100	0.0069	0.0004
50	80	Oracle	-	0.0069	0.0001	-	0.0069	0.0001
100	40	C-Lasso	99.69	0.0109	0.0054	99.73	0.0101	0.0046
100	40	post^{bc}-Lasso	99.69	0.0091	0.0007	99.73	0.0087	0.0004
100	40	post^{fm}-Lasso	99.69	0.0090	0.0007	99.73	0.0086	0.0004
100	40	Cup-Lasso	99.69	0.0090	0.0007	99.73	0.0086	0.0004
100	40	Oracle	-	0.0087	-0.0001	-	0.0087	-0.0001
100	80	C-Lasso	100	0.0062	0.0032	99.99	0.0058	0.0029
100	80	post^{bc}-Lasso	100	0.0046	0.0005	99.99	0.0046	0.0005
100	80	post^{fm}-Lasso	100	0.0046	0.0005	99.99	0.0046	0.0005
100	80	Cup-Lasso	100	0.0046	0.0005	99.99	0.0046	0.0005
100	80	Oracle	-	0.0046	0.0004	-	0.0046	0.0004
DGP2								
50	40	C-Lasso	97.68	0.0654	0.0146	97.53	0.0743	0.0146
50	40	post^{bc}-Lasso	97.68	0.0405	0.0048	97.53	0.0430	0.0048
50	40	post^{fm}-Lasso	97.68	0.0405	0.0042	97.53	0.0430	0.0041
50	40	Cup-Lasso	100	0.0094	0.0004	100	0.0094	0.0004
50	40	Oracle	-	0.0094	0.0004	-	0.0094	0.0004
50	80	C-Lasso	99.21	0.0233	0.0047	99.19	0.0254	0.0047
50	80	post^{bc}-Lasso	99.21	0.0195	-0.0004	99.19	0.0195	-0.0007
50	80	post^{fm}-Lasso	99.21	0.0194	-0.0005	99.19	0.0194	-0.0009
50	80	Cup-Lasso	100	0.0047	-0.0001	100	0.0047	-0.0001
50	80	Oracle	-	0.0047	-0.0001	-	0.0047	-0.0001
100	40	C-Lasso	97.45	0.0500	0.0135	97.37	0.0543	0.0119
100	40	post^{bc}-Lasso	97.45	0.0601	-0.0011	97.37	0.0584	-0.0010
100	40	post^{fm}-Lasso	97.45	0.0601	-0.0016	97.37	0.0585	-0.0015
100	40	Cup-Lasso	100	0.0069	-0.0016	100	0.0069	-0.0016
100	40	Oracle	-	0.0069	-0.0016	-	0.0069	-0.0016
100	80	C-Lasso	99.25	0.0181	0.0061	99.23	0.0194	0.0057
100	80	post^{bc}-Lasso	99.25	0.0172	0.0012	99.23	0.0170	0.0010
100	80	post^{fm}-Lasso	99.25	0.0171	0.0010	99.23	0.0170	0.0010
100	80	Cup-Lasso	100	0.0032	-0.0001	100	0.0032	-0.0001
100	80	Oracle	-	0.0032	-0.0001	-	0.0032	-0.0001

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

		c_λ	0.1			0.2		
N	T		% Correct specification	RMSE	Bias	% Correct specification	RMSE	Bias
DGP3								
50	40	C-Lasso	96.97	0.0563	0.0118	96.87	0.0632	0.0101
50	40	post^{bc}-Lasso	96.97	0.0522	0.0029	96.87	0.0516	0.0022
50	40	post^{fm}-Lasso	96.97	0.0524	0.0023	96.87	0.0519	0.0016
50	40	Cup-Lasso	99.85	0.0145	0.0015	99.81	0.0146	0.0015
50	40	Oracle	-	0.0150	0.0014	-	0.0150	0.0014
50	80	C-Lasso	99.15	0.0297	0.0056	99.11	0.0327	0.0047
50	80	post^{bc}-Lasso	99.15	0.0275	0.0015	99.11	0.0265	0.0013
50	80	post^{fm}-Lasso	99.15	0.0274	0.0015	99.11	0.0265	0.0013
50	80	Cup-Lasso	100	0.0073	0.0010	100	0.0073	0.0010
50	80	Oracle	-	0.0073	0.0006	-	0.0073	0.0006
100	40	C-Lasso	98.65	0.0299	0.0119	98.43	0.0300	0.0110
100	40	post^{bc}-Lasso	98.65	0.0214	0.0028	98.43	0.0222	0.0035
100	40	post^{fm}-Lasso	98.65	0.0213	0.0023	98.43	0.0222	0.0031
100	40	Cup-Lasso	99.93	0.0108	0.0020	99.83	0.0110	0.0021
100	40	Oracle	-	0.0109	0.0018	-	0.0109	0.0018
100	80	C-Lasso	99.05	0.0194	0.0060	99.01	0.0208	0.0053
100	80	post^{bc}-Lasso	99.05	0.0181	0.0007	99.01	0.0183	0.0007
100	80	post^{fm}-Lasso	99.05	0.0180	0.0006	99.01	0.0182	0.0005
100	80	Cup-Lasso	100	0.0054	-0.0002	100	0.0054	-0.0002
100	80	Oracle	-	0.0054	-0.0003	-	0.0054	-0.0003
DGP4								
50	40	C-Lasso	96.92	0.0566	0.0110	96.77	0.0634	0.0099
50	40	post^{bc}-Lasso	96.92	0.0508	0.0018	96.77	0.0498	0.0008
50	40	post^{fm}-Lasso	96.92	0.0511	0.0013	96.77	0.0501	0.0008
50	40	Cup-Lasso	99.91	0.0130	0.0014	99.87	0.0130	0.0015
50	40	Oracle	-	0.0134	0.0014	-	0.0134	0.0014
50	80	C-Lasso	98.99	0.0299	0.0055	98.93	0.0331	0.0045
50	80	post^{bc}-Lasso	98.99	0.0277	0.0009	98.93	0.0263	0.0013
50	80	post^{fm}-Lasso	98.99	0.0277	0.0008	98.93	0.0263	0.0013
50	80	Cup-Lasso	100	0.0066	0.0010	100	0.0066	0.0010
50	80	Oracle	-	0.0065	0.0007	-	0.0065	0.0007
100	40	C-Lasso	98.77	0.0291	0.0123	98.53	0.0295	0.0113
100	40	post^{bc}-Lasso	98.77	0.0205	0.0032	98.53	0.0217	0.0037
100	40	post^{fm}-Lasso	98.77	0.0204	0.0027	98.53	0.0216	0.0032
100	40	Cup-Lasso	99.94	0.0102	0.0020	99.87	0.0103	0.0021
100	40	Oracle	-	0.0103	0.0017	-	0.0103	0.0017
100	80	C-Lasso	99.04	0.0197	0.0059	99.02	0.0211	0.0053
100	80	post^{bc}-Lasso	99.04	0.0181	0.0009	99.02	0.0183	0.0008
100	80	post^{fm}-Lasso	99.04	0.0180	0.0007	99.02	0.0183	0.0007
100	80	Cup-Lasso	100	0.0050	-0.0002	100	0.0050	-0.0002
100	80	Oracle	-	0.0050	-0.0002	-	0.0050	-0.0002

The innovation of Coe and Helpman's model is to explain TFP not only by domestic R&D stock but also foreign R&D stock from trading partners. In growth literature, TFP is the Solow residual, and often regarded as a measure of technology changes, defined by

$$\log(TFP) = \log(Y) - \theta \log(K) - (1-\theta)\log(L), \tag{5.1}$$

where Y is the final output, L is the labor force, K is the capital stock, and θ is the share of capital in GDP. It is well accepted that domestic R&D investment is one of the main sources of the TFP by innovation and improving the qualities of goods. Coe and Helpman (1995) argue that international trade in intermediate goods enables a country to gain access to all inputs available in the rest of the world. In this aspect, the foreign R&D stocks from a country's trading partners also affect this country's TFP. They establish estimation equation of the TFP as follow,

$$\log(F_{it}) = \mu_i^0 + \beta_i^d \log(s_{it}^d) + \beta_i^f \log(s_{it}^f) + u_{it},$$

where i is the country index, F is the total factor productivity, s^d is the real domestic R&D capital stock, s^f is the real foreign R&D capital stock. We follow their specification on the international R&D spillovers model and introduce the unobserved common patterns, such that

$$\log(F_{it}) = \beta_i^d \log(s_{it}^d) + \beta_i^f \log(s_{it}^f) + \lambda_i' f_t + u_{it}, \tag{Eq1}$$

where e_{it} is cross-sectionally dependent with unobserved common patterns. Here we consider (β_i^d, β_i^f) as the long-run cointegrating relations with latent group structures. The unobserved common patterns are modeled by the multi-factor structure as $e_{it} = \lambda_i' f_t + u_{it}$ and the fixed effects μ_i^0 are captured by the factor structure.

In addition, we consider logarithm of human capital (H) as an additional explanatory variable, see (Eq2)

$$\log(F_{it}) = \beta_i^d \log(s_{it}^d) + \beta_i^f \log(s_{it}^f) + \beta_i^h \log(h) + \lambda_i' f_t + u_{it}. \tag{Eq2}$$

The human capital accounts for innovation outside the R&D sector and other aspects of human capital not captured by formal R&D. Engelbrecht (1997) suggests that human capital is found to affect TFP directly as a factor of production, and as a channel for international technology diffusion associated with catch-up effects across countries.

We obtain CH2009 datasets from 1971-2004 for 24 OECD countries. The bilateral import weighted R&D S^{f-biw} from trading partners is a measure of foreign R&D stock. Human capital is measured by year of schooling. See Coe and Helpman's appendix for detailed definition and construction of these variables.

5.1 Estimation Results

Before the PPC-based estimation procedure, we first employ information criteria in Section 3.5 to estimate the number of unobserved factors. We set penalty function as $g_1(N,T) = \frac{N+T}{NT}\log(\min(N,T))$ and $g_2(N,T) = \frac{T}{2\log\log T} \times g_1(N,T)$. The results for both differenced and level data indicates one unob-

Table 5: The information criterion for K_0 (Eq1 & Eq2)

		Eq1			
K/c_λ	0.05	0.1	0.2	0.4	0.6
$K=1$	-4.6315	-4.6584	-4.6812	-4.6834	-4.6794
$K=2$	-4.8073	-4.8760	-4.7356	-4.8319	-4.8332
$K=3$	**-5.0084**	**-5.0942**	**-5.2130**	**-5.2221**	**-5.0992**
$K=4$	-4.8985	-4.9708	-5.0092	-4.6353	-4.9279
$K=5$	-4.8598	-4.8240	-4.4272	-4.9821	-4.8042
$K=6$	-4.4159	-4.2700	-3.6774	-4.8858	-4.6118
		Eq2			
K/c_λ	0.05	0.1	0.2	0.4	0.6
$K=1$	**-4.6011**	-4.6311	-4.6845	-4.6876	-4.6889
$K=2$	-4.5674	**-4.8101**	**-4.8693**	**-4.8138**	**-4.8127**
$K=3$	-3?9180	-4.2002	-4.7259	-4.7467	-4.7045
$K=4$	-2.8630	-3.5698	-4.0314	-4.2412	-4.2497
$K=5$	-2.2351	-4.0434	-1.9373	-3.5935	-4.0737
$K=6$	-2.7073	-3.6627	-3.1292	-3.7489	-2.6413

served nonstationary common factors. We fix $r = r_1 = 1$ in the determination of number groups and the PPC-based estimation procedure.

We set $\rho(N,T) = \frac{2}{3} \times \log(\min(N,T))/\min(N,T)$ and tuning variable $\lambda = c_\lambda \times T^{-3/4}$ where $c_\lambda = (0.05, 0.1, 0.2, 0.4, 0.6)$. Table 5 reports the information criterion as a function of the number of groups under these tuning parameters. The information criterion suggests three groups for (Eq1) and two groups for (Eq2). In our estimation, we first set the number of groups and then specify $c_\lambda = 0.2$, where the information criterion achieves the minimal values.

Table 6 reports the main results of pooled FMOLS and Cup-Lasso estimates with one unobserved nonstationary common factors, where we compare our results to CH2009. In (Eq1), we have two explanatory variables $(\log(S^d), \log(S^{f-biw}))$. First, we compare the result of CH2009 with the pooled FMOLS after controlling cross-sectional dependence. The coefficients of $\log(S^d)$ in CH2009 is qualitatively similar to our pooled FMOLS. The only difference is the slope coefficient of foreign R&D stock, which decrease more than half after considering one unobserved nonstationary common factors. The nonstationary common factor stands for the unobserved global trends in technology changes. It is reasonable that the direct spillovers effects of foreign R&D stock decrease when the unobserved global technology patterns are taken into consideration. Second, we identify quite difference behavior in the group-specific Cup-Lasso estimates. The estimates of group 1 have the largest estimates on the domestic R&D stock and negative one on foreign R&D. For group 2 and group 3, they both have positive estimates on domestic R&D stock and foreign one. In particular, both estimates in group 2 are larger than that of group 3. We summarize the estimation results into three aspects. On the one hand, those results indicate that technology changes of countries in group 1 rely mainly on domestic R&D stock, which stands for the innovation. In addition, the long-run relation between TFP and foreign R&D stock are negative, which suggests that they cannot benefit from international R&D spillovers. Furthermore, it implies that countries in group 1 don't favor convergence through the channel of technology changes. We call it as "Divergence" group. On the other hand, tech-

Table 6: PPC estimation results for (Eq1) and (Eq2)

Eq1					
Slope coefficients	Pooled CH2009	Pooled FM-OLS	Group 1 Cup-Lasso	Group 2 Cup-Lasso	Group 3 Cup-Lasso
$\log(S^d)$	0.095***	0.090***	0.302***	0.102***	0.049***
	(0.0053)	(0.0134)	(0.0300)	(0.0251)	(0.0147)
$log(S^{f-biw})$	0.213***	0.092***	-0.143***	0.161***	0.125***
	(0.0136)	(0.0222)	(0.0336)	(0.0501)	(0.0281)
Eq2					
Slope coefficients	Pooled CH2009	Pooled FM-OLS	Group 1 Post-Lasso	Group 2 Post-Lasso	
$\log(S^d)$	0.098***	0.049***	0.071***	-0.098***	
	(0.0160)	(0.0163)	(0.0174)	(0.0270)	
$log(S^{f-biw})$	0.035***	0.132***	0.063*	0.323***	
	(0.0111)	(0.0316)	(0.0332)	(0.0398)	
$\log(h)$	0.725***	0.644***	0.638***	0.680***	
	(0.0870)	(0.1204)	(0.1302)	(0.1791)	

Note: *** 1% significant; ** 5% significant; * 10% significant.

nology changes of countries in group 2 have balanced sources–innovation effects from domestic R&D stock and catch-up effects from the international R&D spillovers. And the magnitudes of those estimates are similar. In this respect, it favors convergence hypothesis for countries in group 2. Here we refer it as "Balance" group. Then the technology changes in group 3 are mainly determined by foreign R&D stock. They are classified as "Convergence" group.

In (Eq2), we introduce an additional regressor–human capital, which is regarded as a direct sources of technology changes. Our results confirm that human capital is the one of the main sources of productivity growth. In general, similar heterogeneous behavior preserves in (Eq2). First, we can still classify those countries into two groups and define them as "Balance "and "Convergence ". For group 1, the innovation effects and catch-up effects have similar magnitude. For group 2, referred as "Convergence ", where they have significant positive estimates on foreign R&D stocks.

The PPC-based estimation procedure simultaneously determine the group identities and estimate parameters. Table 7 reports the group classification results. We have discussed that the estimation results of countries in group 1 indicate a potential divergence behavior of economies. There are basically two types of countries in "Divergence" group–"Bellwether" and "Loser". The productivity growth much relies on their own innovation, countries like France, Germany, United States are bellwether in global, which own 61.1% proportion of global R&D stock. On the contrary, the rest countries in group 1 only accounts for 1.5% proportion of global R&D stock. Since most OECD countries are classified into group 2 and group 3. it confirms the recent work of Keller (2004) that the major sources of technical changes leading to productivity growth in OECD countries are not domestic, instead, they lie aboard through the channel of international technology diffusion. Furthermore, countries like Israel, South Korea and United Kingdom are classified in "Balance" group. Productivity growth of countries in group 2 relies on both innovation and catch-up effects through the channel of international technology diffusion.

Table 7: Group classification results of Eq1 and Eq2

Eq1				
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
Eq2				
Group 1 "Balance " ($N_1 = 18$)				
Austria	Belgium	Finland	France	Germany
Iceland	Ireland	Israel	Italy	Japan
South Korea	Netherlands	New Zealand	Portugal	Spain
Sweden	Switzerland	United States		
Group 2 "Convergence" ($N_2 = 6$)				
Australia	Canada	Denmark	Greece	Norway
United Kingdom				

Overall, we re-estimate Coe and Helpman's model by both pooled FMOLS and group-specific PPC method with unobserved global trends. Our pooled FMOLS confirms the international R&D spillovers in global after considering unobserved global trend. In addition, our Cup-Lasso estimates show heterogeneous behavior of the long-run relations between domestic R&D and foreign R&D on TFP. It indicates multiple regimes of economies' convergence behaviors. This is also empirically confirms the "Club convergence" theory proposed by Quah (1996, 1997). Countries in the convergence club tend to grow faster and further behind they fall. In our model, we specify the channel of convergence through the technology diffusion. Based on estimation results, we classified those countries into three clubs–"Divergence", "Convergence", and "Balance". We can conclude that international technology diffusion is the major sources of productivity growth of countries in "Convergence" group. The catch-up effect through the channel of technology diffusion is a force towards convergence in income for those countries. On the contrary, countries in "Divergence" group show a opposite story. The productivity growth relies highly on their own R&D stock and they cannot benefit from international R&D spillovers.

6 Conclusion

The main contribution of this paper is to propose a novel approach that handle the unobserved heterogeneity and cross-sectional dependence in nonstationary panel model. We assume that cross-sectional dependence is generated by the unobserved common factors, which can be either stationary or nonstationary. In general, the penalized least square estimators are inconsistent due to the spurious regression induced by unobserved nonstationary factors. We propose the penalized principal component method that jointly estimates the group-specific long-run relations, identifies individuals' membership and unobserved common factors through iterative procedure. The C-Lasso estimators have asymptotic non-negligible biases due to weakly dependent error processes and unobserved stationary factors. We employ the fully modified procedure for bias-correction. Since our PPC-based method allows us to account for the unobserved heterogeneity and cross-sectional dependence simultaneously, it is best fitted method to explain growth convergence puzzle. Our empirical results identify multiple steady states of convergence across economies.

Appendix

In this appendix, we prove the main results, namely, Theorems 3.1-3.9 in the paper. The proofs of these results need some technical lemmas whose proofs are relegated to the online supplementary Appendix B.

A Proof of the Main Results in Section 3

To proceed, we define some notations.

(i) Let $H_1 = \left(\frac{1}{N}\Lambda_1^{0\prime}\Lambda_1^0\right)\left(\frac{1}{T^2}f_1^{0\prime}\hat{f}_1\right)V_{1,NT}^{-1}$, $H_2 = \left(\frac{1}{N}\Lambda_2^{0\prime}\Lambda_2^0\right)\left(\frac{1}{T}f_2^{0\prime}\hat{f}_2\right)V_{2,NT}^{-1}$ and $a_{ij} = \lambda_{1i}^{0\prime}(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N})^{-1}\lambda_{1j}^0$.

(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 $\hat{Q}_{i,xx} = \frac{1}{T^2}x_i' M_{\hat{f}_1} x_i$, $Q_{i,xx} = \frac{1}{T^2}x_i' M_{f_1} x_i$, and $Q_{i,xx}(f_1^0) = \frac{1}{T^2}x_i' M_{f_1^0} x_i$.

(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 need four lemmas.

Lemma A.1 *Suppose that Assumptions 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. Then for any fixed small constant $c \in (0, 1/2)$,*

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

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

(iii) $\limsup_{T\to\infty} \mu_{\max}\left(\frac{x_i' M_{f_1^0} x_i}{d_T T^2}\right) \leq (1 + c)\rho_{\max}$ *a.s.*,

(iv) $\liminf_{T\to\infty} \mu_{\min}\left(\frac{d_T x_i' M_{f_1^0} x_i}{T^2}\right) \geq [(1 + c)\rho_{\max}]^{-1}$ *a.s.*,

where $W_{it} = (x_{it}', f_{1t}^{0\prime})'$ *and* $W_i = (W_{i,1}, W_{i,2}, ..., W_{i,T})'$.

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 f_1 *satisfies* $\frac{1}{T^2}f_1' f_1 = I_{r_1}$ *and* $u_i^* = u_i + f_2^0\lambda_{2i}^0$.

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

(i) $\sup_{f_1}\sup_{N^{-1}\|\mathbf{b}\|^2 \leq M}\left\|\frac{1}{NT^2}\sum_{i=1}^N b_i' x_i' M_{f_1} u_i^*\right\| = o_P(d_T^{-3})$,

(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^* *are 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 $y_i - x_i\beta_i = -x_i b_i + \lambda_{1i}^{0\prime} f_1^{0\prime} + u_i^*$, we have

$$\begin{aligned} Q_{i,NT}(\beta_i, f_1) - Q_{i,NT}(\beta_i^0, f_1^0) =& \frac{1}{T^2}(b_i' x_i' M_{f_1} x_i b_i + \lambda_{1i}^{0\prime} f_1^{0\prime} M_{f_1} f_1^0 \lambda_{1i}^0 - 2b_i' x_i' M_{f_1} f_1^0 \lambda_{1i}^0) \\ &+ \frac{1}{T^2}(2\lambda_{1i}^{0\prime} f_1^{0\prime} M_{f_1} u_i^* - 2b_i' x_i' M_{f_1} u_i^*) - \frac{1}{T^2} u_i^{*\prime}(P_{f_1} - P_{f_1^0}) u_i^*, \end{aligned} \tag{A.1}$$

where $u_i^* = u_i + f_2^0 \lambda_{2i}^0$. Let $S_{i,NT}(\beta_i, f_1) = \frac{1}{T^2}\left(b_i' x_i' M_{f_1} x_i b_i + \lambda_{1i}^{0\prime} f_1^{0\prime} M_{f_1} f_1^0 \lambda_{1i}^0 - 2b_i' x_i' M_{f_1} f_1^0 \lambda_{1i}^0\right)$. Then we have

$$\begin{aligned} Q_{NT}(\boldsymbol{\beta}, f_1) - Q_{NT}(\boldsymbol{\beta}^0, f_1^0) =& \frac{1}{N}\sum_{i=1}^N S_{i,NT}(\beta_i, f_1) \\ &+ \frac{1}{NT^2}\sum_{i=1}^N \left(2\lambda_{1i}^{0\prime} f_1^{0\prime} M_{f_1} u_i^* - 2b_i' x_i' M_{f_1} u_i^* - u_i^{*\prime}(P_{f_1} - P_{f_1^0}) u_i^*\right) \\ =& \frac{1}{N}\sum_{i=1}^N S_{i,NT}(\beta_i, f_1) + o_P(d_T^{-3}), \end{aligned} \tag{A.2}$$

where the last three terms on the right hand 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 $\frac{1}{N}\sum_{i=1}^N \|b_i\|^2 \leq M$ 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

$$\begin{aligned} Q_{NT}^{K,\lambda}(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}, f_1) - Q_{NT}^{K,\lambda}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, f_1^0) =& \frac{1}{N}\sum_{i=1}^N [Q_{NT,i}(\beta_i, f_1) - Q_{NT,i}(\beta_i^0, f_1^0)] + \frac{\lambda}{N}\sum_{i=1}^N \prod_{k=1}^{K^0} \|\beta_i - \hat{\alpha}_k\| \\ \geq& S_{NT}(\boldsymbol{\beta}, f_1) + o_P(d_T^{-3}). \end{aligned} \tag{A.3}$$

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 $Q_{NT}^{K,\lambda}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{f}_1) - Q_{NT}^{K,\lambda}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, f_1^0) \leq 0$, we have

$$S_{NT}(\hat{\boldsymbol{\beta}}, \hat{f}_1) = \frac{1}{NT^2}\sum_{i=1}^N \left[\hat{b}_i' x_i' M_{\hat{f}_1} x_i \hat{b}_i + \lambda_{1i}^{0\prime} f_1^{0\prime} M_{\hat{f}_1} f_1^0 \lambda_{1i}^0 - 2\hat{b}_i' x_i' M_{\hat{f}_1} f_1^0 \lambda_{1i}^0\right] = o_P(d_T^{-3}). \tag{A.4}$$

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

$$\begin{aligned} Q_{NT}^{K,\lambda}(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}, \hat{f}_1) - Q_{NT}^{K,\lambda}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, \hat{f}_1) =& \frac{1}{N}\sum_{i=1}^N [Q_{NT,i}(\beta_i, \hat{f}_1) - Q_{NT,i}(\beta_i^0, \hat{f}_1)] + \frac{\lambda}{N}\sum_{i=1}^N \prod_{k=1}^{K_0} \|\beta_i - \hat{\alpha}_k\| \\ \geq& \frac{1}{NT^2}\sum_{i=1}^N \left[b_i' x_i' M_{\hat{f}_1} x_i b_i - 2b_i' x_i' M_{\hat{f}_1} f_1^0 \lambda_{1i}^0\right] + o_P(d_T^{-3}). \end{aligned} \tag{A.5}$$

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

$$\frac{1}{NT^2}\sum_{i=1}^{N}\left[\hat{b}_i' x_i' M_{\hat{f}_1} x_i \hat{b}_i - 2\hat{b}_i' x_i' M_{\hat{f}_1} f_1^0 \lambda_{1i}^0\right] \leq o_P(d_T^{-3}). \tag{A.6}$$

Combining (A.4) and (A.6) yields that

$$o_P(d_T^{-3}) = \frac{1}{NT^2}\lambda_{1i}^{0\prime} f_1^{0\prime} M_{\hat{f}_1} f_1^0 \lambda_{1i}^0 = \operatorname{tr}\left[\left(\frac{f_1^{0\prime} M_{\hat{f}_1} f_1^0}{T^2}\right)\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)\right] \geq \operatorname{tr}\left(\frac{f_1^{0\prime} M_{\hat{f}_1} f_1^0}{T^2}\right)\mu_{\min}\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right).$$

It follows that $\operatorname{tr}\left(\frac{f_1^{0\prime} M_{\hat{f}_1} f_1^0}{T^2}\right) = o_P(d_T^{-3})$ as $\mu_{\min}\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)$ is bounded away from zero in probability by Assumption 3.2(i). As in Bai (2009, p.1265), this implies that

$$\frac{f_1^{0\prime} M_{\hat{f}_1} f_1^0}{T^2} = \frac{f_1^{0\prime} f_1^0}{T^2} - \frac{f_1^{0\prime}\hat{f}_1}{T^2}\frac{\hat{f}_1{}' f_1^0}{T^2} = o_P(d_T^{-3}), \tag{A.7}$$

and $\frac{f_1^{0\prime}\hat{f}_1}{T^2}$ is asymptotically invertible by the fact that $\frac{f_1^{0\prime} f_1^0}{T^2}$ is asymptotically invertible from Assumption 3.2(ii). (A.7) implies that $\frac{\hat{f}_1' P_{f_1^0}\hat{f}_1}{T^2} - I_{r_1} = o_P(d_T^{-3})$, which further implies that $\left\|P_{\hat{f}_1} - P_{f_1^0}\right\|^2 = 2\operatorname{tr}\left(I_{r_1} - \frac{\hat{f}_1{}' P_{f_1^0}\hat{f}_1}{T^2}\right) = o_P(d_T^{-3})$. By Cauchy-Schwarz inequality and (A.6),

$$o_P(d_T^{-3}) \geq \frac{1}{NT^2}\sum_{i=1}^{N}\hat{b}_i' x_i' M_{\hat{f}_1} x_i \hat{b}_i - 2\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\hat{b}_i' x_i' M_{\hat{f}_1} x_i \hat{b}_i\right\}^{1/2}\left\{\frac{1}{NT^2}\lambda_{1i}^{0\prime} f_1^{0\prime} M_{\hat{f}_1} f_1^0 \lambda_{1i}^0\right\}^{1/2}. \tag{A.8}$$

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

$$\begin{aligned}
o_P(d_T^{-3}) &= \frac{1}{N}\sum_{i=1}^{N}\hat{b}_i'\left(\frac{1}{T^2}x_i' M_{\hat{f}_1} x_i\right)\hat{b}_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\hat{b}_i'\left(\frac{1}{T^2}x_i' M_{f_1^0} x_i\right)\hat{b}_i + \frac{1}{N}\sum_{i=1}^{N}\hat{b}_i'\left(\frac{1}{T^2}x_i'(M_{\hat{f}_1} - M_{f_1^0})x_i\right)\hat{b}_i \\
&\geq \frac{1}{d_T}\min_{1\leq i\leq N}\mu_{\min}\left(\frac{d_T}{T^2}x_i' M_{f_1^0} x_i\right)\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_i\|^2 - \max_{1\leq i\leq N}\frac{\|x_i\|^2}{T^2}\|P_{f_1^0} - P_{\hat{f}_1}\|\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_i\|^2 \\
&\geq \frac{1}{d_T}\left(c\rho_{\min} - o_P(d_T^{-1})\right)\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_i\|^2,
\end{aligned}$$

where the second inequality follows from the fact that $\min_{1\leq i\leq N}\mu_{\min}\left(\frac{d_T}{T^2}x_i' M_{f_1^0} x_i\right) \geq c\rho_{\min} > 0$ a.s. by Lemma A.2(iv), and $\max_{1\leq i\leq N}\frac{\|x_i\|^2}{T^2} \leq \max_{1\leq i\leq N} d_T\mu_{\max}\left(\frac{x_i' x_i}{d_T T^2}\right) = O_P(d_T)$ 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 the heterogeneous slope coefficients and unobserved stationary common factors. We estimate $\hat{f}_1$ from equation (2.9) in Section 2.2 as follows

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

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

$$\begin{aligned}
\hat{f}_1 V_{1,NT} =& \frac{1}{NT^2}\sum_{i=1}^{N} x_i\hat{b}_i\hat{b}_i'x_i'\hat{f}_1 - \frac{1}{NT^2}\sum_{i=1}^{N} x_i\hat{b}_i\lambda_i^{0\prime}f^{0\prime}\hat{f}_1 - \frac{1}{NT^2}\sum_{i=1}^{N} x_i\hat{b}_i u_i'\hat{f}_1 \\
&- \frac{1}{NT^2}\sum_{i=1}^{N} f^0\lambda_i^0\hat{b}_i'x_i'\hat{f}_1 - \frac{1}{NT^2}\sum_{i=1}^{N} u_i\hat{b}_i'x_i'\hat{f}_1 + \frac{1}{NT^2}\sum_{i=1}^{N} f^0\lambda_i^0 u_i'\hat{f}_1 \\
&+ \frac{1}{NT^2}\sum_{i=1}^{N} u_i\lambda_i^{0\prime}f^{0\prime}\hat{f}_1 + \frac{1}{NT^2}\sum_{i=1}^{N} u_i u_i'\hat{f}_1 + \frac{1}{NT^2}\sum_{i=1}^{N} f_2^0\lambda_{2i}^0\lambda_{2i}^{0\prime}f_2^{0\prime}\hat{f}_1 \\
&+ \frac{1}{NT^2}\sum_{i=1}^{N} f_1^0\lambda_{1i}^0\lambda_{2i}^{0\prime}f_2^{0\prime}\hat{f}_1 + \frac{1}{NT^2}\sum_{i=1}^{N} f_2^0\lambda_{2i}^0\lambda_{1i}^{0\prime}f_1^{0\prime}\hat{f}_1 + \frac{1}{NT^2}\sum_{i=1}^{N} f_1^0\lambda_{1i}^0\lambda_{1i}^{0\prime}f_1^{0\prime}\hat{f}_1 \\
\equiv& I_1 + ... + I_{11} + \frac{1}{NT^2}\sum_{i=1}^{N} f_1^0\lambda_{1i}^0\lambda_{1i}^{0\prime}f_1^{0\prime}\hat{f}_1, \text{ say.}
\end{aligned}$$

It follows that $\hat{f}_1 V_{1,NT} - f_1^0\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)\left(\frac{f_1^{0\prime}\hat{f}_1}{T^2}\right) = I_1 + ... + I_{11}$. Let $H_1 = \left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)\left(\frac{f_1^{0\prime}\hat{f}_1}{T^2}\right)V_{1,NT}^{-1}$. Then it is easy to see that $H_1 = O_P(1)$, it is asymptotically nonsingular, and

$$\hat{f}_1 H_1^{-1} - f_1^0 = [I_1 + ... + I_{11}]\left(\frac{f_1^{0\prime}\hat{f}_1}{T^2}\right)^{-1}\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)^{-1}.$$

Note that

$$\frac{1}{T}\left\|\hat{f}_1 H^{-1} - f_1^0\right\| \le \frac{1}{T}(\|I_1\| + ... + \|I_{11}\|)\left\|\left(\frac{f_1^{0\prime}\hat{f}_1}{T^2}\right)^{-1}\right\|\left\|\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)^{-1}\right\|.$$

It remains to analyze $\|I_l\|$ for $l = 1, 2, ..., 11$. For I_1, we have that by the result in (iii),

$$\begin{aligned}
\frac{1}{T}\|I_1\| &= \frac{1}{T}\left\|\frac{1}{NT^2}\sum_{i=1}^{N} x_i\hat{b}_i\hat{b}_i'x_i'\hat{f}_1\right\| \le \frac{1}{N}\sum_{i=1}^{N}\frac{\|x_i\|}{T}\|\hat{b}_i\|^2\frac{\|x_i'\hat{f}_1\|}{T^2} \\
&\le \max_{1\le i\le N}\frac{\|x_i\|^2}{T^2}\frac{\|\hat{f}_1\|}{T}\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_i\|^2 = O_P(d_T\eta_{NT}^2) = o_P(\eta_{NT}),
\end{aligned}$$

where we use the fact that $\max_{1\le i\le N}\frac{\|x_i\|^2}{T^2} = O_P(d_T)$ by Lemma A.2(i) and $\frac{\|\hat{f}_1\|}{T} \le \sqrt{r_1}$. For I_2, we have

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

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,

$$\begin{aligned}
\frac{1}{T}\|I_3\| &= \frac{1}{T}\left\|\frac{1}{NT^2}\sum_{i=1}^N x_i b_i u_i' \hat{f}_1\right\| \leq \max_{1\leq i\leq N}\frac{\|x_i\|}{T}\frac{1}{NT^2}\sum_{i=1}^N \|\hat{b}_i u_i' \hat{f}_1\| \\
&\leq \frac{1}{\sqrt{T}}\frac{\|\hat{f}_1\|}{T}\max_{1\leq i\leq N}\frac{\|x_i\|}{T}\left\{\frac{1}{N}\sum_{i=1}^N\|\hat{b}_i\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\frac{\|u_i\|^2}{T}\right\}^{1/2} = O_P\left(\sqrt{\frac{d_T}{T}}\eta_{NT}\right),
\end{aligned}$$

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

$$\begin{aligned}
\frac{1}{T}\|I_4\| &= \left\|\frac{1}{NT^2}\sum_{i=1}^N f^0\lambda_i^0\hat{b}_i' x_i' \hat{f}_1\right\| \leq \frac{1}{N}\sum_{i=1}^N\frac{\|f^0\|}{T}\left\|\lambda_i^0\hat{b}_i'\right\|\left\|\frac{x_i'\hat{f}_1}{T^2}\right\| \\
&\leq \frac{\|f^0\|}{T}\frac{\|\hat{f}_1\|}{T}\max_{1\leq i\leq N}\frac{\|x_i\|}{T}\left\{\frac{1}{N}\sum_{i=1}^N\|\hat{b}_i\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\|\lambda_i^0\|^2\right\}^{1/2} = O_P(\sqrt{d_T}\eta_{NT}).
\end{aligned}$$

where $\frac{\|f^0\|}{T} \leq \frac{\|f_1^0\|}{T} + \frac{1}{\sqrt{T}}\frac{\|f_2^0\|}{\sqrt{T}} = O_P(1)$. For I_5,

$$\begin{aligned}
\frac{1}{T}\|I_5\| &= \frac{1}{T}\left\|\frac{1}{NT^2}\sum_{i=1}^N u_i\hat{b}_i' x_i' \hat{f}_1\right\| \leq \max_{1\leq i\leq N}\frac{\|x_i'\hat{f}_1\|}{T^2}\frac{1}{NT}\sum_{i=1}^N\|u_i\hat{b}_i'\| \\
&\leq \frac{1}{\sqrt{T}}\frac{\|\hat{f}_1\|}{T}\max_{1\leq i\leq N}\frac{\|x_i\|}{T}\left\{\frac{1}{N}\sum_{i=1}^N\frac{\|u_i\|^2}{T}\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\|\hat{b}_i\|\right\}^{1/2} = O_P\left(\sqrt{\frac{d_T}{T}}\eta_{NT}\right).
\end{aligned}$$

For I_6,

$$\begin{aligned}
\frac{1}{T}\|I_6\| =& \frac{1}{T}\left\|\frac{1}{NT^2}\sum_{i=1}^N f^0\lambda_i^0 u_i'\hat{f}_1\right\| = \frac{1}{T}\left\|\frac{1}{NT^2}f^0\Lambda^{0\prime}u\hat{f}_1\right\| \\
&\leq \frac{1}{\sqrt{NT}}\left(\frac{1}{T}\left\|\hat{f}_1\right\|\right)\left(\frac{1}{T}\left\|f^0\right\|\right)\frac{1}{\sqrt{NT}}\left\|\Lambda^{0\prime}u\right\| = O_P(T^{-1/2}N^{-1/2}),
\end{aligned}$$

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,

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

where $\gamma_N(s,t)$ and ξ_{st} are defined in Assumption 3.2(iii). For $I_8(a)$,

$$\begin{aligned}
\|I_8(a)\|^2 &\leq \sum_{t=1}^T\left\|T^{-3}\sum_{s=1}^T\gamma_N(s,t)\hat{f}_{1s}'\right\|^2 \leq T^{-3}\left(T^{-2}\sum_{s=1}^T\|\hat{f}_{1s}\|^2\right)\left(T^{-1}\sum_{t=1}^T\sum_{s=1}^T\|\gamma_N(s,t)\|^2\right) \\
&= O_P(T^{-3}),
\end{aligned}$$

where $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\gamma_N(s,t)\|^2 \leq M$ by Assumption 3.2(iii) (see also Lemma 1(i) in Bai and Ng (2002)). For $I_8(b)$,

$$\begin{aligned}\|I_8(b)\| &= \sum_{t=1}^{T}\left\|T^{-3}\sum_{s=1}^{T}\xi_{st}\hat{f}_{1s}'\right\|^2 \leq T^{-2}N^{-1}\left(T^{-2}\sum_{s=1}^{T}\left\|\hat{f}_{1s}\right\|^2\right)\left(T^{-2}N\sum_{t=1}^{T}\sum_{s=1}^{T}\|\xi_{st}\|^2\right)\\ &= O_P(T^{-2}N^{-1}),\end{aligned}$$

where we use the fact that $E(\|\xi_{st}\|^2) \leq N^{-2}M$ under Assumption 3.2(iii). Then we have $\frac{1}{T}\|I_8\| = O_P(N^{-1/2}T^{-1}+T^{-3/2})$. For $\|I_9\|$,

$$\frac{1}{T}\|I_9\| = \frac{1}{T}\left\|\frac{1}{NT^2}f_2^0\Lambda_2^{0\prime}\Lambda_2^0 f_2^{0\prime}\hat{f}_1\right\| \leq \frac{1}{T}\frac{\|f_2^0\|^2}{T}\frac{\|\hat{f}_1\|}{T}\left\|\frac{\Lambda_2^{0\prime}\Lambda_2^0}{N}\right\| = O_P(T^{-1}).$$

For $\|I_{10}\|$,

$$\frac{1}{T}\|I_{10}\| = \frac{1}{T}\left\|\frac{1}{NT^2}f_1^0\Lambda_1^{0\prime}\Lambda_2^0 f_2^{0\prime}\hat{f}_1\right\| \leq \frac{1}{\sqrt{NT}}\frac{\|f_1^0\|}{T}\frac{\|f_2^0\|}{\sqrt{T}}\frac{\|\hat{f}_1\|}{T}\frac{\left\|\Lambda_1^{0\prime}\Lambda_2^0\right\|}{\sqrt{N}} = O_P((NT)^{-1/2}),$$

where $\frac{\Lambda_1^{0\prime}\Lambda_2^0}{\sqrt{N}} = O_P(1)$ by Assumption 3.2(i). Analogously, we have $\frac{1}{T}\|I_{11}\| = O_P((NT)^{-1/2})$. In sum, we have shown that

$$\frac{1}{T}\left\|\hat{f}_1 H_1^{-1} - f_1^0\right\| = O_P(\sqrt{d_T}\eta_{NT}) + \frac{1}{\sqrt{T}}O_P(C_{NT}^{-1}).$$

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) $\frac{1}{T}\hat{f}_1'(\hat{f}_1 - f_1^0 H_1) = O_P(T\sqrt{d_T}\eta_{NT} + \delta_{NT}^{-1})$,

(iii) $\|P_{\hat{f}_1} - P_{f_1^0}\|^2 = O_P(\sqrt{d_T}\eta_{NT} + T^{-1}\delta_{NT}^{-1})$,

(iv) $\frac{1}{T}u_k^{*\prime}\left(\hat{f}_1 H_1^{-1} - f_1^0\right) = O_P(\sqrt{Td_T}\eta_{NT} + \delta_{NT}^{-1})$ for each $k = 1, ..., N$.

Lemma A.6 *Suppose that Assumptions 3.1-3.2 hold. Let* $R_{1i} = \frac{1}{T^2}x_i'(P_{f_1^0} - P_{\hat{f}_1})u_i^*$, $R_{2i} = \frac{1}{T^2}x_i'M_{\hat{f}_1}f_1^0\lambda_{1i}^0 - \frac{1}{NT^2}\sum_{j=1}^{N}x_i'M_{\hat{f}_1}x_j a_{ij}\hat{b}_j + \frac{1}{NT^2}\sum_{j=1}^{N}a_{ij}x_i'M_{\hat{f}_1}u_j$, $R_{3i} = \frac{1}{NT^2}\sum_{j=1}^{N}a_{ij}x_i'(P_{f_1^0} - P_{\hat{f}_1})u_j$, *and* $R_{4i} = \frac{1}{T^2}x_i'M_{f_1^0}u_i^* - \frac{1}{NT^2}\sum_{j=1}^{N}a_{ij}x_i'M_{f_1^0}u_j$. *Then*

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

(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 (2.8) is the for each $i = 1, ..., N$, $0_{p\times 1}$ belongs to the sub-differential of

$Q^K_{NT,\lambda}(\beta,\alpha,f_1)$ 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

$$0_{p\times 1}=-\frac{2}{T^2}x_i'M_{\hat f_1}(y_i-x_i\hat\beta_i)+\lambda\sum_{j=1}^{K}\hat e_{ij}\prod_{l=1,l\neq j}^{K}\|\hat\beta_i-\hat\alpha_l\|, \tag{A.10}$$

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

$$\hat Q_{i,xx}\hat b_i=\frac{1}{T^2}x_i'M_{\hat f_1}u_i^*+\frac{1}{T^2}x_i'M_{\hat f_1}f_1^0\lambda_{1i}^0-\frac{\lambda}{2}\sum_{j=1}^{K_0}\hat e_{ij}\prod_{l=1,l\neq j}^{K}\|\hat\beta_i-\hat\alpha_l\|, \tag{A.11}$$

which can be rewritten as

$$\hat Q_{i,xx}\hat b_i=\frac{1}{NT^2}\sum_{j=1}^{N}x_i'M_{\hat f_1}x_j a_{ij}\hat b_j+R_i, \tag{A.12}$$

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

$$\hat Q_2=\begin{pmatrix}\frac{1}{NT^2}x_1'M_{\hat f_1}x_1a_{11} & \frac{1}{NT^2}x_1'M_{\hat f_1}x_2a_{12} & \cdots & \frac{1}{NT^2}x_1'M_{\hat f_1}x_Na_{1N}\\ \frac{1}{NT^2}x_2'M_{\hat f_1}x_1a_{21} & \frac{1}{NT^2}x_2'M_{\hat f_1}x_2a_{22} & \cdots & \frac{1}{NT^2}x_2'M_{\hat f_1}x_Na_{2N}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{1}{NT^2}x_N'M_{\hat f_1}x_1a_{N1} & \frac{1}{NT^2}x_N'M_{\hat f_1}x_2a_{N2} & \cdots & \frac{1}{NT^2}x_N'M_{\hat f_1}x_Na_{NN}\end{pmatrix}.$$

Let $R=(R_1',...,R_N')'$. Then (A.12) implies that $(\hat Q_1-\hat Q_2)\hat{\mathbf{b}}=R$. It follows that

$$\|R\|^2=\mathrm{tr}(\hat{\boldsymbol b}'(\hat Q_1-\hat Q_2)'(\hat Q_1-\hat Q_2)\hat{\boldsymbol b})\geq\|\hat{\boldsymbol b}\|^2\left[\mu_{\min}\left(\hat Q_1-\hat Q_2\right)\right]^2.$$

By Assumption 3.2(v), we have that $\mu_{\min}\left(\hat Q_1-\hat Q_2\right)\geq\rho_{\min}/2>0$ w.p.a.1. Then we have $\frac{1}{N}\|\hat{\boldsymbol b}\|^2\leq\frac{\rho_{\min}^2}{4N}\sum_{i=1}^{N}\|R_i\|^2=O_P(T^{-1}d_T^{1/2}\eta_{NT}+d_T^2\eta_{NT}^4+T^{-2}+\lambda^2)=O_P\left(d_TT^{-2}+\lambda^2\right)$. Consequently, we prove the means square convergence rate of C-Lasso estimators that $\frac{1}{N}\sum_{i=1}^{N}\|\hat b_i\|^2=O_P\left(d_TT^{-2}+\lambda^2\right)$.

Next, we want to strengthen the last result to a stronger one: $\frac{1}{N}\sum_{i=1}^{N}\|\hat b_i\|^2=O_P(d_TT^{-2})$. Let $\boldsymbol\beta=\boldsymbol\beta^0+d_TT^{-1}\boldsymbol v$, where $\boldsymbol v=(v_1,...,v_N)$ is a $p\times N$ matrix. Let $v=$vec$(\boldsymbol v)$. We wan 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 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

$$\begin{aligned}
&T^2\left[Q_{NT}^{\lambda,K}(\boldsymbol{\beta}+d_T^{1/2}T^{-1}\boldsymbol{v},\hat{\boldsymbol{\alpha}},\hat{f}_1)-Q_{NT}^{\lambda,K}(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\hat{f}_1)\right]\\
&\geq \frac{d_T^{1/2}}{N}\sum_{i=1}^{N}\left(\frac{d_T^{1/2}}{T^2}v_i'x_i'M_{\hat{f}_1}x_iv_i-\frac{2}{T}v_i'x_i'M_{\hat{f}_1}(f_1^0-\hat{f}_1H_1)\lambda_{1i}^0-\frac{2}{T}v_i'x_i'M_{\hat{f}_1}u_i^*\right)\\
&=\frac{d_T}{N}\sum_{i=1}^{N}\frac{1}{T^2}v_i'x_i'M_{\hat{f}_1}x_iv_i\\
&\quad-\frac{2d_T^{1/2}}{N}\sum_{i=1}^{N}v_i'\left\{TR_{2i}+\frac{1}{T}x_i'M_{\hat{f}_1}u_i^*+\frac{1}{NT}\sum_{j=1}^{N}a_{ij}x_i'M_{\hat{f}_1}x_j\hat{b}_j-\frac{1}{NT}\sum_{j=1}^{N}a_{ij}x_i'M_{\hat{f}_1}u_j\right\}\\
&\equiv D_{1NT}-2D_{2NT},\ \text{say.}
\end{aligned}$$

where $R_{2i}=\frac{1}{T^2}x_i'M_{\hat{f}_1}f_1^0\lambda_{1i}^0-\frac{1}{NT^2}\sum_{j=1}^{N}x_i'M_{\hat{f}_1}x_ja_{ij}\hat{b}_j+\frac{1}{NT^2}\sum_{j=1}^{N}a_{ij}x_i'M_{\hat{f}_1}u_j$ as defined in Lemma A.6. By Assumption 3.2(v) and Lemma A.5(iii), $D_{1NT}=\frac{d_T}{N}v'\hat{Q}_1v\geq d_T\mu_{\min}\left(\hat{Q}_1\right)N^{-1}\left\|\boldsymbol{v}\right\|^2\geq d_T\rho_{\min}N^{-1}\left\|\boldsymbol{v}\right\|^2/2$ w.p.a.1. By Lemmas A.6(i)-(ii) and A.5(iii),

$$\begin{aligned}
\frac{T^2}{d_TN}\sum_{i=1}^{N}\|R_{2i}\|^2 &= \frac{T^2}{d_T}O_P(T^{-2}d_T\eta_{NT}^2+d_T^2\eta_{NT}^4+T^{-2}\delta_{NT}^{-2})=o_P(1),\\
\frac{1}{d_TNT^2}\sum_{i=1}^{N}\|x_i'M_{\hat{f}_1}u_i^*\|^2 &\leq \frac{2T^2}{d_TN}\sum_{i=1}^{N}\left\|\frac{1}{T^2}x_i'(M_{\hat{f}_1}-M_{f_1^0})u_i^*\right\|^2+\frac{2}{d_TN}\sum_{i=1}^{N}\left\|\frac{1}{T}x_i'M_{f_1^0}u_i^*\right\|^2\\
&= \frac{T^2}{d_T}O_P(T^{-1}d_T\eta_{NT}^2+d_T^2\eta_{NT}^4+T^{-2}C_{NT}^{-2})+\frac{1}{d_T}O_P\left(1\right)=o_P(1).
\end{aligned}$$

Next, we have

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

where we use the fact that $\max_{1\leq j\leq N}\frac{1}{d_TT^2}\left\|x_j\right\|^2=O_P\left(1\right)$ 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 Markov inequality, and $\frac{1}{NT^2}\sum_{i=1}^{N}\left\|\lambda_{1i}^0\right\|^2\|x_i\|^2=O_P\left(1\right)$ by 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),

$$
\begin{aligned}
&\frac{1}{d_T N^3 T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|a_{ij}x_i' M_{\hat{f}_1}u_j\|^2 \\
&\leq \frac{1}{d_T}\frac{1}{N^3T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|a_{ij}\|^2\left\|x_i' M_{\hat{f}_1}u_j\right\|^2 \\
&\leq \frac{1}{d_T}\left[\mu_{\min}\left(\frac{\Lambda_1^{0\prime}\Lambda_1^0}{N}\right)\right]^{-2}\frac{2}{N^3T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\lambda_{1i}^0\|^2\|\lambda_{1j}^0\|^2\left\{\left\|x_i'(M_{\hat{f}_1}-M_{f_1^0})u_j\right\|^2+\left\|x_i' M_{f_1^0}u_j\right\|^2\right\} \\
&= \frac{1}{d_T}O_P\left(N^{-1}Td_T(\sqrt{d_T}\eta_{NT}+\delta_{NT}^{-1})+1\right) = o_P(1).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
|D_{2NT}| \leq\ & d_T\left\{\frac{1}{N}\sum_{i=1}^{N}\|v_i\|^2\right\}^{1/2}\left\{\left(\frac{T^2}{d_T N}\sum_{i=1}^{N}\|\bar{R}_{2i}\|^2\right)^{1/2}+\left(\frac{1}{d_T NT^2}\sum_{i=1}^{N}\|x_i' M_{\hat{f}_1}u_i^*\|^2\right)^{1/2}\right. \\
&\left.+\left(\frac{1}{d_T N^3T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|a_{ij}x_i' M_{\hat{f}_1}\hat{b}_j\|^2\right)^{1/2}+\left(\frac{1}{d_T N^3T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|a_{ij}x_i' M_{\hat{f}_1}u_j\|^2\right)^{1/2}\right\} \\
=\ & d_T N^{-1/2}\|\boldsymbol{v}\|\, o_P(1).
\end{aligned}
$$

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 hand side of (**??**). 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

$$
\begin{aligned}
\left\|\frac{1}{T^2}x_i' M_{\hat{f}_1}f_1^0\lambda_{1i}^0\right\| &\leq \|R_{2i}\|+\left\|\frac{1}{NT^2}\sum_{j=1}^{N}x_i' M_{\hat{f}_1}x_j\hat{b}_j a_{ij}\right\|+\left\|\frac{1}{NT^2}\sum_{j=1}^{N}x_i' M_{\hat{f}_1}u_j a_{ij}\right\| \\
&= O_P(T^{-1}\sqrt{d_T}\eta_{NT}+d_T\eta_{NT}^2+T^{-1}\delta_{NT}^{-1})+O_P(\eta_{NT})+O_P(d_T T^{-1}) = O_P(d_T T^{-1}). \qquad \text{(A.14)}
\end{aligned}
$$

The third term is $O_P(\lambda)$. By Lemma A.5(iii), $\mu_{\min}\left(\frac{1}{T^2}x_i' M_{\hat{f}_1}x_i\right) = \mu_{\min}\left(\frac{1}{T^2}x_i' M_{f_1^0}x_i\right)+o_P(1)$. Noting that $\left(\frac{1}{T^2}x_i' M_{f_1^0}x_i\right)^{-1}$ is the principal $p\times p$ submatrix of $\left(\frac{1}{T^2}W_i' W_i\right)^{-1}$, $\mu_{\min}\left(\frac{1}{T^2}x_i' M_{f_1^0}x_i\right) \geq \mu_{\min}\left(\frac{1}{T^2}W_i' W_i\right)$, 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 $(N,T) \to \infty$, $\left| \prod_{k=1}^K \left\| \hat{\beta}_i - \alpha_k \right\| - \prod_{k=1}^K \left\| \beta_i^0 - \alpha_k \right\| \right| \leq \hat{c}_{iNT}(\alpha) \|\hat{\beta}_i - \beta_i^0\|$, 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 $(N,T) \to \infty$

$$
\begin{aligned}
|P_{NT}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}) - P_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\alpha})| &\leq C_{KNT}(\alpha) \frac{1}{N} \sum_{i=1}^N \|\hat{b}_i\| + 2C_{KNT}(\alpha) \frac{1}{N} \sum_{i=1}^N \|\hat{b}_i\|^2 \\
&\leq C_{KNT}(\alpha) \left\{ \frac{1}{N} \sum_{i=1}^N \|\hat{b}_i\|^2 \right\}^{1/2} + O_P(d_T T^{-2}) = O_P(d_T^{1/2} T^{-1}). \qquad (A.15)
\end{aligned}
$$

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

$$
\begin{aligned}
0 &\geq P_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - P_{NT}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}^0) = P_{NT}(\boldsymbol{\beta}^0, \hat{\boldsymbol{\alpha}}) - P_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0) + O_P(d_T^{1/2} T^{-1}) \\
&= \frac{1}{N} \sum_{i=1}^N \prod_{k=1}^K \|\beta_i^0 - \hat{\alpha}_k\| + O_P(d_T^{1/2} T^{-1}) \\
&= \frac{N_1}{N} \prod_{k=1}^K \|\hat{\alpha}_k - \alpha_1^0\| + \frac{N_2}{N} \prod_{k=1}^K \|\hat{\alpha}_k - \alpha_2^0\| + ... + \frac{N_K}{N} \prod_{k=1}^K \|\hat{\alpha}_k - \alpha_K^0\| + O_P(d_T^{1/2} T^{-1}) \qquad (A.16)
\end{aligned}
$$

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

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

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

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

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

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

(vii) $P\left(\max_{1 \leq i \leq N} \left\| \frac{1}{T^2} x_i' M_{\hat{f}_1} f_1^0 \lambda_{1i}^0 \right\| > c N^{1/2q} (d_T \eta_{NT} + T^{-1/2} d_T^{1/2} C_{NT}^{-1})\right) = o(N^{-1})$.

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

$$
\begin{aligned}
0_{p\times 1} = & -\frac{2}{T}x_i' M_{f_1^0} u_i^* + \frac{2}{T}x_i'(M_{f_1^0} - M_{\hat{f}_1})u_i^* - \frac{2}{T}x_i' M_{\hat{f}_1} f_1^0 \lambda_{1i}^0 + \frac{2}{T^2}x_i' M_{\hat{f}_1} x_i T(\hat{\alpha}_k - \alpha_k^0) \\
& + \left(\frac{2}{T^2}x_i' M_{\hat{f}_1} x_i + \frac{\lambda \hat{c}_{ki}}{\|\hat{\beta}_i - \hat{\alpha}_k\|} I_p\right) T(\hat{\beta}_i - \hat{\alpha}_k) + T\lambda \sum_{j=1, j\neq k}^{K} \hat{e}_{ij} \prod_{l=1, l\neq j}^{K} \|\hat{\beta}_i - \hat{\alpha}_l\| \\
& \equiv -\hat{A}_{1i} + \hat{A}_{2i} - \hat{A}_{3i} + \hat{A}_{4i} + \hat{A}_{5i} + \hat{A}_{6i}, \text{ say,}
\end{aligned}
$$

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

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

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

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

By (A.17)-(A.18) and the fact that $\max_{i\in G_k^0} \frac{1}{T^2} x_i' M_{\hat{f}_1} x_i \leq c d_T \rho_{\max}$ a.s., we have $P\left(\max_{i\in G_k^0}\left\|\hat{A}_{4i}\right\| > c d_T^2 T \psi_{NT}\right)$ $= o(N^{-1})$ and $P\left(\max_{i\in G_k^0}\left\|\hat{A}_{6i}\right\| > c\lambda T \Psi_{NT}\right) = o(N^{-1})$. By Lemma A.7(ii) and Lemma A.8(i),(iii), we can directly claim that

$$
\begin{aligned}
& P\left(\max_{i\in G_k^0} \|\hat{A}_{1i}\| > cT b_T \psi_{NT}\right) = o(N^{-1}), \\
& P\left(\max_{i\in G_k^0} \|\hat{A}_{2i}\| > c\left(T d_T \eta_{NT} + T^{1/2} d_T^{1/2} C_{NT}^{-1}\right)\left(\psi_{NT} + T^{-1/2}(\log T)^3\right)\right) = o(N^{-1}), \text{ and} \\
& P\left(\max_{i\in G_k^0} \|\hat{A}_{3i}\| > c N^{1/2q}(T d_T \eta_{NT} + T^{1/2} d_T^{1/2} C_{NT}^{-1})\right) = o(N^{-1}).
\end{aligned}
$$

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

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

Combing above results together, it follows that $P(\Xi_{k,NT}) = 1 - o(N^{-1})$, where

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

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

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

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

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

where the last inequality follows because that $T\lambda \gg Tb_T\psi_{NT}$ by Assumption 3.3(iv). Consequently, we can conclude that w.p.a.1 $\hat{\beta}_i - \hat{\alpha}_k$ must be in 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

$$\begin{aligned}
P(\cup_{k=1}^K \hat{E}_{kNT}) &\leq \sum_{k=1}^K P(\hat{E}_{kNT}) \leq \sum_{k=1}^K \sum_{i \in G_k^0} P(\hat{E}_{kNT,i}) \\
&\leq N \max_{1 \leq i \leq N} P(\|\hat{A}_{i1}\| \geq cT\lambda c_k^0/4) + o(1) \to 0 \quad \text{as} \quad (N, T) \to \infty.
\end{aligned}$$

This completes the proof of (i). Then the proof of (ii) directly follows SSP and thus omitted. ■

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

Lemma A.9 *Suppose that Assumptions 3.1-3.3 hold and $\sqrt{N} = o(T)$. Then for any $k = 1, ..., K$,*

(i) $\frac{1}{N_kT^2}\sum_{i \in \hat{G}_k} x_i' M_{\hat{f}_1} f_1^0 \lambda_{1i}^0 = \frac{1}{N_kT^2}\sum_{i \in \hat{G}_k} \frac{1}{N}\sum_{j=1}^N x_i' M_{\hat{f}_1} x_j a_{ij} \hat{b}_j - \frac{1}{N_kT^2}\sum_{i \in \hat{G}_k} \frac{1}{N}\sum_{j=1}^N a_{ij} x_i' M_{\hat{f}_1} u_j - \frac{1}{N_kT^2}\sum_{i \in \hat{G}_k} \frac{1}{N}\sum_{j=1}^N a_{ij} x_i' M_{\hat{f}_1} f_2^0 \lambda_{2j}^0 + o_P(N^{-1/2}T^{-1})$,

(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) $\frac{1}{N_kT^2}\sum_{i \in \hat{G}_k} \frac{1}{N}\sum_{j \in \hat{G}_l} x_i' M_{\hat{f}_1} x_j a_{ij} = \frac{1}{N_kT^2}\sum_{i \in G_k^0} \frac{1}{N}\sum_{j \in G_l^0} x_i' M_{f_1^0} x_j a_{ij} + o_P(1)$.

Lemma A.10 *Suppose that Assumptions 3.1-3.3 hold and $\sqrt{N} = o(T)$. Then*

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

(ii) $U_{kNT} = V_{kNT} + B_{kNT} + o_P(1)$,

(iii) $V_{NT} \xrightarrow{d} N(0, \Omega_0)$ conditional on $\mathcal{C}$,

where $\Omega_0 = \lim_{N,T \to \infty} \Omega_{NT}$.

Proof of Theorem 3.4. (i) To study of the oracle property of the C-Lasso estimator, we invoke the sub-differential calculus. A necessary and sufficient condition for $\{\hat{\beta}_i\}$ and $\{\hat{\alpha}_k\}$ to minimize the objective function in (2.8) is that for each $i = 1, ..., N$ (resp. $k = 1, ..., K$), $0_{p\times 1}$ belongs to the sub-differential of $Q^K_{NT,\lambda}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \hat{f}_1)$ 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 $\hat{G}_k$, then $\hat{\beta}_i - \hat{\alpha}_l \to \alpha^0_k - \alpha^0_l \neq 0$ 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.20) and (A.21) we have

$$\frac{2}{N_k T^2}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1}(y_i - x_i\hat{\alpha}_k) + \frac{\lambda}{N}\sum_{i\in\hat{G}_0} \hat{e}_{ik} \prod_{l=1,l\neq k}^{K} \|\hat{\beta}_i - \hat{\alpha}_l\| = 0_{p\times 1}. \tag{A.22}$$

Noting that $1\{i \in \hat{G}_k\} = 1\{i \in G^0_k\} + 1\{i \in \hat{G}_k \setminus G^0_k\} - 1\{i \in G^0_k \setminus \hat{G}_k\}$ and $y_i = x_i\alpha^0_k + f^0_1\lambda^0_{1i} + u^*_i$ when $i \in G^0_k$, we have

$$\begin{aligned}
\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k} x_i M_{\hat{f}_1} y_i =& \frac{1}{N_k T^2}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1} x_i\beta^0_i + \frac{1}{N_k T^2}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1} f^0_1\lambda^0_{1i} + \frac{1}{N_k T^2}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1} u^*_i \\
=& \frac{1}{N_k T^2}\sum_{i\in G^0_k} x_i' M_{\hat{f}_1} x_i\alpha^0_k + \frac{1}{N_k T^2}\sum_{i\in\hat{G}_k\setminus G^0_k} x_i' M_{\hat{f}_1} x_i\beta^0_i - \frac{1}{N_k T^2}\sum_{i\in G^0_k\setminus\hat{G}_k} x_i' M_{\hat{f}_1} x_i\alpha^0_k \\
&+ \frac{1}{N_k T^2}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1} f^0_1\lambda^0_{1i} + \frac{1}{N_k T^2}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1}(u_i + f^0_2\lambda^0_{2i}).
\end{aligned} \tag{A.23}$$

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

$$\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1} x_i(\hat{\alpha}_k-\alpha_k^0)=\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1} f_1^0\lambda_{1i}^0+\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1}\left(u_i+f_2^0\lambda_{2i}^0\right)$$
$$+\hat{C}_{1k}-\hat{C}_{2k}+\hat{C}_{3k}, \tag{A.24}$$

where $\hat{C}_{1k}=\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k\backslash G_k^0} x_i' M_{\hat{f}_1} x_i\beta_i^0$, $\hat{C}_{2k}=\frac{1}{N_k T^2}\sum_{i\in G_k^0\backslash\hat{G}_k} x_i' M_{\hat{f}_1} x_i\alpha_k^0$, 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 SSP, 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 $\frac{\sqrt{N}}{T}\to 0$

$$\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1} f_1^0\lambda_{1i}^0=\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N x_i' M_{\hat{f}_1} x_j a_{ij}\hat{b}_j-\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N a_{ij} x_i' M_{\hat{f}_1} u_j$$
$$-\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^N a_{ij} x_i' M_{\hat{f}_1} f_2^0\lambda_{2j}^0+o_P(N^{-1/2}T^{-1}). \tag{A.25}$$

In addition,

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

by Theorem 3.3. Let $\hat{Q}_{1NT}=$diag$\Big(\frac{1}{N_1 T^2}\sum_{i\in\hat{G}_1} x_i' M_{\hat{f}_1} x_i,\ldots,\frac{1}{N_K T^2}\sum_{i\in\hat{G}_K} x_i' M_{\hat{f}_1} x_i\Big)$ 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

$$\sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)=(\hat{Q}_{1NT}-\hat{Q}_{2NT})^{-1}\sqrt{D_N}\hat{U}_{NT}+o_P(1),$$

where the kth element of $\hat{U}_{NT}$ is

$$\hat{U}_{kNT}=\frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k} x_i' M_{\hat{f}_1}\left(\left(u_i+f_2^0\lambda_{2i}^0\right)-\frac{1}{N}\sum_{j=1}^N a_{ij}\left(u_j+f_2^0\lambda_{2j}^0\right)\right)$$

and $D_N=$diag$(\frac{N}{N_1},...,\frac{N}{N_K})$. 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 $\sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)=Q_{NT}^{-1}\sqrt{D_N}U_{NT}+o_P(1)$. 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,

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

where $V_{NT} = (V_{1NT}, ..., V_{KNT})$ and $B_{NT} = (B_{1NT}, ..., B_{KNT})$. This completes the proof of Theorem 3.4.

(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) and (A.28), we have $\sqrt{N}T\text{vec}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)-\sqrt{D_N}Q_{NT}^{-1}B_{NT} \xrightarrow{d} MN(0, \lim_{N\to\infty} D_N Q_0^{-1}\Omega_0 Q_0^{-1})$. ■

To prove Theorem 3.5, we need the following lemma.

Lemma A.11 *Suppose that Assumptions 3.1-3.3 hold and* $\sqrt{N} = o(T)$. *Then as* $(N,T) \to \infty$,

(i) $\frac{1}{\sqrt{T}}\|\hat{f}_1\hat{\lambda}_{1i} - f_1^0\lambda_{1i}^0\| = O_P(\sqrt{d_T T}\eta_{NT}) + O_P(C_{NT}^{-1})$,

(ii) $\frac{1}{\sqrt{T}}\|\hat{f}_2 - f_2^0 H_2\| = O_P(C_{NT}^{-1})$

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

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

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

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

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

It suffices to show the $\sqrt{N}T\text{vec}\left(\hat{\boldsymbol{\alpha}}_{\hat{G}}^{bc} - \boldsymbol{\alpha}^0\right) = \sqrt{D_N}Q_{NT}^{-1}V_{NT} + o_P(1)$ by (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 (i) 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 as used 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 $\tilde{B}_{kNT,1} = \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \hat{\Delta}_{21,i} - \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \hat{\varkappa}_{ts} \mathbf{1}\{s \leq t\} \hat{\Delta}_{21,i}$. It follows that

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

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 $\left| \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \hat{\varkappa}_{ts} \mathbf{1}\{s \leq t\} \right| \leq \frac{1}{T^3} \sum_{t=1}^T \sum_{s=1}^T \left\| \hat{f}_{1t} \right\| \left\| \hat{f}_{1s} \right\| = O_P(1)$ 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

$$
\begin{aligned}
B_{kNT,2} &= \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} E(x_i' | \mathcal{C}) M_{f_1^0} f_2^0 \left(\lambda_{2i}^0 - \frac{1}{N} \sum_{j=1}^N \lambda_{2j}^0 a_{ij} \right) \\
&= \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} E(x_i' | \mathcal{C}) f_2^0 \bar{\lambda}_{2i}^0 - \frac{1}{\sqrt{N_k}T} \sum_{i \in G_k^0} E(x_i' | \mathcal{C}) P_{f_1^0} f_2^0 \bar{\lambda}_{2i}^0 \equiv B_{kNT,21} - B_{kNT,22}, \text{ say,}
\end{aligned}
$$

where $\bar{\lambda}_{2i}^0 = \lambda_{2i}^0 - \frac{1}{N} \sum_{j=1}^N \lambda_{2j}^0 a_{ij}$. Let $\phi^{f_2, f_1 f_2} = (\phi^{f_2 f_1}(L), \phi^{f_2 f_2}(L)), \phi_i^{\varepsilon, f_1 f_2} = (\phi_i^{\varepsilon f_1}(L), \phi_i^{\varepsilon f_2}(L)) = (\phi^{\varepsilon f_1}(L), \phi^{\varepsilon f_2}(L))$, and $v_t^{f_1 f_2} = (v_t^{f_1 \prime}, v_t^{f_1 \prime})'$. 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

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

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

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

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

where we use the fact that $\phi_i^{\varepsilon, f_1 f_2}(1) \phi^{f_2, f_1 f_2}(1)' = S_2 \phi_i(1) \phi_i(1)' S'_4$ by construction and that $\sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r} \phi'_{i,l} = \phi_i(1) \phi_i(1)' - \sum_{l=0}^{\infty} \phi_{i,l+1} \phi'_{i,l} + \tilde{\phi}_{i,0} \phi_i(1)'$. Following the proof of Lemma A.7 in Huang et al. (2017), we can show that $\frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} S_2 R_{iT,l}^{f_2} S'_4 \bar{\lambda}^0_{2i} = o_P(1)$ for $l = 1, 2, ..., 6$ and $\frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} E(Q_{iT}^{f_2}) = 0$. It follows that $B_{kNT,21} = B^*_{kNT,21} + o_P(1) = \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \Delta_{24,i} \bar{\lambda}^0_{2i} + o_P(1)$. Analogously, we have $B_{kNT,22} = B^*_{kNT,22} + o_P(1)$, where $B^*_{kNT,22} = \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \varkappa_{ts} 1\{s \le t\} \times S_2 \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} \phi_{i,l+r} \phi'_{i,l} S'_4 \bar{\lambda}^0_{2i}$. Let $B^*_{kNT,2} = B^*_{kNT,21} - B^*_{kNT,22}$. Then

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

By Theorem 3.3 and using arguments as used in the proof of Lemma A.9(ii), we can readily show that $\hat{B}_{kNT,2} = \tilde{B}_{kNT,2} + o_P(1)$, where $\tilde{B}_{kNT,2} = \frac{1}{\sqrt{N_k}} \sum_{i \in G_k^0} \hat{\Delta}_{24,i} \hat{\bar{\lambda}}_{2i} - \frac{1}{\sqrt{N_k} T} \sum_{i \in G_k^0} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{\varkappa}_{ts} \mathbf{1}\{s \le t\} \hat{\Delta}_{24,i} \hat{\bar{\lambda}}_{2i}$. 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

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

by Lemma A.11(vii)-(viii). Consequently, $\hat{B}_{kNT,2}-B_{kNT,2}=o_P(1)$.

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

(ii) For the fully-modified post-Lasso estimators $\hat{\alpha}^{fm}_{G_k}$, we first consider the asymptotic distribution for the infeasible version of fully modified post-Lasso estimator $\tilde{\alpha}^{fm}_{G_k}$. Noting that $y_i^{+}=x_i\alpha_k^0+f_1^0\lambda^0_{1i}+f_2^0\lambda^0_{2i}+u_i^{+}$, by (A.24) and (A.25) and Theorem 3.3, we have

$$\begin{aligned}
\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_i'M_{\hat{f}_1}x_i(\tilde{\alpha}^{fm}_{G_k}-\alpha^0_k)=&\frac{1}{N_kT^2}\sum_{i\in G_k^0}x_i'M_{\hat{f}_1}\left(u_i^{+}+f_2^0\lambda^0_{2i}\right)+\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_i'M_{\hat{f}_1}f_1^0\lambda^0_{1i}\\
&-\frac{1}{\sqrt{N_k}T}B^{+}_{kNT,1}-\frac{1}{\sqrt{N_k}T}B_{kNT,2}+o_P(N^{-1/2}T^{-1}). \qquad \text{(A.29)}
\end{aligned}$$

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

$$\begin{aligned}
&\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}x_i'M_{\hat{f}_1}x_i(\tilde{\alpha}^{fm}_{G_k}-\alpha^0_k)-\frac{1}{N_kT^2}\sum_{i\in\hat{G}_k}\frac{1}{N}\sum_{j=1}^{N}x_i'M_{\hat{f}_1}x_ja_{ij}\hat{b}_j\\
=&\ \frac{1}{N_kT^2}\sum_{i\in G_k^0}x_i'M_{f_1^0}\left(u_i^{+}-\frac{1}{N}\sum_{j=1}^{N}u_j^{+}a_{ij}\right)+\frac{1}{N_kT^2}\sum_{i\in G_k^0}x_i'M_{f_1^0}f_2^0\left(\lambda^0_{2i}-\frac{1}{N}\sum_{j=1}^{N}\lambda^0_{2j}a_{ij}\right)\\
&-\frac{1}{\sqrt{N_k}T}B^{+}_{kNT,1}-\frac{1}{\sqrt{N_k}T}B_{kNT,2}+o_P(N^{-1/2}T^{-1})
\end{aligned}$$

By (A.26) and Lemma A.10 (i)-(iii), we have

$$\begin{aligned}
\sqrt{N}T\text{vec}(\tilde{\boldsymbol{\alpha}}^{fm}_{G}-\boldsymbol{\alpha}^0)&=(\hat{Q}_{1NT}-\hat{Q}_{2NT})^{-1}\sqrt{D_N}\left(\left(U^{u+}_{NT}+U^{f_2}_{NT}\right)-B^{+}_{NT,1}-B_{NT,2}\right)+o_P(1)\\
&=\sqrt{D_N}Q^{-1}_{NT}V^{+}_{NT}+o_P(1)
\end{aligned}$$

where

$$\begin{aligned}
U^{u+}_{k,NT}&=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_i'M_{f_1^0}\left(u_i^{+}-\frac{1}{N}\sum_{j=1}^{N}a_{ij}u_j^{+}\right),\\
V^{+}_{kNT,1}&=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}S^{\varepsilon}\phi_i^{\dagger}(1)\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\bar{\varkappa}_{ts}\left(V^{u\varepsilon}_{it}v^{u\varepsilon,+\prime}_{is}\right)-\left[\mathbf{1}\left\{t=s\right\}-\varkappa_{ts}\mathbf{1}\left\{s\le t\right\}\right]I_{1+p}\right\}\phi_i^{\dagger}(1)'S^{u\prime},\\
V^{+}_{kNT,2}&=\frac{1}{\sqrt{N_k}}\sum_{i=1}^{N}\left\{\frac{1}{T}E\left(x_i'|\mathcal{C}\right)\mathbf{1}\left\{i\in G_k^0\right\}-\frac{1}{N}\sum_{j\in G_k^0}a_{ij}\frac{1}{T}E(x_j'|\mathcal{C})\right\}M_{f_1^0}u_i^{+},
\end{aligned}$$

and $U_{k,NT}^{+}=U_{k,NT}^{u+}+U_{k,NT}^{f_2}$ and $V_{kNT}^{+}=V_{kNT,1}^{+}+V_{kNT,2}^{+}+V_{kNT,3}$ are the kth block-element 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), we can show that $V_{NT}^{+}\xrightarrow{d}N(0,\Omega_0^{+})$ condition on $\mathcal{C}$ where $\Omega_0^{+}=\lim_{N,T\to\infty}\Omega_{NT}^{+}$ and $\Omega_{NT}^{+}=\text{Var}\big(V_{NT}^{+}|\mathcal{C}\big)$. Then we have

$$\sqrt{N}T\text{vec}(\tilde{\boldsymbol{\alpha}}_G^{fm}-\boldsymbol{\alpha}^0)\xrightarrow{d}MN(0,\lim_{N\to\infty}D_NQ_0^{-1}\Omega_0^{+}Q_0^{-1}).$$

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

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

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) $\hat{B}_{NT,2}=B_{NT,2}+o_P(1)$. In the proof of bias-correction post-Lasso estimators, we have already prove (ii1) and (ii4). For (ii2), we can apply analogous arguments as used in the proof of Lemma A.11(v) to prove that $E_{\mathcal{C}}\left\|\frac{1}{\sqrt{N_k}}\sum_{i\in\hat{G}_k}(\hat{\Omega}_i-\Omega_i)\right\|=O_P(\frac{H}{T}+\frac{N}{H^{2q}})=o_P(1)$. Since $\Delta_{lm,i}^{+}=\Delta_{lm,i}-\Omega_{lm,i}\Omega_{mi}^{-1}\Delta_{m,i}$, this implies that $\left\|\frac{1}{\sqrt{N_k}}\sum_{i\in\hat{G}_k}(\hat{\Delta}_{21,i}^{+}-\Delta_{21,i}^{+})\right\|^2=o_P(1)$. The latter further implies that $\hat{B}_{NT,1}^{+}=B_{NT,1}^{+}+o_P(1)$. For (ii3) we can apply Theorem 3 to show that

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

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

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

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

Lemma A.12 *Suppose that Assumptions 3.1-3.3 and 3.5 hold. Then*

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

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

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

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

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

(i) For any $1 \leq r_1 \leq r_1^0$, $V_2(r_1, \hat{f}_1^{r_1}) - V_2(r_1, f_1^0 H_1^{r_1}) = O_P(\sqrt{T})$,

(ii) For any $1 \leq r_1 < r_1^0$, we have $\text{plim}_{(N,T)\to\infty} \inf d_T T^{-1}[V_2(r_1, f_1^0 H_1^{r_1}) - V_2(r_1, f_1^0)] = d_{r_1}$ *for some $d_{r_1} > 0$,*

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

where $V_2(r_1, f_1^0 H_1^{r_1})$ is defined analogously to $V_2(r_1, \hat{f}_1^{r_1})$ with $\hat{f}_1^{r_1}$ replaced by $f_1^0 H_1^{r_1}$, and $H_1^{r_1} = (N^{-1}\Lambda^{0\prime}\Lambda^0) \times (T^{-2} f^{0\prime} \hat{f}^{r_1})$.

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:

$$\begin{aligned} V_1(r, \hat{G}^r) - V_1(r^0, \hat{G}^{r^0}) &= [V_1(r, \hat{G}^r) - V(r, G^0 H^r)] + [V_1(r, G^0 H^r) - V_1(r^0, G^0 H^{r^0})] \\ &\quad + [V_1(r^0, G^0 H^{r^0}) - V_1(r^0, \hat{G}^{r^0})]. \end{aligned}$$

Lemma A.12(i) implies that the first and third terms on the right hand side of the last displayed equation are both $O_P(C_{NT}^{-1})$. Noting that $V_1(r^0, G^0 H^{r^0}) = V_1(r^0, G^0)$, the second term has a positive probability limit 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 $C_{NT}^2 \left(V_1(r, \hat{G}^r) - V_1(r^0, \hat{G}^{r^0})\right) = O_P(1)$ and $C_{NT}^2 (r - r^0) g_1(N,T) > C_{NT}^2 g_1(N,T) \to \infty$ 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:

$$\begin{aligned} V_2(r_1, \hat{f}_1^{r_1}) - V_2(r_1^0, \hat{f}_1^{r_1^0}) &= \left[V_2(r_1, \hat{f}^{r_1}) - V_2(r_1, f_1^0 H_1^{r_1})\right] + \left[V_2(r_1, f_1^0 H_1^{r_1}) - V_2(r_1^0, f_1^0 H_1^{r_1^0})\right] \\ &\quad + \left[V(r_1^0, f_1^0 H_1^{r_1^0}) - V(r_1^0, \hat{f}_1^{r_1^0})\right] \\ &\equiv DD_1 + DD_2 + DD_3, \text{ say.} \end{aligned}$$

By Lemma A.13), $DD_1 = O_P(T^{1/2})$, DD_2 is of exact probability order $O_P(T/\log\log T)$, and $DD_3 = O_P(1)$. 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 need the following lemma.

Lemma A.14 *Suppose that Assumptions 3.1-3.3 and 3.7 hold. Then*

$$\max_{K_0 \le K \le K_{\max}} \left| \hat{\sigma}^2_{G_{(K,\lambda)}} - \hat{\sigma}^2_{\hat{G}(K_0,\lambda)} \right| = O_P(\nu_{NT}^2)$$

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 can show that

$$\begin{aligned} IC_3(K_0,\lambda) &= \ln[V_3(K_0)] + pK_0 g_3(N,T) \\ &= \ln\frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k(K_0,\lambda)}\sum_{t=1}^{T}\left[y_{it} - \hat{\alpha}^{fm\prime}_{\hat{G}_k(K_0,\lambda)}x_{it} - \hat{\lambda}_{1i}(K_0,\lambda)'\hat{f}_{1t}(K_0,\lambda)\right]^2 + o(1) \xrightarrow{p} \ln(\sigma_0^2). \end{aligned}$$

We consider the cases of under- and over-fitted models separately. When $1 \le K < K_0$, we have

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

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,

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

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. Econometrica 71, 135-171.

Bai, J., 2004. Estimating cross-section common stochastic trends in nonstationary panel data. Journal of Econometrics 122, 137-183.

Bai, J., 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.

Baltagi, B.H., Pesaran, M.H., 2007. Heterogeneity and cross section dependence in panel data models: theory and applications introduction. Journal of Applied Econometrics 22, 229-232.

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.

Eaton, J., Kortum, S., 2002. Technology, geography, and trade. Econometrica 70, 1741-1779.

Engelbrecht, H.J., 1997. International R&D spillovers, human capital and productivity in OECD economies: An empirical investigation. European Economic Review 41, 1479-1488.

Huang, W., Jin, S., Su, L., 2017. Identifying latent group patterns in cointegrated panels. Working Paper, Singapore Management University.

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

Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899-1925.

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

Lucas, R.E., 1990. Why doesn't capital flow from rich to poor countries. American Economic Review 80, 92-96.

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

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

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

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

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

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, 1045-1055.

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

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

Solow, R.M., 1957. Technical change and the aggregate production function. The Review of Economics and Statistics 39, 312-320.

Su, L., Ju, G., 2017. Identifying latent group patterns in panel data models with interactive fixed effects. Journal of Econometrics, forthcoming.

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., 2016. Identifying latent structures in panel data. Econometrica 84, 2215-2264.

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

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