6-2020

Identifying latent grouped patterns in cointegrated panels

Wenxin HUANG

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

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

Citation

HUANG, Wenxin; JIN, Sainan; and SU, Liangjun. Identifying latent grouped patterns in cointegrated panels. (2020). *Econometric Theory*. 36, (3), 410-456. Research Collection School Of Economics.
Available at: https://ink.library.smu.edu.sg/soe_research/2381

Published in Econometric Theory, Volume 36, Issue 3, June 2020 , pp. 410-456
https://doi.org/10.1017/S0266466619000197

Identifying Latent Grouped Patterns in Cointegrated Panels*

Wenxin Huang[a], Sainan Jin[b], and Liangjun Su[b]

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

[b] School of Economics, Singapore Management University

November 20, 2018

Abstract

We consider a panel cointegration model with latent group structures that allows for heterogeneous long-run relationships across groups. We extend Su, Shi, and Phillips' (2016) classifier-Lasso (C-Lasso) method to the nonstationary panels and allow for the presence of endogeneity in both the stationary and nonstationary regressors in the model. In addition, we allow the dimension of the stationary regressors to diverge with the sample size. We show that we can identify the individuals' group membership and estimate the group-specific long-run cointegrated relationships simultaneously. We demonstrate the desirable property of uniform classification consistency and the oracle properties of both the C-Lasso estimators and their post-Lasso versions. The special case of dynamic penalized least squares is also studied. Simulations show superb finite sample performance in both classification and estimation. In an empirical application, we study the potential heterogeneous behavior in testing the validity of long-run purchasing power parity (PPP) hypothesis in the post-Bretton Woods period from 1975-2014 covering 99 countries. We identify two groups in the period 1975-1998 and three ones in the period 1999-2014. The results confirm that at least some countries favor the long-run PPP hypothesis in the post-Bretton Woods period.

JEL Classification: C13; C33; C51; F31.

Keywords: Classifier Lasso; Dynamic OLS; Heterogeneity; Latent group structure; Nonstationarity; Penalized least squares; Panel cointegration; Purchasing power parity

*The authors sincerely thank the co-editor Anna Mikusheva and two anonymous referees for their many constructive comments on the early version of the paper. They also thank Peter C. B. Phillips and Qiying Wang for discussions on the subject matter and the participants in the 2017 Asian Meeting of the Econometric Society at CUHK and the 2017 Advances in Econometrics Conference at SJTU 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

Recently there has been a growing literature on large dimensional panels with latent group structures; see Lin and Ng (2012), Bonhomme and Manresa (2015, BM hereafter), Sarafidis and Weber (2015), Ando and Bai (2016, 2017), Su, Shi, and Phillips (2016, SSP hereafter), Lu and Su (2017), Su and Ju (2018), Su, Wang, and Jin (2017), Wang, Phillips and Su (2018), among others. In comparison with other approaches to model unobserved heterogeneity in panel data models, an important advantage of the latent group structures is that it offers a flexible way to modeling unobserved heterogeneity while maintaining certain degree of parsimony. Two popular methods have been proposed to identify the unknown group structures. One is based on the celebrated K-means clustering algorithm, and the other is based on the C-Lasso. For example, Lin and Ng (2012) and Sarafidis and Weber (2015) consider a heterogeneous linear regression panel data model where the slope coefficients exhibit an unknown group structure whereas BM consider a homogeneous linear panel data model where the additive fixed effects exhibit group structure. Both groups of authors propose to apply the K-means clustering algorithm to achieve classification. Ando and Bai (2016, 2017) extend BM's approach to allow for group structures among the interactive fixed effects. Motivated by the sparse feature of the slope coefficients under latent group structures, SSP propose a novel variant of the Lasso procedure, i.e., classifier Lasso (C-Lasso), to achieve classification and estimation for both linear and nonlinear panel data models with or without endogeneity. Lu and Su (2017) propose a sequential testing procedure to determine the unknown number of groups; Su and Ju (2018) extend SSP's C-Lasso to panel data models with interactive fixed effects; Su et al. (2018) consider C-Lasso-based sieve estimation of time-varying panel data models with latent structures; Wang et al. (2018) extend the CARDS algorithm of Ke et al. (2015) to the panel data framework to identify the group structures of slope parameters.

In this paper, we consider identifying the latent group structures in nonstationary panels where some regressors are generated from an integrated process. Despite the vast and diverse literature on nonstationary panels, most studies focus on panel unit root or cointegration tests with or without cross-sectional dependence and the literature on formal cointegration analysis is relatively sparse. Depending on whether the cointegrating relationship is allowed to be heterogeneous, one may consider either homogeneous or heterogeneous cointegrating relations. For example, Phillips and Moon (1999) consider a general limit theory for both cases in large dimensional nonstationary panels; Groen and Kleibergen (2003) consider the likelihood-based cointegration analysis for heterogeneous and homogeneous panel vector error-correction models; Kao and Chiang (2000) consider both dynamic OLS (DOLS) and fully-modified OLS (FMOLS) estimation and inference in homogeneous cointegrated panels; Mark and Sul (2003) consider a panel DOLS in homogeneous nonstationary panels; Bai et al. (2009) study homogeneous panel cointegrations with global stochastic trends; Pedroni (2001a) considers FMOLS for heterogeneous cointegrated panels. So the long-run cointegrating relationships can be assumed to be either homogeneous or heterogeneous and we face a trade-off between assuming heterogeneous long-run relationships, which is surely robust and perhaps also close to the reality, and estimating a common or at least an average long-run relationship, which offers efficiency in estimation and inference if the underlying homogeneous assumption is correct.

Despite the different treatments on the long-run relationships, the short-run dynamics, the individual intercepts, or the individual time trends, if exist, are commonly assumed to be heterogeneous across individuals. In this paper, we shall maintain the individual heterogeneity assumption on the individual effects and short-run dynamics and take an intermediate approach to model the long-run relationship. We propose a panel cointegration model with latent group structures where the long-run relationships are homogeneous within a group and heterogeneous across different groups, and the short-run dynamics are allowed to be completely heterogeneous. The key issue is that the individual group membership is unknown and has to be estimated from the data together with the other parameters in the model. We extend SSP's C-Lasso method to the nonstationary panel framework. We consider the SSP's C-Lasso method rather than the K-means clustering algorithm for two reasons. First, the C-Lasso method has a computational advantage over the K-means clustering algorithm. As SSP argue, the C-Lasso problem can be transformed into a sequence of convex problems to be solved easily, while the K-means procedure is NP-hard and tends to be much more computationally involved than the C-Lasso method. Second, the asymptotic theory for the C-Lasso method is well understood for stationary panels. It is natural to extend the theory to nonstationary panels. We will propose a C-Lasso-based penalized least squares (PLS) procedure to identify the unknown group structures and estimate the other parameters in the model jointly.

Nevertheless, the extension of the asymptotic theory from stationary panels to nonstationary panels is technically challenging for two main reasons. First, there is a lack of certain uniform convergence results in the nonstationary panel literature. It is well known that both the K-means clustering algorithm and the C-Lasso method enjoy certain oracle properties, which means the resulting estimators are as asymptotically efficient as if the latent group structures were known. But the establishment of such oracle properties relies on the application of certain exponential inequalities that are available for weakly dependent data as in stationary panels but not available for strongly dependent data as in nonstationary panels. To achieve the extension, we first need to establish some uniform convergence results associated with the nonstationary I(1) variables. Second, we allow for both stationary and nonstationary regressors in our cointegration models. Even though the number of nonstationary regressors is assumed to be fixed, we allow the dimension of stationary regressors to grow with the sample size at a controllable rate. The latter is very important for us to explore the idea of DOLS and develop a panel dynamic PLS procedure. The growing dimension of the stationary regressors does not affect the convergence rate of the estimators of the long-run relationships, but it complicates the asymptotic analysis in several places.

We assume that the number of groups is known and study the asymptotic properties of the PLS estimators. We first establish the preliminary rates of convergence for the coefficient estimators and show that, as expected, the long-run parameters can be estimated consistently at a faster rate than the short-run parameters. Given these preliminary consistency rates, we establish the uniform classification consistency of the C-Lasso method, which essentially means that all parameters within a group can be classified into the same group with probability approaching 1 (w.p.a.1), and all individuals that are classified into the same group indeed belong to the same group w.p.a.1. Such a uniform classification consistency lays down the foundation for the study of the asymptotic distributions of the PLS estimators. We show that both the C-Lasso estimators of the long-run parameters and their post-Lasso versions enjoy the asymptotic oracle properties and then derive their asymptotic distributions under the joint limit theory.[1] We show that such a presence of endogeneity in both nonstationary and stationary regressors does not cause the inconsistency of the long-run parameter estimators but does yield an asymptotic bias in the estimators of both the short-run and long-run parameters. To remove the asymptotic bias in the estimation of the long-run parameters, we explore the idea of DOLS in the time series framework and propose a C-Lasso-based dynamic PLS procedure. When the number of groups is unknown, we propose an information criterion to determine the number of groups. Simulations show superb finite sample performance of the information criterion and C-Lasso-based PLS procedure.

In an empirical application, we apply our method to re-examine the validity of long-run PPP in the post-Bretton Woods period from 1975-2014 for a panel of 99 countries. Due to the establishment of the European Union in 1999, we consider two subperiods, namely, 1975-1998 and 1999-2014. Then we estimate the long-run group-specific relationships by the dynamic PLS method. In general, we observe heterogeneous behavior on the long-run relation between the nominal exchange rate and aggregate price ratio. We find two groups in the 1975-1998 subsample, with one group of countries in favor of the validity of the PPP hypothesis and the other group against the PPP hypothesis. In the 1999-2014 subsample, we identify three groups and find significant evidence in favor of the long-run PPP hypothesis in one group. There are more countries in this group in favor of the validity of the long-run PPP hypothesis in this period. We explain these results by the “Revived Bretton Woods system” (also called Bretton Woods II in the literature) from 2000, see Dooley et al. (2004). These results confirm the belief that at least some selected group of countries obey the long-run PPP rule in the post-Bretton Woods period.

The rest of this paper is organized as follows. We introduce the cointegrated panel data model with latent group structures and propose a C-Lasso-based PLS estimation procedure in Section 2. Section 3 introduces the main assumptions for our asymptotic analysis. We study the asymptotic properties of the PLS estimators. Section 5 reports Monte Carlo simulation results. Section 6 applies the dynamic PLS method to testing the long-run PPP hypothesis. Section 7 concludes. We relegate the proofs of the main results to Appendix A. The online supplement contains of the proofs of technical lemmas, the section on the determination of the number of groups, the section on the practical implementation of the C-Lasso procedure, and some additional simulation and application results.

NOTATION. For any real matrix A, we write the transpose A', the Frobenius norm $||A||$, the spectral norm $||A||_{\text{sp}}$, and the Moore-Penrose inverse as A^{+}. When A is symmetric, we use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote its largest and smallest eigenvalues, respectively. I_a and $\mathbf{0}_{a\times b}$ denote the $a\times a$ identity matrix and $a\times b$ matrix of zeros, and $\mathbf{1}\{\cdot\}$ is the usual indicator function. The operator $\xrightarrow{P}$ denotes convergence in probability, $\Rightarrow$ weak convergence, a.s. almost surely, and plim probability limit. We use $(N,T)\to\infty$ to signify that N and T pass jointly to infinity.

2 Model and Estimation

In this section we introduce the panel cointegration model with latent group structures and then propose a C-Lasso-based penalized least squares method to estimate the model.

2.1 Panel cointegration model with latent group structures

The dependent variable y_{it} is measured for individuals $i = 1, 2, ..., N$ over time $t = 1, 2, ..., T$. We suppose that the nonstationary I(1) variables y_{it} and $x_{1,it}$ are generated according to the following heterogeneous panel cointegration model

$$\begin{cases} y_{it} = \mu_i + \beta_{1,i}' x_{1,it} + \beta_{2,i}' x_{2,it} + u_{it} \\ x_{1,it} = x_{1,it-1} + \varepsilon_{1,it}, \end{cases} , \tag{2.1}$$

where μ_i is the unobserved individual fixed effects (FE), $x_{1,it}$ is a $p_1 \times 1$ vector of nonstationary regressors of order one (I(1) process) for all i, $x_{2,it}$ is a $p_2 \times 1$ vector of stationary regressors (I(0) process) for all i, u_{it} is the idiosyncratic error term with mean zero and finite long-run variance, $\varepsilon_{1,it}$ also has zero mean and finite long-run variance, and $\beta_{1,i}$ and $\beta_{2,i}$ are $p_1 \times 1$ and $p_2 \times 1$ vectors of slope coefficients, respectively.

We assume that p_1 is fixed but allow p_2 to diverge to infinity at certain rate. The latter is very important because we will extend our theory to the panel DOLS framework. In this case, the first equation in (2.1) becomes

$$y_{it} = \mu_i + \beta_{1,i}' x_{1,it} + \sum_{j=-\bar{p}_2}^{\bar{p}_2} \gamma_{i,j}' \Delta x_{1,i,t+j} + v_{it}^{\dagger}, \tag{2.2}$$

where $\Delta x_{1,it} = x_{1,it} - x_{1,i,t-1}$, $x_{2,it}$ only contains the lags and leads of $\Delta x_{1,it}$: $x_{2,it} = (\Delta x_{1,i,t-\bar{p}_2}', ..., \Delta x_{1,i,t+\bar{p}_2}')'$, $\beta_{2,i} = (\gamma_{i,-\bar{p}_2}', ..., \gamma_{i,\bar{p}_2}')'$, $p_2 = (2\bar{p}_2 + 1)p_1$, $\bar{p}_2$ is divergent with T, and $v_{it}^{\dagger}$ is the new error term that typically contains some approximation errors.

In the literature on nonstationary panels, $\beta_{1,i}$, which stands for the long-run cointegrating relationship, can be either homogeneous or heterogeneous, whereas $\beta_{2,i}$, which represents the short-run dynamics, is allowed to be heterogeneous across all individuals in almost all studies. In fact, there is a large literature that imposes a common long-run relationship and allows for individual-specific short-run parameters. For example, in a cross-country study it is possible for different countries or regions to have different dynamics of adjustments towards an equilibrium due to their historical and cultural differences, but they could all converge to the same economic equilibrium in the very long run due to forces of arbitrage and interconnections through international trade and cultural exchanges. See also the concluding remark in Pesaran, Shin and Smith (1999). In this paper we maintain the heterogeneity assumption on $\beta_{2,i}$'s but follow the lead of SSP and assume that $\beta_{1,i}$'s are heterogeneous across groups and homogeneous within a group.

Specifically, we allow the true values of $\beta_{1,i}$, denoted as $\beta_{1,i}^0$, to follow a grouped pattern of the general form

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

where $\alpha_j^0 \neq \alpha_k^0$ for any $j \neq k$, $\cup_{k=1}^K G_k^0 = \{1, 2, \ldots N\}$, and $G_k^0 \cap G_j^0 = \varnothing$ for any $j \neq k$. For now, we assume that the number of groups, K, is known and fixed. But we will study the determination of K in Section C of the online supplement. Let $\boldsymbol{\alpha} \equiv (\alpha_1, \ldots, \alpha_K)$, $\boldsymbol{\beta_1} \equiv (\beta_{1,1}, \ldots, \beta_{1,N})$, and $\boldsymbol{\beta_2} \equiv (\beta_{2,1}, \ldots, \beta_{2,N})$. We denote their true values as $\boldsymbol{\alpha^0}$, $\boldsymbol{\beta_1^0}$, and $\boldsymbol{\beta_2^0}$, respectively. We also use $\beta_{2,i}^0$ and α_k^0 to denote the true coefficients of $\beta_{2,i}$ and α_k. We use $N_k \equiv \#G_k^0$ to denote the cardinality of the set G_k^0. We are interested in identifying each individual's group membership and estimating the long-run cointegrating group-specific coefficients, α_k, $k = 1, ..., K$.

By allowing for the latent group structures for the long-run parameters, we can achieve a right balance between parameter parsimony and model misspecification. Note that the key parameters of interest in nonstationary panels are the coefficients of the nonstationary regressors as they characterize the long-run equilibrium relationship between the dependent variables and the nonstationary regressors. If we allow these parameters to be individual-specific, we can run individual time-series regressions to estimate them but their estimators will have non-standard limiting distributions and can converge to the true values only at the rate T. On the other hand, if we assume these coefficients are common across all individuals, we will have a convenient yet restrictive assumption that facilitates estimation and inference and meanwhile a very large chance of model misspecification. The latent group structure adopted in this paper is an intermediate approach. It allows for certain degree of heterogeneity in the long-run parameters and helps to overcome some problems associated with nonstationary time series analysis too. In particular, under some conditions we can identify the group structure and estimate the group-specific long-run parameters at the rate $\sqrt{N}T$. Moreover, these long-run parameter estimators are asymptotically normal.

Even though we focus only on the linear cointegrating model in this paper, the theory that we are developing is quite different from that in SSP for three main reasons. First, the presence of nonstationary regressors substantially complicates the asymptotic analysis. In particular, we need to establish some uniform convergence rates that are not available in the nonstationary panel literature. Second, the increasing dimension of the stationary regressors in the model also complicates the issue. Third, we allow for endogeneity in both $x_{1,it}$ and $x_{2,it}$. In the time series framework, it is well known that the endogeneity of either the I(1) or I(0) regressors does not cause the inconsistency of the OLS estimator of the long-run relationship. In particular, the estimators of the coefficients of I(1) regressors are still consistent at the rate T despite the fact that it exhibits an endogeneity bias of order $O(1/T)$ (see, e.g., Proposition 19.2 in Hamilton (1994)). We will show that a similar phenomenon occurs in the panel setup.

2.2 Penalized least squares estimation

Without imposing the latent group structures in (2.3), we can estimate $\beta_{1,i}$ and $\beta_{2,i}$ in (2.1) by using the fixed effects estimator. In this case, we consider the within-group transformation

$$\tilde{y}_{it} = \beta_{1,i}' \tilde{x}_{1,it} + \beta_{2,i}' \tilde{x}_{2,it} + \tilde{u}_{it}, \tag{2.4}$$

or in vector form

$$\tilde{y}_i = \tilde{x}_{1,i} \beta_{1,i} + \tilde{x}_{2,i} \beta_{2,i} + \tilde{u}_i, \tag{2.5}$$

where $\tilde{y}_i = (\tilde{y}_{i1}, ..., \tilde{y}_{iT})'$, $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it}$, and $\tilde{x}_{1,it}$, $\tilde{x}_{2,it}$, $\tilde{u}_{it}$, $\bar{x}_{1,i}$, $\bar{x}_{2,i}$, $\bar{u}_i$, $\tilde{x}_{1,i}$, $\tilde{x}_{2,i}$, and $\tilde{u}_i$ are analogously defined. The FE estimators $\tilde{\beta}_{1,i}$ and $\tilde{\beta}_{2,i}$ are obtained as the minimizers of the following least squares criterion function

$$Q_{NT}(\boldsymbol{\beta_1}, \boldsymbol{\beta}_2) = \frac{1}{NT^2}\sum_{i=1}^{N}\left\|\tilde{y}_i - \tilde{x}_{1,i}\beta_{1,i} - \tilde{x}_{2,i}\beta_{2,i}\right\|^2 = \frac{1}{NT^2}\sum_{i=1}^{N}\left\|\tilde{y}_i - \tilde{x}_i\beta_i\right\|^2, \tag{2.6}$$

where $\beta_i = (\beta'_{1,i}, \beta'_{2,i})'$ and $\tilde{x}_i = (\tilde{x}_{1,i}, \tilde{x}_{2,i})$ has a typical row $\tilde{x}'_{it} = (\tilde{x}'_{1,it}, \tilde{x}'_{2,it})$. Let $\tilde{\beta}_i = (\tilde{\beta}'_{1,i}, \tilde{\beta}'_{2,i})'$. Then $\tilde{\beta}_i = (\tilde{x}'_i\tilde{x}_i)^{-1}\tilde{x}'_i\tilde{y}_i$ for each i. As mentioned above, the estimators $\tilde{\beta}_{1,i}$ of the long-run parameters $\beta_{1,i}$ are consistent despite the possible presence of endogeneity bias, but they converge to the true values only at the rate T with nonstandard limiting distributions. When $\beta^0_{1,i}$'s exhibit the latent group structure in (2.3), it is possible to pull over the observations from both the time series and cross-sectional dimensions to obtain more efficient estimators of the group-specific long-run parameters. We will show that these new estimators, possibly after bias correction, converge to the true values at the rate $\sqrt{N}T$ and are asymptotically normally distributed.

To explore the latent group structure of $\beta_{1,i}$'s in (2.3), we propose to estimate $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, and $\boldsymbol{\alpha}$ by minimizing the following C-Lasso-based penalized least squares (PLS) criterion function

$$Q^K_{NT,\lambda}(\boldsymbol{\beta_1}, \boldsymbol{\beta}_2, \boldsymbol{\alpha}) = Q_{NT}(\boldsymbol{\beta_1}, \boldsymbol{\beta}_2) + \frac{\lambda}{N}\sum_{i=1}^{N}(\tilde{\sigma}_i)^{2-K}\prod_{k=1}^{K}\left\|\hat{Q}_{1i}(\beta_{1,i} - \alpha_k)\right\|, \tag{2.7}$$

where $\lambda = \lambda(N,T)$ is a tuning parameter, $\tilde{\sigma}^2_i = \frac{1}{T}\sum_{t=1}^{T}(\tilde{y}_{it} - \tilde{\beta}'_i\tilde{x}_{it})^2$, and $\hat{Q}_{1i} = \frac{1}{T^2}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}'_{1,it}$. When $\tilde{\sigma}_i$ and $\hat{Q}_{1i}$ are replaced by 1 and I_{p_1}, respectively, the penalty term in (2.7) reduces to that in SSP. Here, we introduce these two terms into the penalty to ensure the scale-invariant property of the penalized estimators.

As SSP remark, the second term on the right hand side of (2.7) is a penalty term that takes a novel mixed additive-multiplicative form. It has N additive terms, each of which takes a multiplicative form as the product of K separate penalties. The multiplicative component is needed because for each i we do not know a priori to which point $\beta_{1,i}$ should shrink and must allow $\beta_{1,i}$ to shrink to any one of the K unknown values $\alpha_1, ..., \alpha_K$. Each of the K penalty terms in the multiplicative expression permits $\beta_{1,i}$ to shrink to a particular unknown group-specific parameter vector α_k. The summation component is needed because we need to pull information from all N cross-sectional units in order to identify the group-specific parameters and the individual-specific parameters jointly. Note that the tuning parameter λ is used to control the size of the penalty. A too small value of λ means that the penalty term won't play an important role so that many of $\beta_{1,i}$'s would not shrink toward one of the group-specific values in $\{a_1, ..., \alpha_K\}$; a too large value of λ will force all $\beta_{1,i}$'s to shrink toward one of the group-specific values in $\{a_1, ..., \alpha_K\}$, which may result in misclassification. In theory, we require that λ tend to zero at an appropriate rate as $(N,T) \to \infty$. The exact conditions on λ are stated in Assumption A.3(iv) below.

Minimizing the objective function in (2.7) yields the C-Lasso-based PLS estimates $\hat{\boldsymbol{\beta}}_1$, $\hat{\boldsymbol{\beta}}_2$, and $\hat{\boldsymbol{\alpha}}$. Let $\hat{\beta}_{1,i}$ and $\hat{\alpha}_k$ denote the i^{th} and k^{th} columns of $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\alpha}}$, respectively, i.e., $\hat{\boldsymbol{\beta}}_1 \equiv (\hat{\beta}_{1,1}, ..., \hat{\beta}_{1,N})$ and $\hat{\boldsymbol{\alpha}} \equiv (\hat{\alpha}_1, ..., \hat{\alpha}_K)$. We will study the asymptotic properties of the C-Lasso estimators below.

3 Notations and Assumptions

In this section, we spell out the main notations and assumptions that are needed for the study of the asymptotic properties of our estimators.

Since we include the fixed effects μ_i in (2.1) and assume covariance-stationarity of $x_{2,it}$, we assume without loss of generality that $x_{2,it}$ has zero mean.[2] Let $\varepsilon_{it} = \left(u_{it}, \varepsilon_{1,it}', \varepsilon_{2,it}'\right)'$ where $\varepsilon_{2,it} = x_{2,it}$. The long-run covariance matrix of $\{\varepsilon_{it}\}$ is given by

$$\Omega_i = \sum_{j=-\infty}^{\infty} E(\varepsilon_{ij}\varepsilon_{i0}') = \begin{pmatrix} \Omega_{00,i} & \Omega_{01,i} & \Omega_{02,i} \\ \Omega_{10,i} & \Omega_{11,i} & \Omega_{12,i} \\ \Omega_{20,i} & \Omega_{21,i} & \Omega_{22,i} \end{pmatrix}, \tag{3.1}$$

where, e.g., $\Omega_{00,i} = \sum_{j=-\infty}^{\infty} E(u_{ij}u_{i0}')$, $\Omega_{01,i} = \sum_{j=-\infty}^{\infty} E(u_{ij}\varepsilon_{1,i0}')$, and $\Omega_{02,i} = \sum_{j=-\infty}^{\infty} E(u_{ij}\varepsilon_{2,i0}')$. Following the literature on nonstationary panels, we will make the following decomposition

$$\Omega_i = \Sigma_i + \Lambda_i + \Lambda_i',$$

where $\Sigma_i = E\left(\varepsilon_{it}\varepsilon_{it}'\right)$ denotes the short-run variance of $\{\varepsilon_{it}\}$ and $\Lambda_i = \sum_{j=1}^{\infty} E(\varepsilon_{ij}\varepsilon_{i0}')$. We partition Σ_i and Λ_i conformably with ε_{it} and Ω_i :

$$\Sigma_i = \begin{pmatrix} \Sigma_{00,i} & \Sigma_{01,i} & \Sigma_{02,i} \\ \Sigma_{10,i} & \Sigma_{11,i} & \Sigma_{12,i} \\ \Sigma_{20,i} & \Sigma_{21,i} & \Sigma_{22,i} \end{pmatrix} \text{ and } \Lambda_i = \begin{pmatrix} \Lambda_{00,i} & \Lambda_{01,i} & \Lambda_{02,i} \\ \Lambda_{10,i} & \Lambda_{11,i} & \Lambda_{12,i} \\ \Lambda_{20,i} & \Lambda_{21,i} & \Lambda_{22,i} \end{pmatrix}. \tag{3.2}$$

Let $\Delta_i = \Sigma_i + \Lambda_i$ denote the one-sided long-run covariance of $\{\varepsilon_{it}\}$. Let $p = 1 + p_1 + p_2$ denote the dimension of ε_{it}. Let S_0, S_1, and S_2 denote respectively the $1 \times p$, $p_1 \times p$, and $p_2 \times p$ selection matrices such that $S_0\varepsilon_{it} = u_{it}$ and $S_\ell\varepsilon_{it} = \varepsilon_{\ell,it}$ for $\ell = 1, 2$. In the dynamic DOLS example in (2.2), $\varepsilon_{2,it}$ contains $\varepsilon_{1,it}$. For this reason, we do not require that Ω_i be of full rank. But we will assume that $\Omega_{11,i}$ and $\Sigma_{22,i}$ are of full rank. As in the time series literature, the full rank of $\Omega_{11,i}$ rules out potential cointegration among the variables in $x_{1,it}$ when $p_1 > 1$ and that of $\Sigma_{22,i}$ rules out collinearity among the variables in $x_{2,it}$ when $p_2 > 1$. For more precise conditions, see Assumption A.2 below.

Let $\max_i = \max_{1\leq i\leq N}$ and $\max_t = \max_{1\leq t\leq T}$ unless otherwise stated. Define $\min_i$ and $\min_t$ analogously. We make the following assumptions.

Assumption A.1 (i) For each i, $\{\varepsilon_{it}, t \geq 0\}$ is a linear process such that

$$\varepsilon_{it} = \psi_i(L)\, e_{it} = \sum_{j=0}^{\infty} \psi_{ij} e_{i,t-j},$$

where $\{e_{it}\}$ is an independent process with zero mean and variance-covariance matrix I_p. Each element of e_{it} has finite $2(q+\epsilon)$ moments that are bounded uniformly in (i,t), where $q > 4$ and ϵ is an arbitrarily small positive number.

(ii) $\max_i \sum_{j=0}^{\infty} j^2 \left\|S\psi_{ij}\right\| < \infty$ for any selection matrix S that selects any finite (non-divergent) number of rows in ψ_{ij}.

(iii) For each i, $\{\varepsilon_{it}, t \geq 0\}$ is a strong mixing process with mixing coefficients $\alpha_i(t)$ satisfying $\max_i \alpha_i(\tau) \leq c_\alpha \rho^\tau$ for some $c_\alpha < \infty$ and $\rho \in (0,1)$.

(iv) $\{\varepsilon_{it}, t \geq 0\}$ are independent across i.

Assumption A.2 (i) There exists a constant $\underline{c}_{11}$ such that $\liminf_{T\to\infty} \lambda_{\min}\left(\frac{b_T}{T^2}\sum_{t=1}^T \tilde{x}_{1,it}\tilde{x}_{1,it}'\right) \geq \underline{c}_{11} > 0$ almost surely (a.s.), where $b_T = \log\log T$.

(ii) There exist constants $\underline{c}_{\Omega_{11}}$ and $\bar{c}_{\Omega_{11}}$ such that $0 < \underline{c}_{\Omega_{11}} \leq \min_i \lambda_{\min}(\Omega_{11,i}) \leq \max_i \lambda_{\max}(\Omega_{11,i}) \leq \bar{c}_{\Omega_{11}} < \infty$.

(iii) There exist constants $\underline{c}_{22}$ and $\bar{c}_{22}$ such that $0 < \underline{c}_{22} \leq \min_i \lambda_{\min}(\Sigma_{22,i}) \leq \max_i \lambda_{\max}(\Sigma_{22,i}) \leq \bar{c}_{22} < \infty$.

(iv) Let $\Sigma_{0.2,i}^* = \Sigma_{00,i} - \Sigma_{02,i}\Sigma_{22,i}^{-1}\Sigma_{20,i}$. There exist constants $\underline{c}_{00}$ and $\bar{c}_{00}$ such that $0 < \underline{c}_{00} \leq \min_i \Sigma_{0.2,i}^* \leq \max_i \Sigma_{00,i} \leq \bar{c}_{00} < \infty$.

Assumption A.3 (i) For each $k = 1, ..., K$, $N_k/N \to \tau_k \in (0,1)$ 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 $(N,T) \to \infty$, $N/T^2 \to c_1 \in [0,\infty)$, $T/N^2 \to c_2 \in [0,\infty)$, and $p_2^3 T^{-1}(\log T)^6 \to c_3 \in [0,\infty)$.

(iv) As $(N,T) \to \infty$, $\lambda b_T^2 \to 0$, $\lambda T N^{-\frac{1}{q}} b_T^{-(K+1)}/\log T \to \infty$, $b_T^{K+1} N^{1/q} T^{-1} \log T \to 0$, $b_T N^{2/q} T^{-1/2}/\log T \to 0$, and $b_T p_2^{1/2} N^{1/q} T^{-1/2} \log T = O(1)$.

Assumption A.1(i)-(ii) imposes that the innovation process $\{\varepsilon_{it}\}$ is a linear process that exhibits certain moment and summability conditions. The linearity assumption is weak because of the celebrated Wold decomposition theorem which says that any mean zero covariance stationary process with absolutely summable auto-covariances can be represented as an MA(∞) linear process. See, e.g., Proposition 4.1 in Hamilton (1994). The summability conditions are used to ensure the validity of certain functional central limit theorem. When p_2 is fixed, the selection matrix S is not needed. In our asymptotic analysis, we will frequently call upon the Beveridge and Nelson (1981, BN) decomposition:

$$\varepsilon_{it} = \psi_i(1) e_{it} + \breve{e}_{i,t-1} - \breve{e}_{it}, \tag{3.3}$$

where $\psi_i(1) = \sum_{j=0}^\infty \psi_{ij}$, $\breve{e}_{it} = \sum_{j=0}^\infty \breve{\psi}_{ij} e_{i,t-j}$, and $\breve{\psi}_{ij} = \sum_{k=j+1}^\infty \psi_{ik}$. Following Phillips and Solo (1992, p.989), Assumption A.1(i)-(ii) ensures that

$$\max_i \max_t E\left\|S\breve{e}_{it}\right\|^{2q} < \infty$$

for any selection matrix S such that $S\breve{e}_{it}$ selects only a fixed number of elements in $\breve{e}_{it}$. For example, $S = (S_0', S_1')'$ selects the first $1 + p_1$ elements $\breve{e}_{it}$ that corresponds to $\left(u_{it}, \varepsilon_{1,it}'\right)'$. Assumption A.1(iii) assumes that $\{\varepsilon_{it}, t \geq 0\}$ is a strong mixing process for the convenience of using a Bernstein-type exponential inequality that is available for strong mixing processes. It is satisfied by many well-known processes such as linear stationary autoregressive moving average (ARMA) processes with continuously distributed errors and a large class of processes implied by numerous nonlinear models, including bilinear, nonlinear autoregressive (NLAR), and autoregressive conditional heteroskedastic (ARCH) type models. See Davidson (1994, Ch. 14), Doukhan (1994) and Fan and Yao (2008, Ch. 2.6) for more examples of strong mixing processes. In particular, Davidson (1994, Ch. 14.4) provides some sufficient conditions to verify that a linear process of the type in Assumption A.1(i) is strong mixing, and Andrews (1984) provides an example of autoregressive process that is not strong mixing. The geometric mixing rate can be relaxed to being algebraic with a little bit more involved notation in the proofs. Here we follow SSP and assume the geometric mixing rate condition for simplicity. By White (2001, Theorem 7.18), Assumption A.1(i)-(iii) is far more sufficient to ensure the functional central limit theorem (FCLT) holds for $\{S\varepsilon_{it},\ t \geq 0\}$ for each i provided its long-run variance-covariance matrix is positive definite. Assumption A.1(iv) imposes cross-sectional independence, as was done in the early literature on panel cointegration analyses (see, e.g., Phillips and Moon, 1999; Kao and Chiang, 2000; Mark and Sul, 2003). We do not relax such an assumption in this paper because even under this restrictive assumption, the rigorous asymptotic analysis is already extremely involved.

Assumption A.2(i) requires that $\hat{Q}_{1i} \equiv \frac{1}{T^2}\sum_{t=1}^{T} \tilde{x}_{1,it}\tilde{x}_{1,it}'$ is well behaved uniformly in i. For each i, we can readily apply the results in Park and Phillips (1988, 1989) and show that

$$\hat{Q}_{1i} \Rightarrow \int_0^1 \tilde{B}_{1,i}(r)\,\tilde{B}_{1,i}(r)'\,dr, \tag{3.4}$$

where $\tilde{B}_{1,i} = B_{1,i} - \int_0^1 B_{1,i}(r)\,dr$ and $B_{1,i}$ is a p_1-dimensional Brownian motion with covariance $\Omega_{11,i}$. In this case, as long as $\Omega_{11,i}$ is positive definite, we can ensure that $\hat{Q}_{1i}$ is asymptotically nonsingular for each i. For our asymptotic analysis, we require that both the maximum and minimum eigenvalues of $\hat{Q}_{1i}$ be well behaved uniformly in i. For the maximum eigenvalue, we can call upon the usual law of iterated logarithm (LIL) and show that

$$\limsup_{T\to\infty} \lambda_{\max}(\hat{Q}_{1i}/(2\log\log T)) < \left(\frac{1}{2}+\epsilon\right)\bar{c}_{\Omega_{11}}\,\text{a.s.}, \tag{3.5}$$

where ϵ is an arbitrarily small positive number and $\bar{c}_{\Omega_{11}}$ is a constant defined in Assumption A.2(ii). For the minimum eigenvalue, a sufficient condition for the Assumption A.2(i) to hold is that there exist some positive constants $c_i \in (0,1)$ with $\min_{1\leq i\leq N} c_i \geq c_0 > 0$ such that

$$x_{1,i}'M_{\iota_T}x_{1i} \geq c_i x_{1,i}'x_{1i} \geq c_0 x_{1,i}'x_{1i} \text{ for all } i = 1, ..., N, \tag{3.6}$$

where $M_{\iota_T} = I_T - \iota_T(\iota_T'\iota_T)^{-1}\iota_T' = I_T - \frac{1}{T}\iota_T\iota_T'$ and ι_T is a $T\times 1$ vector of ones. To see the meaning of the above condition, we observe that for any nonrandom vector $\omega \in \mathbb{R}^{p_1}$ such that $\|\omega\| = 1$ and $x_{1,i}\omega$ is nonzero,

$$\begin{aligned}\omega'\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}_{1,it}'\omega &= \omega'x_{1,i}'M_{\iota_T}x_{1i}\omega = \omega'x_{1,i}'x_{1i}\omega - (x_{1,i}\omega)'\iota_T(\iota_T'\iota_T)^{-1}\iota_T'x_{1i}\omega \\ &= \omega'x_{1,i}'x_{1i}\omega\left[1 - \frac{(x_{1,i}\omega)'\iota_T(\iota_T'\iota_T)^{-1}\iota_T'x_{1i}\omega}{(x_{1,i}\omega)'x_{1i}\omega}\right].\end{aligned}$$

So the condition in (3.6) requires the existence of a $c_0 \in (0,1)$ such that $\frac{(x_{1,i}\omega)'\iota_T(\iota_T'\iota_T)^{-1}\iota_T'x_{1i}\omega}{(x_{1,i}\omega)'x_{1i}\omega} \leq 1 - c_0$, which essentially requires that $x_{1i}\omega$ is not ι_T a.s. uniformly in i. Then by the "other" or

Chung-type LIL (see, e.g., Donsker and Varadhan (1977), Lai and Wei (1982a, p.163), Lai and Wei (1982b, p.364), Phillips (1996, p.799), and Bai (2004, pp.140-141)) and the Cramér-Wold device, we have $\liminf_{T\to\infty}\lambda_{\min}\left(\frac{b_T}{T^2}x_{1,i}'x_{1,i}'\right)\geq \underline{c}_1$ for some $\underline{c}_1>0$. This, in conjunction with (3.6), implies that Assumption A.2(i) would be satisfied with $\underline{c}_{11}=\underline{c}_1 c_0$.

Assumption A.2(ii)-(iii) imposes some conditions on the eigenvalues of nonstochastic square matrices. They imply that $\Omega_{11,i}$ and $\Sigma_{22,i}$ have full rank uniformly in i. Assumption A.2(iv) is imposed to ensure nondegenerate limiting distributions. Given Assumption A.2(iii), it implicitly implies that $\Sigma_{20,i}'\Sigma_{20,i}$ is bounded away from the infinity and thus restricts the degree of endogeneity in the stationary regressors.

Assumption A.3(i)-(ii) is commonly assumed in the panel literature with latent group structures; see, e.g., Bonhomme and Manresa (2015), Ando and Bai (2016), SSP, Lu and Su (2017), and Su and Ju (2018). In particular, Assumption A.3(ii) requires the separability of the group-specific parameters. Assumption A.3(iii) imposes conditions on N, T, and p_2. It requires that N should not diverge to infinity at a rate faster than T^2 or slower than $T^{1/2}$. Note that we do not require $N=o(T)$ as in most studies on nonstationary panels under the joint limit theory (see, e.g., Phillips and Moon, 1999; Bai and Ng, 2010). The last condition in Assumption A.3(iii) is analogous to the condition $p_2^3T^{-1}=o(1)$ in the time series framework (e.g., Saikkonen, 1991). Assumption A.3(iv) looks quite complicated but can be simplified to a great deal in the special case where N and T pass to infinity at the same rate as in many macro applications. In this case, noting that $q>4$ as stated in Assumption A.1(i) and $p_2^3T^{-1}=o(1)$ implied by Assumption A.3(iii), we can replace Assumption A.3(iv) by the following assumption:

Assumption A.3(iv*) As $(N,T)\to\infty$, $\lambda b_T^2\to 0$, and $\lambda T^{1-\frac{1}{q}}b_T^{-(K+1)}/\log T\to\infty$.
Then we can find a large range of values for λ satisfying Assumption A.3(iv*). It is sufficient to have

$$\lambda\propto T^{-\alpha}\text{ for }\alpha\in\left(0,\ \frac{q-1}{q}\right).$$

When q is sufficiently large (e.g., the tails of the error terms decay exponentially fast), the upper bound for α is arbitrarily close to 1. If we only require $q>4$, then it is fine to choose $\lambda\propto T^{-3/4}$.

4 Asymptotic Properties

In this section, we first find the preliminary rates of convergence for the coefficient estimators and prove classification consistency. Then we study the oracle properties of C-Lasso estimators and their post-Lasso versions. The special case of panel dynamic PLS is also considered, and an extension to models with incidental time trends is also considered.

4.1 Preliminary rates of convergence

Let $\beta_i^*=(\beta_{1,i}^{0\prime},\beta_{2,i}^{*\prime})'$, where $\beta_{2,i}^*=\beta_{2,i}^0+\Sigma_{22,i}^{-1}\Sigma_{20,i}$. The following theorem establishes the preliminary rates of consistency for both $\hat{\beta}_i$ and $\hat{\alpha}_k$.

Theorem 4.1 *Suppose that Assumptions A.1-A.3 hold. Then*
(i) $||\hat{\beta}_{1,i}-\beta_{1,i}^0||=O_P(T^{-1}+\lambda)$ *and* $||\hat{\beta}_{2,i}-\beta_{2,i}^*||=O_P(p_2^{1/2}(T^{-1/2}+\lambda))$ *for* $i=1,...,N$,

(ii) $\frac{1}{N}\sum_{i=1}^{N}||\hat{\beta}_{1,i}-\beta_{1,i}^{0}||^{2}=O_{P}(b_{T}^{2}T^{-2})$ *and* $\frac{1}{N}\sum_{i=1}^{N}||\hat{\beta}_{2,i}-\beta_{2,i}^{*}||^{2}=O_{P}(p_{2}T^{-1})$,

(iii) $(\hat{\alpha}_{(1)},...,\hat{\alpha}_{(K)})-(\alpha_{1}^{0},...,\alpha_{K}^{0})=O_{P}(b_{T}T^{-1})$ *where* $(\hat{\alpha}_{(1)},...,\hat{\alpha}_{(K)})$ *is a suitable permutation of* $(\hat{\alpha}_{1},...,\hat{\alpha}_{K})$.

Theorems 4.1(i) and (ii) establish the pointwise and mean square convergence of $\hat{\beta}_{i}=(\hat{\beta}_{1i}^{\prime},\hat{\beta}_{2,i}^{\prime})^{\prime}$, respectively; Theorem 4.1(iii) indicates that $\hat{\alpha}_{1},...,\hat{\alpha}_{K}$ consistently estimate the true group-specific coefficients, $\alpha_{1}^{0},...,\alpha_{K}^{0}$, subject to a suitable permutation. We summarize some interesting findings. First, despite the presence of endogeneity in both the nonstationary and stationary regressors, we can estimate the true coefficients $(\beta_{1,i}^{0})$ of the nonstationary regressors consistently. Second, when $\Sigma_{20,i}$ is nonzero, we cannot estimate the true coefficients $(\beta_{2,i}^{0})$ of the stationary regressors consistently. Instead, $\hat{\beta}_{2,i}$ is consistent with the pseudo true value $\beta_{2,i}^{*}=\beta_{2,i}^{0}+\Sigma_{22,i}^{-1}\Sigma_{20,i}$, where $\Sigma_{22,i}^{-1}\Sigma_{20,i}$ signifies the endogeneity bias. Third, the effect of increasing dimension (p_{2}) appears in the rates of convergence for $\hat{\beta}_{2,i}$ but not in those for $\hat{\beta}_{1,i}$. Apparently, $\hat{\beta}_{1,i}$'s converge to their true values faster than $\hat{\beta}_{2,i}$'s to their pseudo-true values. Fourth, as in SSP, the pointwise convergence of $\hat{\beta}_{i}$ depends on λ while the mean square convergence of $\{\hat{\beta}_{1,i},\hat{\beta}_{2,i}\}$ and the convergence of $\hat{\alpha}_{k}$'s do not. As we have shown in the proof of the above theorem, the convergence of $\hat{\alpha}_{k}$ only depends on the mean square convergence of $\{\hat{\beta}_{1,i}\}$.

For notational simplicity, hereafter we will write $\hat{\alpha}_{(k)}$ as $\hat{\alpha}_{k}$. We define the estimated groups

$$\hat{G}_{k}=\{i\in\{1,2,...,N\}:\hat{\beta}_{1,i}=\hat{\alpha}_{k}\}\text{ for }k=1,...,K. \tag{4.1}$$

To study the classification consistency, we need to establish the uniform consistency of $\hat{\beta}_{1,i}$ and $\hat{\beta}_{2,i}$. This is reported in the next theorem.

Theorem 4.2 *Suppose that Assumptions A.1-A.3 hold. Then for any fixed* $c>0$,

(i) $P(\max_{1\leq i\leq N}||\hat{\beta}_{1,i}-\beta_{1,i}^{0}||\geq cb_{T}a_{1NT})=o\left(N^{-1}\right)$,

(ii) $P(\max_{1\leq i\leq N}||\hat{\beta}_{2,i}-\beta_{2,i}^{*}||\geq cp_{2}^{1/2}a_{2NT})=o\left(N^{-1}\right)$,

where $a_{1NT}=T^{-1}N^{1/q}\left(\log T\right)^{(1+\epsilon)/2}$ *for some arbitrarily small* $\epsilon>0$, *and* $a_{2NT}=T^{-1/2}\left(\log T\right)^{3}$.

The uniform convergence rate of $\hat{\beta}_{1,i}$ is not affected by p_{2} but is slower than the time series convergence rate T^{-1}. The higher q is (which means the higher order moments for the error terms), the closer a_{1NT} is to T^{-1}. When the error terms have exponentially decaying tails as assumed in Bonhomme and Manresa (2015), we can make a_{1NT} arbitrarily close to T^{-1} subject to a logarithm factor.

4.2 Classification consistency

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

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

where $i=1,...,N$ and $k=1,...,K$. Let $\hat{E}_{kNT}=\cup_{i\in\hat{G}_{k}}\hat{E}_{kNTi}$ and $\hat{F}_{kNT}=\cup_{i\in\hat{G}_{k}}\hat{F}_{kNTi}$. $\hat{E}_{kNT}$ denotes the error event of not classifying an element of G_{k}^{0} into the estimated group $\hat{G}_{k}$; and $\hat{F}_{kNT}$ denotes the error event of classifying an element that does not belong to G_{k}^{0} into the estimated group $\hat{G}_{k}$. Following SSP, we say that a classification method is *individually consistent* if $P(\hat{E}_{kNT,i})\to0$ and $P(\hat{F}_{kNT,i})\to0$

as $(N,T)\to\infty$ for each $i\in G_k^0$ and $k=1,...,K$, and 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 4.3 *Suppose that Assumptions A.1-A.3 hold. Then as* $(N,T)\to\infty$
(i) $P(\cup_{k=1}^K \hat{E}_{kNT})\leq \sum_{k=1}^K P(\hat{E}_{kNT})\to 0,$
(ii) $P(\cup_{k=1}^K \hat{F}_{kNT})\leq \sum_{k=1}^K P(\hat{F}_{kNT})\to 0.$

Theorem 4.3 implies that all individuals within certain group, say G_k^0, can be simultaneously correctly classified into the same group (denoted as $\hat{G}_k$) w.p.a.1. Conversely, all individuals that are classified into the same group, say $\hat{G}_k$, simultaneously correctly belong to the same group (G_k^0) w.p.a.1. The result implies that in large samples, we can virtually take the estimated group as the true group. In particular, let $\hat{N}_k=\#\hat{G}_k$. One can easily show that $P(\hat{G}_k=G_k^0)\to 1$ so that $P(\hat{N}_k=N_k)\to 1$.

Note that Theorem 4.3 is an asymptotic result and it does not ensure that all individuals can be classified into one of the estimated groups in finite samples. Indeed, when T is not large, some units might not be classified if λ is not sufficiently big and we stick to the classification rule in (4.1). In practice, we classify $i\in\hat{G}_k$ if $\hat{\beta}_i=\hat{\alpha}_k$ for some $k=1,...,K$, and $i\in\hat{G}_l$ for some $l=1,...,K$ if $||\hat{\beta}_i-\hat{\alpha}_l||=\min\{||\hat{\beta}_i-\hat{\alpha}_1||,...,||\hat{\beta}_i-\hat{\alpha}_K||\}$ and $\sum_{k=1}^K \mathbf{1}\{\hat{\beta}_i=\hat{\alpha}_k\}=0$. Since Theorem 4.3 ensures $\sum_{k=1}^K P(\hat{\beta}_i=\hat{\alpha}_k)\to 1$ as $(N,T)\to\infty$ uniformly in i, we can ignore such a modification in large samples in subsequent theoretical analyses and restrict our attention to the classification rule in (4.1) to avoid confusion.

4.3 Oracle properties and post-Lasso estimators

To study the oracle property of the C-Lasso-based PLS estimators, we add some notations:

$$\begin{aligned}
\mathbb{Q}_{(k)} &\equiv \lim_{N_k\to\infty}\frac{1}{6N_k}\sum_{i\in G_k^0} S_1\psi_i(1)\psi_i(1)' S_1' = \lim_{N_k\to\infty}\frac{1}{6N_k}\sum_{i\in G_k^0}\Omega_{11,i},\\
\mathbb{B}_{k,NT} &\equiv \mathbb{B}_{1k,NT}+\mathbb{B}_{2k,NT},\\
\mathbb{B}_{1k,NT} &= \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0} S_1\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\psi_{i,s+r}\psi_{i,s}' s_i,\\
\mathbb{B}_{2k,NT} &= \frac{-1}{\sqrt{N_k}}\frac{T+1}{2T}\sum_{i\in G_k^0} S_1\psi_i(1)\psi_i(1)' s_i,\\
\mathbb{V}_{(k)} &\equiv \lim_{N_k\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0}\left(\frac{1}{6}s_i'\Omega_i s_i S_1\Omega_i S_1' - \frac{1}{12}\left(s_i'\Omega_i S_1'\otimes S_1\Omega_i s_i\right)K_{p_1,1}\right),\\
\mathbb{V}_{22,i} &= \left(\Sigma_{22,i}^{-1}J_{1,i}\otimes J_{2,i}\right)V_i^0\left(J_{1,i}'\Sigma_{22,i}^{-1}\otimes J_{2,i}'\right),
\end{aligned}$$

where $s_i=S_0'-S_2'\Sigma_{22,i}^{-1}\Sigma_{20,i}$, $J_{1,i}=\left(\mathbf{0}_{p_2\times 1},\mathbf{0}_{p_2\times p_1},I_{p_2}\right)$, $J_{2,i}=\left(1,\mathbf{0}_{1\times p_1},-\Sigma_{20,i}'\Sigma_{22,i}^{-1}\right)$, $K_{p_1,1}$ is the $p_1\times p_1$ commutation matrix,[3] and $V_i^0=\lim_{T\to\infty}\mathrm{Var}(T^{-1/2}\sum_{t=1}^T \mathrm{vec}(\varepsilon_{it}\varepsilon_{it}'-\Sigma_i))$.

The following theorem reports the asymptotic properties of $\hat{\alpha}_k$ and $\hat{\beta}_{2,i}$.

Theorem 4.4 *Suppose that Assumptions A.1-A.3 hold. Let* $\mathbb{S}_2$ *denote an* $l \times p_2$ *selection matrix such that* $\mathbb{S}_2 \beta_{2,i}$ *selects only* l *elements in* $\beta_{2,i}$, *where* l *is a fixed integer that does not grow with* (N, T). *Then*

(i) $\sqrt{N_k} T(\hat{\alpha}_k - \alpha_k^0) - \mathbb{Q}_{(k)}^{-1} \mathbb{B}_{k,NT} \Rightarrow N(0, \mathbb{Q}_{(k)}^{-1} \mathbb{V}_{(k)} \mathbb{Q}_{(k)}^{-1})$ *as* $(N, T) \to \infty$ *for* $k = 1, ..., K$,

(ii) $\sqrt{T} \mathbb{S}_2 \left(\hat{\beta}_{2,i} - \beta_{2,i}^* \right) \Rightarrow N\left(0, \mathbb{S}_2 \mathbb{V}_{22,i} \mathbb{S}_2'\right)$ *as* $T \to \infty$ *for each* $i = 1, ..., N$.

To understand the above results, we consider the case where the group membership is known. In this case, the oracle estimators of α_k and $\beta_{2,i}$ are respectively given by

$$\hat{\alpha}_k^{\text{oracle}} = \left(\sum_{i \in G_k^0} \tilde{x}_{1,i}' M_{2,i} \tilde{x}_{1,i} \right)^{-1} \sum_{i \in G_k^0} \tilde{x}_{1,i}' M_{2,i} \tilde{y}_i \quad \text{for } k = 1, ..., K,$$
$$\hat{\beta}_{2,i}^{\text{oracle}} = \left(\tilde{x}_{2,i}' \tilde{x}_{2,i} \right)^{-1} \tilde{x}_{2,i}' (\tilde{y}_i - \tilde{x}_{1,i} \hat{\alpha}_k^{\text{oracle}}) \text{ for } i \in G_k^0,$$

where $M_{2,i} = I_T - \tilde{x}_{2,i} \left(\tilde{x}_{2,i}' \tilde{x}_{2,i} \right)^{-1} \tilde{x}_{2,i}'$. One can readily show that $\hat{\alpha}_k$ shares the same asymptotic bias and variance as $\hat{\alpha}_k^{\text{oracle}}$, and similarly, $\hat{\beta}_{2,i}$ shares the same asymptotic bias and variance as $\hat{\beta}_{2,i}^{\text{oracle}}$. In this case, we say that our C-Lasso estimators $\hat{\alpha}_k$ and $\hat{\beta}_{2,i}$ are asymptotically *oracle efficient*. As expected, $\hat{\alpha}_k$ may have an asymptotic bias of order $O\left(T^{-1}\right)$ in the presence of endogeneity, but it converges to its true value at the usual $\sqrt{N_k} T$-rate after bias correction.

A close examination of the asymptotic bias of $\hat{\alpha}_k$ indicates that $\mathbb{B}_{k,NT}$ can be rewritten as the summation of two terms, $\mathbb{B}_{1k,NT}$ and $\mathbb{B}_{2k,NT}$. $\mathbb{B}_{1k,NT}$ appears even without the within-group transformation as in Phillips and Moon (1999); $\mathbb{B}_{2k,NT}$ is simply due to the time-demeaning operator. As mentioned above, we allow for both sources of endogeneity. When $\Sigma_{20,i} \neq 0$, we have a contemporaneous correlation between the stationary regressor $x_{2,it}$ and the error term u_{it} in the cointegrating regression model. When $S_1 \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \psi_{i,s+r} \psi_{i,s}' S_0' \neq 0$ or $S_1 \psi_i(1) \psi_i(1)' S_0' \neq 0$, we allow the correlation of u_{it} with some leads or current values of $\varepsilon_{1,it}$. When both types of correlations vanish, $\mathbb{B}_{k,NT} = 0$, so that there is no endogeneity bias in this special case.

Note that we specify a selection matrix $\mathbb{S}_2$ in Theorem 4.4 that is not needed if p_2 is fixed. When p_2 diverges to infinity, we cannot derive the asymptotic normality of $\hat{\beta}_{2,i}$ directly. Instead, we follow the literature on inferences with a diverging number of parameters (e.g., Fan and Peng, 2004; Lam and Fan, 2008; Lu and Su, 2015; Qian and Su, 2016a and 2016b) and prove the asymptotic normality for any arbitrary finite linear combinations of elements of $\hat{\beta}_{2,i}$.

Given the estimated groups, $\{\hat{G}_k, k = 1, ..., K\}$, we can obtain the post-Lasso estimators of α_k and $\beta_{2,i}$ as

$$\hat{\alpha}_k^{\text{post}} = \left(\sum_{i \in \hat{G}_k} \tilde{x}_{1,i}' M_{2,i} \tilde{x}_{1,i} \right)^{-1} \sum_{i \in \hat{G}_k} \tilde{x}_{1,i}' M_{2,i} \tilde{y}_i \quad \text{for } k = 1, ..., K,$$
$$\hat{\beta}_{2,i}^{\text{post}} = \left(\tilde{x}_{2,i}' \tilde{x}_{2,i} \right)^{-1} \tilde{x}_{2,i}' (\tilde{y}_i - \tilde{x}_{1,i} \hat{\alpha}_k^{\text{post}}) \text{ for } i \in \hat{G}_k.$$

We show in the proof of Theorem 4.4 that the C-Lasso estimators $\hat{\alpha}_k$ and $\hat{\beta}_{2,i}$ are asymptotically equivalent to their post-Lasso versions $\hat{\alpha}_k^{\text{post}}$ and $\hat{\beta}_{2,i}^{\text{post}}$, respectively. The following theorem reports the limiting distributions of $\hat{\alpha}_k^{\text{post}}$ and $\hat{\beta}_{2,i}^{\text{post}}$.

Theorem 4.5 *Suppose that Assumptions A.1-A.3 hold. Then*

(i) $\sqrt{N_k}T(\hat{\alpha}_k^{post} - \alpha_k^0) - \mathbb{Q}_{(k)}^{-1}\mathbb{B}_{k,NT} \Rightarrow N(0, \mathbb{Q}_{(k)}^{-1}\mathbb{V}_{(k)}\mathbb{Q}_{(k)}^{-1})$ *for* $k = 1, ..., K$,

(ii) $\sqrt{T}\mathbb{S}_2(\hat{\beta}_{2,i}^{post} - \beta_{2,i}^*) \Rightarrow N(0, \mathbb{S}_2\mathbb{V}_{22,i}\mathbb{S}_2')$ *for* $i = 1, ..., N$,

where $\mathbb{Q}_{(k)}$, $\mathbb{B}_{k,NT}$, $\mathbb{V}_{(k)}$, *and* $\mathbb{V}_{22,i}$ *are as defined before Theorem 4.4 and* $\mathbb{S}_2$ *is as defined in Theorem 4.4.*

Given the asymptotic results in Theorems 4.4 and 4.5, one can make inference as if the true group membership is known. Despite the asymptotic equivalence of the C-Lasso estimators and their post-Lasso versions, it is well known that the post-Lasso estimators tend to have a smaller finite sample bias in simulations and are thus recommended for practical uses. Despite this, in order to make inference on the long-run cointegrating relationship, we have to remove the bias. There are two standard ways to correct the endogeneity bias in the time series literature, namely, fully-modified least squares (FMOLS) and dynamic OLS (DOLS). In principle, one can consider either the panel DOLS or panel FMOLS method as in Kao and Chiang (2000) and Mark and Sul (2003) based on the estimated groups. The procedure is standard and thus omitted. Alternatively, we can consider the use of the DOLS idea in the C-Lasso procedure, which yields the C-Lasso-based dynamic PLS (DPLS) estimation procedure. See the next subsection for details.

4.4 The case of dynamic PLS

In this subsection, we focus on the dynamic PLS (DPLS) estimation of the panel cointegration model with latent group structures. We show that the results in Theorems 4.4 and 4.5 continue to be valid with little modification.

For notational clarity, we now assume that $\{y_{it}, x_{1it}\}$ are generated by

$$\begin{cases} y_{it} = \mu_i + \beta_{1,i}' x_{1,it} + u_{it} \\ x_{1,it} = x_{1,it-1} + \varepsilon_{1,it} \end{cases}, \tag{4.2}$$

where μ_i, u_{it}, and $\varepsilon_{1,it}$ are defined as before, and $\beta_{1,i}$'s exhibit the latent structures in (2.3).

To consider the panel DPLS estimation method, we follow Saikkonen (1991) and Stock and Watson (1993) and make the following assumption.

Assumption A.4. (i) The process $\{u_{it}\}$ can be projected on to $\{\varepsilon_{1,it}\}$ as follows: $u_{it} = \sum_{j=-\infty}^{\infty} \gamma_{ij}\varepsilon_{1,i,t+j} + v_{it}$, where $\sum_{j=-\infty}^{\infty} \|\gamma_{ij}\| < \infty$, v_{it} is an error term with mean zero and finite $2q^{\text{th}}$ moment where $q > 4$, and v_{it} and $\varepsilon_{1,it}$ are uncorrelated for all lags and leads.

(ii) As $(N, T) \to \infty$, there exists $a > 1/2$ such that $T^a \sum_{|j|>\bar{p}_2} ||\gamma_{ij}|| \to 0$, $N^{1/2}T^{1/2-a} \to 0$, and $N^{1/2}\bar{p}_2 T^{-a} \to 0$.

Assumption A.4(i) ensures that $E(\varepsilon_{1,it}v_{it+k}) = 0$ for $k = 0, \pm 1, \pm 2, \ldots$ and Assumption A.4(ii) ensures that the values of $\varepsilon_{1,it}$ in the very remote past and future have only negligible impacts on u_{it}. Therefore, we can truncate the leads and lags and run the following DOLS regression model

$$y_{it} = \mu_i + \beta_{1,i}' x_{1,it} + \sum_{j=-\bar{p}_2}^{\bar{p}_2} \gamma_{ij}' \Delta x_{1,i,t+j} + v_{it}^{\dagger}, \tag{4.3}$$

where $v_{it}^{\dagger}=v_{it}^{a}+v_{it}$, and $v_{it}^{a}=\sum_{|j|>\bar{p}_2}\gamma_{ij}'\Delta x_{1,i,t+j}$ signifies the approximation/truncation error. Let $x_{2,it}$ denote a collection of the lags and leads of $\Delta x_{1,it}:x_{2,it}=(\Delta x_{1,i,t-\bar{p}_2}',...,\Delta x_{1,i,t+\bar{p}_2}')'$. Let $\beta_{2,i}=(\gamma_{i,-\bar{p}_2}',...,\gamma_{i\bar{p}_2}')'$ and $p_2=(2\bar{p}_2+1)p_1$. After the within-group transformation, we have the following model

$$\tilde{y}_{it}=\beta_{1,i}'\tilde{x}_{1,it}+\sum_{j=-\bar{p}_2}^{\bar{p}_2}\gamma_{ij}'\widetilde{\Delta x}_{1,i,t+j}+\tilde{v}_{it}^{\dagger}=\beta_{1,i}'\tilde{x}_{1,it}+\beta_{2,i}'\tilde{x}_{2,it}+\tilde{v}_{it}^{\dagger}, \tag{4.4}$$

where $\tilde{v}_{it}^{\dagger}=v_{it}^{\dagger}-\bar{v}_i^{\dagger}$, $\bar{v}_i^{\dagger}=\frac{1}{T-2\bar{p}_2}\sum_{t=\bar{p}_2+1}^{T-\bar{p}_2}v_{it}^{\dagger}$, and $\tilde{y}_{it}$ and $\tilde{x}_{2,i}$ are analogously defined.

As before, we can continue to consider the C-Lasso-based PLS regression and obtain the Lasso estimators of $\beta_{1,i}$, $\beta_{2,i}$, and α_k. We denote these estimators as $\hat{\beta}_{1,i}^{D}$, $\hat{\beta}_{2,i}^{D}$, and $\hat{\alpha}_k^{D}$, where D abbreviates DPLS. Let $\hat{G}_k$ denote the estimated group as before. The corresponding post-Lasso estimators of α_k and $\beta_{2,i}$ take the form

$$\begin{aligned}\hat{\alpha}_k^{D,\text{ post}} &= \left(\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i}\right)^{-1}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{y}_i \text{ for } k=1,...,K,\\ \hat{\beta}_{2,i}^{D,\text{ post}} &= \left(\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\tilde{x}_{2,i}'\left(\tilde{y}_i-\tilde{x}_{1,i}\hat{\alpha}_k^{D,\text{ post}}\right) \text{ for } i\in\hat{G}_k,\end{aligned}$$

where $\tilde{x}_{1,i}=(\tilde{x}_{1,i,\bar{p}_2+1},...,\tilde{x}_{1,i,T-\bar{p}_2})'$, $\tilde{y}_i$ and $\tilde{x}_{2,i}$ are analogously defined, and $M_{2,i}=I_{T-2\bar{p}_2}-\tilde{x}_{2,i}\left(\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\tilde{x}_{2,i}'$.

The following theorem shows the asymptotic properties of $\hat{\alpha}_k^{D,\text{ post}}$ and $\hat{\beta}_{2,i}^{D,\text{ post}}$ where expressions for both $\mathbb{V}_{(k)}$ and $\mathbb{V}_{22,i}$ are greatly simplified.

Theorem 4.6 *Suppose that Assumptions A.1, A.2(i)-(iii) and A.3-A.4 hold. Suppose that there exists a constant* $\underline{c}_{00}$ *such that* $\min_{1\le i\le N}\Sigma_{00,i}\ge\underline{c}_{00}>0$. *Then*

(i) $\sqrt{N_k}T(\hat{\alpha}_k^{D,post}-\alpha_k^0)\Rightarrow N(0,\,\mathbb{Q}_{(k)}^{-1}\mathbb{V}_{(k)}^{\dagger}\mathbb{Q}_{(k)}^{-1})$ *for* $k=1,...,K$,

(ii) $\sqrt{T}\mathbb{S}_2(\hat{\beta}_{2,i}^{D,post}-\beta_{2,i}^0)\Rightarrow N\left(0,\mathbb{S}_2\mathbb{V}_{22,i}\mathbb{S}_2'\right)$ *for* $i=1,...,N$,

where $\mathbb{Q}_{(k)}\equiv\lim_{N_k\to\infty}\frac{1}{6N_k}\sum_{i\in G_k^0}\Omega_{11,i}$, $\mathbb{V}_{(k)}^{\dagger}\equiv\lim_{N_k\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0}\frac{1}{6}\Omega_{00,i}^{\dagger}\Omega_{11,i}$, $\Omega_{00,i}^{\dagger}=\Omega_{00,i}-\Omega_{01,i}\Omega_{11,i}^{-1}\Omega_{10,i}$ *and* $\mathbb{V}_{22,i}=\Sigma_{22,i}^{-1}V_{22,i}\Sigma_{22,i}^{-1}$ *with* $V_{22,i}=\lim_{T\to\infty}Var(T^{-1/2}\sum_{t=1}^{T}x_{2,it}u_{it})$.

Even though we have not stated in the above theorem, $\hat{\alpha}_k^{D}$ and $\hat{\beta}_{2,i}^{D}$ are asymptotically equivalent to $\hat{\alpha}_k^{D,\text{post}}$ and $\hat{\beta}_{2,i}^{D,\text{post}}$, respectively. Thus both C-Lasso-based DPLS estimators and their post-Lasso versions have asymptotic normal distributions and are asymptotically oracle efficient. One can readily construct the usual t-statistics and F-statistics to make inference. For example, to make inference on the group-specific long-run cointegrating relationship, we can estimate $\mathbb{Q}_{(k)}$ and $\mathbb{V}_{(k)}^{\dagger}$, respectively by[4]

$$\hat{\mathbb{Q}}_{(k)}=\frac{1}{\hat{N}_kT^2}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i}\text{ and }\hat{\mathbb{V}}_{(k)}^{\dagger}\equiv\frac{1}{\hat{N}_k}\sum_{i\in\hat{G}_k}\frac{1}{6}\hat{\Omega}_{00,i}^{\dagger}\hat{\Omega}_{11,i},$$

where $\hat{\Omega}_{00,i}^{\dagger}=\hat{\Omega}_{00,i}-\hat{\Omega}_{01,i}\hat{\Omega}_{11,i}^{-1}\hat{\Omega}_{10,i}$, and $\hat{\Omega}_{00,i}$, $\hat{\Omega}_{11,i}$, $\hat{\Omega}_{01,i}$ and $\hat{\Omega}_{10,i}$ denote the HAC estimator of the long-run variance-covariance components $\Omega_{00,i}$, $\Omega_{11,i}$, $\Omega_{01,i}$ and $\Omega_{10,i}$ in Ω_i. In practice, we recommend the use of $\hat{\alpha}_k^{D,\text{post}}$ and $\hat{\beta}_{2,i}^{D,\text{post}}$ because the post-Lasso estimators typically outperform the C-Lasso ones.

4.5 The case of incidental time trends

Our panel cointegration model can be extended to models with both individual fixed effects and incidental time trends:

$$\begin{cases} y_{it} = \mu_i + \rho_i t + \beta'_{1,i} x_{1,it} + \beta'_{2,i} x_{2,it} + u_{it} \\ x_{1,it} = \mu_{1,i} + x_{1,it-1} + \varepsilon_{1,it}, \end{cases}, \tag{4.5}$$

where $i = 1, ..., N$ and $t = 1, ..., T$, $\rho_i t$ denotes the incidental time trend, we allow for the presence of an intercept term $\mu_{1,i}$ in the I(1) process $\{x_{1,it}\}$, and the other variables are defined as before. The above model reduces to model (2.1) when $\rho_i = 0$ and $\mu_{1i} = 0$ for all i. In that case, we have employed the within-group *demeaned* transformation to eliminate the individual fixed effects. In the presence of both individual effects and incidental time trends in the above model, we can similarly employ the within-group *detrended* data to eliminate both individual fixed effects and incidental time trends. Specifically, we consider the detrended model:

$$\dot{y}_{it} = \beta'_{1,i} \dot{x}_{1,it} + \beta'_{2,i} \dot{x}_{2,it} + \dot{u}_{it}, \tag{4.6}$$

where $\dot{y}_{it} = y_{it} - \sum_{t=1}^{T} y_{it} g'_t \left(\sum_{s=1}^{T} g_s g'_s \right)^{-1} g_t$ with $g_t = (1, t)'$, and $\dot{x}_{1,it}$, $\dot{x}_{2,it}$, and $\dot{u}_{it}$ are analogously defined. Then we can apply the same estimation procedure as used in Section 2.2 with the dotted variables replacing the tilded variables. The asymptotic properties of the resulting C-Lasso estimators and their post-Lasso versions will be modified by changing the demeaned Brownian motion to the detrended one in the limiting distributions.

To see this point clearly, we observe that

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

where $x^0_{1,it} = \sum_{s=1}^{t} \varepsilon_{1,is}$ is a purely random walk process. Define $\kappa_T =$diag$(1, T^{-1})$ and $g(r) = (1, r)'$. Let $t = \lfloor Tr \rfloor$, the integer part of Tr for $r \in [0, 1]$. Then as $T \to \infty$, $\kappa_T g_t \to g(r)$ uniformly in $r \in [0, 1]$. By the functional central limit theorem and continuous mapping theorem, we have

$$\begin{aligned} \frac{1}{\sqrt{T}} \dot{x}_{1,i\lfloor Tr \rfloor} =& \frac{1}{\sqrt{T}} \left[x_{1,i\lfloor Tr \rfloor} - \sum_{t=1}^{T} x_{1,it} g'_t \left(\sum_{s=1}^{T} g_s g'_s \right)^{-1} g_t \right] \\ =& \frac{1}{\sqrt{T}} \left[x^0_{1,i\lfloor Tr \rfloor} - \sum_{s=1}^{T} x^0_{1,it} g'_t \left(\sum_{s=1}^{T} g_s g'_s \right)^{-1} g_t \right] \\ =& \frac{x^0_{1,i\lfloor Tr \rfloor}}{\sqrt{T}} - \frac{1}{T} \sum_{t=1}^{T} \frac{x^0_{1,it}}{\sqrt{T}} \kappa_T g'_t \left(\frac{1}{T} \sum_{t=1}^{T} \kappa_T g_t g'_t \kappa_T \right)^{-1} \kappa_T g_T \\ \Rightarrow& B_{1,i}(r) - \int B_{1,i}(r) g(r)' dr \left(\int g(r) g(r)' \right)^{-1} g(r) \equiv B^{\tau}_{1,i}(r), \end{aligned}$$

where $B_{1,i}(\cdot)$ is as defined below (3.4), and $B^{\tau}_{1,i}(\cdot)$ is a detrended Brownian motion and independent across i. Following the analysis in Sections 4.1-4.4, we can show that Theorems 4.1-4.3 continue to hold with the demeaned data replaced by the detrended data, and that the limiting distributions in Theorems 4.4-4.6 can be modified accordingly to account for different asymptotic moments on $\mathbb{Q}_k$ and $\mathbb{V}_k$ with the demeaned Brownian motion replaced by the detrended Brownian motion. For brevity, we do not report the details here.

5 Monte Carlo Simulation

In this section, we evaluate the finite sample performance of both PLS-based and DPLS-based C-Lasso estimates and their post-Lasso versions.

5.1 Data generating processes

We consider five data generating processes (DGPs). The observations in DGPs 1-3 are drawn from three groups with $N_1 : N_2 : N_3 = 0.3 : 0.4 : 0.3$. DGPs 4-5 try to mimic the estimates and estimated group structures in the empirical application, where observations in DGP 4 are drawn from two groups with $N_1 : N_2 = 0.9 : 0.1$, and those in DGP 5 are drawn from three groups with $N_1 : N_2 : N_3 = 0.5 : 0.3 : 0.2$. There are four combinations of the sample sizes with $N = 50$, 100 and $T = 40$, 80.

DGP 1 (Strictly Exogenous Nonstationary Regressors) The observations (y_{it}, x'_{it}) are generated from the following cointegrated panel

$$\begin{cases} y_{it} = \mu_i + \beta_i^{0\prime} x_{it} + u_{it} = \mu_i + \beta_{1,i}^{0\prime} x_{1,it} + u_{it} \\ x_{1,it} = x_{1,it-1} + \varepsilon_{1,it} \end{cases}, \tag{5.1}$$

where $\mu_i \sim$ IID $N(0,1)$, $x_{it} = x_{1,it}$ is a 2×1 vector, $\varepsilon_{it} = (u_{it},\ \varepsilon'_{1,it})'$ follows a multivariate standard normal distribution, and $\beta_i^0 = \beta_{1,i}^0$ exhibits the group structures in (2.3) for $K = 3$ and

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

DGP 2 (Weakly Dependent Nonstationary Regressors) The observations (y_{it}, x'_{it}) are generated via (5.1) but we now allow for correlations between the two nonstationary regressors in $x_{1,it}$, the correlations between $x_{1,it}$ and μ_i, and the correlations between u_{it} and $\varepsilon_{1,it}$. Specifically, for each i we generate a 4-dimensional time series $\left\{\varepsilon_{it}^{\dagger}, t \geq 1\right\}$ via a linear process $\varepsilon_{it}^{\dagger} = \sum_{j=1}^{\infty} \psi_{ij} e_{i,t-j}$, where e_{it} are IID $N(0, I_4)$, $\psi_{ij} = 0.5 \cdot j^{-3.5} \cdot \Omega_1^{1/2}$, and $\Omega_1^{1/2}$ is the symmetric square root of $\Omega_1 \equiv \begin{pmatrix} 1 & 0.3 & 0.2 & 0 \\ 0.3 & 1 & 0.2 & 0.2 \\ 0.2 & 0.3 & 1 & 0.2 \\ 0 & 0.2 & 0.2 & 1 \end{pmatrix}$. Then we set $u_{it} = S_0 \varepsilon_{it}^{\dagger}$, $\varepsilon_{1,it} = S_1 \varepsilon_{it}^{\dagger}$, and $\mu_i = S_\mu \varepsilon_{i1}^{\dagger}$, where $S_0 = (1,0,0,0)$, $S_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$, and $S_\mu = (0,0,0,1)$.

DGP 3 (Weakly Dependent Nonstationary and Stationary Regressors) The observations (y_{it}, x'_{it}) are generated from the following cointegrated panel

$$\begin{cases} y_{it} = \mu_i + \beta_i^{0\prime} x_{it} + u_{it} = \mu_i + \beta_{1,i}^{0\prime} x_{1,it} + \beta_{2i}^0 x_{2,it} + u_{it} \\ x_{1,it} = x_{1,it-1} + \varepsilon_{1,it} \end{cases},$$

where $x_{1,it}$ is a 2×1 vector, $\beta_{1,i}^0$ exhibits the group structures and preserves the setting in DGP 1, and $x_{2,it} = \varepsilon_{2,it}$ contains a scalar stationary regressor. The coefficients of the stationary regressors are heterogeneous across all i such that $\beta_{2,i} \sim$ IID $N(0.5, 1)$. To allow correlation between μ_i and x_{it}, for each i we first generate a 5-dimensional time series $\left\{\varepsilon_{it}^\dagger, t \geq 1\right\}$ via a linear process $\varepsilon_{it}^\dagger = \sum_{j=1}^\infty \psi_{ij} e_{i,t-j}$, where e_{it} are IID $N(0, I_5)$, $\psi_{ij} = 0.5 \cdot j^{-3.5} \cdot \Omega_2^{1/2}$, and $\Omega_2^{1/2}$ is the symmetric square root of $\Omega_2 \equiv \begin{pmatrix} 1 & 0.3 & 0.2 & 0.2 & 0 \\ 0.3 & 1 & 0.2 & 0 & 0.2 \\ 0.2 & 0.2 & 1 & 0 & 0.2 \\ 0.2 & 0 & 0 & 1 & 0.2 \\ 0 & 0.2 & 0.2 & 0.2 & 1 \end{pmatrix}$. Then we set $u_{it} = S_0 \varepsilon_{it}^\dagger$, $\varepsilon_{1,it} = S_1 \varepsilon_{it}^\dagger$, $\varepsilon_{2,it} = S_2 \varepsilon_{it}^\dagger$, and $\mu_i = S_\mu \varepsilon_{i1}^\dagger$, where $S_0 = (1, 0, 0, 0, 0)$, $S_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$, $S_2 = (0, 0, 0, 1, 0)$ and $S_\mu = (0, 0, 0, 0, 1)$.

DGP 4 (Mimicking the first subsample in Table 5) The observations (y_{it}, x_{it}') are generated via (5.1), where $x_{it} = x_{1,it}$ contains one nonstationary regressor. For each i, we first generate a 3-dimensional time series $\left\{\varepsilon_{it}^\dagger, t \geq 1\right\}$ via a linear process $\varepsilon_{it}^\dagger = \sum_{j=1}^\infty \psi_{ij} e_{i,t-j}$, e_{it} are IID $N(0, I_3)$, $\psi_{ij} = 0.5 \cdot j^{-3.5} \cdot \Omega_1^{1/2}$, and $\Omega_1^{1/2}$ is the symmetric square root of $\Omega_1 \equiv \begin{pmatrix} 1 & 0.3 & 0 \\ 0.3 & 1 & 0.2 \\ 0 & 0.2 & 1 \end{pmatrix}$. Then we set $u_{it} = S_0 \varepsilon_{it}^\dagger$, $\varepsilon_{1,it} = S_1 \varepsilon_{it}^\dagger$, and $\mu_i = S_\mu \varepsilon_{i1}^\dagger$, where $S_0 = (1, 0, 0)$, $S_1 = (0, 1, 0)$, and $S_\mu = (0, 0, 1)$. $\beta_i^0 = \beta_{1,i}^0$ exhibits the group structures in (2.3) for $K = 2$ with $(\alpha_1^0, \alpha_2^0) = (0.9, -0.7)$, which is the collection of the group-specific estimates for the first subsample in Table 5. Note that we set $N_1 : N_2 = 0.9 : 0.1$ for this DGP.

DGP 5 (Mimicking the second subsample in Table 5) The observations (y_{it}, x_{it}') are generated via (5.1). The innovation processes are generated via the same processes in DGP 4. Now, $\beta_i^0 = \beta_{1,i}^0$ exhibits the group structures in (2.3) for $K = 3$ with $(\alpha_1^0, \alpha_2^0, \alpha_3^0) = (0.9, 0.2, -0.6)$, the collection of the group-specific estimates for the second subsample in Table 5. Note that for this DGP we set $N_1 : N_2 : N_3 = 0.5 : 0.3 : 0.2$, which is close to $49 : 27 : 23$, the ratios of estimated numbers of elements in the three estimated groups.

In all cases, the number of replications is 10,000.

5.2 Classification and estimation

For the moment, we assume that the number of groups is known and examine the performance of classification and estimation. When the number of groups is unknown, we can apply the information criterion (IC) introduced in Section C of the online supplement to determine the number of groups. We also examine the finite sample performance of the IC in Section E of the online supplement.

For classification, we consider the PLS-based C-Lasso classification results for DGPs 1, 2, 4 and 5, and

Table 1: Empirical classification errors in percentage

	c_λ		0.1		0.2	
	N	T	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$
DGP1	50	40	0.212	0.221	0.515	0.410
	50	80	0.000	0.000	0.001	0.001
	100	40	0.218	0.226	0.475	0.384
	100	80	0.000	0.000	0.001	0.001
DGP2	50	40	0.483	0.506	0.875	0.728
	50	80	0.000	0.000	0.003	0.002
	100	40	0.500	0.518	0.796	0.667
	100	80	0.000	0.000	0.004	0.003
DGP3	50	40	0.535	0.563	0.799	0.684
(PLS)	50	80	0.001	0.001	0.005	0.004
	100	40	0.532	0.562	0.745	0.640
	100	80	0.000	0.000	0.003	0.002
DGP3	50	40	6.337	5.630	12.255	9.700
(DPLS)	50	80	0.038	0.031	0.186	0.141
	100	40	6.027	5.432	11.453	9.138
	100	80	0.033	0.026	0.157	0.120
DGP4	50	40	1.234	0.834	0.821	0.543
	50	80	0.014	0.008	0.004	0.002
	100	40	1.225	0.823	0.801	0.527
	100	80	0.011	0.007	0.004	0.003
DGP5	50	40	0.000	0.000	0.040	0.004
	50	80	0.000	0.000	0.000	0.000
	100	40	0.000	0.000	0.032	0.004
	100	80	0.000	0.000	0.001	0.000

both the PLS- and DPLS-based C-Lasso classification results for DGP 3. For the DPLS-based classification in DGP 3, we introduce the lags and leads of $\Delta x_{1,it}$ in our penalized estimation by setting $\bar{p}_2 = \lfloor T^{1/4} \rfloor$. We follow Section 4.2 and define two types of average classification errors: $\bar{P}(\hat{E}) = \frac{1}{N}\sum_{i=1}^{N} \hat{P}(\cup_{k=1}^{K} \hat{E}_{kNT,i})$ and $\bar{P}(\hat{F}) = \frac{1}{N}\sum_{i=1}^{N} \hat{P}(\cup_{k=1}^{K} \hat{F}_{kNT,i})$, where $\hat{P}$ is the empirical mean over 10,000 replications. Table 1 reports the classification errors by setting $\lambda = c_\lambda T^{-3/4}$ with $c_\lambda = 0.1$ and 0.2.[5] We summarize some important findings from Table 1. First, both types of classification errors vary over c_λ. The smaller value of c_λ, the smaller percentage of the classification errors. This means that a larger value of penalty term tends to lead to a higher rate of misclassification. Second, as T increases, the percentage of classification errors drops significantly. In fact, when T is 80, we have less than 1% of individuals misclassified in all cases under investigation. Third, for DGP 3, the performance of the DPLS-based C-Lasso classification is not as good as that of the PLS-based C-Lasso estimation. Despite this fact, the former performance becomes acceptable when $T = 80$ for both choices of c_λ.

For the estimation, we consider both the C-Lasso estimates and their post-Lasso versions. Specifically, for all DGPs we consider the PLS-based C-Lasso estimates, the OLS-based post-Lasso estimates, the DOLS-based post-Lasso estimates, and the oracle estimates that are obtained by using the true group structures. For DGP 3, we also consider the DPLS-based C-Lasso estimates, their post-Lasso versions, and the oracle estimates. For all DOLS-based estimates, we set $\bar{p}_2$ as above. We report the bias, root-mean-square error (RMSE), and coverage probability of the two-sided nominal 95% confidence interval for the estimate $\hat{\beta}_{1,i}(1)$ of the first parameter $\beta_{1,i}(1)$ in $\beta_{1,i}$ for each DGP in Tables 2-3, where all criteria are averaged over different groups and across 10,000 replications. For example, we calculate the RMSE of $\hat{\beta}_{1,i}(1)$'s as $\frac{1}{N}\sum_{k=1}^{K_0} N_k \text{RMSE}(\hat{\alpha}_{k,1})$ with $\hat{\alpha}_{k,1}$ denoting the first element in $\hat{\alpha}_k$ for one replication and then average them across all replications for each case.

Table 2 reports the estimation results for DGPs 1-2 and 4-5 based on the PLS method. Table 3 reports the estimation results for DGP 3 based on both the PLS and DPLS methods.[6] These tables reveal some general patterns. First, the bias and RMSE of the C-Lasso estimates and their post-Lasso versions always decrease as either N or T increases, and they decrease faster when T increases than when N increases. This is as expected due to faster convergence rate of the estimates along the time dimension than along the cross-sectional dimension. Second, when there is no endogeneity issue in DGP1, the finite sample performance of the post-Lasso OLS estimates is close to that of the oracle ones and dominates that of the DOLS-based post-Lasso estimates. This indicates that the DOLS may hurt in finite samples when there is no endogeneity issue in the model. Third, when endogeneity is present in DGPs 2-5, the post-Lasso DOLS estimators are distinctly superior to the C-Lasso and post-Lasso OLS ones for all cases and their performance is very close to that of the oracle ones. Since the endogeneity issue is not well accounted for the C-Lasso and post-Lasso OLS estimates, their coverage probabilities may deteriorate when N or T increases. Fourth, for DGP 3 the DPLS-based C-Lasso estimates outperform the PLS-based C-Lasso estimates to a great margin, but the post-Lasso estimates are not quite distinct from each other in terms of bias and RMSE. Fifth, the coverage probabilities of the DOLS-based post-Lasso estimates are generally quite close to the nominal level (95%) in all cases (except for DGP 1 in the absence of endogeneity). For DGP3, the coverage probabilities of DPLS-based C-Lasso estimates are closer to the nominal level compared to those of the PLS-based C-Lasso estimates. These two facts suggest that the DOLS bias correction yields good coverage probability when endogeneity is present. Lastly, in general the post-Lasso DOLS estimates outperform the C-Lasso estimates (except for DGP 1 in the absence of endogeneity) and thus are recommended for practical uses.

6 Application: Testing the PPP hypothesis

In this section we apply our method to reinvestigate the purchasing power parity (PPP) hypothesis in international economics.

6.1 PPP hypothesis

The PPP hypothesis assumes that in the absence of transaction costs and trade barriers, a basket of identical goods will have the same price in different markets when the prices are expressed in the same currency. Unlike the law of one price for one particular good, the PPP is built on a "basket of goods", indicating that the nominal exchange rate is adjusted by the relative general price index for international comparison. The long-run PPP hypothesis was broadly accepted in the post-war period before the breakdown of the Bretton Woods system in the early 1970s. In the post-Bretton Woods period, most applied work fails to

Table 2: RMSEs, Biases and Coverage probabilities for various estimates

N	T	$c_\lambda = 0.2$	RMSE	Bias	Coverage %	RMSE	Bias	Coverage %
				DGP1-PLS			DGP2-PLS	
50	40	C-Lasso	0.0174	0.0001	93.05	0.0287	0.0223	85.67
		Post-Lasso^{OLS}	0.0173	0.0001	93.24	0.0276	0.0211	87.47
		Post-Lasso^{DOLS}	0.0226	0.0000	84.58	0.0215	0.0001	94.72
		Oracle	0.0172	0.0001	93.30	0.0215	0.0000	94.73
50	80	C-Lasso	0.0082	0.0001	93.51	0.0138	0.0107	75.74
		Post-Lasso^{OLS}	0.0082	0.0001	93.55	0.0135	0.0105	76.98
		Post-Lasso^{DOLS}	0.0091	0.0001	90.27	0.0088	0.0000	94.15
		Oracle	0.0082	0.0001	93.55	0.0088	0.0000	94.15
100	40	C-Lasso	0.0122	0.0001	93.75	0.0252	0.0218	73.51
		Post-Lasso^{OLS}	0.0121	0.0001	94.01	0.0240	0.0205	77.62
		Post-Lasso^{DOLS}	0.0155	0.0001	85.75	0.0148	0.0001	95.82
		Oracle	0.0120	0.0001	94.08	0.0148	0.0001	95.85
100	80	C-Lasso	0.0056	0.0000	94.42	0.0120	0.0105	59.63
		Post-Lasso^{OLS}	0.0056	0.0000	94.42	0.0117	0.0101	62.00
		Post-Lasso^{DOLS}	0.0063	0.0001	91.57	0.0060	0.0001	95.26
		Oracle	0.0056	0.0000	94.42	0.0060	0.0001	95.27
				DGP4-PLS			DGP5-PLS	
50	40	C-Lasso	0.0290	0.0233	73.88	0.0263	0.0226	52.22
		Post-Lasso^{OLS}	0.0285	0.0226	76.03	0.0263	0.0226	52.21
		Post-Lasso^{DOLS}	0.0188	-0.0001	93.57	0.0139	0.0001	94.18
		Oracle	0.0188	0.0001	93.70	0.0139	0.0001	94.18
50	80	C-Lasso	0.0140	0.0114	68.02	0.0128	0.0110	44.90
		Post-Lasso^{OLS}	0.0139	0.0112	68.76	0.0128	0.0110	44.89
		Post-Lasso^{DOLS}	0.0081	0.0000	94.06	0.0061	-0.0001	94.31
		Oracle	0.0081	0.0000	94.06	0.0061	-0.0001	94.31
100	40	C-Lasso	0.0259	0.0229	53.83	0.0242	0.0223	24.31
		Post-Lasso^{OLS}	0.0252	0.0221	58.16	0.0243	0.0223	24.27
		Post-Lasso^{DOLS}	0.0130	-0.0002	94.22	0.0097	0.0000	94.31
		Oracle	0.0130	0.0000	94.32	0.0097	0.0000	94.31
100	80	C-Lasso	0.0126	0.0113	46.00	0.0119	0.0109	18.40
		Post-Lasso^{OLS}	0.0124	0.0110	47.67	0.0119	0.0109	18.40
		Post-Lasso^{DOLS}	0.0057	0.0000	94.49	0.0043	0.0000	94.45
		Oracle	0.0057	0.0000	94.49	0.0043	0.0000	94.45

Table 3: RMSEs, Biases and Coverage probabilities for various estimates

N	T		RMSE	Bias	Coverage %		RMSE	Bias	Coverage %
			DGP3-PLS				DGP3-DPLS		
50	40	C-Lasso	0.0275	0.0206	88.14	C-Lasso	0.0232	0.0000	93.31
		Post-Lasso^{OLS}	0.0318	0.0193	81.74				
		Post-Lasso^{DOLS}	0.0215	0.0000	94.90	Post-Lasso	0.0227	0.0000	93.91
		Oracle	0.0214	0.0000	95.02	Oracle	0.0214	0.0000	95.02
50	80	C-Lasso	0.0126	0.0094	80.41	C-Lasso	0.0087	0.0000	94.24
		Post-Lasso^{OLS}	0.0156	0.0091	71.00				
		Post-Lasso^{DOLS}	0.0086	0.0000	94.29	Post-Lasso	0.0086	0.0000	94.28
		Oracle	0.0086	0.0000	94.29	Oracle	0.0086	0.0000	94.29
100	40	C-Lasso	0.0237	0.0200	78.75	C-Lasso	0.0162	0.0000	94.67
		Post-Lasso^{OLS}	0.0254	0.0184	75.25				
		Post-Lasso^{DOLS}	0.0148	-0.0001	96.02	Post-Lasso	0.0157	-0.0001	95.24
		Oracle	0.0147	-0.0001	96.11	Oracle	0.0150	-0.0005	96.11
100	80	C-Lasso	0.0108	0.0091	67.20	C-Lasso	0.0060	0.0000	95.11
		Post-Lasso^{OLS}	0.0121	0.0088	63.49				
		Post-Lasso^{DOLS}	0.0060	0.0000	95.01	Post-Lasso	0.0060	0.0000	95.11
		Oracle	0.0060	0.0000	95.01	Oracle	0.0059	0.0000	95.16

support the validity of the long-run PPP; see, e.g., Frenkel (1981) and Adler and Lehmann (1983). Some researchers attribute this to the low power of time series unit root tests when T is short and advocate the use of panel unit root tests. Indeed, some panel unit root testing results favor the PPP hypothesis in the post-Bretton Woods period; see, e.g., Oh (1996) and Papell (1997). Even so, the empirical findings are still mixed. There remain two main issues in testing the validity of the PPP hypothesis by using panel data. One is the sample selection issue and the other is the unobserved heterogeneity issue. Our cointegrated panel model with latent group structures can provide a data-driven method to address these two issues simultaneously and is expected to offer some new insights into the PPP hypothesis.

6.2 Model and data

The PPP hypothesis has two versions: strong and weak. We first consider the strong PPP hypothesis. Denote the domestic price index as P_{it}, the corresponding foreign price index as P_{jt}, and E_{it} as the nominal exchange rate. If the strong PPP hypothesis holds, we have the equation $E_{it} = \frac{P_{it}}{P_{jt}}$ where we suppress the dependence of E_{it} on j, which is typically fixed in panel studies. In the logarithmic form, we have $e_{it} = p_{it} - p_{jt}$, where $e_{it} = \log(E_{it})$, $p_{it} = \log(P_{it})$, and $p_{jt} = \log(P_{jt})$. Previous panel unit root tests are built on the equation

$$e_{it} = (p_{it} - p_{jt}) + u_{it}, \tag{6.1}$$

where u_{it} stands for the real exchange rate. The rejection of the null hypothesis that the processes $\{u_{it}, t \geq 1\}$ are all nonstationary is regarded as evidence in favor of the validity of the long-run PPP or mean-reversion in real exchange rates. The most important assumption in the strong PPP hypothesis is that there exists a one-to-one relationship between the nominal exchange rates and aggregate price ratios. In practice, the movements may not be directly proportional, leading to the cointegrating slopes deviating away from the unity. Pedroni (2004) modifies (6.1) by allowing for heterogeneous coefficients across individuals and estimating the following long-run PPP hypothesis in weak version

$$e_{it} = \mu_i + \beta_i(p_{it} - p_{jt}) + u_{it} = \mu_i + \beta_i \Delta p_{ij,t} + u_{it}, \tag{6.2}$$

where β_i is allowed to vary across countries and is expected to be positive, $\Delta p_{ij,t} = p_{it} - p_{jt}$, and μ_i is the unobserved fixed effect for country i.

In our weak PPP model, we assume that β_i exhibits the latent group structures studied in this paper. By pooling the slope coefficients within a group altogether, we can obtain more efficient estimates than those obtained from a fully heterogeneous cointegrated panel model. In addition, since our C-Lasso method is a data-driven method, we do not manually assign different countries to different groups, which alleviates the sample selection problem.

We obtain monthly and quarterly data of the nominal exchange rate and consumer price index (CPI) from January 1975 to July 2014 covering 99 countries from International Financial Statistics. Here, we use the CPI to represent the general price index. We choose the time span from 1975 to 2014 to cover the post-Bretton Woods period. Given the fact that Euro dollar was introduced to the global financial markets as an accounting currency on 1 January 1999, we consider two subsamples. We obtain a balanced panel with 67 countries in the period 1975-1998 and another balanced panel with 99 countries in the period 1999-2014. For the quarterly data, we have 91 time series periods in 1975Q.1-1998.Q4 and 55 times series periods in 1999.Q1-2014.Q2. For the monthly data, we have 283 time series periods in period 1975.M1-1998.M12 and 172 times series periods in 1999.M1-2014.M7.

6.3 Group and estimation results

In this section, we present the classification and estimation results for the quarterly data. The results for the monthly data are relegated to Section F in the online supplement. We determine the number of groups by using the information criterion (IC) proposed in Section C of the online supplement. Table A.2 in the online supplement reports the information criterion with different tuning parameter values: $\lambda = c_\lambda \times T^{-3/4}$ where $c_\lambda = 0.025$, 0.05, 0.1, and 0.2. Obviously, the IC is robust to the choice of tuning parameters. Following the majority rule, we decide to select $K = 2$ groups for the period 1975.Q1-1998.Q4 and $K = 3$ groups for the period 1999.Q1-2014.Q2. Note that the IC is minimized at $c_\lambda = 0.1$ and 0.05 for the first and second, subsamples respectively. We will choose $c_\lambda = 0.1$ and 0.05 for these two subsamples, respectively and report the estimation results.

Table 4 reports the DPLS estimation results for the subsamples 1975.Q1-1998.Q4 and 1999.Q1-2014.Q2 by using $c_\lambda = 0.1$ and 0.05, respectively. We summarize some important findings from Table 4. First, the group-specific estimates vary a lot across groups, which indicates strong unobserved heterogeneities in both subsamples. Second, both C-Lasso estimate and its post-Lasso one for Group 1 are reasonably close to the unity in both the first and second subsamples, which lends some positive supports to the weak-form long-run PPP hypothesis. But the estimates in Group 2 in either subsample suggest a negative long-run relationship between the price index difference and the exchange rate, which contradicts the long-run PPP hypothesis. The estimate for Group 3 in the second subsample is positive and quite small in comparison

Table 4: Estimation results for the quarterly data

	Panel A: From 1975.Q1-1998.Q4						
	Pool	Group 1			Group 2		
	DOLS	C-Lasso	post-Lasso		C-Lasso	post-Lasso	
β_i	0.7465	0.8609	0.8608		-0.7007	-0.6992	
	(0.0207)	(0.0190)	(0.0190)		(0.0857)	(0.0857)	
	Panel B: From 1999.Q1-2014.Q2						
	Pool	Group 1		Group 2		Group 3	
	DOLS	C-Lasso	post-Lasso	C-Lasso	post-Lasso	C-Lasso	post-Lasso
β_i	0.3623	0.8667	0.8681	-0.5732	-0.5775	0.1986	0.1960
	(0.0184)	(0.0189)	(0.0189)	(0.0227)	(0.0228)	(0.0296)	(0.0296)

with the unity, which suggests a quite weak proportional relation between the change in the price index difference and that in the exchange rate. Third, similar results are also observed for the monthly data, and the long-run relationship between the nominal exchange rate and general price index presents similar patterns in either subsample period. This indicates the robustness of our findings.

Table 5 summarizes the group classification results for the two subsamples; see also Figure 1 for the classification results for the second subsample. Interestingly, we find that the majority of the countries in the first subsample are classified into Group 1, which indicates the long-run PPP holds for most countries in the period 1975.Q1-1998.Q4. During this time span, we have only 68 countries in the dataset, and some developing countries like Argentina, Brazil, and Russia are excluded from our subsample due to the fact that they have experienced hyperinflation. For the second subsample, we find even more interesting results. Figure 1 suggests that those countries that support the long-run PPP equilibrium are mainly located in Europe, Africa, Middle East, and North American. The members of Group 1 suggest a polarization of economic development. Further, we observe that most countries in Groups 2 and 3 are either fast-growing or middle-income countries (e.g., South Korea, Singapore, and Brazil) in the last decades in East Asia and South America. It confirms the Balassa-Samuelson effect, where the productivity differentials are one of the most important factors behind the PPP deviation, see Balassa (1964) and Samuelson (1964). In this case, countries with rapidly expanding economies should tend to have more rapidly appreciating exchange rates. In general, our results suggest heterogeneous behavior in the long-run PPP hypothesis.

7 Conclusion

In this paper, we propose a C-Lasso-based PLS procedure to estimate a cointegrated panel with latent group structures on the long-run cointegrating relationships. We allow for completely heterogeneous short-run dynamics but assume that long-run relationships are homogeneous within a group and heterogeneous across different groups. Our method can determine the individual's group membership consistently and estimate the parameters efficiently. To remove the asymptotic bias in the estimators of the long-run parameters, we also consider the dynamic PLS procedure. Simulation results confirm the asymptotic studies. An application to testing the validity of the long-run PPP hypothesis suggests strong evidence of

Table 5: Classification results for the quarterly data

Panel A: From 1975.Q1-1998.Q4				
Group 1 ($N_1 = 62$)				
Algeria	Australia	**Austria**	**Bahrain**	**Belgium**
Bolivia	**Botswana**	**Canada**	**Colombia**	**Costa Rica**
Cyprus	**Denmark**	Dominican	**Egypt**	El Salvador
Finland	**France**	**Ghana**	**Greece**	Guatemala
Honduras	**Hungary**	Iceland	**India**	**Indonesia**
Iran	Ireland	**Israel**	**Italy**	**Ivory Coast**
Jamaica	**Japan**	**Jordan**	**Kenya**	**South Korea**
Luxembourg	**Malta**	**Mauritius**	**Mexico**	**Morocco**
Nepal	**Netherlands**	New Zealand	**Nigeria**	**Norway**
Pakistan	**Paraguay**	**Philippines**	**Portugal**	**Singapore**
South Africa	**Spain**	**Sri Lanka**	**Sudan**	**Sweden**
Switzerland	Tanzania	**Thailand**	**Trinidad and Tobago**	**Turkey**
Uruguay	**Venezuela**			
Group 2 ($N_2 = 5$)				
Ecuador	**Kuwait**	Malaysia	**Myanmar**	Saudi Arabia
Panel B: From 1999.Q1-2014.Q2				
Group 1 ($N_1 = 49$)				
Angola	**Argentina**	**Austria**	**Bangladesh**	**Belgium**
Botswana	Brunei	**Canada**	**Costa Rica**	**Denmark**
Dominican	**Europe**	**Finland**	**France**	**Germany**
Ghana	**Honduras**	**Iceland**	**Iran**	**Italy**
Jamaica	**Japan**	**Jordan**	**Luxembourg**	**Malawi**
Mexico	**Mongolia**	**Morocco**	**Mozambique**	**Netherlands**
Nigeria	**Norway**	**Pakistan**	**Romania**	**Saudi Arbia**
Sri Lanka	**Sudan**	**Sweden**	**Switzerland**	**Tanzania**
Trinidad and Tobago	**Tunisia**	**Turkey**	**Uganda**	**United Kingdom**
Ukraine	**Venezuela**	**Viet Nam**	Zambia	
Group 2 ($N_2 = 23$)				
Albania	**Armenia**	Australia	Bolivia	**Brazil**
Bulgaria	**Colombia**	**Congo**	**Croatia**	El Salvador
Georgia	**Hungary**	**Ireland**	**Ivory Coast**	**Kuwait**
Latvia	**Macau**	**Moldova**	New Zealand	**Peru**
Philippines	**Spain**	**Thailand**		
Group 3 ($N_3 = 27$)				
Algeria	Cambodia	**Czech Republic**	Egypt	**Guatemala**
Hong Kong	India	**Indonesia**	**Israel**	**Kazakhstan**
Kenya	**South Korea**	**Kyrgyzstan**	**Laos**	Lithuania
Macedonia	**Malaysia**	Mauritius	**Myanmar**	Nepal
Paraguay	**Poland**	**Portugal**	**Russia**	**Singapore**
South Africa	Uruguay			

Note: Countries in bold denote coincidences of the classification results based on the monthly and quarterly datasets.

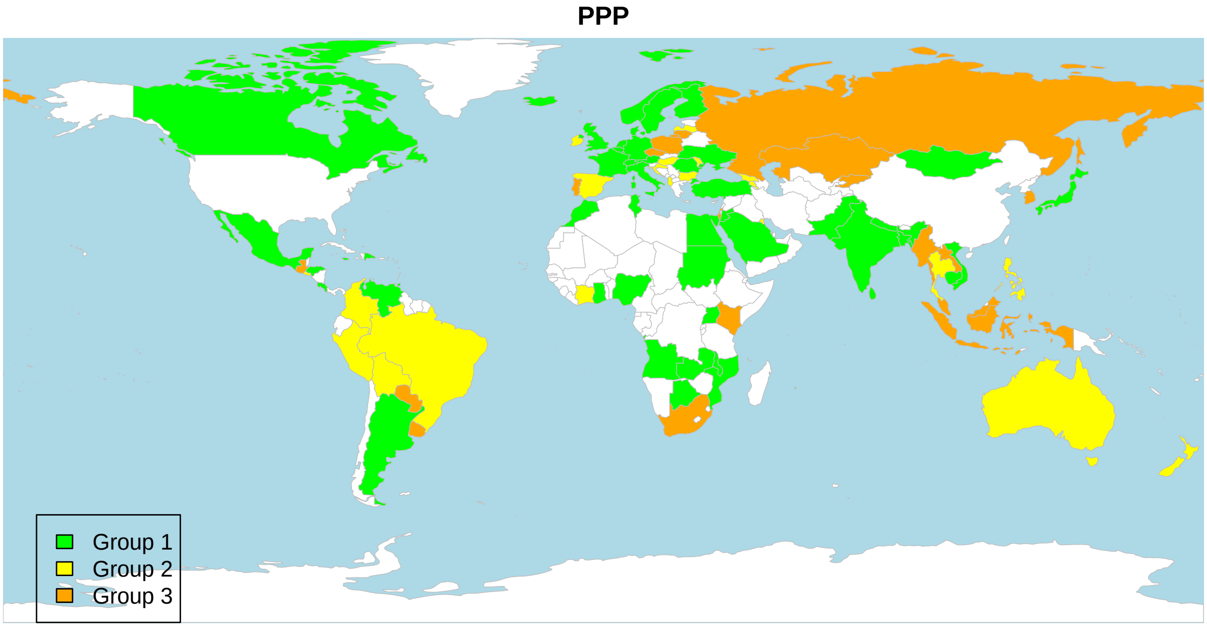

Figure 1: The geographic features of countries in the three groups in subsample 2 (1999-2014)

latent group structures.

There are several interesting topics for future research. First, we do not allow for cross-sectional dependence in our model. In macro-econometrics, cross-sectional dependence is frequently modelled via the multi-factor error structure (Pesaran, 2006) or interactive fixed effects (Bai, 2009). Depending on whether we allow for unit-root behavior in the factors, different methods can be called upon (see, e.g., Bai and Ng, 2004; Bai and Kao, 2006; Bai et al., 2009; Bai and Ng, 2010). But this certainly complicates the asymptotic analysis and deserves a separate treatment. Second, when the dimension of the nonstationary variables is higher than 2, multiple cointegrating relationships may exist. It is worthwhile to consider the panel vector error-correction model or likelihood-based panel cointegration analysis in this case. Third, as an anonymous referee insightfully points out, in practice it is worthwhile to allow for the presence of a single group, e.g., the Kth group, that contains individuals with heterogeneous slope coefficients. As one can imagine, both the C-Lasso and K-means algorithms fail in this case and one has to design a new algorithm to pin down the elements in the first $K-1$ groups. One possible way is to consider a sequential testing procedure based on some preliminary consistent estimates of the slope coefficients as in Wang and Su (2018). We leave these topics for future research.

Notes

[1]Most asymptotic theories in the panel cointegration analysis have been established under the sequential limit theory. A few exceptions include Phillips and Moon (1999), Sun (2004), and Bai and Ng (2010).

[2]If $E(x_{2,it}) = \nu_{2i} \neq 0$, we can rewrite the first equation in (2.1) as $y_{it} = \mu_i^* + \beta_{1,i}' x_{1,it} + \beta_{2,i}' x_{2,it}^* + u_{it}$,

where $x^*_{2,it} = x_{2,it} - \nu_{2i}$ has zero mean and $\mu^*_i = \mu_i + \beta'_{2,i}\nu_{2i}$.

[3]The commutation matrix is used for transforming the vectorized form of a matrix into the vectorized form of its transpose. For any $m \times n$ matrix A, $K_{m,n}$ is the $mn \times mn$ matrix which transforms vec(A) into vec(A') : $K_{m,n}$ vec(A) =vec(A').

[4]Noting that by Lemma A.4(i)

$$\begin{aligned} Q_{k,NT} &= \frac{1}{N_k T^2} \sum_{i \in G^0_k} \tilde{x}'_{1,i}\tilde{x}_{1,i} - \frac{1}{N_k T^2} \sum_{i \in G^0_k} \left(\tilde{x}'_{1,i}\tilde{x}_{2,i}\right)\left(\tilde{x}'_{2,i}\tilde{x}_{2,i}\right)^{-1}\left(\tilde{x}'_{2,i}\tilde{x}_{1,i}\right) \\ &\equiv \frac{1}{N_k T^2} \sum_{i \in G^0_k} \tilde{x}'_{1,i}\tilde{x}_{1,i} + O_P\left(b_T^{-1}\right) = \frac{1}{N_k T^2} \sum_{i \in G^0_k} \tilde{x}'_{1,i}\tilde{x}_{1,i} + o_P(1), \end{aligned}$$

we can also consistently estimate $\mathbb{Q}_{(k)}$ by $\tilde{\mathbb{Q}}_{(k)} = \frac{1}{\hat{N}_k T^2} \sum_{i \in \hat{G}_k} \tilde{x}'_{1,i}\tilde{x}_{1,i}$.

[5]See Section D in the online supplement for more details on the determination of λ in practice.

[6]The estimation results for $c_\lambda = 0.1$ are available upon request.

REFERENCES

Adler, M. & B. Lehmann (1983) Deviations from purchasing power parity in the long run. *Journal of Finance* 38(5), 1471-1487.

Ando, T. & J. Bai (2016) Panel data model with grouped factor structure under unknown group membership. *Journal of Applied Econometrics* 31(1), 163-191.

Ando, T. & J. Bai (2017) Clustering huge number of financial time series: a panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association* 112(519), 1182-1198.

Andrews, D.W. (1984) Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21(4), 930-934.

Bai, J. (2004) Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics* 122(1), 137-183.

Bai, J. (2009) Panel data models with interactive fixed effects. *Econometrica* 77(4), 1229-1279.

Bai, J. & C. Kao (2006) On the estimation and inference of a panel cointegration model with cross-sectional dependence. In: B. Baltagi (ed.), *Contributions to Economic Analysis*, pp.3-30, Elsevier, Amsterdam.

Bai, J., C. Kao, & S. Ng (2009) Panel cointegration with global stochastic trends. *Journal of Econometrics* 149(1), 82-99.

Bai, J. & S. Ng (2004) A PANIC attack on unit roots and cointegration. *Econometrica* 72(4), 1127-1177.

Bai, J. & S. Ng (2010) Panel unit root tests with cross-section dependence: a futher investigation. *Econometric Theory* 26(4), 1088-1114.

Balassa, B. (1964) The purchasing-power parity doctrine: a reappraisal. *Journal of Political Economy* 72(6), 584-596.

Beveridge, S. & C.R. Nelson (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics* 7(2), 151-174.

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

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press, New York.

Donsker, M.D. & S.R.S. Varadhan (1977) On laws of the iterated logarithm for local times. *Communications on Pure and Applied Mathematics* 30(6), 707-753.

Dooley, M.P., D. Folkerts-Landau, & P. Garber (2004) The revived Bretton woods system. *International Journal of Finance & Economics* 9(4), 307-313.

Doukhan, P. (1994) *Mixing*. Springer, New York.

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

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.

Frenkel, J.A. (1981) The collapse of purchasing power parities during the 1970's. *European Economic Review* 16(1), 145-165.

Groen, J.J.J. & F. Kleibergen (2003) Likelihood-based cointegration analysis in panels of vector error-correction models. *Journal of Business & Economic Statistics* 21(2), 295-318.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press, New Jersey.

Kao, C. & M.H. Chiang (2000). On the estimation and inference of a cointegrated regression in panel data. *Advances in econometrics* 20, 179-222.

Lai, T.L. & C.Z. Wei (1982a) Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics* 10(1), 154-166.

Lai, T.L. & C.Z. Wei (1982b) Asymptotic properties of projections with applications to stochastic regression problems. *Journal of Multivariate Analysis* 12(3), 346-370.

Lam, C. & J. Fan (2008) Profile-kernel likelihood inference with diverging number of parameters. *Annals of Statistics* 36(5), 2232-2260.

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

Lu, X. & L. Su (2015) Jackknife model averaging for quantile regressions. *Journal of Econometrics* 188(1), 40-58.

Lu, X. & L. Su (2017) Determining the number of groups in latent panel structures with an application to income and democracy. *Quantitative Economics* 8(3), 729-760.

Oh, K.Y. (1996) Purchasing power parity and unit root tests using panel data. *Journal of International Money and Finance* 15(3), 405-418.

Mark, N. C. & D. Sul (2003) Cointegration vector estimation by panel DOLS and long-run money demand. *Oxford Bulletin of Economics and Statistics* 65(5), 655-680.

Papell, D.H. (1997) Searching for stationarity: purchasing power parity under the current float. *Journal of International Economics* 43(3-4), 313-332.

Park, J.Y. & P.C.B. Phillips (1988) Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4(3), 468-497.

Park, J.Y. & P.C.B. Phillips (1989) Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory* 5(1), 95-131.

Pedroni, P. (2001) Fully modified OLS for heterogeneous cointegrated panels. *Advances in Econometrics* 15, 93-130.

Pedroni, P. (2004) Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric Theory* 20(3), 597-625.

Pesaran, M.H. (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74(4), 967-1012.

Pesaran, M.H., Y. Shin, & R.P. Smith (1999) Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association* 94(446), 621-634.

Phillips, P.C.B. (1996) Econometric model determination. *Econometrica* 64, 763-812.

Phillips, P.C.B. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67(5), 1057-1111.

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

Qian, J. & L. Su (2016a) Shrinkage estimation of regression models with multiple structural changes. *Econometric Theory* 32(6), 1376-1433.

Qian, J. & L. Su (2016b) Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics* 191(1), 86–109.

Saikkonen, P. (1991) Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7(1), 1-21.

Samuelson, P.A. (1964) Theoretical notes on trade problems. *Review of Economics and Statistics* 46, 145-154.

Sarafidis, V. & N. Weber (2015) A partially heterogeneous framework for analyzing panel data. *Oxford Bulletin of Economics and Statistics* 77(2), 274-296.

Stock, J. H. & M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 783-820

Su, L. & G. Ju (2018) Identifying latent grouped effects in panel data models with interactive fixed effects. *Journal of Econometrics*, 206(2), 554-573.

Su, L., Z. Shi, & P.C.B. Phillips (2016) Identifying latent structures in panel data. *Econometrica* 84(6), 2215-2264.

Su, L., X. Wang, & S. Jin (2018) Sieve estimation of time-varying panel data models with latent structures. *Journal of Business & Economic Statistics*, forthcoming.

Sun, Y. (2004) Estimation of the long-run average relationship in nonstationary panel time series. *Econometric Theory* 20(6), 1227-1260.

Wang, W., P.C.B. Phillips, & L. Su (2018) Homogeneity pursuit in panel data models: theory and applications. *Journal of Applied Econometrics* 33(6), 797-815.

Wang, W., & L. Su (2018) Identifying latent group structures in nonlinear panels. Working Paper, Singapore Management University.

White, H. (2001) *Asymptotic Theory for Econometricians*. Emerald, UK.

Appendix

A Proofs of the Main Results in Section 4

In this appendix, we first state some technical lemmas that are used in the proofs of Theorems 4.1-4.6 and then prove these main results. The proofs of the technical lemmas are relegated to the online supplementary Appendix B.

Let $x_{1,it}^0 = \sum_{s=1}^t \varepsilon_{1,it}$. Noting that $x_{1,it} = x_{1,i0} + \sum_{s=1}^t \varepsilon_{1,it}$ and $\tilde{x}_{1i,t} = x_{1,it} - \frac{1}{T}\sum_{s=1}^T x_{1,is} = x_{1,it}^0 - \frac{1}{T}\sum_{s=1}^T x_{1,is}^0$, the initial value $x_{1,i0}$ does not play a role in our analysis. Without loss of generality, we assume that $x_{1,i0} = 0$ and write $x_{1,it}$ for $\sum_{s=1}^t \varepsilon_{1,it}$ hereafter. Recall that

$$\begin{aligned}
\hat{Q}_{i,\tilde{x}\tilde{x}} &= \begin{pmatrix} \frac{1}{T^2}\sum_{t=1}^T \tilde{x}_{1,it}\tilde{x}_{1,it}' & \frac{1}{T^2}\sum_{t=1}^T \tilde{x}_{1,it}\tilde{x}_{2,it}' \\ \frac{1}{T^2}\sum_{t=1}^T \tilde{x}_{2,it}\tilde{x}_{1,it}' & \frac{1}{T}\sum_{t=1}^T \tilde{x}_{2,it}\tilde{x}_{2,it}' \end{pmatrix} = \begin{pmatrix} \hat{Q}_{i,\tilde{x}_1\tilde{x}_1} & \hat{Q}_{i,\tilde{x}_1\tilde{x}_2} \\ \hat{Q}_{i,\tilde{x}_2\tilde{x}_1} & \hat{Q}_{i,\tilde{x}_2\tilde{x}_2} \end{pmatrix}, \\
\hat{Q}_{i,\tilde{x}\tilde{u}} &= \begin{pmatrix} \frac{1}{T^2}\sum_{t=1}^T \tilde{x}_{1,it}\tilde{u}_{it} \\ \frac{1}{T^2}\sum_{t=1}^T \tilde{x}_{2,it}\tilde{u}_{it} \end{pmatrix} = \begin{pmatrix} \hat{Q}_{i,\tilde{x}_1\tilde{u}} \\ \hat{Q}_{i,\tilde{x}_2\tilde{u}} \end{pmatrix}, \\
\hat{Q}_{i,\tilde{x}\tilde{u}^*} &= \begin{pmatrix} \frac{1}{T^2}\sum_{t=1}^T \tilde{x}_{1,it}\tilde{u}_{it}^* \\ \frac{1}{T^2}\sum_{t=1}^T \tilde{x}_{2,it}\tilde{u}_{it}^* \end{pmatrix} = \begin{pmatrix} \hat{Q}_{i,\tilde{x}_1\tilde{u}^*} \\ \hat{Q}_{i,\tilde{x}_2\tilde{u}^*} \end{pmatrix},
\end{aligned}$$

where $\tilde{u}_{it}^* = \tilde{u}_{it} - \tilde{x}_{2it}'\Sigma_{22,i}^{-1}\Sigma_{20,i}$. Let $\tilde{x}_{1,i} = (\tilde{x}_{1,i1}, ..., \tilde{x}_{1,iN})'$. Define $\tilde{x}_{2,i}$, $\tilde{u}_i$, and $\tilde{u}_i^*$ analogously. Let $M_{\ell,i} = I_T - \tilde{x}_{\ell,i}(\tilde{x}_{\ell,i}'\tilde{x}_{\ell,i})^{-1}\tilde{x}_{\ell,i}'$ for $\ell = 1,2$, where I_T is a $T \times T$ identity matrix. Recall that $D_T = \begin{pmatrix} I_{p_1} & 0 \\ 0 & \sqrt{T}I_{p_2} \end{pmatrix}$. We shall abbreviate $\hat{Q}_{i,\tilde{x}_1\tilde{x}_1}$ as $\hat{Q}_{1i}$ frequently for notational simplicity.

To prove the main results in the paper, we need the following lemmas.

Lemma A.1 *Let $\mathbb{S} = (\mathbb{S}_1, \mathbb{S}_2)$ be a selection matrix, where $\mathbb{S}_1$ and $\mathbb{S}_2$ are $l \times p_1$ and $l \times p_2$ matrices, respectively, and l is a fixed integer. Suppose that Assumptions A.1-A.3 hold. Then for each $i = 1, ..., N$,*

(i) $\mathbb{S}D_T\hat{Q}_{i,\tilde{x}\tilde{x}}D_T\mathbb{S}' \Rightarrow \mathbb{S}\begin{pmatrix} \int_0^1 \tilde{B}_{1,i}\tilde{B}_{1,i}' & 0 \\ 0 & \Sigma_{22,i} \end{pmatrix}\mathbb{S}'$,

(ii) $T\hat{Q}_{i,\tilde{x}_1\tilde{u}^*} \Rightarrow \int_0^1 \tilde{B}_{1,i}dB_{0,i}' + \Delta_{10,i} - \left(\int_0^1 \tilde{B}_{1,i}dB_{2,i}' + \Delta_{12,i}\right)\Sigma_{22,i}^{-1}\Sigma_{20,i}$,

(iii) $T^{3/2}\mathbb{S}_2\hat{Q}_{i,\tilde{x}_2\tilde{u}^*} \Rightarrow \mathbb{S}_2 (J_{1,i} \otimes J_{2,i}) N\left(0, V_i^0\right)$,

(iv) $T\left(\tilde{\beta}_{1,i} - \beta_{1,i}^0\right) \Rightarrow \left(\int_0^1 \tilde{B}_{1,i}\tilde{B}_{1,i}'\right)^{-1}\left[\int_0^1 \tilde{B}_{1,i}dB_{0,i}' + \Delta_{10,i} - \left(\int_0^1 \tilde{B}_{1,i}dB_{2,i}' + \Delta_{12,i}\right)\Sigma_{22,i}^{-1}\Sigma_{20,i}\right]$,

(v) $\sqrt{T}\mathbb{S}_2\left(\tilde{\beta}_{2,i} - \beta_{2,i}^*\right) \Rightarrow \mathbb{S}_2\left(\Sigma_{22,i}^{-1}J_{1,i} \otimes J_{2,i}\right) N\left(0, V_i^0\right)$,

where $\tilde{B}_{1,i} = B_{1,i} - \int_0^1 B_{1,i}(r)dr$, $\Delta_{10,i} = \Sigma_{10,i} + \Lambda_{10,i}$, $J_{1,i} = \left(\mathbf{0}_{p_2\times 1}, \mathbf{0}_{p_2\times p_1}, I_{p_2}\right)$, $J_{2,i} = \left(1, \mathbf{0}_{1\times p_1}, -\Sigma_{20,i}'\Sigma_{22,i}^{-1}\right)$, *and* $V_i^0 = \lim_{T\to\infty} Var(T^{-1/2}\sum_{t=1}^T vec(\varepsilon_{it}\varepsilon_{it}' - \Sigma_i))$.

Lemma A.2 *Suppose that Assumptions A.1-A.3 hold. Then for any fixed constant $c > 0$,*

(i) $P\left(\max_{1\le i\le N}\frac{1}{T^2}\left\|\tilde{x}_{1,i}'\tilde{u}_i\right\| \ge ca_{1NT}\right) = o\left(N^{-1}\right)$,

(ii) $P\left(\max_{1\le i\le N}\left\|\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i - \Sigma_{20,i}\right\| \ge cp_2^{1/2}a_{2NT}\right) = o\left(N^{-1}\right)$,

(iii) $P\left(\max_{1\le i\le N}\frac{1}{T^2}\left\|\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\| \ge cp_2^{1/2}a_{1NT}\right) = o\left(N^{-1}\right)$,

(iv) $P\left(\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^T \tilde{x}_{2,it}\tilde{x}_{2,it}' - \Sigma_{22,i}\right\| \ge cp_2a_{2NT}\right) = o\left(N^{-1}\right)$,

(v) $P\left(\max_{1\le i\le N}\left\|\hat{Q}_{i,\tilde{x}_1\tilde{u}^*}\right\| \ge ca_{1NT}\right) = o\left(N^{-1}\right)$,

(vi) $P\left(\max_{1\le i\le N}\left\|T\hat{Q}_{i,\tilde{x}_2\tilde{u}^*}\right\| \ge cp_2^{1/2}a_{2NT}\right) = o\left(N^{-1}\right)$.

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

(i) $\lim\sup_{T\to\infty} \lambda_{\max}\left(\frac{1}{2T^2\log\log T}\tilde{x}_{1,i}\tilde{x}_{1,i}'\right) \leq \left(\frac{1}{2}+c\right)\bar{c}_{\Omega_{11}}$ *a.s. for any fixed small constant* $c>0$,

(ii) $P\left(\min_{1\leq i\leq N}\lambda_{\min}(T\hat{Q}_{i,\tilde{x}_2\tilde{x}_2}) \geq \underline{c}_{22}/2\right) = 1-o\left(N^{-1}\right)$,

(iii) $P\left(\min_{1\leq i\leq N}\lambda_{\min}(D_T\hat{Q}_{i,\tilde{x}\tilde{x}}D_T) \geq \underline{c}_{11}/(2b_T)\right) = 1-o\left(N^{-1}\right)$.

Lemma A.4 *Suppose that Assumptions A.1-A.3 hold. Then for any constant* $c>0$,

(i) $P\left(\max_{1\leq i\leq N}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i} - \frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{1,i}\right\| > cb_T^{-1}\right) = o\left(N^{-1}\right)$,

(ii) $P\left(\max_{1\leq i\leq N}\left\|\frac{1}{T}\tilde{x}_{2,i}'M_{1,i}\tilde{x}_{2,i} - \Sigma_{22,i}\right\| > c\ p_2 a_{2NT}\right) = o\left(N^{-1}\right)$,

(iii) $P\left(\max_{1\leq i\leq N}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i^*\right\| > ca_{1NT}\right) = o\left(N^{-1}\right)$,

(iv) $P\left(\max_{1\leq i\leq N}\left\|\frac{1}{T}\tilde{x}_{2,i}'M_{1,i}\tilde{u}_i^*\right\| > cp_2^{1/2}a_{2NT}\right) = o\left(N^{-1}\right)$.

Lemma A.5 *Suppose that Assumptions A.1-A.3 hold. Then for any* $\epsilon>0$,

(i) $P\left(\max_{1\leq i\leq N}\left\|\tilde{\beta}_{1,i}-\beta_{1,i}^0\right\| > c\ b_T a_{1NT}\right) = o\left(N^{-1}\right)$,

(ii) $P\left(\max_{1\leq i\leq N}\left\|\tilde{\beta}_{2,i}-\beta_{2,i}^*\right\| > cp_2^{1/2}a_{2NT}\right) = o\left(N^{-1}\right)$,

(iii) $P\left(\max_{1\leq i\leq N}\left\|\tilde{\sigma}_i^2-\Sigma_{0.2,i}^*\right\| > \epsilon\right) = o\left(N^{-1}\right)$,

where recall that $\Sigma_{0.2,i}^* = \Sigma_{00,i}-\Sigma_{02,i}\Sigma_{22,i}^{-1}\Sigma_{20,i}$.

Lemma A.6 *Suppose that Assumptions A.1-A.3 hold. Then*

(i) $\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{u}_i^*\right\|^2 = O_P\left(T^{-2}\right)$,

(ii) $\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T^{3/2}}\tilde{x}_{2,i}'\tilde{u}_i^*\right\|^2 = O_P\left(p_2T^{-2}\right)$,

(iii) $\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{1,i}\right\|^2 = O_P\left(1\right)$,

(iv) $\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\|^2 = O_P\left(p_2T^{-2}\right)$

(v) $\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T^2}\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i^*\right\|^2 = O_P\left(T^{-2}\right)$.

To study the asymptotic distributions of the post-Lasso estimators $\hat{\alpha}_k^{\text{post}}$, we let $Q_{k,NT} = \frac{1}{N_kT^2}\sum_{i\in G_k^0}\tilde{x}_{1,i}' \times M_{2,i}\tilde{x}_{1,i}$ and $V_{k,NT} = \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i$ for $k=1,...,K$. We make the following decomposition for $V_{k,NT} = \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i$:

$$\begin{aligned}
V_{k,NT} &= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'\left(\tilde{u}_i-\tilde{x}_{2,i}\Sigma_{22,i}^{-1}\Sigma_{20,i}\right) + \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'\tilde{x}_{2,i}\Sigma_{22,i}^{-1}(\Sigma_{20,i}-\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i) \\
&\quad + \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'\tilde{x}_{2,i}\left[\Sigma_{22,i}^{-1}-\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\right]\Sigma_{20,i} \\
&\quad + \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'\tilde{x}_{2,i}\left[\Sigma_{22,i}^{-1}-\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\right]\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i-\Sigma_{20,i}\right) \\
&\equiv V_{1k,NT}+V_{2k,NT}+V_{3k,NT}+V_{4k,NT}.
\end{aligned}$$

The following lemma studies the asymptotic properties of $Q_{k,NT}$, $V_{\ell k,NT}$ for $\ell=1,2,3,4$, and $V_{k,NT}$.

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

(i) $Q_{k,NT}\xrightarrow{P}\mathbb{Q}_{(k)}$,

(ii) $V_{1k,NT}-\mathbb{B}_{k,NT}\Rightarrow N\left(0,\mathbb{V}_{(k)}\right)$,

(iii) $V_{2k,NT} = o_P\left(1\right)$,

(iv) $V_{3k,NT} = o_P\left(1\right)$

(v) $V_{4k,NT} = o_P(1)$,

(vi) $V_{k,NT} - \mathbb{B}_{k,NT} \Rightarrow N\left(0, \mathbb{V}_{(k)}\right)$,

where $\mathbb{Q}_{(k)}$, $\mathbb{B}_{k,NT}$, *and* $\mathbb{V}_{(k)}$ *are as defined before Theorem 4.4.*

To consider the DOLS estimator. Let $\tilde{v}_i^a = \left(\tilde{v}_{i,\bar{p}_2+1}^a, ..., \tilde{v}_{i,T-\bar{p}_2}^a\right)'$, $\tilde{v}_{it}^a = v_{it}^a - \frac{1}{T-2\bar{p}_2}\sum_{t=\bar{p}_2+1}^{T-\bar{p}_2} v_{it}^a$, where $v_{it}^a = \sum_{|j|\geq \bar{p}_2} \gamma_{i,j}' \Delta x_{1,i,t-j}$ signifies the approximation error. Adjust the definitions of $\tilde{x}_{1,i}$ and $M_{2,i}$ to use the time series observations $x_{\ell,i} = (x_{\ell,i,\bar{p}_2+1}, ..., x_{\ell,i,T-\bar{p}_2})'$, $\ell = 1, 2$, where recall that $x_{2,it} = (\Delta x_{1,i,t-\bar{p}_2+1}', ..., \Delta x_{1,i,t+\bar{p}_2}')'$.

Lemma A.8 *Let the conditions in Theorem 4.6 hold. Then* $\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0} \tilde{x}_{1,i}' M_{2,i}\tilde{v}_i^a = o_P(1)$.

Proof of Theorem 4.1 (i) First, noting that $\beta_{2,i}$'s do not enter the penalty term in the PLS objective function in (2.7), we can concentrate them out to obtain the following objective function

$$Q_{NT,\lambda}^{K,c}(\boldsymbol{\beta}_1, \boldsymbol{\alpha}) = \frac{1}{N}\sum_{i=1}^{N} Q_{NT,i}^c(\beta_{1,i}) + \frac{\lambda}{N}\sum_{i=1}^{N}(\tilde{\sigma}_i)^{2-K}\prod_{k=1}^{K} \parallel \hat{Q}_{1i}(\beta_{1,i} - \alpha_k) \parallel . \tag{A.1}$$

where $Q_{NT,i}^c(\beta_{1,i}) = \frac{1}{T^2}\left\|M_{2,i}(\tilde{y}_i - \tilde{x}_{1,i}\beta_{1,i})\right\|^2$. Let $Q_{NTi,\lambda}^{K,c}(\beta_{1i}, \boldsymbol{\alpha}) = Q_{NT,i}^c(\beta_{1,i}) + \lambda(\tilde{\sigma}_i)^{2-K}\prod_{k=1}^K ||\hat{Q}_{1i}(\beta_{1,i} - \alpha_k)||$. Then $Q_{NT,\lambda}^{K,c}(\boldsymbol{\beta}_1, \boldsymbol{\alpha}) = \frac{1}{N}\sum_{i=1}^N Q_{NTi,\lambda}^{K,c}(\beta_{1,i}, \alpha)$. Let $\hat{b}_{1,i} = \hat{\beta}_{1,i} - \beta_{1,i}^0$ and $\hat{b}_{2,i} = \hat{\beta}_{2,i} - \beta_{2,i}^*$. Noting that $M_{2,i}(\tilde{y}_i - \tilde{x}_{1,i}\beta_{1,i}) = M_{2,i}\left[\tilde{u}_i - \tilde{x}_{1,i}(\beta_{1,i} - \beta_{1,i}^0)\right]$, we have

$$\begin{aligned} Q_{NT,i}(\hat{\beta}_{1,i}) - Q_{NT,i}(\beta_{1,i}^0) &= \frac{1}{T^2}\left\|M_{2,i}(\tilde{u}_i - \tilde{x}_{1,i}\hat{b}_{1,i})\right\|^2 - \frac{1}{T^2}\left\|M_{2,i}\tilde{u}_i\right\|^2 \\ &= \hat{b}_{1,i}'\breve{Q}_{i,\tilde{x}_1\tilde{x}_1}\hat{b}_{1,i} - 2\hat{b}_{1,i}'\breve{Q}_{i,\tilde{x}_1\tilde{u}}, \end{aligned} \tag{A.2}$$

where $\breve{Q}_{i,\tilde{x}_1\tilde{x}_1} = \frac{1}{T^2}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i}$ and $\breve{Q}_{i,\tilde{x}_1\tilde{u}} = \frac{1}{T^2}\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i$. By the triangle and reverse triangle inequalities, the fact that $\|Ab\| \leq \|A\|_{\rm sp}\|b\|$ for conformable matrix A and vector b, we have

$$\begin{aligned} &\left|\prod_{k=1}^{K}\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \alpha_k)\| - \prod_{k=1}^{K}\|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_k)\|\right| \\ \leq &\left|\prod_{k=1}^{K-1}\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \alpha_k)\|\{\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \alpha_K)\| - \|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_K)\|\}\right| \\ &+\left|\prod_{k=1}^{K-2}\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \alpha_k)\|\|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_K)\|\{\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \alpha_{K-1})\| - \|\hat{Q}_{1i}(\beta_{1i}^0 - \alpha_{K-1})\|\}\right| \\ &+ ... \\ &+\left|\prod_{k=2}^{K}\|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_k)\|\{\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \alpha_1)\| - \|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_1)\|\}\right| \\ \leq &\hat{c}_{i,NT}(\boldsymbol{\alpha})\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \beta_{1,i}^0)\| \leq \hat{c}_{i,NT}(\boldsymbol{\alpha})\|\hat{Q}_{1i}\|_{\rm sp}\|\hat{b}_{1,i}\|, \end{aligned} \tag{A.3}$$

where $\hat{c}_{i,NT}(\boldsymbol{\alpha}) = \prod_{k=1}^{K-1}\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \alpha_k)\| + \prod_{k=1}^{K-2}\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \alpha_k)\|\|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_K)\| + ... + |\prod_{k=2}^K\|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_k)\| = O_p(1)$ as $||\hat{Q}_{1i}||_{\rm sp} = O_P(1)$. Since $\hat{\beta}_{1,i}$ minimize $Q_{NTi,\lambda}^{K,c}$, we have $Q_{NTi,\lambda}^{K,c}(\hat{\beta}_{1,i}, \hat{\alpha}) - Q_{NTi,\lambda}^{K,c}(\beta_{1,i}^0\hat{\alpha}) \leq 0$. Combining with (A.2)-(A.3), we have

$$\hat{b}_{1,i}'\breve{Q}_{i,\tilde{x}_1\tilde{x}_1}\hat{b}_{1,i} \leq 2\hat{b}_{1,i}'\breve{Q}_{i,\tilde{x}_1\tilde{u}} + \lambda(\tilde{\sigma}_i)^{2-K}\hat{c}_{i,NT}(\alpha)\|\hat{Q}_{1i}\|_{\rm sp}\|\hat{b}_{1,i}\|.$$

Then

$$\underline{c}_{i,\tilde{x}_1\tilde{x}_1}\|\hat{b}_{1,i}\| \leq \|2\check{Q}_{i,\tilde{x}_1\tilde{u}}\| + \lambda(\tilde{\sigma}_i)^{2-K}\hat{c}_{i,NT}(\hat{\alpha})\|\hat{Q}_{1i}\|_{\text{sp}}, \tag{A.4}$$

where $\underline{c}_{i,\tilde{x}_1\tilde{x}_1} = \lambda_{\min}(\check{Q}_{i,\tilde{x}_1\tilde{x}_1}) = \lambda_{\min}(\hat{Q}_{i,\tilde{x}_1\tilde{x}_1} - T^{1/2}\hat{Q}_{i,\tilde{x}_1\tilde{x}_2}(T\hat{Q}_{i,\tilde{x}_2\tilde{x}_2})^{-1}T^{1/2}\hat{Q}_{i,\tilde{x}_2\tilde{x}_1}) \geq \lambda_{\min}(\hat{Q}_{i,\tilde{x}_1\tilde{x}_1}) - o_P(1)$ is bounded away from zero in probability by Lemma A.1(i). In fact, we can apply Lemmas A.2(iii)-(iv) and Assumptions A.2(i), A.2(iii), and A.3(iii)-(iv) and show that

$$P\left(\min_i b_T\underline{c}_{i,\tilde{x}_1\tilde{x}_1} \geq \underline{c}_{11}/2\right) = 1 - o\left(N^{-1}\right). \tag{A.5}$$

Then, by Lemmas A.1(i), A.2(iv), A.5(iii), and Assumption A.2(iii),

$$\|\hat{b}_{1,i}\| \leq \underline{c}_{i,\tilde{x}_1\tilde{x}_1}^{-1}\left(2\|\check{Q}_{i,\tilde{x}_1\tilde{u}}\| + \lambda(\tilde{\sigma}_i)^{2-K}\hat{c}_{i,NT}(\hat{\alpha})\|\hat{Q}_{1i}\|_{\text{sp}}\right) = O_P(T^{-1}+\lambda), \tag{A.6}$$

because

$$\begin{aligned}
\|\check{Q}_{i,\tilde{x}_1\tilde{u}}\| &= \frac{1}{T^2}\left\|\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i\right\| = \frac{1}{T^2}\left\|\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i^*\right\| \\
&= \left\|\hat{Q}_{i,\tilde{x}_1\tilde{u}^*} - \hat{Q}_{i,\tilde{x}_1\tilde{x}_2}(\hat{Q}_{i,\tilde{x}_2\tilde{x}_2})^{-1}\hat{Q}_{i,\tilde{x}_1\tilde{u}^*}\right\| \\
&\leq \left\|\hat{Q}_{i,\tilde{x}_1\tilde{u}^*}\right\| + T^{-1}\left\|T\hat{Q}_{i,\tilde{x}_1\tilde{x}_2}\right\|\left\|T\hat{Q}_{i,\tilde{x}_2\tilde{u}^*}\right\|\left\|(T\hat{Q}_{i,\tilde{x}_2\tilde{x}_2})^{-1}\right\| = O_P\left(T^{-1}\right).
\end{aligned}$$

Now, noting that $\tilde{y}_i - \tilde{x}_{1,i}\hat{\beta}_{1,i} = \tilde{u}_i^* + \tilde{x}_{2,i}\beta_{2,i}^* - \tilde{x}_{1,i}\hat{b}_{1,i}$ and $\hat{\beta}_{2,i} = \left(\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\tilde{x}_{2,i}'(\tilde{y}_i - \tilde{x}_{1,i}\hat{\beta}_{1,i}) = \beta_{2,i}^* + \left(\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\tilde{x}_{2,i}'(\tilde{u}_i^* - \tilde{x}_{1,i}\hat{b}_{1,i})$, we have

$$\begin{aligned}
\left\|\hat{b}_{2,i}\right\| &= \left\|\hat{\beta}_{2,i} - \beta_{2,i}^*\right\| \leq \left\|\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\right\|_{\text{sp}}\left\{\frac{1}{T}\left\|\tilde{x}_{2,i}'\tilde{u}_i^*\right\| + \frac{1}{T}\left\|\tilde{x}_{2,i}'\tilde{x}_{1,i}\right\|\left\|\hat{b}_{1,i}\right\|\right\} \\
&= O_P(1)\left\{O_P(p_2^{1/2}T^{-1/2}) + O_P(p_2^{1/2})O_P(T^{-1}+\lambda)\right\} = O_P(p_2^{1/2}(T^{-1/2}+\lambda)), \qquad \text{(A.7)}
\end{aligned}$$

as we can readily show that $\|(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i})^{-1}\|_{\text{sp}} = O_P(1)$ given Lemma A.2(iv) and Assumption A.2(iii), and that $\frac{1}{T}\left\|\tilde{x}_{2,i}'\tilde{u}_i^*\right\| = O_P(p_2^{1/2}T^{-1/2})$ and $\frac{1}{T}\left\|\tilde{x}_{2,i}'\tilde{x}_{1,i}\right\| = O_P(p_2^{1/2})$ as in the proof of Lemma A.1(i)-(iii).

(ii) By the Minkowski's inequality, as $(N,T) \to \infty$ we have

$$\begin{aligned}
\hat{c}_{i,NT}(\boldsymbol{\alpha}) &\leq \prod_{k=1}^{K-1}\left\{\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \beta_{1,i}^0)\| + \|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_k)\|\right\} \\
&\quad + \prod_{k=1}^{K-2}\left\{\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \beta_{1,i}^0)\| + \|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_k)\|\right\}\|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_K)\| + \ldots + \prod_{k=2}^{K}\|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_k)\| \\
&= \sum_{s=0}^{K-1}\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \beta_{1,i}^0)\|^s\prod_{k=1}^{s}a_{ks}\|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_k)\|^{K-1-s} \\
&\leq C_{K,NT}(\alpha)\sum_{s=0}^{K-1}\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \beta_{1,i}^0)\|^s \leq C_{K,NT}(\alpha)(1 + 2\|\hat{Q}_{1i}\|_{\text{sp}}\|\hat{b}_{1,i}\|), \qquad \text{(A.8)}
\end{aligned}$$

where a_{ks}'s are finite integers and $C_{K,NT}(\boldsymbol{\alpha}) = \max_i \max_{1\leq s\leq k\leq K-1}\prod_{k=1}^{s}a_{ks}\|\hat{Q}_{1i}(\beta_{1,i}^0 - \alpha_k)\|^{K-1-s}$ $= \max_{1\leq l\leq K}\max_{1\leq s\leq k\leq K-1}\prod_{k=1}^{s}a_{ks}\|\hat{Q}_{1i}(\alpha_l^0 - \alpha_k)\|^{K-1-s} = O(1)$ as K is finite. Let $\hat{C}_K = C_{KNT}(\hat{\alpha})$. By Lemmas A.3(i) and (iii) and Assumption A.3(iv) , $2\lambda\hat{C}_K(\tilde{\sigma}_i)^{2-K}\underline{c}_{i,\tilde{x}_1\tilde{x}_1}^{-1}\|\hat{Q}_{1i}\|_{\text{sp}}^2 = O_P(\lambda b_T \log\log T) =$

$o_P(1)$ uniformly in i. Combining (A.6) and (A.8) yields

$$\|\hat{b}_{1,i}\| \leq \frac{\underline{c}_{i,\tilde{x}_1\tilde{x}_1}^{-1}}{1-c_{NT}} \left\{ \|2\breve{Q}_{i,\tilde{x}_1\tilde{u}}\| + \lambda \hat{C}_K (\tilde{\sigma}_i)^{2-K} \|\hat{Q}_{1i}\|_{\text{sp}} \right\},$$

where $c_{NT} = 2\lambda\hat{C}_K \max_i (\tilde{\sigma}_i)^{2-K} \underline{c}_{i,\tilde{x}_1\tilde{x}_1}^{-1} \|\hat{Q}_{1i}\|_{\text{sp}}^2 = o_P(1)$. Then by Lemmas A.5(iii) and A.6(v),

$$\frac{1}{N}\sum_{i=1}^{N} \|\hat{b}_{1,i}\|^2 \leq (\frac{\hat{\underline{c}}_{\tilde{x}_1\tilde{x}_1}}{1-c_{NT}})^2 \frac{1}{N}\sum_{i=1}^{N} [\|2\breve{Q}_{i,\tilde{x}_1\tilde{u}}\| + \lambda\hat{C}_K(\tilde{\sigma}_i)^{2-K}\|\hat{Q}_{1i}\|]^2 = O_P(b_T^2(T^{-2}+\lambda^2)), \tag{A.9}$$

where $\hat{\underline{c}}_{\tilde{x}_1\tilde{x}_1} = \left[\min_i \underline{c}_{i,\tilde{x}_1\tilde{x}_1}\right]^{-1} = O_P(b_T)$ by (A.5).

To refine the result in (A.9), we shall prove that $\frac{1}{N}\sum_{i=1}^{N} ||\hat{b}_{1,i}||^2 = O_P(b_T^2 T^{-2})$. Let $\boldsymbol{\beta}_1^0 = \left(\beta_{1,1}^{0\prime}, ..., \beta_{1,N}^{0\prime}\right)'$ and $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0 + b_T T^{-1}\boldsymbol{\nu}_1$, where $\boldsymbol{\nu}_1 = (v_{1,1}', ..., v_{1,N}')'$ and $\nu_{1,i}$ is a p_1-vector. We want to show that for any given $\epsilon^* > 0$, there exists a large constant $L = L(\epsilon^*)$ such that, for sufficiently large N and T we have

$$P\left\{ \inf_{N^{-1}\sum_{i=1}^{N}\|\nu_{1,i}\|^2 = L} Q_{NT,\lambda}^{K,c}(\boldsymbol{\beta}_1^0 + b_T T^{-1}\boldsymbol{\nu}_1, \hat{\boldsymbol{\alpha}}) > Q_{NT,\lambda}^{K,c}(\boldsymbol{\beta}_1^0, \boldsymbol{\alpha}^0) \right\} \geq 1 - \epsilon^*. \tag{A.10}$$

This implies that w.p.a.1 there is a local minimum $\{\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\alpha}}\}$ such that $\frac{1}{N}\sum_{i=1}^{N} ||\hat{b}_{1,i}||^2 = O_p(b_T^2 T^{-2})$ regardless of the property of $\hat{\boldsymbol{\alpha}}$. By (A.2), Lemma A.3(iii), and the Cauchy-Schwarz inequality, with probability $1 - o\left(N^{-1}\right)$ we have

$$\begin{aligned}
&T^2\left[Q_{NT,\lambda}^{K,c}(\boldsymbol{\beta}_1^0 + b_T T^{-1}\boldsymbol{\nu}_1, \hat{\boldsymbol{\alpha}}) - Q_{NT,\lambda}^{K,c}(\boldsymbol{\beta}_1^0, \boldsymbol{\alpha}^0)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} b_T^2 \nu_{1,i}' \breve{Q}_{i,\tilde{x}_1\tilde{x}_1}\nu_{1,i} - \frac{2T}{N}\sum_{i=1}^{N} b_T \nu_{1,i}'\breve{Q}_{i,\tilde{x}_1\tilde{u}} + \frac{\lambda T^2}{N}\sum_{i=1}^{N}(\tilde{\sigma}_i)^{2-K}\prod_{k=1}^{K}\|\hat{Q}_{1i}(\beta_{1,i}^0 + b_T T^{-1}\nu_{1,i} - \hat{\alpha}_k)\| \\
&\geq \left[\frac{1}{2}\underline{c}_{11}\frac{1}{N}\sum_{i=1}^{N}\|b_T\nu_{1,i}\|^2 - 2\left(\frac{1}{N}\sum_{i=1}^{N}\|b_T\nu_{1,i}\|^2\right)^{1/2}\left(\frac{T^2}{N}\sum_{i=1}^{N}\|\breve{Q}_{i,\tilde{x}_1\tilde{u}}\|^2\right)^{1/2}\right] \\
&\equiv D_{1NT} - D_{2NT}, \text{ say.}
\end{aligned}$$

By Lemma A.6(v), $\frac{T^2}{N}\sum_{i=1}^{N}\|\breve{Q}_{i,\tilde{x}_1\tilde{u}}\|^2 = O_P(1)$. So D_{1NT} dominates D_{2NT} for sufficiently large L. That is, $T^2[Q_{NT,\lambda}^{K,c}(\boldsymbol{\beta}_1^0 + b_T T^{-1}\boldsymbol{\nu}_1, \hat{\boldsymbol{\alpha}}) - Q_{NT,\lambda}^{K,c}(\boldsymbol{\beta}_1^0, \boldsymbol{\alpha}^0)] > 0$ for sufficiently large L. Consequently, we must have $N^{-1}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^2 = O_P(b_T^2 T^{-2})$.

Note that $\left\|(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i})^{-1}\right\|_{\text{sp}} = [\lambda_{\min}(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i})]^{-1}$ and

$$\min_i \lambda_{\min}(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}) \geq \min_i \lambda_{\min}(\Sigma_{22,i}) - \max_i \left\|\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i} - \Sigma_{22,i}\right\| \geq \frac{\underline{c}_{22}}{2} \text{ with probability } 1 - o\left(N^{-1}\right) \tag{A.11}$$

by Lemma A.2(iv) and Assumption A.2(iii). Then we have by (A.7), Lemmas A.2(iii)-(iv) and A.6(ii), and Assumptions A.2(iii) and A.3(iv) that

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{b}_{2,i}\right\|^{2} &\leq 2\max_{i}\left\|\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\right\|_{\rm sp}^{2}\frac{1}{NT^{2}}\sum_{i=1}^{N}\left\{\left\|\tilde{x}_{2,i}'\tilde{u}_{i}^{*}\right\|^{2}+\left\|\tilde{x}_{2,i}'\tilde{x}_{1,i}\right\|^{2}\left\|\hat{b}_{1,i}\right\|^{2}\right\}\\ &\leq O_{P}(1)\left\{\frac{1}{NT^{2}}\sum_{i=1}^{N}\left\|\tilde{x}_{2,i}'\tilde{u}_{i}^{*}\right\|^{2}+\max_{i}\frac{1}{T^{2}}\left\|\tilde{x}_{2,i}'\tilde{x}_{1,i}\right\|^{2}\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{b}_{1,i}\right\|^{2}\right\}\\ &= O_{P}\left(p_{2}T^{-1}\right)+O_{P}(p_{2}a_{1NT}^{2})O_{P}(b_{T}^{2}T^{-2})=O_{P}\left(p_{2}T^{-1}\right).\end{aligned}$$

(iii) Let $P_{NT}(\boldsymbol{\beta}_1,\boldsymbol{\alpha})=\frac{1}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\beta_{1,i}-\alpha_k\|$. By (A.3) and (A.8), as $(N,T)\to\infty$,

$$\begin{aligned}|P_{NT}(\hat{\boldsymbol{\beta}}_1,\boldsymbol{\alpha})-P_{NT}(\boldsymbol{\beta}_1^0,\boldsymbol{\alpha})| &\leq C_{K,NT}(\boldsymbol{\alpha})\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|+2C_{K,NT}(\boldsymbol{\alpha})\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^{2}\\ &\leq C_{K,NT}(\boldsymbol{\alpha})\left(\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{1,i}\|^{2}\right)^{1/2}+O_{P}(b_{T}^{2}T^{-2})=O_{P}(b_{T}T^{-1}). \qquad \text{(A.12)}\end{aligned}$$

By (A.12), and the fact that $P_{NT}(\boldsymbol{\beta}_1^0,\boldsymbol{\alpha}^0)=0$ and that $P_{NT}(\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\alpha}})-P_{NT}(\hat{\boldsymbol{\beta}}_1,\boldsymbol{\alpha}^0)\leq 0$, we have

$$\begin{aligned}0 &\geq P_{NT}(\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\alpha}})-P_{NT}(\hat{\boldsymbol{\beta}}_1,\boldsymbol{\alpha}^0)=P_{NT}(\boldsymbol{\beta}_1^0,\hat{\boldsymbol{\alpha}})-P_{NT}(\boldsymbol{\beta}_1^0,\boldsymbol{\alpha}^0)+O_P(b_TT^{-1})\\ &=\frac{1}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\beta_{1,i}^0-\hat{\alpha}_k\|+O_P(b_TT^{-1})\\ &=\frac{N_1}{N}\prod_{k=1}^{K}\|\hat{\alpha}_k-\alpha_1^0\|+\frac{N_2}{N}\prod_{k=1}^{K}\|\hat{\alpha}_k-\alpha_2^0\|+...+\frac{N_K}{N}\prod_{k=1}^{K}\|\hat{\alpha}_k-\alpha_K^0\|+O_P(b_TT^{-1}). \qquad \text{(A.13)}\end{aligned}$$

By Assumption A.3(i), $N_k/N\to\tau_k\in(0,1)$ for each $k=1,...K$. So (A.13) implies that $\prod_{k=1}^{K}\|\hat{\alpha}_k-\alpha_j^0\|=O_P(b_TT^{-1})$ for $j=1,...,K$. It follows that $(\hat{\alpha}_{(1)},...,\hat{\alpha}_{(K)})-(\alpha_1^0,...,\alpha_K^0)=O_P(b_TT^{-1})$. ■

Proof of Theorem 4.2. (i) By Lemma A.3(i), $\limsup_{T\to\infty}\left\|\hat{Q}_{1i}\right\|_{\rm sp}\leq 2\bar{c}_{\Omega_{11}}\log\log T$ a.s. By Lemma A.3(iii), $P(\min_{1\leq i\leq N}b_T\underline{c}_{i,\tilde{x}_1\tilde{x}_1}\geq\underline{c}_{11}/2)=1-o\left(N^{-1}\right)$. By Lemma A.5(iii) and Assumption A.2(iv), $P\left(\min_{1\leq i\leq N}\tilde{\sigma}_i^2\geq\underline{c}_{00}/2\right)=1-o\left(N^{-1}\right)$. Noting that

$$\left\|\breve{Q}_{i,\tilde{x}_1\tilde{u}}\right\|^2\leq 2\left\|\hat{Q}_{i,\tilde{x}_1\tilde{u}^*}\right\|^2+2\left\|\hat{Q}_{i,\tilde{x}_1\tilde{x}_2}\right\|^2\left\|T\hat{Q}_{i,\tilde{x}_2\tilde{u}^*}\right\|^2\left\|(T\hat{Q}_{i,\tilde{x}_2\tilde{x}_2})^{-1}\right\|_{\rm sp}^2,$$

we can readily apply Lemma A.2(iii)-(v) and Assumptions A.2(iii) and A.3(iii)-(iv) and show that $P(\max_i\left\|\breve{Q}_{i,\tilde{x}_1\tilde{u}}\right\|\geq ca_{1NT})=o(N^{-1})$. Then by (A.4) and (A.8) we can show that $P(\max_{1\leq i\leq N}||\hat{b}_{1,i}||\geq cb_Ta_{1NT})=o(N^{-1})$.

(ii) By (A.7) and (A.11), Lemma A.2(vi), the result in part (i), and Assumption A.3(iii)-(iv)

$$
\begin{aligned}
P\left(\max_i \left\|\hat{b}_{2,i}\right\| \geq c p_2^{1/2} a_{2NT}\right) &\leq P\left(\max_i \left\|\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\right\|_{\text{sp}} \frac{1}{T}\left\{\left\|\tilde{x}_{2,i}'\tilde{u}_i^*\right\| + \left\|\tilde{x}_{2,i}'\tilde{x}_{1,i}\right\|\left\|\hat{b}_{1,i}\right\|\right\} \geq c p_2^{1/2} a_{2NT}\right) \\
&\leq P\left(\max_i \frac{1}{T}\left(\left\|\tilde{x}_{2,i}'\tilde{u}_i^*\right\| + \left\|\tilde{x}_{2,i}'\tilde{x}_{1,i}\right\|\left\|\hat{b}_{1,i}\right\|\right) \geq c p_2^{1/2} a_{2NT}\underline{c}_{22}/2\right) \\
&\quad + P\left(\min_i \lambda\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right) \leq \underline{c}_{22}/2\right) \\
&\leq P\left(\max_i \frac{1}{T}\left\|\tilde{x}_{2,i}'\tilde{u}_i^*\right\| \geq c p_2^{1/2} a_{2NT}\underline{c}_{22}/4\right) \\
&\quad + P\left(\max_i \frac{1}{T}\left\|\tilde{x}_{2,i}'\tilde{x}_{1,i}\right\|\left\|\hat{b}_{1,i}\right\| \geq c p_2^{1/2} a_{2NT}\underline{c}_{22}/4\right) + o\left(N^{-1}\right) \\
&= o\left(N^{-1}\right),
\end{aligned}
$$

where we also use the fact $\max_i \frac{1}{T}\left\|\tilde{x}_{2,i}'\tilde{x}_{1,i}\right\| ||\hat{b}_{1,i}|| = o\left(Ta_{1NT}\right) o\left(b_T a_{1NT}\right) = o(p_2^{1/2} a_{2NT})$ with probability $1 - o\left(N^{-1}\right)$. ■

Proof of Theorem 4.3. We fix $k \in \{1, ..., K\}$. By the consistency of $\hat{\alpha}_k$ and $\hat{\beta}_{1,i}$, we have $\hat{\beta}_{1,i} - \hat{\alpha}_l \to \alpha_k^0 - \alpha_l^0 \neq 0$ for all $i \in G_k^0$ and $l \neq k$. It follows that w.p.a.1 $\left\|\hat{\beta}_{1,i} - \hat{\alpha}_l\right\| \neq 0$ for all $i \in G_k^0$ and $l \neq k$. Note that $\tilde{y}_{it} - \tilde{x}_{1,it}'\hat{\beta}_{1,i} - \tilde{x}_{2,it}'\hat{\beta}_{2,i} = \tilde{u}_{it}^* - \tilde{x}_{1,it}'\hat{b}_{1,i} - \tilde{x}_{2,it}'\hat{b}_{2,i}$.

Now, suppose that $\left\|\hat{\beta}_{1,i} - \hat{\alpha}_k\right\| \neq 0$ for some $i \in G_k^0$. Then the first order condition (with respect to $\beta_{1,i}$) for the minimization problem in (2.7) implies that

$$
\begin{aligned}
0 &= T\frac{\partial Q_{iNT,\lambda}^K(\hat{\boldsymbol{\beta}}_{\mathbf{1}}, \hat{\boldsymbol{\beta}}_{\mathbf{2}}, \hat{\boldsymbol{\alpha}})}{\partial \beta_{1,i}} \\
&= -2\frac{1}{T}\sum_{t=1}^T \tilde{x}_{1,it}(\tilde{y}_{it} - \tilde{x}_{1,it}'\hat{\beta}_{1,i} - \tilde{x}_{2,it}'\hat{\beta}_{2,i}) + T\lambda(\tilde{\sigma}_i)^{2-K}\sum_{j=1}^K \hat{Q}_{1i}\hat{\varrho}_{ij}\prod_{l=1,l\neq j}^K \left\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_l)\right\| \\
&= -\frac{2}{T}\sum_{t=1}^T \tilde{x}_{1,it}\tilde{u}_{it}^* + \left(2 + \frac{\lambda(\tilde{\sigma}_i)^{2-K}\hat{c}_{1,ik}}{\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_k)\right\|}\hat{Q}_{1i}\right) T\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_k) \\
&\quad + 2T\hat{Q}_{i,x_1x_2}\hat{b}_{2,i} + 2T\hat{Q}_{1i}(\hat{\alpha}_k - \alpha_k^0) + T\lambda(\tilde{\sigma}_i)^{2-K}\sum_{j=1,j\neq k}^K \hat{Q}_{1i}\hat{\varrho}_{ij}\prod_{l=1,l\neq j}^K \left\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_l)\right\| \\
&\equiv -\hat{B}_{i1} + \hat{B}_{i2} + \hat{B}_{i3} + \hat{B}_{i4} + \hat{B}_{i5},
\end{aligned} \tag{A.14}
$$

where $\hat{\varrho}_{ij} = \hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_j)/\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_j)\right\|$ if $\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_j)\right\| \neq 0$ and $\left\|\hat{\varrho}_{ij}\right\| \leq 1$ otherwise, $\hat{c}_{1,ik} = \prod_{l=1,l\neq k}^K \left\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_l)\right\| \asymp c_{1,ik}^0 \equiv \prod_{l=1,l\neq k}^K \left\|\hat{Q}_{1i}(\alpha_k^0 - \alpha_l^0)\right\|$ for $i \in G_k^0$ by Theorem 4.1, where $a \asymp b$ signifies that a and b are of the same probability order.

By Theorem 4.2(ii), we can readily show that $P\left(\left\|\hat{\alpha}_k - \alpha_k^0\right\| \geq cb_T a_{1NT}\right) = o\left(N^{-1}\right)$ for any fixed $c > 0$. This, in conjunction with Lemma A.3(i) and Theorem 4.2(i)-(ii), implies that

$$
\left\|\hat{Q}_{1i}\right\|_{\text{sp}} \leq 2\overline{c}_{\Omega_{11}} \log\log T \text{ and } c_k^0\left(\underline{c}_{11}/b_T\right)^{K-1} \leq \hat{c}_{1,ik} \leq c_k^0\left(2\overline{c}_{\Omega_{11}}\log\log T\right)^{K-1} \text{ a.s.,} \tag{A.15}
$$

where $c_k^0 \equiv \prod_{l=1,l\neq k}^{K} \left\|\alpha_k^0 - \alpha_l^0\right\| > 0$ by Assumption A.3(ii). Then

$$P\left(\max_{i\in G_k^0}\left\|\hat{B}_{i5}\right\| \geq CT\lambda\left(\log\log T\right)^K b_T a_{1NT}\right) = o\left(N^{-1}\right) \tag{A.16}$$

for some large constant $C > 0$. By Lemma A.3(i) and Theorem 4.2(iii),

$$\begin{aligned}
& P\left(\max_{i\in G_k^0}\left\|\hat{B}_{i4}\right\| \geq Cb_T T a_{1NT}\log\log T\right) \\
\leq\ & P\left(\max_{i\in G_k^0}\left\|2\hat{Q}_{1i}(\hat{\alpha}_k - \alpha_k^0)\right\| \geq Cb_T a_{1NT}\log\log T,\ \max_{i\in G_k^0}\left\|\hat{Q}_{1i}\right\|_{\text{sp}} \leq 2\overline{c}_{\Omega_{11}}\log\log T\right) \\
& + P\left(\max_{i\in G_k^0}\left\|\hat{Q}_{1i}\right\|_{\text{sp}} \geq 2\overline{c}_{\Omega_{11}}\log\log T\right) \\
\leq\ & P\left(\max_{i\in G_k^0}\left\|\hat{\alpha}_k - \alpha_k^0\right\| \geq Cb_T a_{1NT}/(4\overline{c}_{\Omega_{11}})\right) + 0 = o\left(N^{-1}\right)
\end{aligned} \tag{A.17}$$

for any constant $C > 0$. By Lemma A.2(iii) and Theorem 4.2(ii)

$$P\left(\max_{i\in G_k^0}\left\|\hat{B}_{i3}\right\| \geq CTb_T p_2 a_{1NT} a_{2NT}\right) = P\left(\max_{i\in G_k^0}\left\|2T\hat{Q}_{i,x_1x_2}\hat{b}_{2,i}\right\| \geq Cb_T p_2 a_{1NT} a_{2NT}\right) = o\left(N^{-1}\right). \tag{A.18}$$

By Lemma A.5(iii), Assumptions A.2(i) and A.2(iv), we have with probability $1 - o\left(N^{-1}\right)$

$$\begin{aligned}
\left(\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_k)\right)'\hat{B}_{i2} &= (\hat{\beta}_{1,i} - \hat{\alpha}_k)'\hat{Q}_{1i}\left(2 + \frac{\lambda(\tilde{\sigma}_i)^{2-K}\hat{c}_{1,ik}}{\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_k)\right\|}\hat{Q}_{1i}\right)T\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_k) \\
&\geq T\lambda(\hat{\beta}_{1,i} - \hat{\alpha}_k)'\hat{Q}_{1i}\frac{(\tilde{\sigma}_i)^{2-K}\hat{c}_{1,ik}}{\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_k)\right\|}\hat{Q}_{1i}\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_k) \\
&\geq T\lambda b_T^{-1}\lambda_{\min}\left(b_T\hat{Q}_{1i}\right)(\tilde{\sigma}_i)^{2-K}\hat{c}_{1,ik}\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_k)\right\| \\
&\geq \underline{c}_{11}^K c_k^0(2\overline{c}_{00})^{1-K/2}T\lambda b_T^{-K}\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i} - \hat{\alpha}_k)\right\|.
\end{aligned} \tag{A.19}$$

Define

$$\begin{aligned}
\Gamma_{kNT} \equiv\ & \left\{\underline{c}_{11}c_k^0/b_T \leq \min_{i\in G_k^0}\hat{c}_{1,ik} \leq \max_{i\in G_k^0}\hat{c}_{1,ik} \leq 2c_k^0\overline{c}_{\Omega_{11}}\log\log T\right\} \\
& \cap\left\{\max_{i\in G_k^0}\left\|\hat{B}_{i5}\right\| \leq CT\lambda\left(\log\log T\right)^K b_T a_{1NT}\right\} \cap \left\{\max_{i\in G_k^0}\left\|\hat{B}_{i4}\right\| \leq Cb_T T a_{1NT}\log\log T\right\} \\
& \cap\left\{\max_{i\in G_k^0}\left\|\hat{B}_{i3}\right\| \leq CTb_T p_2 a_{1NT} a_{2NT}\right\}.
\end{aligned}$$

Then $P\left(\Gamma_{kNT}\right) = 1 - o\left(N^{-1}\right)$ by (A.15)-(A.18). Let Γ_{kNT}^c denote the complement of Γ_{kNT}. Conditional on Γ_{kNT}, we have, uniformly in $i \in G_k^0$,

$$
\begin{aligned}
& \left|\left(\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_k)\right)'\left(\hat{B}_{i2}+\hat{B}_{i3}+\hat{B}_{i4}+\hat{B}_{i5}\right)\right| \\
\geq\ & \left|\left(\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_k)\right)'\hat{B}_{i2}\right|-\left|\left(\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_k)\right)'\left(\hat{B}_{i3}+\hat{B}_{i4}+\hat{B}_{i5}\right)\right| \\
\geq\ & \left\{\underline{c}_{11}^K c_k^0(2\bar{c}_{00})^{1-K/2}T\lambda b_T^{-K}-C\left[Tb_Tp_2a_{1NT}a_{2NT}+b_TTa_{1NT}\log\log T+T\lambda\left(\log\log T\right)^K b_Ta_{1NT}\right]\right\} \\
& \times\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_k)\right\| \\
\geq\ & \frac{1}{2}\underline{c}_{11}^K c_k^0(2\bar{c}_{00})^{1-K/2}T\lambda b_T^{-K}\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_k)\right\| \text{ for sufficiently large } (N,T),
\end{aligned}
$$

where the last equality follows because $Tb_Tp_2a_{1NT}a_{2NT}+b_TTa_{1NT}\log\log T+T\lambda\left(\log\log T\right)^K b_Ta_{1NT}=o\left(T\lambda b_T^{-K}\right)$ by Assumption A.3(iv). It follows that for all $i \in G_k^0$,

$$
\begin{aligned}
P(\hat{E}_{kNT,i}) &= P\left(i\notin\hat{G}_k|i\in G_k^0\right)=P\left(\hat{B}_{i1}=\hat{B}_{i2}+\hat{B}_{i3}+\hat{B}_{i4}+\hat{B}_{i5}\right) \\
&\leq P\left(\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_k)\hat{B}_{i1}\right\|\geq\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_k)\left(\hat{B}_{i2}+\hat{B}_{i3}+\hat{B}_{i4}+\hat{B}_{i5}\right)\right\|\right) \\
&\leq P\left(\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_k)\hat{B}_{i1}\right\|\geq\left\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_k)\left(\hat{B}_{i2}+\hat{B}_{i3}+\hat{B}_{i4}+\hat{B}_{i5}\right)\right\|,\Gamma_{kNT}\right) \\
&\quad +P\left(\Gamma_{kNT}^c\right) \\
&\leq P\left(\left\|\hat{B}_{i1}\right\|\geq\frac{1}{2}\underline{c}_{11}^K c_k^0(2\bar{c}_{00})^{1-K/2}T\lambda b_T^{-K}\right)+o\left(N^{-1}\right) \\
&= o\left(1\right),
\end{aligned}
$$

where the last line follows by the fact that $||\hat{B}_{i1}||=O_P\left(1\right)$ by Lemma A.1(ii) and that $T\lambda b_T^{-K}\to\infty$ under Assumption A.3(iv).

In addition, by Lemma A.2(v) and the fact that $a_{1NT}=o\left(\lambda b_T^{-K}\right)$ under Assumption A.3(iv),

$$
\begin{aligned}
P(\cup_{k=1}^K\hat{E}_{kNT}) &\leq \sum_{k=1}^K P(\hat{E}_{kNT})\leq\sum_{k=1}^K\sum_{i\in G_k}P(\hat{E}_{kNT,i}) \\
&\leq \sum_{k=1}^K\sum_{i\in G_k}P\left(\left\|\hat{B}_{i1}\right\|\geq\frac{1}{2}\underline{c}_{11}^K c_k^0(2\bar{c}_{00})^{2-K}T\lambda b_T^{-K}\right)+o\left(1\right) \\
&\leq N\max_{1\leq i\leq N}P\left(\left\|\hat{B}_{i1}\right\|\geq\frac{1}{2}\underline{c}_{11}^K c_k^0(2\bar{c}_{00})^{2-K}T\lambda b_T^{-K}\right)+o(1)=o\left(1\right).
\end{aligned}
\tag{A.20}
$$

We have completed the proof of Theorem 4.3(i).

Given (i), the proof of (ii) is similar to Theorem 4.2(ii) in SSP and thus omitted. ■

Proof of Theorem 4.4. We first write our mixed panel model in vector form: $\tilde{y}_i=\tilde{x}_{1,i}\beta_{1,i}+\tilde{x}_{2,i}\beta_{2,i}+\tilde{u}_i$, where $\tilde{x}_{l,i}=(\tilde{x}_{l,i1},...,\tilde{x}_{l,iT})'$ for $l=1,2$, and $\tilde{y}_i$ and $\tilde{u}_i$ are similarly defined. Recall that $M_{2,i}=I_T-\tilde{x}_{2,i}(\tilde{x}_{2,i}'\tilde{x}_{2,i})^{-1}\tilde{x}_{2,i}'$. Then we rewrite the objective function $Q_{NT,\lambda}^K(\boldsymbol{\beta_1},\boldsymbol{\beta}_2,\boldsymbol{\alpha})$ as follows

$$
Q_{NT,\lambda}^K(\boldsymbol{\beta_1},\boldsymbol{\beta}_2,\boldsymbol{\alpha})=Q_{NT}(\boldsymbol{\beta_1},\boldsymbol{\beta}_2)+\frac{\lambda}{N}\sum_{i=1}^N(\tilde{\sigma}_i)^{2-K}\prod_{k=1}^K\|\hat{Q}_{1i}(\beta_{1,i}-\alpha_k)\|, \tag{A.21}
$$

where

$$Q_{NT}(\boldsymbol{\beta_1},\boldsymbol{\beta}_2) = \frac{1}{NT^2}\sum_{i=1}^{N}(\tilde{y}_i - \tilde{x}_{1,i}\beta_{1,i} - \tilde{x}_{2,i}\beta_{2,i})'(\tilde{y}_i - \tilde{x}_{1,i}\beta_{1,i} - \tilde{x}_{2,i}\beta_{2,i}). \tag{A.22}$$

The first order conditions are

$$\mathbf{0}_{p_1\times 1} = \frac{-2}{T^2}\tilde{x}'_{1,i}(\tilde{y}_i - \tilde{x}_{1,i}\hat{\beta}_{1,i} - \tilde{x}_{2,i}\hat{\beta}_{2,i}) + \lambda(\tilde{\sigma}_i)^{2-K}\sum_{j=1}^{K}\hat{Q}_{1i}\hat{\varrho}_{ij}\prod_{l=1,l\neq j}^{K}\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_l)\|\ \forall\ i=1,...,N, \tag{A.23}$$

$$\mathbf{0}_{p_2\times 1} = \frac{-2}{T^2}\tilde{x}'_{2,i}(\tilde{y}_i - \tilde{x}_{1,i}\hat{\beta}_{1,i} - \tilde{x}_{2,i}\hat{\beta}_{2,i})\ \forall\ i=1,...,N,\ \text{and} \tag{A.24}$$

$$\mathbf{0}_{p_1\times 1} = \frac{\lambda}{N}\sum_{i=1}^{N}(\tilde{\sigma}_i)^{2-K}\hat{Q}_{1i}\hat{\varrho}_{ik}\prod_{l=1,l\neq k}^{K}\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_l)\|\ \forall\ k=1,...,K, \tag{A.25}$$

where $\hat{\varrho}_{ij}$ is defined after (A.14). Let $k\in\{1,...,K\}$ be fixed. We observe that (a) $\|\hat{\beta}_{1,i}-\hat{\alpha}_k\|=0$ for any $i\in\hat{G}_k$ by the definition of $\hat{G}_k$, and (b) $\hat{\beta}_{1,i}-\hat{\alpha}_l\xrightarrow{p}\alpha_k^0-\alpha_l^0\neq 0$ for any $i\in\hat{G}_k$ and $l\neq k$. It follows that $\|\hat{\varrho}_{ij}\|\leq\|1\|$ for any $i\in\hat{G}_k$ and $\hat{\varrho}_{ij}=\hat{Q}_{1i}(\hat{\alpha}_k-\hat{\alpha}_j)/\|\hat{Q}_{1i}(\hat{\alpha}_k-\hat{\alpha}_j)\|$ for any $i\in\hat{G}_k$ and $j\neq k$. Let $\hat{G}_0$ denote the set of unclassified individuals. Given Theorem 4.3, it is easy to show that $P(\#\hat{G}_0>0)=o(1)$. Noting that $\prod_{l=1}^{K}\|\hat{Q}_{1i}(\hat{\alpha}_k-\hat{\alpha}_l)\|=0$ for any l, we have

$$\begin{aligned}&\sum_{i\in\hat{G}_k}\sum_{j=1,j\neq k}^{K}(\tilde{\sigma}_i)^{2-K}\hat{Q}_{1i}\hat{\varrho}_{ij}\prod_{l=1,l\neq j}^{K}\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_l)\|\\&=\sum_{i\in\hat{G}_k}\sum_{j=1,j\neq k}^{K}(\tilde{\sigma}_i)^{2-K}\frac{\hat{Q}_{1i}^2(\hat{\alpha}_k-\hat{\alpha}_j)}{\|\hat{Q}_{1i}(\hat{\alpha}_k-\hat{\alpha}_j)\|}\prod_{l=1,l\neq j}^{K}\|\hat{Q}_{1i}(\hat{\alpha}_k-\hat{\alpha}_l)\|=\mathbf{0}_{p_1\times 1}.\end{aligned} \tag{A.26}$$

It follows that by (A.25) and (A.26)

$$\begin{aligned}\mathbf{0}_{p_1\times 1} &= \sum_{i=1}^{N}(\tilde{\sigma}_i)^{2-K}\hat{Q}_{1i}\hat{\varrho}_{ik}\prod_{l=1,l\neq k}^{K}\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_l)\|\\&=\sum_{i\in\hat{G}_k}(\tilde{\sigma}_i)^{2-K}\hat{Q}_{1i}\hat{\varrho}_{ik}\prod_{l=1,l\neq k}^{K}\|\hat{Q}_{1i}(\hat{\alpha}_k-\hat{\alpha}_l)\|+\sum_{i\in\hat{G}_0}(\tilde{\sigma}_i)^{2-K}\hat{Q}_{1i}\hat{\varrho}_{ik}\prod_{l=1,l\neq k}^{K}\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_l)\|\\&\quad+\sum_{j=1,j\neq k}^{K}\sum_{i\in\hat{G}_j}(\tilde{\sigma}_i)^{2-K}\frac{\hat{Q}_{1i}^2(\hat{\alpha}_j-\hat{\alpha}_k)}{\|\hat{Q}_{1i}(\hat{\alpha}_j-\hat{\alpha}_k)\|}\prod_{l=1,l\neq j}^{K}\|\hat{Q}_{1i}(\hat{\alpha}_j-\hat{\alpha}_l)\|\\&=\sum_{i\in\hat{G}_k}(\tilde{\sigma}_i)^{2-K}\hat{Q}_{1i}\hat{\varrho}_{ik}\prod_{l=1,l\neq k}^{K}\|\hat{Q}_{1i}(\hat{\alpha}_k-\hat{\alpha}_l)\|+\sum_{i\in\hat{G}_0}(\tilde{\sigma}_i)^{2-K}\hat{Q}_{1i}\hat{\varrho}_{ik}\prod_{l=1,l\neq k}^{K}\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_l)\|.\end{aligned} \tag{A.27}$$

Averaging both sides of (A.23) over $i\in\hat{G}_k$ and using (A.26) and (A.27), we have

$$\mathbf{0}_{p_1\times 1} = \frac{2}{N_kT^2}\sum_{i\in\hat{G}_k}\tilde{x}'_{1,i}(\tilde{y}_i-\tilde{x}_{1,i}\hat{\alpha}_k-\tilde{x}_{2,i}\hat{\beta}_{2,i})+\frac{\lambda}{N_k}\sum_{i\in\hat{G}_0}(\tilde{\sigma}_i)^{2-K}\hat{Q}_{1i}\hat{\varrho}_{ik}\prod_{l=1,l\neq k}^{K}\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_l)\|. \tag{A.28}$$

Solving $\hat{\beta}_{2,i}$ from (A.24) as a function of $\hat{\beta}_{1,i}$ and replacing $\hat{\beta}_{1,i}$ by $\hat{\alpha}_k$ for $i \in \hat{G}_k$ yields

$$\hat{\beta}_{2,i} = \left(\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1} \tilde{x}_{2,i}'(\tilde{y}_i - \tilde{x}_{1,i}\hat{\alpha}_k). \tag{A.29}$$

Plugging (A.29) into (A.28) yields

$$\begin{aligned}\hat{\alpha}_k &= \left(\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}' M_{2,i}\tilde{x}_{1,i}\right)^{-1}\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}' M_{2,i}\tilde{y}_i \\ &\quad + \left(\frac{1}{N_k T^2}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}' M_{2,i}\tilde{x}_{1,i}\right)^{-1}\frac{\lambda}{2N_k}\sum_{i\in\hat{G}_0}(\tilde{\sigma}_i)^{2-K}\hat{e}_{ik}\prod_{l=1,l\neq k}^{K}\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_l)\| \\ &\equiv \hat{\alpha}_k^{\text{post}} + \hat{R}_k, \text{ say.}\end{aligned}$$

Noting that $\hat{Q}_{1i}\hat{\varrho}_{ik}\prod_{l=1,l\neq k}^{K}\|\hat{Q}_{1i}(\hat{\beta}_{1,i}-\hat{\alpha}_l)\| \neq 0$ only if $i \in \hat{G}_0$ and by (A.20), we have that for any $\epsilon > 0$

$$P\left(\sqrt{N}T||\hat{R}_k|| \geq \epsilon\right) \leq \sum_{k=1}^{K}\sum_{i\in G_k^0} P(i \in \hat{G}_0 | i \in G_k^0) \leq \sum_{k=1}^{K}\sum_{i\in G_k^0} P(i \notin \hat{G}_k | i \in G_k^0) = o(1).$$

That is, $\sqrt{N}T||\hat{R}_k|| = o_P(1)$ and $\hat{\alpha}_k$ is asymptotically equivalent to its post-Lasso estimator $\hat{\alpha}_{\hat{G}_k}$. Similarly, given the fast convergence rate of $\hat{\alpha}_{\hat{G}_k}$, $\hat{\beta}_{2,i}$ in (A.29) is also asymptotically equivalent to its post-Lasso version $\hat{\beta}_{2,i}^{\text{post}}$, where $\hat{\beta}_{2,i}^{\text{post}} = \left(\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\tilde{x}_{2,i}'(\tilde{y}_i - \tilde{x}_{1,i}\hat{\alpha}_k^{\text{post}})$ for each $i \in \hat{G}_k$. We formally study the asymptotic properties of $\hat{\alpha}_k^{\text{post}}$ and $\hat{\beta}_{2,i}^{\text{post}}$ in the proof of Theorem 4.5 below. ■

Proof of Theorem 4.5. (i) Noting that $\tilde{y}_i = \tilde{x}_{1,i}\beta_{1,i}^0 + \tilde{x}_{2,i}\beta_{2,i}^0 + \tilde{u}_i$, we have

$$\sqrt{N_k}T(\hat{\alpha}_k^{\text{post}} - \alpha_k^0) = \hat{Q}_{(k)}^{-1}\hat{V}_{(k)} + \hat{Q}_{(k)}^{-1}\hat{R}_{(k)},$$

where $\hat{Q}_{(k)} = \frac{1}{N_k T^2}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i}$, $\hat{V}_{(k)} = \frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i$, and $\hat{R}_{(k)} = \frac{1}{\sqrt{N_k}T}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}$ $\times\tilde{x}_{1,i}\left(\beta_{1,i}^0 - \alpha_k^0\right)$. Noting that $\mathbf{1}\{i \in \hat{G}_k\} = \mathbf{1}\{i \in G_k^0\} + \mathbf{1}\{i \in \hat{G}_k \backslash G_k^0\} - \mathbf{1}\{i \in G_k^0 \backslash \hat{G}_k\}$, we have

$$\begin{aligned}\hat{Q}_{(k)} &= \frac{1}{N_k T^2}\sum_{i\in G_k^0}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i} + \frac{1}{N_k T^2}\sum_{i\in\hat{G}_k\backslash G_k^0}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i} - \frac{1}{N_k T^2}\sum_{i\in G_k^0\backslash\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i} \\ &\equiv Q_{k,NT} + \hat{Q}_{1(k)} + \hat{Q}_{2(k)}, \text{ say.}\end{aligned}$$

By Theorem 4.3 $P(||\hat{Q}_{1(k)}|| \geq \epsilon N^{-1/2}T^{-1}) \leq P(\hat{F}_{kNT}) = o(1)$ and $P(||\hat{Q}_{2(k)}|| \geq \epsilon N^{-1/2}T^{-1}) \leq P(\hat{E}_{kNT}) = o(1)$ for any $\epsilon > 0$. It follows that $\hat{Q}_{(k)} = Q_{k,NT} + o_P\left(N^{-1/2}T^{-1}\right)$. Similarly, we can show that $\hat{V}_{(k)} = V_{k,NT} + o_P\left(N^{-1/2}T^{-1}\right)$ and $\hat{R}_{(k)} = o_P\left(N^{-1/2}T^{-1}\right)$, where $V_{k,NT} = \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i$. It follows that $\sqrt{N_k}T(\hat{\alpha}_k^{\text{post}} - \alpha_k^0) = Q_{k,NT}^{-1}V_{k,NT} + o_P(1)$. Then the conclusion in (i) follows from Lemmas A.7(i)-(vi).

(ii) Noting that $\hat{\beta}_{2,i}^{\text{post}} = \left(\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\tilde{x}_{2,i}'(\tilde{y}_i - \tilde{x}_{1,i}\hat{\alpha}_k^{\text{post}})$ and $\tilde{y}_i = \tilde{x}_{1,i}\alpha_k^0 + \tilde{x}_{2,i}\beta_{2,i}^* + \tilde{u}_i^*$ for $i \in G_k^0$, we have for $i \in G_k^0$ and $l \times p_2$ selection matrix $\mathbb{S}_2$,

$$\begin{aligned}\sqrt{T}\mathbb{S}_2\left(\hat{\beta}_{2,i}^{\text{post}}-\beta_{2,i}^*\right) &= \mathbb{S}_2\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\frac{1}{\sqrt{T}}\tilde{x}_{2,i}'\tilde{u}_i^*+\frac{1}{\sqrt{T}}\mathbb{S}_2\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{1,i}T\left(\alpha_k^0-\hat{\alpha}_k^{\text{post}}\right)\\ &= \mathbb{S}_2\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\frac{1}{\sqrt{T}}\tilde{x}_{2,i}'\tilde{u}_i^*+O_P\left(T^{-1/2}\right)\\ &\Rightarrow N\left(0,\mathbb{S}_2\mathbb{V}_{22,i}\mathbb{S}_2'\right)\end{aligned}$$

by (i) and Lemmas A.1(i) and (iii). Here $\mathbb{V}_{22,i}=\left(\Sigma_{22,i}^{-1}J_{1,i}\otimes J_{2,i}\right)V_i^0\left(J_{1,i}'\Sigma_{22,i}^{-1}\otimes J_{2,i}'\right)$. ■

Proof of Theorem 4.6. (i) In vector form, we have the regression model:

$$\tilde{y}_i=\tilde{x}_{1,i}\beta_{1,i}+\tilde{x}_{2,i}\beta_{2,i}+\tilde{v}_i^{\dagger}, \tag{A.30}$$

where $\tilde{x}_{2,i}=\left(\tilde{x}_{2,i,\bar{p}_2+1},...,\tilde{x}_{2,i,T-\bar{p}_2}\right)'$, $\tilde{x}_{2,it}=x_{2,it}-\frac{1}{T-2\bar{p}_2}\sum_{t=\bar{p}_2+1}^{T-\bar{p}_2}x_{2,it}$, $x_{2,it}=(\Delta x_{1,i,t-\bar{p}_2}',...,\Delta x_{1,it}',...,\Delta x_{1,i,t+\bar{p}_2}')'$, and $\tilde{x}_{1,i}$ and $\tilde{v}_i^{\dagger}$ are similarly defined. In particular, a typical element of $\tilde{v}_i$ is given by $\tilde{v}_{it}^{\dagger}=v_{it}^{\dagger}-\frac{1}{T-2\bar{p}_2}\sum_{t=\bar{p}_2+1}^{T-\bar{p}_2}v_{it}^{\dagger}$, where $v_{it}^{\dagger}=v_{it}^a+v_{it}$ and $v_{it}^a=\sum_{|j|\geq\bar{p}_2}\gamma_{i,j}'\Delta x_{1,i,t-j}$ signifies the approximation error.

Assumption A4 ensures the approximation error term v_{it}^a is asymptotically negligible in our asymptotic analysis. Following the proofs of Theorems 4.1-4.4, we can prove that the C-Lasso estimator $\hat{\alpha}_k^D$ of α_k is asymptotically equivalent to its post-Lasso version $\hat{\alpha}_k^{D,\text{post}}$, where

$$\hat{\alpha}_k^{D,\text{post}}=\left(\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i}\right)^{-1}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{y}_i.$$

As in the proof Theorem 4.5, we can show that $\sqrt{N_k}T(\hat{\alpha}_k^{\text{post}}-\alpha_k^0)=Q_{k,NT}^{-1}V_{k,NT}+o_P(1)$, where $Q_{k,NT}=\frac{1}{N_kT^2}\sum_{i\in G_k^0}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i}$ and $V_{k,NT}=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'M_{2,i}\tilde{v}_i^{\dagger}$. Lemma A.7(i) continues to apply: $Q_{k,NT}=\mathbb{Q}_{(k)}+o_P(1)$. Now

$$V_{k,NT}=\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'M_{2,i}\tilde{v}_i+\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'M_{2,i}\tilde{v}_i^a\equiv\mathcal{V}_{k,NT}+\mathcal{V}_{k,NT}^a,\text{ say.}$$

Lemma A.7(ii)-(vi) continues to apply to $V_{k,NT}(1)$ with little modification. Now, v_{it} plays the role of u_{it}^* in the lemma. But since v_{it} is uncorrelated to all lags and leads of $\Delta x_{1,it}=\varepsilon_{1,it}$, s_i defined in Theorem 4.4 becomes $s_i=S_0'-S_2'\Sigma_{22,i}^{-1}\Sigma_{20,i}=S_0'$ as $\Sigma_{20,i}$ is now zero. Then

$$\begin{aligned}\mathbb{B}_{1k,NT} &= \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}S_1\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\psi_{i,s+r}\psi_{i,s}'S_0'=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\sum_{t=0}^{\infty}E\left(\varepsilon_{1,it}v_{i0}\right)=0,\\ \mathbb{B}_{2k,NT} &= \frac{-1}{\sqrt{N_k}}\frac{T+1}{2T}\sum_{i\in G_k^0}S_1\psi_i(1)\psi_i(1)'S_0=\frac{-1}{\sqrt{N_k}}\frac{T+1}{2T}\sum_{i\in G_k^0}\sum_{t=-\infty}^{\infty}E\left(\varepsilon_{1,it}v_{i0}\right)=0.\end{aligned}$$

It follows that $\mathcal{V}_{k,NT}\Rightarrow N(0,\mathbb{V}_{(k)}^{\dagger})$, where $\mathbb{V}_{(k)}^{\dagger}\equiv\lim_{N_k\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0}\frac{1}{6}\Omega_{00,i}^{\dagger}\Omega_{11,i}$,and $\Omega_{00,i}^{\dagger}=\Omega_{00,i}-\Omega_{01,i}\Omega_{11,i}^{-1}\Omega_{10,i}$. In addition, $\mathcal{V}_{k,NT}^a=o_P(1)$ by Lemma A.8. Consequently, $\sqrt{N_k}T(\hat{\alpha}_k^{\text{post}}-\alpha_k^0)\Rightarrow N(0,\mathbb{Q}_{(k)}^{-1}\mathbb{V}_{(k)}^{\dagger}\mathbb{Q}_{(k)}^{-1})$.

(ii) This follows from Theorem 4.5(ii) and the fact that $\Sigma_{20,i}=0$ so that $\beta_{2,i}^*=\beta_{2,i}^0$. ■

Online Supplement to
"Identifying Latent Grouped Patterns in Cointegrated Panels"

Wenxin Huang[a], Sainan Jin[b], and Liangjun Su[b]
[a] Antai College of Economics and Management, Shanghai Jiao Tong University
[b] School of Economics, Singapore Management University

This supplement is composed of four parts. Section B contains the proofs of the technical lemmas in the above paper. Section C studies the determination of the number of groups. Section D provides some details on the practical implementation of the C-Lasso procedure. Sections E and F contain some additional simulation and application results, respectively.

B Proofs of the Technical Lemmas

Proof of Lemma A.1. (i) By Park and Phillips (1988, 1989), we can readily show that

$$\begin{aligned}
\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{1,i} &= \frac{1}{T^2}\sum_{t=1}^{T}x_{1,it}x_{1,it}' - \frac{1}{\sqrt{T}}\bar{x}_{1,i}\frac{1}{\sqrt{T}}\bar{x}_{1,i}' \\
&\Rightarrow \int_0^1 B_{1,i}B_{1,i}' - \int_0^1 B_{1,i}\int_0^1 B_{1,i} = \int_0^1 \tilde{B}_{1,i}\tilde{B}_{1,i}', \\
\frac{1}{T}\tilde{x}_{1,i}'\tilde{x}_{2,i}\mathbb{S}_2' &= \frac{1}{T}\sum_{t=1}^{T}x_{1,it}x_{2,it}'\mathbb{S}_2' - \frac{1}{\sqrt{T}}\bar{x}_{1,i}\sqrt{T}\bar{x}_{2,i}'\mathbb{S}_2 \\
&\Rightarrow \left(\int_0^1 B_{1,i}dB_{2,i}' + \Delta_{12,i}\right)\mathbb{S}_2' - \int_0^1 B_{1,i}B_{2,i}(1)\,\mathbb{S}_2' = \left(\int_0^1 \tilde{B}_{1,i}dB_{2,i}' + \Delta_{12,i}\right)\mathbb{S}_2', \\
\frac{1}{T}\mathbb{S}_2\tilde{x}_{2,i}'\tilde{x}_{2,i}\mathbb{S}_2' &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{S}_2 x_{2,it}x_{2,it}'\mathbb{S}_2' - \mathbb{S}_2\bar{x}_{2,i}\bar{x}_{2,i}'\mathbb{S}_2' = \mathbb{S}_2\Sigma_{22,i}\mathbb{S}_2' + O_P\left(T^{-1/2}\right).
\end{aligned}$$

It follows that $\mathbb{S}D_T\hat{Q}_{i,\tilde{x}\tilde{x}}D_T\mathbb{S}' = \mathbb{S}\begin{pmatrix} \frac{1}{T^2}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}_{1,it}' & \frac{1}{T^{3/2}}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}_{2,it}' \\ \frac{1}{T^{3/2}}\sum_{t=1}^{T}\tilde{x}_{2,it}\tilde{x}_{1,it}' & \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{2,it}\tilde{x}_{2,it}' \end{pmatrix}\mathbb{S}' \Rightarrow \mathbb{S}\begin{pmatrix} \int_0^1 \tilde{B}_{1,i}\tilde{B}_{1,i}' & 0 \\ 0 & \Sigma_{22,i} \end{pmatrix}\mathbb{S}'.$

(ii) and (iii). By Park and Phillips (1988, 1989),

$$\begin{aligned}
\frac{1}{T}\tilde{x}_{1,i}'\tilde{u}_i &= \frac{1}{T}\sum_{t=1}^{T}x_{1,it}u_{it} - \frac{1}{\sqrt{T}}\bar{x}_{1,i}\sqrt{T}\bar{u}_i \\
&\Rightarrow \left(\int_0^1 B_{1,i}dB_{0,i} + \Delta_{10,i}\right) - \int_0^1 B_{1,i}B_{0,i}(1) = \int_0^1 \tilde{B}_{1,i}dB_{0,i}' + \Delta_{10,i},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{S}_2\left(\frac{1}{\sqrt{T}}\tilde{x}_{2,i}'\tilde{u}_i - \sqrt{T}\Sigma_{20,i}\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{S}_2\left(x_{2,it}u_{it} - \Sigma_{20,i}\right) - \frac{1}{\sqrt{T}}\sqrt{T}\mathbb{S}_2\bar{x}_{2,i}\sqrt{T}\bar{u}_i \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{S}_2\left(x_{2,it}u_{it} - \Sigma_{20,i}\right) - O_P\left(T^{-1/2}\right) \\
&\Rightarrow N(0, \mathbb{S}_2 V_{20,i}^0\mathbb{S}_2'),
\end{aligned}$$

where we allow that $\Sigma_{20,i}=E\left(x_{2,it}u_{it}\right)$ to be nonzero and $V^0_{20,i}$ denotes the long-run covariance of $x_{2,it}u_{it}-\Sigma_{20,i}$. It follows that

$$\begin{aligned}
T\hat{Q}_{i,\tilde{x}_1\tilde{u}^*} &= \frac{1}{T}\tilde{x}'_{1,i}\tilde{u}^*_i=\frac{1}{T}\tilde{x}'_{1,i}\tilde{u}_i-\frac{1}{T}\tilde{x}'_{1,i}\tilde{x}_{2,i}\Sigma^{-1}_{22,i}\Sigma_{20,i} \\
&\Rightarrow \int_0^1\tilde{B}_{1,i}dB_{0,i}+\Delta_{10,i}-\left(\int_0^1\tilde{B}_{1,i}dB'_{2,i}+\Delta_{12,i}\right)\Sigma^{-1}_{22,i}\Sigma_{20,i},\text{ and} \\
T^{3/2}\mathbb{S}_2\hat{Q}_{i,\tilde{x}_2\tilde{u}^*} &= \frac{1}{\sqrt{T}}\mathbb{S}_2\tilde{x}'_{2,i}\tilde{u}^*_i=\frac{1}{\sqrt{T}}\mathbb{S}_2\left(\tilde{x}'_{2,i}\tilde{u}_i-\tilde{x}'_{2,i}\tilde{x}_{2,i}\Sigma^{-1}_{22,i}\Sigma_{20,i}\right) \\
&= \mathbb{S}_2\left(\frac{1}{\sqrt{T}}x'_{2,i}u_i-\sqrt{T}\Sigma_{20,i}\right)-\sqrt{T}\mathbb{S}_2\left(\frac{1}{T}x'_{2,i}x_{2,i}-\Sigma_{22,i}\right)\Sigma^{-1}_{22,i}\Sigma_{20,i}+O_P\left(T^{-1/2}\right) \\
&= \mathbb{S}_2J_{1,i}\left(T^{-1/2}\sum_{t=1}^{T}\left(w_{it}w'_{it}-\Sigma_i\right)\right)J'_{2,i}+o_P\left(1\right) \\
&\Rightarrow \mathbb{S}_2\left(J_{1,i}\otimes J_{2,i}\right)N\left(0,V^0_i\right),
\end{aligned}$$

where $J_{1,i}=(\mathbf{0}_{p_2\times 1},\mathbf{0}_{p_2\times p_1},I_{p_2})$ and $J_{2,i}=\left(1,\mathbf{0}_{1\times p_1},-\Sigma'_{20,i}\Sigma^{-1}_{22,i}\right)$. Combining the above results yields the results in (ii)-(iii).

(iv) and (v). Our conditions ensure that $\left\|\frac{1}{T}\tilde{x}'_{2,i}\tilde{x}_{2,i}-\Sigma_{22,i}\right\|=O_P\left(p_2T^{-1/2}\right)=o_P\left(1\right)$ and $\lambda_{\min}(\frac{1}{T}\tilde{x}'_{2,i}\tilde{x}_{2,i})\geq\lambda_{\min}\left(\Sigma_{22,i}\right)-\left\|\frac{1}{T}\tilde{x}'_{2,i}\tilde{x}_{2,i}-\Sigma_{22,i}\right\|\geq\underline{c}_{22}/2$ with probability approaching 1 (w.p.a.1). [See the proof of Lemma A.2(iv) which is related to the former claim.] By (i)-(iii),

$$\begin{aligned}
\frac{1}{T^2}\tilde{x}'_{1,i}M_{2,i}\tilde{x}_{1,i} &= \frac{1}{T^2}\tilde{x}'_{1,i}\tilde{x}_{1,i}-\frac{1}{T}\left(\frac{1}{T}\tilde{x}'_{1,i}\tilde{x}_{2,i}\right)\left(\frac{1}{T}\tilde{x}'_{2,i}\tilde{x}_{2,i}\right)^{-1}\left(\frac{1}{T}\tilde{x}'_{2,i}\tilde{x}_{1,i}\right)\Rightarrow\int_0^1\tilde{B}_{1,i}\tilde{B}'_{1,i}, \\
\frac{1}{T}\mathbb{S}_2\tilde{x}'_{2,i}M_{1,i}\tilde{x}_{2,i}\mathbb{S}'_2 &= \frac{1}{T}\mathbb{S}_2\tilde{x}'_{2,i}\tilde{x}_{2,i}\mathbb{S}'_2-\frac{1}{T}\left(\frac{1}{T}\mathbb{S}_2\tilde{x}'_{2,i}\tilde{x}_{1,i}\right)\left(\frac{1}{T^2}\tilde{x}'_{1,i}\tilde{x}_{1,i}\right)^{-1}\left(\frac{1}{T}\tilde{x}'_{1,i}\tilde{x}_{2,i}\mathbb{S}'_2\right) \\
&= \mathbb{S}_2\Sigma_{22,i}\mathbb{S}'_2+O_P\left(T^{-1}\right), \\
\frac{1}{T}\tilde{x}'_{1,i}M_{2,i}\tilde{u}_i &= \frac{1}{T}\tilde{x}'_{1,i}\tilde{u}_i-\left(\frac{1}{T}\tilde{x}'_{1,i}\tilde{x}_{2,i}\right)\left(\frac{1}{T}\tilde{x}'_{2,i}\tilde{x}_{2,i}\right)^{-1}\left(\frac{1}{T}\tilde{x}'_{2,i}\tilde{u}_i\right) \\
&\Rightarrow \left(\int_0^1\tilde{B}_{1,i}dB'_{0,i}+\Delta_{10,i}\right)-\left(\int_0^1\tilde{B}_{1,i}dB'_{2,i}+\Delta_{12,i}\right)\Sigma^{-1}_{22,i}\Sigma_{20,i},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{S}_2\left(\frac{1}{\sqrt{T}}\tilde{x}'_{2,i}M_{1,i}\tilde{u}_i-\sqrt{T}\Sigma_{20,i}\right) &= \mathbb{S}_2\left(\frac{1}{\sqrt{T}}\tilde{x}'_{2,i}\tilde{u}_i-\sqrt{T}\Sigma_{20,i}\right) \\
&\quad -\frac{1}{\sqrt{T}}\left(\frac{1}{T}\mathbb{S}_2\tilde{x}'_{2,i}\tilde{x}_{1,i}\right)\left(\frac{1}{T^2}\tilde{x}'_{1,i}\tilde{x}_{1,i}\right)^{-1}\left(\frac{1}{T}\tilde{x}'_{1,i}\tilde{u}_i\right) \\
&= \mathbb{S}_2\left(\frac{1}{\sqrt{T}}x'_{2,i}u_i-\sqrt{T}\Sigma_{20,i}\right)-O_P\left(T^{-1/2}\right)\Rightarrow N(0,\mathbb{S}_2V^0_{20,i}\mathbb{S}'_2).
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
T\left(\tilde{\beta}_{1,i}-\beta^0_{1,i}\right) &= \left(\frac{1}{T^2}\tilde{x}'_{1,i}M_{2,i}\tilde{x}_{1,i}\right)^{-1}\frac{1}{T}\tilde{x}'_{1,i}M_{2,i}\tilde{u}_i \\
&\Rightarrow \left(\int_0^1\tilde{B}_{1,i}\tilde{B}'_{1,i}\right)^{-1}\left[\left(\int_0^1\tilde{B}_{1,i}dB'_{0,i}+\Delta_{10,i}\right)-\left(\int_0^1\tilde{B}_{1,i}dB'_{2,i}+\Delta_{12,i}\right)\Sigma^{-1}_{22,i}\Sigma_{20,i}\right],
\end{aligned}$$

and

$$\begin{aligned}
\sqrt{T}\mathbb{S}_2\left(\tilde{\beta}_{2,i}-\beta^*_{2,i}\right) &= \mathbb{S}_2\left(\frac{1}{T}\tilde{x}'_{2,i}M_{1,i}\tilde{x}_{2,i}\right)^{-1}\frac{1}{\sqrt{T}}\tilde{x}'_{2,i}M_{1,i}\tilde{u}_i-\sqrt{T}\Sigma^{-1}_{22,i}\Sigma_{20,i} \\
&= \mathbb{S}_2\left(\frac{1}{T}\tilde{x}'_{2,i}M_{1,i}\tilde{x}_{2,i}\right)^{-1}\left(\frac{1}{\sqrt{T}}\tilde{x}'_{2,i}M_{1,i}\tilde{u}_i-\sqrt{T}\Sigma_{20,i}\right) \\
&\quad -\mathbb{S}_2\left(\frac{1}{T}\tilde{x}'_{2,i}M_{1,i}\tilde{x}_{2,i}\right)^{-1}\sqrt{T}\left(\frac{1}{T}\tilde{x}'_{2,i}M_{1,i}\tilde{x}_{2,i}-\Sigma_{22,i}\right)\Sigma^{-1}_{22,i}\Sigma_{20,i} \\
&= \mathbb{S}_2\Sigma^{-1}_{22,i}\left(\frac{1}{\sqrt{T}}x'_{2,i}u_i-\sqrt{T}\Sigma_{20,i}\right)-\mathbb{S}_2\Sigma^{-1}_{22,i}\sqrt{T}\left(\frac{1}{T}x'_{2,i}x_{2,i}-\Sigma_{22,i}\right)\Sigma^{-1}_{22,i}\Sigma_{20,i}+o_P(1) \\
&= \mathbb{S}_2\Sigma^{-1}_{22,i}J_{1,i}\left(T^{-1/2}\sum_{t=1}^{T}\left(w_{it}w'_{it}-\Sigma_i\right)\right)J'_{2,i}+o_P(1) \\
&\Rightarrow \mathbb{S}_2\left(\Sigma^{-1}_{22,i}J_{1,i}\otimes J_{2,i}\right)N\left(0,V^0_i\right)\equiv N\left(0,\mathbb{S}_2V_i\mathbb{S}'_2\right),
\end{aligned}$$

where $\beta^*_{2,i}=\beta^0_{2,i}+\Sigma^{-1}_{22,i}\Sigma_{20,i}$, $J_{1,i}=\left(\mathbf{0}_{p_2\times 1},\mathbf{0}_{p_2\times p_1},\Sigma^{-1}_{22,i}\right)$, $J_{2,i}=\left(1,\mathbf{0}_{1\times p_1},-\Sigma'_{20,i}\Sigma^{-1}_{22,i}\right)$, and $V_i=(J_{1.i}\otimes J_{2.i})V^0_i\left(J'_{1,i}\otimes J'_{2,i}\right)$. ■

Proof of Lemma A.2. (i) Noting that $\frac{1}{T^2}\tilde{x}'_{1,i}\tilde{u}_i=\frac{1}{T^2}x'_{1,i}u_i-\frac{1}{T}\bar{x}_{1,i}\bar{u}_i$, it suffices to prove (i) by showing that (i1) $P(\max_{1\le i\le N}\frac{1}{T^2}\left\|x'_{1,i}u_i\right\|\ge ca_{1NT})=o\left(N^{-1}\right)$ and (i2) $P(\max_{1\le i\le N}\frac{1}{T}\left\|\bar{x}_{1,i}\bar{u}_i\right\|\ge ca_{1NT})=o\left(N^{-1}\right)$ for any fixed constant $c>0$. Recall that $\varepsilon_{it}=(u_{it},\varepsilon'_{1,it},\varepsilon'_{2,it})'$. Let S_0 and S_1 be $1\times(1+p_1+p_2)$ and $p_1\times(1+p_1+p_2)$ selection matrices such that $S_0\varepsilon_{it}=u_{it}$ and $S_1\varepsilon_{it}=\varepsilon_{1,it}$. Noting that $x_{1,it}=\sum_{s=1}^{t-1}\varepsilon_{1,is}+\varepsilon_{1,it}$ and $\varepsilon_{it}=\psi_{i0}e_{it}+\varepsilon^{\dagger}_{i,t-1}$ where $\varepsilon^{\dagger}_{i,t-1}=\sum_{j=1}^{\infty}\psi_{ij}e_{i,t-j}$, we have

$$\begin{aligned}
\frac{1}{T^2}x'_{1,i}u_i &= \frac{1}{T^2}\sum_{t=1}^{T}x_{1,it}u_{it}=\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\varepsilon_{1,is}+\varepsilon_{1,it}\right)u_{it} \\
&= S_1\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{is}\varepsilon'_{it}S'_0+S_1\frac{1}{T^2}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon'_{it}S'_0 \\
&= S_1\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{is}e'_{it}\psi'_{i0}S'_0+S_1\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{is}\varepsilon^{\dagger\prime}_{i,t-1}S'_0+S_1\frac{1}{T^2}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon'_{it}S'_0 \\
&\equiv b_{1i}+b_{2i}+b_{3i},\ \text{say.}
\end{aligned}\tag{B.1}$$

We prove (i1) by showing that $N\cdot P\left(\max_{1\le i\le N}|b_{li}|\ge ca_{1NT}\right)=o(1)$ for any fixed constant $c>0$ and $l=1,2,3$. For notational simplicity, we assume that $p_1=1$.

We first study b_{1i}. Let $z_{it}=S_1\sum_{s=1}^{t-1}\varepsilon_{is}e'_{it}\psi'_{i0}S'_0$ and $\mathcal{F}_{i,t}=\sigma\left(e_{it},e_{i,t-1},...\right)$, the sigma-field generated by the series $\{e_{it}\}$. Then $b_{1i}=\frac{1}{T^2}\sum_{t=1}^{T}z_{it}$. Noting that $E\left(z_{it}|\mathcal{F}_{i,t-1}\right)=0$ by construction, we want to apply the exponential inequality for martingales (see, e.g., Freedman (1975, Proposition 2.1)). Let $c_{1NT}=N^{2/q}T^{1/2}$. We make the following decomposition:

$$b_{1i}=\frac{1}{T^2}\sum_{t=1}^{T}z_{1it}+\frac{1}{T^2}\sum_{t=1}^{T}z_{2it}-\frac{1}{T^2}\sum_{t=1}^{T}E\left[z_{2it}|\mathcal{F}_{i,t-1}\right]\equiv b_{1i,1}+b_{1i,2}-b_{1i,3},$$

where $z_{1it}=z_{it}\mathbf{1}_{it}-E\left[z_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\right]$, $z_{2it}=z_{it}\bar{\mathbf{1}}_{it}$, $\mathbf{1}_{it}=\mathbf{1}\{|z_{it}|\le c_{1NT}\}$, and $\bar{\mathbf{1}}_{it}=1-\mathbf{1}_{it}$. It suffices to show that $N\cdot P\left(\max_{1\le i\le N}|b_{1i,l}|\ge ca_{1NT}\right)=o(1)$ for $l=1,2,3$.

Let $V_{iT}=\sum_{t=1}^{T}E\left[z^2_{1i,t}|\mathcal{F}_{i,t-1}\right]$, $v_{NT}=N^{2/q}T^2$, and $\mathbf{1}_{it}=\mathbf{1}\{|z_{it}|\le c_{1NT}\}$. Then by the Hölder's and

Jensen's inequalities and the law of iterated expectations,

$$\begin{aligned} E\left(V_{iT}^{q}\right) &= E\left(\sum_{t=1}^{T} E\left[z_{1it}^{2}|\mathcal{F}_{i,t-1}\right]\right)^{q} \leq T^{q-1}\sum_{t=1}^{T} E\left[\left\{E\left[z_{1it}^{2}|\mathcal{F}_{i,t-1}\right]\right\}^{q}\right] \\ &\leq T^{q-1}\sum_{t=1}^{T} E\left|z_{1it}\right|^{2q} \leq 2^{2q}T^{q-1}\sum_{t=1}^{T} E\left|z_{it}\right|^{2q} \leq CT^{q-1}\sum_{t=1}^{T} t^{q} \leq CT^{2q}, \end{aligned}$$

where the fourth inequality follows because

$$E\left|z_{it}\right|^{2q} = E\left|S_1\sum_{s=1}^{t-1}\varepsilon_{is}e_{it}'\psi_{i0}'S_0'\right|^{2q} \leq CE\left\|\sum_{s=1}^{t-1}\varepsilon_{is}\right\|^{2q} E\left\|e_{it}\right\|^{2q} \leq Ct^{q}.$$

Here we use the independence between e_{it} and $\sum_{s=1}^{t-1}\varepsilon_{is}$ and the fact that $E\left\|\sum_{s=1}^{t-1}\varepsilon_{is}\right\|^{2q} \leq Ct^{q}$. [Recall that we allow the constant C to vary across places.] It follows that $N^2\max_{1\leq i\leq N}P\left(V_{iT}>v_{NT}\right)\leq N^2\max_{1\leq i\leq N}v_{NT}^{-q}E\left(V_{iT}^{q}\mathbf{1}\left\{V_{iT}>v_{NT}\right\}\right)=o\left(N^2T^{2q}v_{NT}^{-q}\right)=o\left(1\right).$ By Proposition 2.1 in Freedman (1975), we have

$$\begin{aligned} N\cdot P\left(\max_{1\leq i\leq N}\left|b_{1i,1}\right|\geq ca_{1NT}\right) &\leq N^2\max_{1\leq i\leq N}P\left(\left|\frac{1}{T^2}\sum_{t=1}^{T}z_{1it}\right|\geq ca_{1NT}\right) \\ &\leq N^2\max_{1\leq i\leq N}P\left(\left|\sum_{t=1}^{T}z_{1it}\right|\geq CT^2a_{1NT},V_{iT}\leq v_{NT}\right) \\ &\quad +N^2\max_{1\leq i\leq N}P\left(V_{iT}>v_{NT}\right) \\ &= N^2\cdot\exp\left(-\frac{c^2T^4a_{1NT}^2}{2v_{NT}+4cT^2a_{1NT}c_{1NT}}\right)+o\left(N^2T^{2q}v_{NT}^{-q}\right) \\ &= o\left(1\right)+o\left(1\right)=o\left(1\right) \end{aligned}$$

as $T^4a_{1NT}^2/v_{NT}=\frac{T^4a_{1NT}^2}{N^{2/q}T^2}=T^2N^{-2/q}a_{1NT}^2\geq\left(\log N\right)^{1+\epsilon}$ and $T^2a_{1NT}/c_{1NT}\geq\left(\log N\right)^{1+\epsilon}$ for some $\epsilon>0$ by Assumption A.3(iii).

$$\begin{aligned} N\cdot P\left(\max_{1\leq i\leq N}\left|b_{1i,2}\right|\geq ca_{1NT}\right) &\leq N\cdot P\left(\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left|z_{it}\right|\geq c_{1NT}\right) \\ &\leq \frac{N^2T}{c_{1NT}^{2q}}\cdot\max_{1\leq i\leq N}\max_{1\leq t\leq T}E\left[\left|z_{it}\right|^{2q}\mathbf{1}\left\{\left|z_{it}\right|\geq c_{1NT}\right\}\right] \\ &= o\left(N^2T^{q+1}c_{1NT}^{-2q}\right)=o\left(N^{-2}T\right)=o\left(1\right). \end{aligned}$$

Similarly, we can show that $N\cdot P\left(\max_{1\leq i\leq N}\left|b_{1i,3}\right|\geq ca_{1NT}\right)=o\left(1\right).$ Consequently, we have $N\cdot P\left(\max_{1\leq i\leq N}\left|b_{1i}\right|\geq ca_{1NT}\right)=o\left(1\right).$

To study b_{2i}, we apply the Beveridge-Nelson (BN) decomposition (see, e.g., Lemma 2.1 in Phillips and Solo (1992)) to obtain

$$\varepsilon_{it}=\psi_i\left(1\right)e_{it}+\breve{e}_{i,t-1}-\breve{e}_{it}$$

where $\psi_i\left(1\right)=\sum_{j=0}^{\infty}\psi_{ij}$, $\breve{e}_{it}=\breve{\psi}_i\left(L\right)e_{it}=\sum_{j=0}^{\infty}\breve{\psi}_{ij}e_{i,t-j}$, and $\breve{\psi}_{ij}=\sum_{k=j+1}^{\infty}\psi_{ik}$. Then $\sum_{s=1}^{t-1}\varepsilon_{is}=\psi_i\left(1\right)\sum_{s=1}^{t-1}e_{is}+\breve{e}_{i0}-\breve{e}_{i,t-1}$ and $\varepsilon_{i,t-1}^{\dagger}=\varepsilon_{it}-\psi_{i0}e_{it}=\psi_i^{\dagger}\left(1\right)e_{it}+\breve{e}_{i,t-1}-\breve{e}_{it}$, where $\psi_i^{\dagger}\left(1\right)=\sum_{j=1}^{\infty}\psi_{ij}$.

It follows that

$$
\begin{aligned}
b_{2i} &= S_1 \frac{1}{T^2} \sum_{t=1}^{T} \left(\sum_{s=1}^{t-1} \varepsilon_{is} \right) \varepsilon_{i,t-1}^{\dagger\prime} S_0' \\
&= S_1 \frac{1}{T^2} \sum_{t=1}^{T} \left(\psi_i (1) \sum_{s=1}^{t-1} e_{is} + \breve{e}_{i,0} - \breve{e}_{i,t-1} \right) \left(\psi_i^{\dagger} (1) e_{it} + \breve{e}_{i,t-1} - \breve{e}_{it} \right)' S_0' \\
&= S_1 \frac{1}{T^2} \sum_{t=1}^{T} \psi_i (1) \sum_{s=1}^{t-1} e_{is} e_{it}' \psi_i^{\dagger} (1)' S_0' + S_1 \frac{1}{T^2} \sum_{t=1}^{T} \psi_i (1) \sum_{s=1}^{t-1} e_{is} \left(\breve{e}_{i,t-1} - \breve{e}_{it} \right)' S_0' \\
&\quad + S_1 \frac{1}{T^2} \sum_{t=1}^{T} \left(\breve{e}_{i0} - \breve{e}_{i,t-1} \right) e_{it}' \psi_i^{\dagger} (1)' S_0' + S_1 \frac{1}{T^2} \sum_{t=1}^{T} \left(\breve{e}_{i0} - \breve{e}_{i,t-1} \right) \left(\breve{e}_{i,t-1} - \breve{e}_{it} \right)' S_0' \\
&\equiv b_{2i,1} + b_{2i,2} + b_{2i,3} + b_{2i,4}.
\end{aligned}
$$

Noting that $b_{2i,1} = \frac{1}{T^2} \sum_{t=1}^{T} z_{it}^{\dagger}$ and $b_{2i,3} = \frac{1}{T^2} \sum_{t=1}^{T} z_{it}^{\ddagger}$ where $z_{it}^{\dagger} = S_1 \psi_i (1) \sum_{s=1}^{t-1} e_{is} e_{it}' \psi_i^{\dagger} (1)' S_0'$ and $z_{it}^{\ddagger} = S_1 \left(\breve{e}_{i,0} - \breve{e}_{i,t-1} \right) e_{it}' \psi_i^{\dagger} (1)' S_0'$ satisfy $E \left\{ z_{it}^{\dagger} | \mathcal{F}_{i,t-1} \right\} = 0$ and $E \left\{ z_{it}^{\ddagger} | \mathcal{F}_{i,t-1} \right\} = 0$, we can also follow the analysis of b_{1i} and show that $N \cdot P \left(\max_{1 \le i \le N} |b_{2i,l}| \ge c a_{1NT} \right) = o(1)$ for $l = 1, 3$. It remains to show that $N \cdot P \left(\max_{1 \le i \le N} |b_{2i,l}| \ge c a_{1NT} \right) = o(1)$ for $l = 2, 4$. For $b_{2i,2}$, we have

$$
\begin{aligned}
b_{2i,2} &= S_1 \frac{1}{T^2} \sum_{s=1}^{T-1} \psi_i (1) e_{is} \sum_{t=s+1}^{T} \left(\breve{e}_{i,t-1} - \breve{e}_{it} \right)' S_0' \\
&= S_1 \frac{1}{T^2} \sum_{s=1}^{T-1} \psi_i (1) e_{is} \left(\breve{e}_{is} - \breve{e}_{iT} \right)' S_0' \\
&= S_1 \frac{1}{T^2} \sum_{s=1}^{T-1} \psi_i (1) e_{is} \breve{e}_{is}' S_0' - S_1 \frac{1}{T^2} \sum_{s=1}^{T-1} \psi_i (1) e_{is} \breve{e}_{iT}' S_0' \equiv b_{2i,2a} - b_{2i,2b}.
\end{aligned}
$$

For $b_{2i,2b}$, we have

$$
\begin{aligned}
N \cdot P \left(\max_{1 \le i \le N} |b_{2i,2b}| \ge c a_{1NT} \right) &\le N^2 \max_{1 \le i \le N} P \left(|b_{2i,2b}| \ge c a_{1NT} \right) \\
&\le N^2 T^{-2q} (c a_{1NT})^{-q} \max_{1 \le i \le N} E \left\| S_1 \sum_{s=1}^{T-1} \psi_i (1) e_{is} \breve{e}_{iT}' S_0' \right\|^q \\
&\le C N^2 T^{-2q} (c a_{1NT})^{-q} \max_{1 \le i \le N} \left\{ E \left\| \sum_{s=1}^{T-1} S_1 \psi_i (1) e_{is} \right\|^{2q} E \left\| S_0 \breve{e}_{iT} \right\|^{2q} \right\}^{1/2} \\
&\le C N^2 T^{-2q} (c a_{1NT})^{-q} O(T^{q/2}) = O(N^2 T^{-3q/2} a_{1NT}^{-q}) = o(1).
\end{aligned}
$$

For $b_{2i,2a}$, we can make futher decomposition

$$\begin{aligned}
b_{2i,2a} &= S_1\psi_i(1)\frac{1}{T^2}\sum_{s=1}^{T-1}\Omega_i\breve{\psi}_{i0}'S_0' + S_1\psi_i(1)\frac{1}{T^2}\sum_{s=1}^{T-1}\left(e_{is}e_{is}'-\Omega_i\right)\breve{\psi}_{i0}'S_0' \\
&\quad + S_1\psi_i(1)\frac{1}{T^2}\sum_{s=1}^{T-1}e_{is}\left(\breve{e}_{is}-\breve{\psi}_{i0}e_{is}\right)'S_0' \\
&\equiv b_{2i,2a}(1)+b_{2i,2a}(2)+b_{2i,2a}(3).
\end{aligned}$$

Apparently, $\max_{1\le i\le N}|b_{2i,2a}(1)| \le \frac{1}{T}\max_{1\le i\le N}\left\{\|S_1\psi_i(1)\|\left\|S_0\breve{\psi}_{i0}\right\|\right\}\max_{1\le i\le N}\|\Omega_i\|_{\text{sp}} \le \frac{C}{T}$. Noting that $E\left(e_{is}e_{is}'-\Omega_i|\mathcal{F}_{i,s-1}\right)=0$, we can apply the Markov and Burkholder's inequalities to obtain

$$\begin{aligned}
N\cdot P\left(\max_{1\le i\le N}|b_{2i,2a}(2)|\ge ca_{1NT}\right) &\le N^2\max_{1\le i\le N}P\left(|b_{2i,2a}(2)|\ge ca_{1NT}\right) \\
&\le N^2T^{-2q}(ca_{1NT})^{-q}\max_{1\le i\le N}E\left\|S_1\psi_i(1)\sum_{s=1}^{T-1}\left(e_{is}e_{is}'-\Omega_i\right)\breve{\psi}_{i0}'S_0'\right\|^q \\
&\le CN^2T^{-2q}(ca_{1NT})^{-q}T^{q/2}=O(N^2T^{-3/2q}a_{1NT}^{-q})=o(1).
\end{aligned}$$

Similarly, noting that $E\left[e_{is}(\breve{e}_{is}-\breve{\psi}_{i0}e_{is})|\mathcal{F}_{i,s-1}\right]=0$, we can apply the Markov and Burkholder's inequalities (e.g., Hall and Heyde, 1980, p.23) to obtain $N\cdot P\left(\max_{1\le i\le N}|b_{2i,2a}(3)|\ge ca_{1NT}\right)=o(1)$. Consequently, we have $N\cdot P(\max_{1\le i\le N}|b_{2i,2}|\ge ca_{1NT})=o(1)$.

For $b_{2i,4}$, we make the following decomposition

$$b_{2i,4}=S_1\frac{1}{T^2}\breve{e}_{i0}\left(\breve{e}_{i0}-\breve{e}_{iT}\right)'S_0' - S_1\frac{1}{T^2}\sum_{t=1}^{T}\breve{e}_{i,t-1}\left(\breve{e}_{i,t-1}-\breve{e}_{it}\right)'S_0' \equiv b_{2i,4a}-b_{2i,4b}.$$

By the Markov inequality

$$\begin{aligned}
N\cdot P\left(\max_{1\le i\le N}|b_{2i,4a}|\ge ca_{1NT}\right) &\le N^2\max_{1\le i\le N}P\left(|b_{2i,4a}|\ge ca_{1NT}\right) \\
&\le N^2T^{-2q}(ca_{1NT})^{-q}E\left\|S_1\breve{e}_{i0}\left(\breve{e}_{i0}-\breve{e}_{iT}\right)S_0'\right\| \\
&= O(N^2T^{-2q}a_{1NT}^{-q})=o(1).
\end{aligned}$$

As in the analysis of $b_{2i,2a}$, we can show that $N\cdot P\left(\max_{1\le i\le N}|b_{2i,4b}|\ge ca_{1NT}\right)=o(1)$. Then $N\cdot P\left(\max_{1\le i\le N}|b_{2i,4}|\ge ca_{1NT}\right)=o(1)$. In sum, we have $N\cdot P\left(\max_{1\le i\le N}|b_{2i}|\ge ca_{1NT}\right)=o(1)$.

For b_{3i}, we make the following decomposition

$$\begin{aligned}
b_{3i} &= S_1\frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\psi_{ij}e_{i,t-j}e_{i,t-l}'\psi_{il}'S_0' \\
&= S_1\frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{\infty}\psi_{ij}\Omega_i\psi_{ij}'S_0' + S_1\frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{\infty}\psi_{ij}\left[e_{i,t-j}e_{i,t-j}'-\Omega_i\right]\psi_{ij}'S_0' \\
&\quad + S_1\frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{\infty}\sum_{l=j+1}^{\infty}\psi_{ij}e_{i,t-j}e_{i,t-l}'\psi_{il}'S_0' + S_1\frac{1}{T^2}\sum_{t=1}^{T}\sum_{l=1}^{\infty}\sum_{j=l+1}^{\infty}\psi_{ij}e_{i,t-j}e_{i,t-l}'\psi_{il}'S_0' \\
&\equiv b_{3i,1}+b_{3i,2}+b_{3i,3}+b_{3i,4}.
\end{aligned}$$

It is easy to show that $\max_{1\le i\le N}|b_{3i,1}| \le C/T$. For $b_{3i,l}$ with $l=2,3,4$, by tedious calculations we can show that $E(b_{3,l})^4 \le CT^{-6}$. With this, we can apply the Markov inequality to show that $N\cdot P(\max_{1\le i\le N}|b_{3i,l}|\ge ca_{1NT}) \le N^2(ca_{1NT})^{-4}\max_{1\le i\le N}E|b_{3i,l}|^4 = O\left(N^2T^{-6}a_{1NT}^{-4}\right)=o(1)$ for $l=2,3,4$. Then $N\cdot P(\max_{1\le i\le N}|b_{3i}|\ge ca_{1NT})=o(1)$ for each fixed constant $c>0$.

Consequently, we have shown (i1).

We now show (i2) $P\left(\max_{1\le i\le N}\frac{1}{T}\|\bar{x}_{1,i}\bar{u}_i\|\ge ca_{1NT}\right)=o\left(N^{-1}\right)$ for any fixed constant $c>0$. Noting that by Lemma S1.2 in Su, Shi, and Phillips (2016b, hereafter SSPb),

$$P\left(\max_{1\le i\le N}|\bar{u}_i|\ge cT^{-1/2}(\log T)^3\right)=o\left(N^{-1}\right).$$

Using $\varepsilon_{it}=\psi_i(1)e_{it}+\breve{e}_{i,t-1}-\breve{e}_{it}$,

$$\begin{aligned}\frac{1}{T}\bar{x}_{1,i} &= \frac{1}{T^2}\sum_{t=1}^{T}x_{1,it}=\frac{1}{T^2}\sum_{t=1}^{T}S_1\sum_{s=1}^{t}\varepsilon_{is}\\ &= \frac{1}{T^2}S_1\psi_i(1)\sum_{t=1}^{T}\sum_{s=1}^{t}e_{is}+\frac{1}{T}S_1\breve{e}_{i0}-\frac{1}{T^2}S_1\sum_{t=1}^{T}\breve{e}_{it}\\ &\equiv c_{1,i}+c_{2,i}+c_{3,i},\ \text{say.}\end{aligned}$$

As in the analysis of b_{1i}, we can show that $N\cdot P(\max_{1\le i\le N}|c_{1,i}|\ge ca_{1NT})=o(1)$. By the Markov inequality, we can show that

$$\begin{aligned}N\cdot P\left(\max_{1\le i\le N}|c_{2,i}|\ge ca_{1NT}T^{1/2}(\log T)^{-3}\right) &\le N^2\max_{1\le i\le N}P\left(\|S_1\breve{e}_{i0}\|\ge a_{1NT}T^{3/2}(\log T)^{-3}\right)\\ &= N^2o\left(a_{1NT}^{-2q}T^{-3q}(\log T)^{6q}\right)=o(1).\end{aligned}$$

For $c_{3,i}$, we use the fact that $E\left\|\sum_{t=1}^{T}S_1\breve{e}_{it}\right\|^{2q}\le CT^q$ and the Markov inequality to obtain

$$\begin{aligned}N\cdot P\left(\max_{1\le i\le N}|c_{3,i}|\ge ca_{1NT}T^{1/2}(\log T)^{-3}\right) &\le N^2\max_{1\le i\le N}P\left(\frac{1}{T^2}\left\|\sum_{t=1}^{T}S_1\breve{e}_{it}\right\|\ge ca_{1NT}T^{1/2}(\log T)^{-3}\right)\\ &\le \frac{N^2T^{-4q}}{\left(ca_{1NT}T^{1/2}(\log T)^{-3}\right)^{2q}}O(T^q)\\ &= O\left(N^2a_{1NT}^{-2q}T^{-4q}(\log T)^{6q}\right)=o(1).\end{aligned}$$

Consequently, $P(\max_{1\le i\le N}\frac{1}{T}\|\bar{x}_{1,i}\|\ge ca_{1NT}T^{1/2}(\log T)^{-3})=o\left(N^{-1}\right)$. Let $z_{it}=\sum_{s=1}^{t}S_1\psi_i(1)e_{is}$. Then $c_{1,i}=\frac{1}{T^2}\sum_{t=1}^{T}z_{it}$ where $E(z_{it}|\mathcal{F}_{i,t-1})=z_{i,t-1}$. We can readily follow the analysis of b_{2i} and show that $P(\max_{1\le i\le N}|c_{1,i}|\ge ca_{1NT})=o\left(N^{-1}\right)$. It follows that $N\cdot P(\max_{1\le i\le N}|\bar{x}_{1,i}|\ge ca_{1NT}T^{1/2}(\log T)^{-3})=$

$o(1)$. Then

$$\begin{aligned}
& P\left(\max_{1\le i\le N}\frac{1}{T}\|\bar{x}_{1,i}\bar{u}_i\|\ge ca_{1NT}\right)\\
\le\ & P\left(\max_{1\le i\le N}\frac{1}{T}\|\bar{x}_{1,i}\bar{u}_i\|\ge ca_{1NT},\ \max_{1\le i\le N}|\bar{u}_i|\le T^{-1/2}(\log T)^3\right)+P\left(\max_{1\le i\le N}|\bar{u}_i|\ge T^{-1/2}(\log T)^3\right)\\
\le\ & P\left(\max_{1\le i\le N}\frac{1}{T}\|\bar{x}_{1,i}\|\ge ca_{1NT}T^{1/2}(\log T)^{-3}\right)+o\left(N^{-1}\right)\\
=\ & o\left(N^{-1}\right)+o\left(N^{-1}\right)=o\left(N^{-1}\right).
\end{aligned}$$

(ii) Noting that $\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i=\frac{1}{T}x_{2,i}'u_i-\bar{x}_{2,i}\bar{u}_i$, it suffices to prove (ii) by showing that (i1) $P(\max_{1\le i\le N}||\frac{1}{T}x_{2,i}'u_i-\Sigma_{20,i}||\ge cp_2^{1/2}a_{2NT})=o\left(N^{-1}\right)$ and (i2) $P(\max_{1\le i\le N}\frac{1}{T}\|\bar{x}_{2,i}\bar{u}_i\|\ge cp_2^{1/2}a_{2NT})=o\left(N^{-1}\right)$ for any fixed constant $c>0$, where $a_{2NT}=T^{-1/2}(\log T)^3$. (i1) follows directly from a modification of the proof of Lemma S.1.2 in SSPb. Noting that both $x_{2,it}$ and u_{it} have zero mean, we can follow SSPb and show that $P\left(\max_{1\le i\le N}\|\bar{x}_{2,i}\|\ge cp_2^{1/2}a_{2NT}\right)=o\left(N^{-1}\right)$ and $P\left(\max_{1\le i\le N}|\bar{u}_i|\ge ca_{2NT}\right)=o\left(N^{-1}\right)$, implying that $P\left(\max_{1\le i\le N}\|\bar{x}_{2,i}\bar{u}_i\|\ge ca_{2NT}^2\right)=o\left(N^{-1}\right)$. Consequently, we have $P(\max_{1\le i\le N}||\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i-\Sigma_{20,i}||\ge cp_2^{1/2}a_{2NT})=o\left(N^{-1}\right)$.

(iii) Noting that $E(x_{2,it})=0$, the proof is analogous to that of (i) and thus omitted.

(iv) Note that $\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}-\Sigma_{22,i}=\frac{1}{T}\sum_{t=1}^T\left(x_{2,it}x_{2,it}'-\Sigma_{22,i}\right)-\bar{x}_{2,i}\bar{x}_{2,i}'$. Using Lemma S1.2 in SSPb, we can readily show that $P(\max_{1\le i\le N}||\frac{1}{T}\sum_{t=1}^T\left(x_{2,it}x_{2,it}'-\Sigma_{22,i}\right)||\ge cp_2a_{2NT})=o\left(N^{-1}\right)$ and $P(\max_{1\le i\le N}\|\bar{x}_{2,i}\|\ge cp_2^{1/2}a_{2NT})=o\left(N^{-1}\right)$ for any $c>0$. Thus (iv) follows.

(v) Note that $\hat{Q}_{i,\tilde{x}_1\tilde{u}^*}=\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{u}_i-\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i}\Sigma_{22,i}^{-1}\Sigma_{20,i}$. The condition in Assumption A.2(ii)-(iii) ensures that $\tilde{x}_{2,it}'\Sigma_{22,i}^{-1}\Sigma_{20,i}$ behaves like $\tilde{u}_{it}$ despite the possible divergence of p_2. As a result, part (i) also holds when $\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{u}_i$ is replaced by $\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i}\Sigma_{22,i}^{-1}\Sigma_{20,i}$. Then

$$\begin{aligned}
P\left(\max_{1\le i\le N}\left\|\hat{Q}_{i,\tilde{x}_1\tilde{u}^*}\right\|\ge ca_{1NT}\right) &\le P\left(\max_{1\le i\le N}\left\|\frac{1}{T^2}\tilde{x}_{i,1}'\tilde{u}_i\right\|\ge ca_{1NT}/2\right)\\
&\quad+P\left(\max_{1\le i\le N}\left\|\frac{1}{T^2}\tilde{x}_{i,1}'\tilde{x}_{i,2}\Sigma_{22,i}^{-1}\Sigma_{20,i}\right\|\ge ca_{1NT}/2\right)\\
&= o\left(N^{-1}\right)+o\left(N^{-1}\right)=o\left(N^{-1}\right).
\end{aligned}$$

(vi) Note that $T\hat{Q}_{i,\tilde{x}_2\tilde{u}^*}=\frac{1}{T}\tilde{x}_{2,i}'\left(\tilde{u}_i-\tilde{x}_{2,i}'\Sigma_{22,i}^{-1}\Sigma_{20,i}\right)=\frac{1}{T}x_{2,i}'\left(u_i-x_{2,i}'\Sigma_{22,i}^{-1}\Sigma_{20,i}\right)-\bar{x}_{2,i}'\left(\bar{u}_i-\bar{x}_{2,i}'\Sigma_{22,i}^{-1}\Sigma_{20,i}\right)$. Since $E\left(u_{it}-x_{2,it}'\Sigma_{22,i}^{-1}\Sigma_{20,i}\right)=0$, we can use Lemma S1.2 in SSPb and show that $P(\max_{1\le i\le N}||\frac{1}{T}x_{2,i}'(u_i-x_{2,i}'\Sigma_{22,i}^{-1}\Sigma_{20,i})||\ge cp_2^{1/2}a_{2NT}/2)=o\left(N^{-1}\right)$ for any fixed $c>0$. In addition, $P(\max_{1\le i\le N}||\bar{x}_{2,i}||\ge cp_2^{1/2}a_{2NT})=o\left(N^{-1}\right)$ and $P(\max_{1\le i\le N}||\bar{u}_i||\ge ca_{2NT})=o\left(N^{-1}\right)$, from which we can readily show that $P(\max_{1\le i\le N}||\frac{1}{T}\bar{x}_{2,i}'\left(\bar{u}_i-\bar{x}_{2,i}'\Sigma_{22,i}^{-1}\Sigma_{20,i}\right)||\ge cp_2^{1/2}a_{2NT}/2)=o\left(N^{-1}\right)$. Then (vi) follows. ∎

Proof of Lemma A.3. (i) Let $v\in\mathbb{R}^{p_1}$ be an arbitrary vector such that $\|v\|=1$. Let $d_T=\sqrt{2T\log\log T}$. By arguments used in the proof of Lemma 2.1 of Corradi (1999), we can verify the conditions in Theorem 2 of Eberlein (1986) and obtain

$$\begin{aligned}
\frac{1}{d_{[Tr]}}\Omega_{11,i}^{-1/2}\tilde{x}_{1,i,[Tr]} &= \frac{1}{d_{[Tr]}}\Omega_{11,i}^{-1/2}\left[x_{1,i,[Tr]}-\bar{x}_{1,i}\right]=\frac{1}{d_{[Tr]}}\Omega_{11,i}^{-1/2}\left[\sum_{s=1}^{[Tr]}\varepsilon_{1,is}-\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^t\varepsilon_{1,is}\right]\\
&= \frac{1}{d_{[Tr]}}\tilde{B}_i(Tr)+o_{a.s.}(1)
\end{aligned}$$

for each $r \in [0,1]$. This result can be strengthened to be hold uniform in $r \in [0,1]$ with $d_{[Tr]}$ replaced by d_T. Let $t = [Tr]$ and $\Omega_{11,i}^{-1/2} \sum_{s=1}^{[Tr]} \varepsilon_{1,is} = S_{i,[Tr]} = S_{i,T}(r)$. Define

$$\eta_{i,T}(r) = ([Tr] + 1 - Tr) S_{i,T}(r) + (Tr - [Tr]) S_{i,T+1}(r).$$

Then $\sup_{r\in[0,1]} \frac{1}{d_T} \|S_{i,T}(r) - B_i(Tr)\| = o_{a.s.}(1)$, $\sup_{r\in[0,1]} \frac{1}{d_T} ||S_{i,T}(r) - \eta_{i,T}(r)|| = o_{a.s.}(1)$ as $T \to \infty$, and the set of norm limit points of $\{d_T^{-1} S_{i,T}\}$ and $\{d_T^{-1} \eta_{i,T}\}$ coincides with the set of norm limit points of $\{d_T^{-1} B_{i,T}\}$ with probability one. By Theorem 1 in Strassen (1964), the latter is relatively norm compact with the set of limit points coinciding a.s. with $\mathcal{K}$, where

$$\mathcal{K} = \left\{ f : [0,1] \to \mathbb{R}^{p_1}, f(0) = 0,\ f \text{ is absolutely continuous, } \int_0^1 \left\| \dot{f}(r) \right\|^2 dr \le 1 \right\}.$$

Here $\dot{f}(r) = \partial f(r) / \nabla r$. First, observe that

$$P\left(\limsup_{T\to\infty} \max_{\|v\|=1} v' \Omega_{11,i}^{-1/2} \frac{1}{T d_T^2} \sum_{t=1}^{T} \tilde{x}_{1,it} \tilde{x}_{1,it}' \Omega_{11,i}^{-1/2} v = \limsup_{T\to\infty} \max_{\|v\|=1} \frac{1}{d_T^2} v' \int_0^1 \tilde{\eta}_{i,T}(r) \tilde{\eta}_{i,T}(r)' dr v \right) = 1,$$

where $\tilde{\eta}_{i,T}(r) = \eta_{i,T}(r) - \int_0^1 \eta_{i,T}(r) dr$.

Now, let ϕ_v denote the continuous map from the space of p_1-dimensional continuous functions on $[0,1]$, closed with respect to the sup norm, to the Euclidean space such that $\phi_v(f) = v' \int_0^1 \tilde{f}(r) \tilde{f}(r)' dr v$ where $\tilde{f}(r) = f(r) - \int_0^1 f(r) dr$. By the Corollary of Theorem 3 in Strassen (1964), with probability one $\{\phi_v(d_T^{-1} \eta_{i,T})\}$ is relatively norm compact with the set of norm limit points coinciding almost surely with $\phi_v(\mathbb{K})$. This implies that

$$P\left(\limsup_{T\to\infty} \max_{\|v\|=1} \frac{1}{d_T^2} v' \int_0^1 \tilde{\eta}_{i,T}(r) \tilde{\eta}_{i,T}(r)' dr v = \sup_{f\in\mathcal{K}} \max_{\|v\|=1} \phi_v(f) \right).$$

By the definition of ϕ_v and K,

$$\begin{aligned}
\sup_{f\in\mathcal{K}} \max_{\|v\|=1} \phi_v(f) &= \sup_{f\in\mathcal{K}} \max_{\|v\|=1} v' \int_0^1 \tilde{f}(r) \tilde{f}(r)' dr v \\
&\le \sup_{f\in\mathcal{K}} \max_{\|v\|=1} v' \left(\int_0^1 f(r) f(r)' dr - \int_0^1 f(r) dr \int_0^1 f(r)' dr \right) v \\
&\le \sup_{f\in\mathcal{K}} \max_{\|v\|=1} \int_0^1 \left(\int_0^r v' \dot{f}(s) ds \right)^2 dr \le \sup_{f\in\mathcal{K}} \max_{\|v\|=1} \int_0^1 \left(\int_0^r 1^2 ds \right) \int_0^r \left(v' \dot{f}(s) \right)^2 ds \\
&= \sup_{f\in\mathcal{K}} \max_{\|v\|=1} \int_0^1 r \left(v' \int_0^r \dot{f}(s) \dot{f}(s)' ds v \right) dr \le \int_0^1 r dr = \frac{1}{2},
\end{aligned}$$

where the second inequality follows from the Hölder's inequality, and the third follows from the fact that

$$\max_{\|v\|=1} v' \int_0^r \dot{f}(s) \dot{f}(s)' ds v \le \lambda_{\max} \left(\int_0^r \dot{f}(s) \dot{f}(s)' ds \right) \le \operatorname{tr} \left(\int_0^r \dot{f}(s) \dot{f}(s)' ds \right) \le \int_0^1 \left\| \dot{f}(s) \right\|^2 ds = 1,$$

for any $r \in [0,1]$ and $f \in \mathcal{K}$. It follows that

$$\limsup_{T\to\infty} \lambda_{\max}\left(\Omega_{11,i}^{-1/2}\frac{1}{Td_T^2}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}_{1,it}'\Omega_{11,i}^{-1/2}\right) \leq \frac{1}{2}+c \text{ a.s. for any } c>0,$$

and

$$\begin{aligned}
\limsup_{T\to\infty} \lambda_{\max}\left(\frac{1}{Td_T^2}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}_{1,it}'\right) &= \limsup_{T\to\infty} \lambda_{\max}\left(\Omega_{11,i}^{-1/2}\frac{1}{Td_T^2}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}_{1,it}'\Omega_{11,i}^{-1/2}\Omega_{11,i}\right) \\
&\leq \limsup_{T\to\infty} \lambda_{\max}\left(\Omega_{11,i}^{-1/2}\frac{1}{Td_T^2}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}_{1,it}'\Omega_{11,i}^{-1/2}\right)\max_{1\leq i\leq N}\lambda_{\max}\left(\Omega_{11,i}\right) \\
&\leq \left(\frac{1}{2}+c\right)\bar{c}_{\Omega_{11}} \text{ a.s.,}
\end{aligned}$$

where recall that $\bar{c}_{\Omega_{11}}$ denotes the upper bound for $\lambda_{\max}\left(\Omega_{11,i}\right)$.

(ii) Let v be a $p_2 \times 1$ vector such that $\|v\| = 1$. By Lemma A.2(iv), with probability $1 - o\left(N^{-1}\right)$ we have

$$\begin{aligned}
\min_{1\leq i\leq N}\inf_{\|v\|=1} v'T\hat{Q}_{i,\tilde{x}_2\tilde{x}_2}v &= \min_{1\leq i\leq N}\inf_{\|v\|=1}\left(v'\Sigma_{22,i}v + v'\left(T\hat{Q}_{i,\tilde{x}_2\tilde{x}_2}-\Sigma_{22,i}\right)v\right) \\
&\geq \min_{1\leq i\leq N}\inf_{\|v\|=1} v'\Sigma_{22,i}v_1 - \max_{1\leq i\leq N}\left\|T\hat{Q}_{i,\tilde{x}_2\tilde{x}_2}-\Sigma_{22,i}\right\| \\
&\geq \min_{1\leq i\leq N}\lambda_{\min}(\Sigma_{22,i}) - o(1) \geq \underline{c}_{22}/2.
\end{aligned}$$

(iii) Note that $D_T\hat{Q}_{i,\tilde{x}\tilde{x}}D_T = \begin{pmatrix} \frac{1}{T^2}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}_{1,it}' & \frac{1}{T^{3/2}}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}_{2,it}' \\ \frac{1}{T^{3/2}}\sum_{t=1}^{T}\tilde{x}_{2,it}\tilde{x}_{1,it}' & \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{2,it}\tilde{x}_{2,it}' \end{pmatrix} = \begin{pmatrix} \hat{Q}_{i,x_1x_1} & \sqrt{T}\hat{Q}_{i,x_1x_2} \\ \sqrt{T}\hat{Q}_{i,x_1x_2}' & T\hat{Q}_{i,x_2x_2} \end{pmatrix}$.
Let $v = (v_1', v_2')'$ be a $(p_1+p_2)\times 1$ vector such that $\|v\| = 1$. Then by Lemmas A.2(iii)-(iv) and Assumptions A.2(i), A.2(iii), and A.3(iv), with probability $1 - o\left(N^{-1}\right)$

$$\begin{aligned}
\min_{1\leq i\leq N}\inf_{\|v\|=1} v'D_T\hat{Q}_{i,\tilde{x}\tilde{x}}D_Tv &= \min_{1\leq i\leq N}\inf_{\|v\|=1}\left(v_1'\hat{Q}_{i,\tilde{x}_1\tilde{x}_1}v_1 + v_2'\hat{Q}_{i,\tilde{x}_2\tilde{x}_2}v_2 + 2v_1'\hat{Q}_{i,\tilde{x}_1\tilde{x}_2}v_2\right) \\
&\geq \min_{1\leq i\leq N}\inf_{\|v\|=1}\left(v_1'\hat{Q}_{i,\tilde{x}_1\tilde{x}_1}v_1 + v_2'T\hat{Q}_{i,\tilde{x}_2\tilde{x}_2}v_2\right) - 2\max_{1\leq i\leq N}\left\|\sqrt{T}\hat{Q}_{i,\tilde{x}_1\tilde{x}_2}\right\| \\
&\geq \min_{1\leq i\leq N}\left[\min\lambda_{\min}(\hat{Q}_{i,\tilde{x}_1\tilde{x}_1}),\ \lambda_{\min}(T\hat{Q}_{i,\tilde{x}_2\tilde{x}_2})\right] - 2\max_{1\leq i\leq N}\left\|\sqrt{T}\hat{Q}_{i,\tilde{x}_1\tilde{x}_2}\right\| \\
&\geq \underline{c}_{11}/(2b_T).
\end{aligned}$$

Then (iii) follows. ■

Proof of Lemma A.4. (i) Noting that $\frac{1}{T^2}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i} - \frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{1,i} = -T(\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i})(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i})^{-1} \times(\frac{1}{T^2}\tilde{x}_{2,i}'\tilde{x}_{1,i})$, it suffices to show that

$$\max_{1\leq i\leq N}\left\|T(\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i})(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i})^{-1}(\frac{1}{T^2}\tilde{x}_{2,i}'\tilde{x}_{1,i})\right\|$$

is $o(1)$ with probability $1-o\left(N^{-1}\right)$. This follows because by Lemma A.2(iii)-(iv) and Assumptions A.2(iii) and A.3(iv), with probability $1-o\left(N^{-1}\right)$ we have

$$\begin{aligned}\max_{1\leq i\leq N}\left\|T(\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i})(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i})^{-1}(\frac{1}{T^2}\tilde{x}_{2,i}'\tilde{x}_{1,i})\right\| &\leq T\max_{1\leq i\leq N}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\|^2\left[\min_{1\leq i\leq N}\lambda_{\min}\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)\right]^{-1} \\ &= To\left(p_2a_{1NT}^2\right)O\left(1\right)=o\left(b_T^{-1}\right).\end{aligned}$$

(ii) Noting that $\frac{1}{T}\tilde{x}_{2,i}'M_{1,i}\tilde{x}_{2,i}=\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}-Tb_T(\frac{1}{T^2}\tilde{x}_{2,i}'\tilde{x}_{1,i})(\frac{b_T}{T^2}\tilde{x}_{1,i}'\tilde{x}_{1,i})^{-1}(\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i})$, the result follows from Lemmas A.2(iii)-(iv) and Assumption A.2(i) and the fact that $Tb_T(\sqrt{p_2}a_{1NT})^2=O(p_2a_{2NT})$. The detailed arguments are analogous to those used in the proof of (iii) below.

(iii) Note that $\frac{1}{T}\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i^*=\frac{1}{T}\tilde{x}_{1,i}'\tilde{u}_i^*-\frac{1}{T}\tilde{x}_{1,i}'\tilde{x}_{2,i}(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i})^{-1}\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i^*$. By Lemma A.2(v), $P(\max_{1\leq i\leq N}||\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{u}^*||>ca_{1NT}/2)=o\left(N^{-1}\right)$. Define the following two events:

$$E_{1NT}=\left\{\min_{1\leq i\leq N}\lambda_{\min}\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)\geq\underline{c}_{22}/2\right\}\text{ and }E_{2NT}=\left\{\max_{1\leq i\leq N}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\|_{\text{sp}}\leq p_2^{1/2}a_{1NT}\right\}.$$

By Lemma A.2(iii)-(iv), $P\left(E_{lNT}\right)=1-o\left(N^{-1}\right)$ for $l=1,2$. Denote the complement of $E_{\ell NT}$ as $E_{\ell NT}^c$ for $\ell=1,2$. Then, in view of the fact that $\|AB\|\leq\|A\|_{\text{sp}}\|B\|$ and $\|A\|_{\text{sp}}\leq\|A\|$ for any two conformable matrices A and B, we have

$$\begin{aligned}&P\left(\max_{1\leq i\leq N}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i}\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i^*\right)\right\|>ca_{1NT}/2,\right)\\ \leq\ &P\left(\max_{1\leq i\leq N}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\|_{\text{sp}}\left\|\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i^*\right\|\left[\min_{1\leq i\leq N}\lambda_{\min}\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)\right]^{-1}>ca_{1NT}/2,E_{1NT}\cap E_{2NT}\right)\\ &+P\left(E_{1NT}^c\cup E_{2NT}^c\right)\\ \leq\ &P\left(\max_{1\leq i\leq N}\left\|\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i^*\right\|>c\cdot\underline{c}_{22}p_2^{-1/2}/4\right)+o\left(N^{-1}\right)\\ =\ &o\left(N^{-1}\right)+o\left(N^{-1}\right)=o\left(N^{-1}\right),\end{aligned}$$

where the first equality follows by Lemma A.2(vi) and the fact that $p_2^{1/2}a_{2NT}=o(p_2^{-1/2})$. Consequently, the result in (iii) follows.

(iv) Note that $\frac{1}{T}\tilde{x}_{2,i}'M_{1,i}\tilde{u}_i^*=\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i^*-T(\frac{1}{T^2}\tilde{x}_{2,i}'\tilde{x}_{1,i})(\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{1,i})^{-1}(\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{u}_i^*)\equiv I_{1i}-I_{2i}$, say. By Lemma A.2(vi), $P(\max_{1\leq i\leq N}\|I_{1i}\|\geq cp_2^{1/2}a_{2NT}/2)=o\left(N^{-1}\right)$. Noting that

$$\|I_{2i}\|\leq b_T\left\|\frac{1}{T^2}\tilde{x}_{2,i}'\tilde{x}_{1,i}\right\|_{\text{sp}}\left[\lambda_{\min}\left(\frac{b_T}{T^2}\tilde{x}_{1,i}'\tilde{x}_{1,i}\right)\right]^{-1}\left\|\frac{1}{T}\tilde{x}_{1,i}'\tilde{u}_i^*\right\|$$

and $b_Ta_{1NT}=o\left(1\right)$, we can readily apply Lemmas A.2(iii), A.2(v), and A.3(i) to show that $P(\max_{1\leq i\leq N}\|I_{2i}\|\geq cp_2^{1/2}a_{2NT}/2)=o\left(N^{-1}\right)$. ■

Proof of Lemma A.5. (i) Noting that $\tilde{\beta}_{1,i}-\beta_{1,i}^0=\left(\frac{b_T}{T^2}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i}\right)^{-1}\frac{b_T}{T^2}\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i^*$, the result follows from Lemmas A.4(i) and (iii), and Assumption A.2(i).

(ii) Noting that

$$
\begin{aligned}
\tilde{\beta}_{2,i}-\beta^*_{2,i} &= \left(\frac{1}{T}\tilde{x}'_{2,i}M_{1,i}\tilde{x}_{2,i}\right)^{-1}\frac{1}{T}\tilde{x}'_{2,i}M_{1,i}\tilde{u}^*_i \\
&= \left[\left(\frac{1}{T}\tilde{x}'_{2,i}M_{1,i}\tilde{x}_{2,i}\right)^{-1}-\Sigma^{-1}_{22,i}\right]\frac{1}{T}\tilde{x}'_{2,i}M_{1,i}\tilde{u}^*_i+\Sigma^{-1}_{22,i}\frac{1}{T}\tilde{x}'_{2,i}M_{1,i}\tilde{u}^*_i,
\end{aligned}
$$

the result follows from Lemma A.4(ii) and (iv) and Assumption A.2(iii).

(iii) Let $\beta^*_i=(\beta^{0\prime}_{1,i},\beta^{*\prime}_{2,i})'$, which is $\beta^0_i=(\beta^{0\prime}_{1,i},\beta^{0\prime}_{2,i})'$ if $\Sigma_{20,i}=0$. Noting that $\tilde{y}_{it}=\tilde{x}'_{1,it}\beta^0_{1,i}+\tilde{x}'_{2,it}\beta^0_{2,i}+\tilde{u}_{it}=\tilde{x}'_{it}\beta^*_i+\tilde{u}^*_{it}$ with $\tilde{u}^*_{it}=\tilde{u}_{it}-\tilde{x}'_{2,it}\Sigma^{-1}_{22,i}\Sigma_{20,i}$, we have

$$
\begin{aligned}
\tilde{\sigma}^2_i &= \frac{1}{T}\sum_{t=1}^{T}[\tilde{y}_{it}-\tilde{\beta}'_i\tilde{x}_{it}]^2=\frac{1}{T}\sum_{t=1}^{T}[\tilde{u}^*_{it}+\tilde{x}'_{it}(\beta^*_i-\tilde{\beta}_i)]^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}(\tilde{u}^*_{it})^2+(\beta^*_i-\tilde{\beta}_i)'\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}'_{it}(\beta^*_i-\tilde{\beta}_i)+2(\beta^*_i-\tilde{\beta}_i)'\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{u}^*_{it} \\
&\equiv D_{1i}+D_{2i}+2D_{3i},\ \text{say.}
\end{aligned}
$$

We prove (i) by showing that (i1) $P\left(\max_{1\le i\le N}\left|D_{1i}-\Sigma^*_{0.2,i}\right|>\epsilon\right)=o\left(N^{-1}\right)$, (i2) $P\left(\max_{1\le i\le N}|D_{2i}|>\epsilon\right)=o\left(N^{-1}\right)$, and (i3) $P\left(\max_{1\le i\le N}|D_{3i}|>\epsilon\right)=o\left(N^{-1}\right)$ for any $\epsilon>0$. Noting that

$$
\begin{aligned}
D_{1i}-\Sigma^*_{0.2,i} &= \frac{1}{T}\sum_{t=1}^{T}\left(\tilde{u}_{it}-\tilde{x}'_{2,it}\Sigma^{-1}_{22,i}\Sigma_{20,i}\right)^2-\left(\Sigma_{00,i}-\Sigma_{02,i}\Sigma^{-1}_{22,i}\Sigma_{20,i}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[u^2_{it}-E\left(u^2_{it}\right)\right]-\overline{u}^2_i+\Sigma_{02,i}\Sigma^{-1}_{22,i}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{2,it}\tilde{x}'_{2,it}-\Sigma_{22,i}\right)\Sigma^{-1}_{22,i}\Sigma_{20,i} \\
&\quad -2\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{x}'_{2,it}-\Sigma_{02,i}\right)\Sigma^{-1}_{22,i}\Sigma_{20,i} \\
&\equiv D_{1i,1}+D_{1i,2}+D_{1i,3}+D_{1i,4}.
\end{aligned}
$$

By a simple application of Lemma S1.2 in SSPb, we can show that $P\left(\max_{1\le i\le N}|D_{1i,\ell}|>\epsilon/4\right)=o\left(N^{-1}\right)$ for $\ell=1,2$. By Lemma A.2(iv) and Assumption A.2(iv), $P\left(\max_{1\le i\le N}|D_{1i,3}|>\epsilon/4\right)=o\left(N^{-1}\right)$. By Lemma A.2(ii) and Assumption A.2(iv), $P\left(\max_{1\le i\le N}|D_{1i,4}|>\epsilon/4\right)=o\left(N^{-1}\right)$. It follows that

$$
P\left(\max_{1\le i\le N}\left|D_{1i}-\Sigma^*_{0.2,i}\right|>\epsilon\right)=o(N^{-1}).
$$

For D_{2i}, we have by the Cauchy-Schwarz inequality

$$
\begin{aligned}
D_{2i} &\le 2(\beta^0_{1,i}-\tilde{\beta}_{1,i})'\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}'_{1,it}(\beta^0_{1,i}-\tilde{\beta}_{1,i})+2(\beta^*_{2,i}-\tilde{\beta}_{2,i})'\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{2,it}\tilde{x}'_{2,it}(\beta^*_{2,i}-\tilde{\beta}_{2,i}) \\
&\equiv 2D_{2i,1}+2D_{2i,2}.
\end{aligned}
$$

With probability $1-o\left(N^{-1}\right)$, $D_{2i,1}$ is bounded above by

$$2T\log\log T\max_{1\leq i\leq N}\left\|\beta_{1,i}^{0}-\tilde{\beta}_{1,i}\right\|^{2}\max_{1\leq i\leq N}\left\|\frac{1}{2T^{2}\log\log T}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{x}_{1,it}^{\prime}\right\|_{\text{sp}}=o\left(T\log\log Tb_{T}^{2}a_{1NT}^{2}\right)=o\left(1\right),$$

by Lemma A.3(i) and part (i) and Assumption A.3(iii). And $D_{2i,2}$ bounded above by

$$\max_{1\leq i\leq N}\left\|\beta_{2,i}^{*}-\tilde{\beta}_{2,i}\right\|^{2}\max_{1\leq i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{2,it}\tilde{x}_{2,it}^{\prime}\right\|_{\text{sp}}=o(p_{2}a_{2NT}^{2})=o\left(1\right),$$

by Lemma A.2(iv), Assumption A.2(iii), and part (ii). It follows that $P\left(\max_{1\leq i\leq N}|D_{2i}|>\epsilon\right)=o\left(N^{-1}\right)$. Similarly, with probability $1-o\left(N^{-1}\right)$,

$$\begin{aligned}|D_{3i}| &\leq \left|(\beta_{1,i}^{0}-\tilde{\beta}_{1,i})^{\prime}\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{u}_{it}^{*}\right|+\left|(\beta_{2,i}^{*}-\tilde{\beta}_{2,i})^{\prime}\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{2,it}\tilde{u}_{it}^{*}\right| \\ &\leq T\left\|\beta_{1,i}^{0}-\tilde{\beta}_{1,i}\right\|\left\|\frac{1}{T^{2}}\sum_{t=1}^{T}\tilde{x}_{1,it}\tilde{u}_{it}^{*}\right\|+\left\|\beta_{2,i}^{*}-\tilde{\beta}_{2,i}\right\|\left\|\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{2,it}\tilde{u}_{it}^{*}\right\| \\ &= To\left(b_{T}a_{1NT}\right)o\left(a_{1NT}\right)+o(p_{2}^{1/2}a_{2NT})o(p_{2}^{1/2}a_{2NT})=o\left(1\right),\end{aligned}$$

by Lemma A.2(v)-(iv), parts (i)-(ii), and Assumption A.3(iii). It follows that $P\left(\max_{1\leq i\leq N}|D_{2i}|>\epsilon\right)=o\left(N^{-1}\right)$. ■

Proof of Lemma A.6. (i) Noting that $\frac{1}{T^{2}}\tilde{x}_{1,i}^{\prime}\tilde{u}_{i}^{*}=\frac{1}{T^{2}}\tilde{x}_{1,i}^{\prime}\tilde{u}_{i}-\frac{1}{T^{2}}\tilde{x}_{1,i}^{\prime}\tilde{x}_{2,i}\Sigma_{22,i}^{-1}\Sigma_{20,i}$, it suffices to show $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^{2}}\tilde{x}_{1,i}^{\prime}\tilde{u}_{i}\right\|^{2}=O_{P}\left(T^{-2}\right)$ and $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^{2}}\tilde{x}_{1,i}^{\prime}\tilde{x}_{2,i}\Sigma_{22,i}^{-1}\Sigma_{20,i}\right\|^{2}=O_{P}\left(T^{-2}\right)$. We only show the former one as the proof of the latter claim is similar under the side condition $\left\|\Sigma_{22,i}^{-1}\Sigma_{20,i}\right\|\leq C<\infty$, which is ensured by Assumption A.2(ii)-(iii). By equation (B.1) and the Cauchy-Schwarz inequality

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^{2}}\tilde{x}_{1,i}^{\prime}\tilde{u}_{i}\right\|^{2} &\leq \frac{2}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^{2}}x_{1,i}^{\prime}u_{i}\right\|^{2}+\frac{2}{NT^{2}}\sum_{i=1}^{N}\left\|\bar{x}_{1,i}\bar{u}_{i}\right\|^{2} \\ &\leq \frac{6}{N}\sum_{i=1}^{N}\left(\left\|b_{1i}\right\|^{2}+\left\|b_{2i}\right\|^{2}+\left\|b_{3i}\right\|^{2}\right)+\frac{2}{NT^{2}}\sum_{i=1}^{N}\left\|\bar{x}_{1,i}\bar{u}_{i}\right\|^{2} \\ &\equiv 6d_{1}+6d_{2}+6d_{3}+2d_{4},\text{ say.}\end{aligned}$$

For d_1, we have

$$\begin{aligned}E(d_{1}) &= \frac{1}{N}\sum_{i=1}^{N}E\left(\left\|S_{1}\frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{is}e_{it}^{\prime}\psi_{i0}^{\prime}S_{0}^{\prime}\right\|^{2}\right) \\ &= \frac{1}{NT^{4}}\sum_{i=1}^{N}E\left(\left\|\sum_{t=1}^{T}z_{it}\right\|^{2}\right)=\frac{1}{NT^{4}}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left(z_{it}^{2}\right) \\ &\leq \frac{C}{NT^{4}}\sum_{i=1}^{N}\sum_{t=1}^{T}t=O\left(T^{-2}\right),\end{aligned}$$

where $z_{it} = S_1 \sum_{s=1}^{t-1} \varepsilon_{is} e'_{it} \psi'_{i0} S'_0$ satisfies $E(z_{it}|\mathcal{F}_{i,t-1}) = 0$ and the first inequality follows because we can show that $E(z_{it}^2) \leq Ct$. For d_2, we can follow the analysis of $b_{2,i}$ in the proof of Lemma A.2 and show that $E(d_2) \leq \frac{4}{N} \sum_{i=1}^N E[\|b_{2i,1}\|^2 + \|b_{2i,2}\|^2 + \|b_{2i,3}\|^2 + \|b_{2i,4}\|^2] = O(T^{-2})$. In addition,

$$\begin{aligned} E(d_3) &= \frac{1}{N}\sum_{i=1}^N E\left\|S_1 \frac{1}{T^2}\sum_{t=1}^T \varepsilon_{it}\varepsilon'_{it} S'_0\right\|^2 \leq \frac{C}{NT^2}\sum_{i=1}^N E\left\{\frac{1}{T}\sum_{t=1}^T \|\varepsilon_{it}\|^2\right\}^2 \\ &\leq \frac{C}{NT^2}\sum_{i=1}^N \frac{1}{T}\sum_{t=1}^T E\|\varepsilon_{it}\|^4 = O\left(T^{-2}\right), \text{ and} \\ E(d_4) &= \frac{1}{NT^2}\sum_{i=1}^N E\|\bar{x}_{1,i}\bar{u}_i\|^2 \leq \frac{1}{NT^2}\sum_{i=1}^N \left\{E(\|\bar{x}_{1,i}\|^4)E\left(\bar{u}_i^4\right)\right\}^{1/2} = O\left(T^{-2}\right), \end{aligned}$$

where the last equality follows from the fact that $E(\bar{u}_i^4) \leq CT^{-2}$ and $E(\|\bar{x}_{1,i}\|^4) \leq CT^2$. Consequently, $\frac{1}{N}\sum_{i=1}^N E\left\|\frac{1}{T^2}\tilde{x}'_{1,i}\tilde{u}_i\right\|^2 = O\left(T^{-2}\right)$ and $\frac{1}{N}\sum_{i=1}^N \left\|\frac{1}{T^2}\tilde{x}'_{1,i}\tilde{u}_i\right\|^2 = O_P\left(T^{-2}\right)$ by the Markov inequality.

(ii) Noting that

$$\begin{aligned} \frac{1}{T^{3/2}}\tilde{x}'_{2,i}\tilde{u}_i^* &= \frac{1}{T^{3/2}}\tilde{x}'_{2,i}\tilde{u}_i - \frac{1}{T^{3/2}}\tilde{x}'_{2,i}\tilde{x}_{2,i}\Sigma_{22,i}^{-1}\Sigma_{20,i} \\ &= \frac{1}{T^{3/2}}\left(\tilde{x}'_{2,i}\tilde{u}_i - \Sigma_{20,i}\right) - \frac{1}{T^{3/2}}\left(\tilde{x}'_{2,i}\tilde{x}_{2,i} - \Sigma_{22,i}\right)\Sigma_{22,i}^{-1}\Sigma_{20,i} \\ &= \frac{1}{T^{3/2}}\left(x'_{2,i}u_i - \Sigma_{20,i}\right) - \frac{1}{T^{3/2}}\left(x'_{2,i}x_{2,i} - \Sigma_{22,i}\right)\Sigma_{22,i}^{-1}\Sigma_{20,i} - \frac{1}{T^{1/2}}\bar{x}'_{2,i}\bar{u}_i \\ &\quad + \frac{1}{T^{1/2}}\bar{x}'_{2,i}\bar{x}_{2,i}\Sigma_{22,i}^{-1}\Sigma_{20,i} \equiv d_{1i} + d_{2i} + d_{3i} + d_{4i}, \text{ say,} \end{aligned}$$

it suffices to show $\frac{1}{N}\sum_{i=1}^N \|d_{\ell i}\|^2 = O_P\left(p_2 T^{-2}\right)$ for $\ell = 1, 2, 3, 4$. We can prove these by the Markov inequality. Then (ii) follows.

(iii) By the BN decomposition $x_{1,it} = S_1[\psi_i(1)\sum_{s=1}^t e_{it} + \breve{e}_{i,0} - \breve{e}_{i,t}]$ and the Cauchy-Schwarz inequality,

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^N E\left\|\hat{Q}_{1i}\right\|^2 &\leq \frac{1}{N}\sum_{i=1}^N E\left\|\frac{1}{T^2}\sum_{t=1}^T x_{1,it}x'_{1,it}\right\|^2 \\ &\leq \frac{1}{N}\sum_{i=1}^N E\left\|\frac{1}{T^2}\sum_{t=1}^T S_1\psi_i(1)\sum_{s=1}^t e_{it}\sum_{r=1}^t e'_{ir}\psi_i(1)' S'_1\right\|^2 \\ &\quad + \frac{1}{N}\sum_{i=1}^N E\left\|\frac{1}{T^2}\sum_{t=1}^T S_1[(\breve{e}_{i,0} - \breve{e}_{i,t})(\breve{e}_{i,0} - \breve{e}_{i,t})' S'_1\right\|^2. \end{aligned}$$

By straightforward moment calculation, we can bound the first term in the last expression by $O(1)$ and the second term by $O\left(T^{-2}\right)$. Then $\frac{1}{N}\sum_{i=1}^N \left\|\hat{Q}_{1i}\right\|^2 = O_P(1)$ by the Markov inequality.

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

(v) Noting that $\tilde{x}'_{1,i}M_{2,i}\tilde{u}_i^* = \tilde{x}'_{1,i}\tilde{u}_i^* - \tilde{x}'_{1,i}\tilde{x}_{2,i}(\tilde{x}'_{2,i}\tilde{x}_{2,i})^{-1}\tilde{x}'_{2,i}\tilde{u}_i^*$, we have by Lemmas A.2(ii), A.2(iv), part (iv), and Assumption A.3(iii)

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'M_{2,i}\tilde{u}_i^*\right\|^2 &\leq \frac{2}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{u}_i^*\right\|^2+\frac{2}{NT^2}\sum_{i=1}^{N}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i}\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i^*\right\|^2\\ &\leq O_P\left(T^{-2}\right)+\max_i\left\|\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\right\|_{\rm sp}\max_i\left\|\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i^*\right\|^2\frac{2}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\|^2\\ &\leq O_P\left(T^{-2}\right)+O_P\left(1\right)o_P(p_2^{1/2}a_{2NT})O_P\left(p_2T^{-2}\right)=O_P\left(T^{-2}\right).\ \blacksquare\end{aligned}$$

Proof of Lemma A.7. (i) Note that $Q_{k,NT}=\frac{1}{N_kT^2}\sum_{i\in G_k^0}\tilde{x}_{1,i}'\tilde{x}_{1,i}-\frac{1}{N_kT^2}\sum_{i\in G_k^0}\left(\tilde{x}_{1,i}'\tilde{x}_{2,i}\right)\left(\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}$ $\times\left(\tilde{x}_{2,i}'\tilde{x}_{1,i}\right)\equiv Q_{1k,NT}-Q_{2k,NT}$. By Lemmas A.2(ii)-(iii), $\|Q_{2k,NT}\|\leq T\max_{i\in G_k^0}\left\|\frac{1}{T^2}\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\|^2\left\|(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i})^{-1}\right\|_{\rm sp}$ $=To_P\left(p_2a_{1NT}^2\right)=o_P\left(1\right)$. By the arguments used in Phillips and Moon (1999, Section 4), we can show that

$$\begin{aligned}Q_{1k,NT} &= \frac{1}{N_k}\sum_{i\in G_k^0}E\left(\int_0^1\tilde{B}_{1,i}\tilde{B}_{1,i}'\right)+o_P\left(1\right)=\frac{1}{N_k}\sum_{i\in G_k^0}S_1\psi_i\left(1\right)E\left(\int_0^1\tilde{W}_{1,i}\tilde{W}_{1,i}'\right)\psi_i\left(1\right)'S_1'+o_P\left(1\right)\\ &= \frac{1}{6N_k}\sum_{i\in G_k^0}S_1\psi_i\left(1\right)\psi_i\left(1\right)'S_1'+o_P\left(1\right),\end{aligned}$$

where we use the fact that $E(\int_0^1\tilde{W}_{1,i}\tilde{W}_{1,i}')=E(\int_0^1 W_{1,i}W_{1,i}')-E(\int_0^1 W_{1,i}\int_0^1 W_{1,i}')=\frac{1}{2}I_{p_1}-\frac{1}{3}I_{p_1}=\frac{1}{6}I_{p_1}$.Thus (i) follows.

(ii) Let $E_{it}\equiv\sum_{s=1}^{t}e_{is}$, $e_{it}^*\equiv e_{it}-\frac{1}{T}\sum_{s=t}^{T}e_{is}$, and $p=1+p_1+p_2$. Note that $\tilde{e}_{it}=e_{it}-\bar{e}_i=e_{it}^*-\frac{1}{T}E_{i,t-1}$. We apply the arguments as used in the proof of Theorem 16 in Phillips and Moon (1999, PM hereafter) and derive the limiting distribution of $V_{1k,NT}$ below.[7]

First, we apply the BN decompositions. Noting that $\varepsilon_{it}=\psi_i\left(1\right)e_{it}+\breve{e}_{i,t-1}-\breve{e}_{i,t}$, we have $x_{1,it}=S_1\left[\psi_i\left(1\right)E_{it}+\breve{e}_{i,0}-\breve{e}_{i,t}\right]$ and

$$u_{it}-x_{2,it}'\Sigma_{22,i}^{-1}\Sigma_{20,i}=\left[\psi_i\left(1\right)e_{it}+\breve{e}_{i,t-1}-\breve{e}_{i,t}\right]'\left[S_0'-S_2'\Sigma_{22,i}^{-1}\Sigma_{20,i}\right]=\left[e_{it}'\psi_i\left(1\right)'+\breve{e}_{i,t-1}'-\breve{e}_{i,t}'\right]s_i,$$

where $s_i=S_0'-S_2'\Sigma_{22,i}^{-1}\Sigma_{20,i}$ is a $p\times 1$ vector. It follows that we can write the demeaned versions of $x_{1,it}$ and $u_{it}-x_{2,it}'\Sigma_{22,i}^{-1}\Sigma_{20,i}$ as

$$\tilde{x}_{1,it}=S_1\left[\psi_i\left(1\right)\tilde{E}_{it}+\tilde{\breve{e}}_{i,0}-\tilde{\breve{e}}_{i,t}\right]\text{ and }\tilde{u}_{it}-\tilde{x}_{2,it}'\Sigma_{22,i}^{-1}\Sigma_{20,i}=\left[\tilde{e}_{it}'\psi_i\left(1\right)'+\tilde{\breve{e}}_{i,t-1}'-\tilde{\breve{e}}_{i,t}'\right]s_i,$$

where $\tilde{E}_{it}=E_{it}-\frac{1}{T}\sum_{s=1}^{T}E_{is}$ and $\tilde{\breve{e}}_{i,t}=\breve{e}_{i,t}-\frac{1}{T}\sum_{s=1}^{T}\tilde{\breve{e}}_{i,s}$. Let $S_{i,t}^{\varepsilon}=\sum_{s=1}^{t}\varepsilon_{is}$ and $\tilde{S}_{i,t}^{\varepsilon}=S_{i,t}^{\varepsilon}-\frac{1}{T}\sum_{s=1}^{T}S_{i,t}^{\varepsilon}$.

As in PM (p.1105), we can obtaining the following decomposition,

$$\begin{aligned}\bar{V}_{k,NT}-\mathbb{B}_{k,NT}=&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{1,it}\left(\tilde{u}_{it}-\tilde{x}_{2,it}'\Sigma_{22,i}^{-1}\Sigma_{20,i}\right)-\mathbb{B}_{k,NT}\\ =&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\{\frac{1}{T}\sum_{t=1}^{T}S_1\psi_i(1)\left[\tilde{E}_{it}\tilde{e}_{it}'-\left(1-\frac{T+1}{2T}\right)I_{p+1}\right]\psi_i(1)'s_i\right.\\ &+\frac{1}{T}\sum_{t=1}^{T-1}S_1\left(\tilde{\varepsilon}_{i,t+1}\tilde{\breve{\varepsilon}}_{it}'+\sum_{s=0}^{\infty}\psi_{i,s+1}\breve{\psi}_{i,s}'\right)s_i-\frac{1}{T}\sum_{s=0}^{\infty}S_1\psi_{i,s+1}\breve{\psi}_{i,s}'s_i\\ &-\frac{1}{T}\sum_{t=1}^{T}S_1\psi_i(1)\left[\tilde{\breve{\varepsilon}}_{it}\tilde{e}_{it}'\psi_i(1)'-\breve{\psi}_{i,0}\psi_i(1)'\right]s_i+\frac{1}{T}\sum_{t=1}^{T}S_1\tilde{\breve{\varepsilon}}_{i0}\tilde{e}_{it}'\psi_i(1)'s_i\\ &\left.-\frac{1}{T}S_1\tilde{S}_{i,T}^{\varepsilon}\tilde{\breve{\varepsilon}}_{iT}'s_i+\frac{1}{T}S_1\tilde{S}_{i,1}^{\varepsilon}\tilde{\breve{\varepsilon}}_{i0}'s_i\right\}\\ =&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\{Q_{i,T}+R_{1i,T}+R_{2i,T}+R_{3i,T}+R_{4i,T}+R_{5i,T}+R_{6i,T}\right\},\ \text{say.}\end{aligned}$$

Note that our terms $Q_{i,T}$ and $R_{\ell i,T}$ $(\ell=1,2,...,6)$ parallel to the corresponding term in PM. There are three main differences: (1) all variables involved here are time-demeaned versions of those in PM; (2) we need to center $\tilde{E}_{it}\tilde{e}_{it}'$ around its expectation $(1-\frac{T+1}{2T})I_{p+1}$ while PM center the non-demeaned version of $\tilde{E}_{it}\tilde{e}_{it}'$ around its expectation I_{p+1}, and the difference between the two centering terms, namely, $-\frac{T+1}{2T}I_{p+1}$, enters the bias term $\mathbb{B}_{2k,NT}$ and reflects the contribution of time-demeaning of random variables in the regression; (3) the sign $R_{2i,T}$ is negative rather positive. One can verify that $\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\psi_{i,s+r}\psi_{i,s}'=\psi_i(1)\psi_i(1)'+\sum_{s=0}^{\infty}\psi_{i,s+1}\breve{\psi}_{i,s}'-\breve{\psi}_{i,0}\psi_i(1)'$.

Second, we study the asymptotic distribution of $N_k^{-1/2}\sum_{i\in G_k^0}Q_{i,T}$. Noting that $\frac{1}{T}\sum_{t=1}^{T}\tilde{E}_{it}\tilde{e}_{it}'=\frac{1}{T}\sum_{t=1}^{T}E_{it}\tilde{e}_{it}'$, and $\tilde{e}_{it}=e_{it}-\frac{1}{T}E_{iT}$, we have

$$\begin{aligned}\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}Q_{i,T}=&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^{T}S_1\psi_i(1)\left[\tilde{E}_{it}\tilde{e}_{it}'-\left(1-\frac{T+1}{2T}\right)I_p\right]\psi_i(1)'s_i\\ =&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^{T}S_1\psi_i(1)E_{it-1}e_{it}'\psi_i(1)'s_i-\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T^2}\sum_{t=1}^{T}S_1\psi_i(1)\left(E_{it}E_{iT}'-tI_p\right)\psi_i(1)'s_i\\ &+\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^{T}S_1\psi_i(1)\left[e_{it}e_{it}'-I_p\right]\psi_i(1)'s_i\\ =&\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\{Q_{1i,T}-Q_{2i,T}+Q_{3i,T}\right\},\ \text{say.}\end{aligned}$$

By direct moment calculations, we can readily show that

$$\left\|E\left(\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}Q_{3i,T}\right)\right\|\leq\frac{1}{T}\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\|S_1\psi_i(1)\psi_i(1)'s_i\right\|=O\left(\sqrt{N_kp_2}/T\right)=o(1)$$

and $\left\|\text{Var}\left(\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}Q_{3i,T}\right)\right\| = \left\|\frac{1}{N_k}\sum_{i\in G_k^0}\text{Var}\left(Q_{3i,T}\right)\right\| \leq \frac{1}{N_k}\sum_{i\in G_k^0}E\left\|Q_{3i,T}\right\|^2 = O\left(p_2T^{-2}\right) = o\left(1\right).$ Then $\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}Q_{3i,T} = o_P\left(1\right)$ by the Chebyshev inequality. It follows that

$$\text{Var}\left(\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}Q_{i,T}\right) = \frac{1}{N_k}\sum_{i\in G_k^0}\left\{\text{Var}\left(Q_{1i,T}\right)+\text{Var}\left(Q_{2i,T}\right)-\text{Cov}\left(Q_{1i,T},Q_{2i,T}\right)-\text{Cov}\left(Q_{1i,T},Q_{2i,T}\right)'\right\}$$

Then we study the asymptotic variance by terms. For $Q_{1i,T}$,

$$\begin{aligned}
\text{Var}\left(Q_{1i,T}\right) =& \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[S_1\psi_i\left(1\right)E_{i,t-1}e_{it}'\psi_i\left(1\right)'s_i\ s_i'\psi_i\left(1\right)e_{is}E_{i,s-1}'\psi_i\left(1\right)'S_1'\right]\\
=& \frac{1}{T^2}\sum_{t=1}^{T}E\left\{S_1\psi_i\left(1\right)E_{i,t-1}e_{it}'\psi_i\left(1\right)'s_i\ s_i'\psi_i\left(1\right)e_{it}E_{i,t-1}'\psi_i\left(1\right)'S_1'\right\}\\
=& \frac{1}{T^2}\sum_{t=1}^{T}\left(s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right)E\left\{\text{vec}\left(E_{i,t-1}e_{it}'\right)\quad\left[\text{vec}\left(E_{i,t-1}e_{it}'\right)\right]'\ \right\}\left(s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right)'\\
=& \frac{1}{T^2}\sum_{t=1}^{T}\left(s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right)E\left(e_{it}e_{it}'\otimes E_{i,t-1}E_{i,t-1}'\right)\left(s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right)'\\
=& \frac{1}{T^2}\sum_{t=1}^{T}\left(t-1\right)\left(s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right)\left(s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right)'\\
=& \frac{1}{2}s_i'\psi_i\left(1\right)\psi_i\left(1\right)'s_i\otimes S_1\psi_i\left(1\right)\psi_i\left(1\right)'S_1'+O\left(T^{-1}\right)\\
=& \frac{1}{2}s_i'\Omega_is_iS_1\Omega_iS_1'+O\left(T^{-1}\right),
\end{aligned}$$

where the second equality follows from the fact that $\{E_{i,t-1}e_{it}', \mathcal{F}_{i,t}\}$ is an martingale difference sequence (m.d.s.), the third equality holds because $\text{vec}(A_1A_2A_3) = (A_3'\otimes A_1)\text{vec}(A_2)$ with $A_1 = S_1\psi_i\left(1\right)$, $A_2 = E_{i,t-1}e_{it}'$, and $A_3 = \psi_i\left(1\right)'s_i$, the fourth equality follows from the fact that $\text{vec}(a_1a_2') = a_2\otimes a_1$ and $\left(a_2\otimes a_1\right)\left(a_2\otimes a_1\right)' = a_2a_2'\otimes a_1a_1'$, the fifth equality holds because $E\left(e_{it}e_{it}'\otimes E_{i,t-1}E_{i,t-1}'\right) = E\left(e_{it}e_{it}'\right)\otimes E\left(E_{i,t-1}E_{i,t-1}'\right) = \left(t-1\right)I_{(1+p)^2}$. Similarly, we have for $Q_{2i,T}$,

$$\begin{aligned}
&\text{Var}\left(Q_{2i,T}\right)\\
=&\ \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[S_1\psi_i\left(1\right)\left(E_{i,t}E_{i,T}'-t\right)\psi_i\left(1\right)'s_i\ s_i'\psi_i\left(1\right)\left(E_{i,s}E_{i,T}'-s\right)\psi_i\left(1\right)'S_1'\right]\\
=&\ \left(s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right)\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\text{vec}\left(E_{i,t}E_{i,T}'\right)\text{vec}\left(E_{i,s}E_{i,T}'\right)'-t\text{vec}\left(I_{1+p}\right)\text{vec}\left(E_{i,s}E_{i,T}'\right)'\\
&-s\text{vec}\left(E_{i,t}E_{i,T}'\right)\text{vec}\left(I_{1+p}\right)'+ts\text{vec}\left(I_{1+p}\right)\text{vec}\left(I_{1+p}\right)'](s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right))'\\
=&\ [s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)]\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\text{vec}\left(E_{i,t}E_{i,T}'\right)\text{vec}\left(E_{i,s}E_{i,T}'\right)'-ts\text{vec}\left(I_{1+p}\right)\text{vec}\left(I_{1+p}\right)'][s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)]'\\
=&\ [s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)]\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\left(E_{i,T}E_{i,T}'\otimes E_{i,t}E_{i,s}'\right)-ts\text{vec}\left(I_{1+p}\right)\text{vec}\left(I_{1+p}\right)'][s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)]'
\end{aligned}$$

and

$$\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(E_{i,T}E_{i,T}'\otimes E_{i,t}E_{i,s}'\right)=\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[T\left(t\wedge s\right)I_{(p+1)^2}+ts\left(K_{1+p}+\text{vec}(I_{p+1})\text{vec}(I_{p+1})'\right)\right]$$
$$=\left\{\frac{1}{3}I_{(1+p)^2}+\frac{1}{4}\left(K_{p+1}+\text{vec}(I_{p+1})\text{vec}(I_{p+1})'\right)\right\}+O\left(T^{-1}\right).$$

where K_{1+p} is the $(p+1)^2\times(p+1)^2$ commutation matrix such that $K_{1+p}\text{vec}(A)$ $=\text{vec}(A')$ for any $(p+1)\times(p+1)$ matrix A. It follows that

$$\text{Var}\left(Q_{2i,T}\right)=\left(s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right)\left(\frac{1}{3}I_{(1+p)^2}+\frac{1}{4}K_{1+p}\right)\left(s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right)'+O\left(T^{-1}\right).$$

For $\text{Cov}(Q_{1i,T},Q_{2i,T})$,

$$\begin{aligned}
\text{Cov}\left(Q_{1i,T},Q_{2i,T}\right) &= \frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[S_1\psi_i\left(1\right)E_{i,t-1}e_{it}'\psi_i\left(1\right)'s_i\ s_i'\psi_i\left(1\right)E_{i,s}E_{i,T}'\psi_i\left(1\right)'S_1'\right]\\
&= \left[s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right]\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\{\text{vec}\left(E_{i,t-1}e_{it}'\right)\text{vec}\left(E_{i,s}E_{i,T}'\right)'\right\}\left[s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right]'\\
&= \left[s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right]\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(e_{it}E_{i,T}'\otimes E_{i,t-1}E_{i,s}'\right)\left[s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right]'\\
&= \left[s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right]\left\{\frac{1}{3}I_{(1+p)^2}+\frac{1}{6}K_{1+p}\right\}\left[s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)\right]'+O\left(T^{-1}\right),
\end{aligned}$$

where we use the fact that

$$\begin{aligned}
&\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(e_{it}E_{i,T}'\otimes E_{i,t-1}E_{i,s}'\right)\\
=\ &\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{t-1}E\left(e_{it}E_{i,T}'\otimes E_{i,t-1}E_{i,s}'\right)+\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=t}^{T}E\left(e_{it}E_{i,T}'\otimes E_{i,t-1}E_{i,s}'\right)\\
=\ &\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{t-1}E\left(e_{it}E_{i,T}'\otimes E_{i,t-1}E_{i,s}'\right)\\
&+\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=t}^{T}\left\{E\left(e_{it}E_{i,T}'\otimes E_{i,t-1}E_{i,t-1}'\right)+E[e_{it}E_{i,T}'\otimes E_{i,t-1}\left(E_{i,s}-E_{i,t-1}\right)']\right\}\\
=\ &\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{t-1}sI_{(1+p)^2}+\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=t}^{T}(t-1)\left(I_{(1+p)^2}+K_{1+p}\right)\\
=\ &\frac{1}{3}I_{(1+p)^2}+\frac{1}{6}K_{1+p}+O\left(T^{-1}\right).
\end{aligned}$$

Thus we have

$$
\begin{aligned}
\operatorname{Var}\left(\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0} Q_{i,T}\right) &= \frac{1}{N_k}\sum_{i\in G_k^0}\left\{\operatorname{Var}\left(Q_{1i,T}\right)+\operatorname{Var}\left(Q_{2i,T}\right)-2\operatorname{Cov}\left(Q_{1i,T},Q_{2i,T}\right)\right\}\\
&=\frac{1}{N_k}\sum_{i\in G_k^0}\left\{\frac{1}{6}s_i'\Omega_i s_i S_1\Omega_i S_1' - \frac{1}{12}[s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)]K_{1+p}[s_i'\psi_i\left(1\right)\otimes S_1\psi_i\left(1\right)]'\right\}\\
&\quad +O\left(T^{-1}\right)\\
&=\frac{1}{N_k}\sum_{i\in G_k^0}\left\{\frac{1}{6}s_i'\Omega_i s_i S_1\Omega_i S_1' - \frac{1}{12}\left(s_i'\Omega_i S_1'\otimes S_1\Omega_i s_i\right)K_{p_1,1}\right\}+O\left(T^{-1}\right),
\end{aligned}
$$

where $K_{p_1,1}$ is the $p_1\times p_1$ commutation matrix. It follows that $\operatorname{Var}(N_K^{-1/2}\sum_{i\in G_k^0}Q_{i,T})\rightarrow\lim_{N_k\rightarrow\infty}\frac{1}{N_k}\sum_{i\in G_k^0}$ $[\frac{1}{6}s_i'\Omega_i s_i S_1\Omega_i S_1'-\frac{1}{12}\left(s_i'\Omega_i S_1'\otimes S_1\Omega_i s_i\right)K_{p_1,1}]\equiv V_{(k)}$. This limit contributes to the asymptotic variance of our estimator. In addition, we can verify that $\sum_{i=1}^N E\left\|N_K^{-1/2}Q_{i,T}\right\|^4=O\left(N_k^{-1}\right)$, which verifies the Lyapunov condition for the central limit theorem for independent but non-identically distributed (i.n.i.d.) observations. Consequently, we have shown that $Q_{i,T}\Rightarrow N\left(0,\mathbb{V}_{(k)}\right)$. Third, we study $R_{1i,T}$:

$$
\begin{aligned}
\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{1i,T} &= \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^{T-1}S_1\left(\tilde{\varepsilon}_{i,t+1}\breve{\tilde{\varepsilon}}_{it}'-\sum_{s=0}^{\infty}\psi_{i,s+1}\breve{\psi}_{i,s}'\right)s_i\\
&= \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\{\frac{1}{T}\sum_{t=1}^{T-1}S_1\left(\varepsilon_{i,t+1}\breve{\varepsilon}_{it}'-\sum_{s=0}^{\infty}\psi_{i,s+1}\breve{\psi}_{i,s}'\right)s_i-\frac{1}{T}\sum_{t=1}^{T-1}S_1\frac{1}{T}\sum_{s=1}^{T}\varepsilon_{is}\breve{\varepsilon}_{it}'s_i\right.\\
&\quad\left.-\frac{1}{T}\sum_{t=1}^{T-1}S_1\varepsilon_{i,t+1}\frac{1}{T}\sum_{r=1}^{T}\breve{\varepsilon}_{ir}'s_i+\frac{T-1}{T}S_1\left(\frac{1}{T}\sum_{s=1}^{T}\varepsilon_{is}\frac{1}{T}\sum_{r=1}^{T}\breve{\varepsilon}_{ir}'\right)s_i\right\}\\
&\equiv \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\{R_{1i,T1}-R_{1i,T2}-R_{1i,T3}+R_{1i,T4}\right\}.
\end{aligned}
$$

Following PM, we can show that $E\left\|N_k^{-1/2}\sum_{i\in G_k^0}R_{1i,T1}\right\|^2=O\left(p_2T^{-1}\right)$. For $R_{1i,T2}$, we apply the Cauchy-Schwarz and Markov inequalities

$$
\begin{aligned}
\left\|\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{1i,T2}\right\| &\leq \sqrt{N_k}\left\{\frac{1}{N_k}\sum_{i\in G_k^0}\left\|S_1\frac{1}{T}\sum_{s=1}^{T}\varepsilon_{is}\right\|^2\right\}^{1/2}\left\{\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\frac{1}{T}\sum_{r=1}^{T-1}\breve{\varepsilon}_{ir}'s_i\right\|^2\right\}^{1/2}\\
&= \sqrt{N_k}O_P(T^{-1/2})O_P(p_2^{1/2}T^{-1/2})=o_P\left(1\right),
\end{aligned}
$$

where we use the fact $\frac{1}{N_k}\sum_{i\in G_k^0}E\left\|S_1\frac{1}{T}\sum_{s=1}^{T}\varepsilon_{is}\right\|^2\leq CT^{-1}\frac{1}{N_k}\sum_{i\in G_k^0}\operatorname{tr}(S_1\Omega_i S_1')\leq CT^{-1}\operatorname{tr}(S_1S_1')=O\left(T^{-1}\right)$. Similarly, we can show that $\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{1i,T\ell}=o_P\left(1\right)$ for $\ell=3,4$. Thus we have $\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{1i,T}$ $=o_P\left(1\right)$.

Fourth,

$$\begin{aligned}\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{3i,T} &= \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\{\frac{1}{T}\sum_{t=1}^{T}S_1\psi_i\left(1\right)\left[\breve{\varepsilon}_{it}e_{it}'\psi_i\left(1\right)'-\breve{\psi}_{i,0}\psi_i\left(1\right)'\right]s_i-\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}S_1\psi_i\left(1\right)\breve{\varepsilon}_{it}e_{is}'\psi_i\left(1\right)'\right\}\\ &\equiv \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\{R_{3i,T1}-R_{3i,T2}\right\}.\end{aligned}$$

It is easy to show that $E\left\|N_k^{-1/2}\sum_{i\in G_k^0}R_{3i,T1}\right\|^2=O\left(p_2T^{-1}\right)$, implying that $N_k^{-1/2}\sum_{i\in G_k^0}R_{3i,T1}=o_P\left(1\right).$ As in the analysis of $R_{1i,T2}$, we can show that $N_k^{-1/2}\sum_{i\in G_k^0}R_{3i,T2}=O_P(\sqrt{N_kp_2}/T)=o_P\left(1\right).$ Thus $N_k^{-1/2}\sum_{i\in G_k^0}R_{3i,T}=o_P\left(1\right).$

Fifth,

$$\begin{aligned}\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{4i,T} &= \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\{\frac{1}{T}\sum_{t=1}^{T}S_1\breve{\varepsilon}_{i0}e_{it}'\psi_i\left(1\right)'s_i-\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}S_1\breve{\varepsilon}_{it}e_{is}'\psi_i\left(1\right)'s_i\right\}\\ &\equiv \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\{R_{4i,T1}-R_{4i,T2}\right\}.\end{aligned}$$

Noting that $R_{4i,T2}=R_{3i,T2}$, $N_k^{-1/2}\sum_{i\in G_k^0}R_{4i,T2}=o_P\left(1\right).$ For $R_{4i,T1}$, in view of the fact $\breve{\varepsilon}_{i0}=\sum_{s=0}^{\infty}\tilde{\psi}_{i,s}e_{i,-s}$ and $\{e_{it},t\geq 1\}$ are mutually independent, we can readily show that $E\left(R_{4i,T1}\right)=0$ and

$$\begin{aligned}E\left\|\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{4i,T1}\right\|^2 &= \frac{1}{N_k}\sum_{i\in G_k^0}E\left\|\frac{1}{T}\sum_{t=1}^{T}S_1\breve{\varepsilon}_{i0}e_{it}'\psi_i\left(1\right)'s_i\right\|^2\\ &= \frac{1}{T^2N_k}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(s_i'\psi_i\left(1\right)e_{is}\breve{\varepsilon}_{i0}'S_1'S_1\breve{\varepsilon}_{i0}e_{it}'\psi_i\left(1\right)'s_i\right)\\ &= \frac{1}{T^2N_k}\sum_{i\in G_k^0}\sum_{t=1}^{T}\mathrm{tr}\left[E\left(\breve{\varepsilon}_{i0}'S_1'S_1\breve{\varepsilon}_{i0}\right)E\left(e_{it}'\psi_i\left(1\right)'s_is_i'\psi_i\left(1\right)e_{it}\right)\right]\\ &= \frac{1}{TN_k}\sum_{i\in G_k^0}E\left(\breve{\varepsilon}_{i0}'S_1'S_1\breve{\varepsilon}_{i0}\right)s_i'\psi_i\left(1\right)\psi_i\left(1\right)'s_i=O\left(p_2^2T^{-1}\right),\end{aligned}$$

where we use the fact that

$$\begin{aligned}s_i'\psi_i\left(1\right)\psi_i\left(1\right)'s_i &= s_i'\Omega_is_i\leq\lambda_{\max}\left(\Omega_i\right)s_i's_i\leq 2\lambda_{\max}\left(\Omega_i\right)\left(S_0S_0'+\Sigma_{20,i}'\Sigma_{22,i}^{-1}S_2S_2'\Sigma_{22,i}^{-1}\Sigma_{20,i}\right)\\ &\leq 2\lambda_{\max}\left(\Omega_i\right)\left[1+\lambda_{\max}\left(S_2S_2'\right)\left[\lambda_{\min}\left(\Sigma_{22,i}\right)\right]^{-2}\Sigma_{20,i}'\Sigma_{20,i}\right]\leq Cp_2\end{aligned}$$

and $E\left(\breve{\varepsilon}_{i0}'S_1'S_1\breve{\varepsilon}_{i0}\right)\leq E\left(\breve{\varepsilon}_{i0}'\breve{\varepsilon}_{i0}\right)\leq Cp_2$. Then $N_k^{-1/2}\sum_{i\in G_k^0}R_{4i,T1}=O_P\left(p_2T^{-1/2}\right)$ and $N_k^{-1/2}\sum_{i\in G_k^0}R_{4i,T}$ $=o_P\left(1\right).$

Sixth, we can show that

$$\begin{aligned}
N_k^{-1/2}\sum_{i\in G_k^0} R_{5i,T} &= N_k^{-1/2}\sum_{i\in G_k^0}\frac{1}{T}S_1\left\{S_{i,T}^{\varepsilon}\breve{\varepsilon}_{iT}'-\frac{1}{T}\sum_{t=1}^{T}S_{i,t}^{\varepsilon}\breve{\varepsilon}_{iT}'-\frac{1}{T}S_{i,T}^{\varepsilon}\sum_{t=1}^{T}\breve{\varepsilon}_{it}'+\frac{1}{T^2}\sum_{t=1}^{T}S_{i,t}^{\varepsilon}\sum_{t=1}^{T}\breve{\varepsilon}_{it}'\right\}s_i \\
&= N_k^{-1/2}\sum_{i\in G_k^0}\frac{1}{T}S_1S_{i,T}^{\varepsilon}\breve{\varepsilon}_{iT}'s_i+o_P(1)=N_k^{-1/2}\sum_{i\in G_k^0}\bar{R}_{5i,T}+o_P(1)\,,\text{ say.}
\end{aligned}$$

Note that

$$\begin{aligned}
\left\|E\left(N_k^{-1/2}\sum_{i\in G_k^0}\bar{R}_{5i,T}\right)\right\| &= \left\|N_k^{-1/2}\sum_{i\in G_k^0}\frac{1}{T}S_1\psi_i(1)\sum_{t=1}^{T}\sum_{r=1}^{\infty}E\left(e_{it}e_{i,T-r}'\right)\breve{\psi}_{i,r}'s_i\right\| \\
&= \left\|N_k^{-1/2}\sum_{i\in G_k^0}\frac{1}{T}S_1\psi_i(1)\sum_{t=1}^{T}\breve{\psi}_{i,T-t}'s_i\right\| \\
&\le N_k^{-1/2}\sum_{i\in G_k^0}\frac{1}{T}\left\|S_1\psi_i(1)\right\|_{\text{sp}}\sum_{r=1}^{\infty}\left\|\breve{\psi}_{i,r}'s_i\right\|_{\text{sp}}=O\left(\sqrt{N_k}p_2/T\right)=o(1)\,.
\end{aligned}$$

Similarly, we can verify that $\left\|\text{Var}\left(N_k^{-1/2}\sum_{i\in G_k^0}\bar{R}_{5i,T}\right)\right\|\le N_k^{-1}\sum_{i\in G_k^0}\left\|\text{Var}\left(\bar{R}_{5i,T}\right)\right\|=O\left(p_2^2/T\right)=o(1)\,.$ It follows that $N_k^{-1/2}\sum_{i\in G_k^0}R_{5i,T}=o_P(1)\,.$

Last, it is trivial to show $\left\|N_k^{-1/2}\sum_{i\in G_k^0}R_{2i,T}\right\|=O\left(\sqrt{N_k}p_2/T\right)=o(1)$ and $\left\|N_k^{-1/2}\sum_{i\in G_k^0}R_{6i,T}\right\|=O_P\left(\sqrt{N_k}p_2/T\right)=o_P(1)\,.$

In sum, we have shown $V_{1k,NT}-\mathbb{B}_{1k,NT}\Rightarrow N\left(0,\mathbb{V}_{(k)}\right)$. This completes the proof of (ii).

(iii) For $V_{2k,NT}$, we have

$$\begin{aligned}
V_{2k,NT} &= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'\tilde{x}_{2,i}\Sigma_{22,i}^{-1}\left(\Sigma_{20,i}-\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i\right) \\
&= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_{1,i}'x_{2,i}\Sigma_{22,i}^{-1}\left(\Sigma_{20,i}-\frac{1}{T}x_{2,i}'u_i\right)+\frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_{1,i}'x_{2,i}\Sigma_{22,i}^{-1}\bar{x}_{2,i}\bar{u}_i \\
&\quad+\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\bar{x}_{1,i}\bar{x}_{2,i}'\Sigma_{22,i}^{-1}\left(\frac{1}{T}x_{2,i}'u_i-\Sigma_{20,i}\right)-\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\bar{x}_{1,i}\bar{x}_{2,i}'\Sigma_{22,i}^{-1}\bar{x}_{2,i}\bar{u}_i \\
&\equiv V_{2k,NTa}+V_{2k,NTb}+V_{2k,NTc}+V_{2k,NTd},\text{ say.}
\end{aligned}$$

Noting that $\frac{1}{N_kT}\sum_{i\in G_k^0}\left\|\bar{x}_{1,i}\right\|^2=O_P(1)\,,\ \frac{1}{N_k}\sum_{i\in G_k^0}\left\|\bar{x}_{2,i}\right\|^2=O_P\left(p_2T^{-1}\right)$, and $\frac{1}{N_kT^2}\sum_{i\in G_k^0}\left\|x_{1,i}'x_{2,i}\right\|^2=O_P(p_2)$ by direct moment calculation and Markov inequality and $\max_{1\le i\le N}\left\|\bar{x}_{2,i}\right\|=o_P(p_2^{1/2}a_{2NT})$ and $\max_{1\le i\le N}\left|\bar{u}_i\right|=o_P(a_{2NT})$ by a simple application of Lemma S.1.2 in Su, Shi and Phillips (2016b, SSPb hereafter), we have

$$\begin{aligned}
\|V_{2k,NTb}\| &\leq \sqrt{N_k}\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\left\|x_{1,i}'x_{2,i}\right\|^2\right\}^{1/2}\max_{i\in G_k^0}\|\bar{x}_{2,i}\|\max_{i\in G_k^0}\|\bar{u}_i\|\left\|\Sigma_{22,i}^{-1}\right\|_{\text{sp}} \\
&= \sqrt{N_k}O_P(p_2^{1/2})o_P(p_2^{1/2}a_{2NT})o_P(a_{2NT}) = o_P(1), \\
\|V_{2k,NTc}\| &\leq \sqrt{N_k}\left\{\frac{1}{N_kT}\sum_{i\in G_k^0}\|\bar{x}_{1,i}\|^2\right\}^{1/2}\left\{\frac{1}{N_k}\sum_{i\in G_k^0}\|\bar{x}_{2,i}\|^2\right\}^{1/2}\max_{i\in G_k^0}\left\|\frac{1}{T}x_{2,i}'u_i-\Sigma_{20,i}\right\|\left\|\Sigma_{22,i}^{-1}\right\|_{\text{sp}} \\
&= \sqrt{N_k}O_P(1)O_P(p_2^{1/2}T^{-1/2})o_P(p_2^{1/2}a_{2NT}) = o_P(1),
\end{aligned}$$

and

$$\begin{aligned}
\|V_{2k,NTd}\| &\leq \sqrt{N_kT}\left\{\frac{1}{N_kT}\sum_{i\in G_k^0}\|\bar{x}_{1,i}\|^2\right\}^{1/2}\max_{i\in G_k^0}\|\bar{x}_{2,i}\|^2\max_{i\in G_k^0}\|\bar{u}_i\|\left\|\Sigma_{22,i}^{-1}\right\|_{\text{sp}} \\
&= \sqrt{N_kT}o_P(p_2a_{2NT}^2)o_P(a_{2NT}) = o_P(1).
\end{aligned}$$

For $V_{2k,NTa}$, it is easy to see that

$$\begin{aligned}
\|V_{2k,NTa}\| &\leq \sqrt{N_k}\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\left\|x_{1,i}'x_{2,i}\right\|^2\right\}^{1/2}\left\{\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\Sigma_{20,i}-\frac{1}{T}x_{2,i}'u_i\right\|^2\right\}^{1/2}\max_{1\leq i\leq N}\left\|\Sigma_{22,i}^{-1}\right\|_{\text{sp}} \\
&= \sqrt{N_k}O_P\left(p_2^{1/2}\right)O_P\left(T^{-1/2}p_2^{1/2}\right)O(1) = O_P\left(p_2\sqrt{N_k/T}\right)
\end{aligned}$$

which is $o_P(1)$ if we assume that $p_2^2N_k/T = o(1)$. But this is a very strong assumption that we try to avoid. To do this, we can employ the BN decomposition and write $x_{1,it} = S_1\left[\psi_i(1)\sum_{s=1}^t e_{is} + \breve{e}_{i,0} - \breve{e}_{i,t}\right]$ and $x_{2,it} = S_2\left[\psi_i(1)e_{it} + \breve{e}_{i,t-1} - \breve{e}_{i,t}\right]$. Let $B_{iT} = \Sigma_{22,i}^{-1}\left(\Sigma_{20,i} - \frac{1}{T}x_{2,i}'u_i\right)$. As in the analysis of $V_{1k,NT}$, we can show that

$$\begin{aligned}
V_{2k,NTa} &= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_{1,i}'x_{2,i}B_{iT} \\
&= \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^T S_1\psi_i(1)E_{it}e_{it}'\psi_i(1)'S_2'B_{iT} + o_P(1) \\
&= \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\left\{S_1\psi_i(1)\psi_i(1)'S_2'B_{iT} + \frac{1}{T}\sum_{t=1}^T S_1\psi_i(1)(e_{it}e_{it}'-I_p)\psi_i(1)'S_2'B_{iT}\right. \\
&\quad \left.\frac{1}{T}\sum_{t=1}^T S_1\psi_i(1)E_{i,t-1}e_{it}'\psi_i(1)'S_2'B_{iT}\right\} + o_P(1) \\
&\equiv \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\{R_{2iT,1} + R_{2iT,2} + R_{2iT,3}\} + o_P(1).
\end{aligned}$$

Noting that $E(R_{2iT,1})=0$ and

$$\begin{aligned}\left\|\mathrm{Var}\left(\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{2iT,1}\right)\right\| &\leq \frac{1}{N_k}\sum_{i\in G_k^0}\left\|\mathrm{Var}\left(R_{2iT,1}\right)\right\|=\frac{1}{N_k}\sum_{i\in G_k^0}\left\|S_1\Omega_iS_2'E\left(B_{iT}B_{iT}'\right)S_2\Omega_iS_1'\right\|\\ &\leq \frac{1}{N_k}\sum_{i\in G_k^0}\left\|E\left(B_{iT}B_{iT}'\right)\right\|_{\mathrm{sp}}\left\|S_1\Omega_iS_2'S_2\Omega_iS_1'\right\|\\ &\leq \frac{C}{N_k}\sum_{i\in G_k^0}\left\|E\left(B_{iT}B_{iT}'\right)\right\|=O\left(p_2T^{-1}\right)=o\left(1\right),\end{aligned}$$

we have $\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{2iT,1}=o_P\left(1\right)$. Next,

$$\begin{aligned}\left\|\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{2iT,2}\right\| &\leq \sqrt{N_k}\left\{\frac{1}{N_k}\sum_{i\in G_k^0}\left\|\frac{1}{T}\sum_{t=1}^{T}S_1\psi_i\left(1\right)\left(e_{it}e_{it}'-I_p\right)\psi_i\left(1\right)'S_2'\right\|^2\right\}^{1/2}\left\{\frac{1}{N_k}\sum_{i\in G_k^0}\left\|B_{iT}\right\|^2\right\}^{1/2}\\ &= \sqrt{N_k}O_P(p_2^{1/2}T^{-1/2})O_P(p_2^{1/2}T^{-1/2})=o_P\left(1\right).\end{aligned}$$

Let $z_{it}=S_1\psi_i\left(1\right)E_{i,t-1}e_{it}'\psi_i\left(1\right)'S_2'$. Then $\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{2iT,3}=\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^{T}z_{it}B_{iT}$. Noting that $E(z_{it}|\mathcal{F}_{i,t-1})=0$, by the Burkholder's and Hölder's inequalities, for any $r\geq 2$,

$$E\left\|\sum_{t=1}^{T}z_{it}\right\|^r\leq CE\left\{\sum_{t=1}^{T}\left\|z_{it}\right\|^2\right\}^{r/2}\leq C_1\left\{\sum_{t=1}^{T}E\left(\left\|z_{it}\right\|^r\right)\right\}^{r/2}\leq C_2p_2^{r/2}\sum_{t=1}^{T}t^{r/2}\leq C_2p_2^{r/2}T^{r/2+1}.$$

where C_1 and C_2 are constants that depend on r. Then by the Hölder's inequality

$$\begin{aligned}E\left\|\frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}R_{2iT,3}\right\| &\leq \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}E\left\|\sum_{t=1}^{T}z_{it}B_{iT}\right\|\\ &\leq \frac{1}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\left\{E\left\|\sum_{t=1}^{T}z_{it}\right\|^r\right\}^{1/r}\left\{E\left\|B_{iT}\right\|^q\right\}^{1/q}\\ &\leq \frac{C}{\sqrt{N_k}}\sum_{i\in G_k^0}\frac{1}{T}\left\{p_2^{r/2}T^{r/2+1}\right\}^{1/r}p_2^{1/2}T^{-1/2}\\ &\leq C\sqrt{N_k}p_2T^{-1+1/r}=o\left(1\right).\end{aligned}$$

where $\frac{1}{q}+\frac{1}{r}=1$.

Consequently, we have shown that $V_{2k,NT}=o_P\left(1\right)$.

(iv) Following the analysis of $V_{4k,NT}$ below, we can readily show that

$$\begin{aligned}V_{3k,NT} &= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'\tilde{x}_{2,i}\Sigma_{22,i}^{-1}\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}-\Sigma_{22,i}\right)\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\Sigma_{20,i}\\ &= \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}x_{1,i}'x_{2,i}\Sigma_{22,i}^{-1}\left(\frac{1}{T}x_{2,i}'x_{2,i}-\Sigma_{22,i}\right)\Sigma_{22,i}^{-1}\Sigma_{20,i}+o_P\left(1\right)\\ &\equiv V_{3k,NTa}+o_P\left(1\right).\end{aligned}$$

Following the analysis of $V_{2k,NT}$, we can show that $\|V_{3k,NTa}\| = o_P(1)$ by resorting to the BN decomposition, moment calculations, and Chebyshev inequality.

(v) For $V_{4k,NT}$, by the Cauchy-Schwarz inequality and Lemmas A.2(iii)-(iv) and A.6(iv),

$$\begin{aligned}
\|V_{4k,NT}\| &\leq \sqrt{N_k}\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\left\|\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\|^2\right\}^{1/2}\max_{i\in G_k^0}\left\|\left(\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}-\Sigma_{22,i}^{-1}\right\|_{\text{sp}}\max_{i\in G_k^0}\left\|\frac{1}{T}\tilde{x}_{2,i}'\tilde{u}_i-\Sigma_{20,i}\right\| \\
&= \sqrt{N_k}O_P(p_2^{1/2})O_P\left(p_2a_{2NT}\right)O_P(p_2^{1/2}a_{2NT}) = o_P(1).
\end{aligned}$$

(vi) This follows from (i)-(v). ■

Proof of Lemma A.8. Let $\mathcal{V}_{k,NT}^a = \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'M_{2,i}\tilde{v}_i^a$. We make the following decomposition

$$\mathcal{V}_{k,NT}^a = \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'\tilde{v}_i^a - \frac{1}{\sqrt{N_k}T}\sum_{i\in G_k^0}\tilde{x}_{1,i}'\tilde{x}_{2,i}\left(\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\tilde{x}_{2,i}'\tilde{v}_i^a \equiv \mathcal{V}_{1k,NT}^a - \mathcal{V}_{2k,NT}^a.$$

Noting that $v_{it}^a = \sum_{|j|\geq\bar{p}_2}\gamma_{i,j}'\varepsilon_{1,t+j}$, we have

$$\max_{i,t}E\left[(v_{it}^a)^2\right] = \max_{i,t}\sum_{|j|\geq\bar{p}_2}\sum_{|l|\geq\bar{p}_2}\gamma_{i,j}'E\left(\varepsilon_{1,t+j}\varepsilon_{1,t+j}'\right)\gamma_{i,l} \leq C\max_i\left(\sum_{|j|\geq\bar{p}_2}\|\gamma_{i,j}\|\right)^2 \leq CT^{-2a},$$

and

$$\max_i E\left[(\bar{v}_i^a)^2\right] = \max_{i,t}E\left(\frac{1}{T_{\bar{p}_2}}\sum_{t=\bar{p}_2+1}^{T-\bar{p}_2}v_{it}^a\right)^2 \leq \frac{1}{T_{\bar{p}_2}}\sum_{t=\bar{p}_2+1}^{T-\bar{p}_2}\max_{i,t}E\left(v_{it}^a\right)^2 \leq CT^{-2a}.$$

Then $\max_{i,t}E\left[(\tilde{v}_{it}^a)^2\right] \leq 2\max_{i,t}E\left[(v_{it}^a)^2\right] + 2\max_i E\left[(\bar{v}_i^a)^2\right] \leq 4CT^{-2a}$. Analogously, we can show that

$$\max_{i,t}E\left[(\tilde{v}_{it}^a)^4\right] = C\max_i\left(\sum_{|j|\geq\bar{p}_2}\|\gamma_{i,j}\|\right)^4 \leq CT^{-4a}.$$

It follows that

$$\begin{aligned}
E\left\|V_{1k,NT}^a\right\| &= E\left\|\frac{1}{\sqrt{N_k}T_{\bar{p}_2}}\sum_{i\in G_k^0}\sum_{t=\bar{p}_2+1}^{T-\bar{p}_2}x_{1,it}\tilde{v}_{it}^a\right\| \leq \frac{1}{\sqrt{N_k}T_{\bar{p}_2}}\sum_{i\in G_k^0}\sum_{t=\bar{p}_2+1}^{T-\bar{p}_2}E\left\|x_{1,it}\tilde{v}_{it}^a\right\| \\
&\leq \frac{1}{\sqrt{N_k}T_{\bar{p}_2}}\sum_{i\in G_k^0}\sum_{t=\bar{p}_2+1}^{T-\bar{p}_2}\left\{E\left\|x_{1,it}\right\|^2\right\}^{1/2}\left\{E\left\|\tilde{v}_{it}^a\right\|^2\right\}^{1/2} \\
&\leq C\sqrt{N_k}T_{p_2}^{-a-1}\sum_{t=\bar{p}_2+1}^{T-\bar{p}_2}t^{1/2} = O\left(N_k^{1/2}T^{-a+\frac{1}{2}}\right) = o(1),
\end{aligned}$$

where we use the fact that $\max_{1\leq i\leq N}E\left\|x_{1,it}\right\|^2 \leq Ct$. Then $V_{1k,NT}^a = o_P(1)$ by the Markov inequality.

Next, noting that $\left\|\frac{1}{T}\tilde{x}_{2,i}'\tilde{x}_{2,i}-\Sigma_{22,i}\right\|_{\text{sp}}=o_P(1)$ by Lemma A.2(iv) and

$$\begin{aligned}\max_i E\left\|x_{2,i}'\tilde{v}_i^a\right\|^2 &\leq \max_i \sum_{t=1}^T\sum_{s=1}^T E\left(x_{2,is}'x_{2,it}\tilde{v}_{it}^a\tilde{v}_{is}^{a2}\right)\\ &\leq C\max_i \sum_{t=1}^T\sum_{s=1}^T\left\{E\left\|x_{2,it}\right\|^4\right\}^{1/2}\left\{E\left(\tilde{v}_{it}^a\right)^4\right\}^{1/2}\\ &\leq C\sum_{t=1}^T\sum_{s=1}^T p_2T^{-2a}=Cp_2T^{-2a+2},\end{aligned}$$

we have

$$\begin{aligned}\left\|\mathcal{V}_{k,NT}^a\right\| &\leq \frac{1+o_P(1)}{\sqrt{N_k}T^2}\sum_{i\in G_k^0}\left\|\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\|\left\|\Sigma_{22,i}^{-1}\right\|_{\text{sp}}\left\|\tilde{x}_{2,i}'\tilde{v}_i^a\right\|\\ &\leq \underline{c}_{22}^{-1}\frac{1+o_P(1)}{\sqrt{N_k}T^2}\sum_{i\in G_k^0}\left\|\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\|\left\|x_{2,i}'\tilde{v}_i^a\right\|\\ &\leq \underline{c}_{22}^{-1}\left(1+o_P(1)\right)\sqrt{N_k}\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\left\|\tilde{x}_{1,i}'\tilde{x}_{2,i}\right\|^2\right\}^{1/2}\left\{\frac{1}{N_kT^2}\sum_{i\in G_k^0}\left\|x_{2,i}'\tilde{v}_i^a\right\|^2\right\}^{1/2}\\ &= \sqrt{N_k}O_P\left(p_2^{1/2}\right)O_P\left(p_2^{1/2}T^{-a}\right)=O_P\left(p_2N_k^{1/2}T^{-a}\right)=o_P(1).\end{aligned}$$

In sum, we have shown that $\mathcal{V}_{k,NT}^a=o_P(1)$. ■

C Determination of the Number of Groups

In this section, we now propose a BIC-type information criterion to choose K, the number of groups. We now use K_0 to denote the true number of groups and K a generic number of groups. We assume that the true number of groups is bounded from above by a finite integer $K_{\max}$ and $1\leq K_0\leq K_{\max}$. By minimizing the objective function in (2.7), we obtain the C-Lasso estimators $\{\hat{\alpha}_k(K,\lambda),\hat{\beta}_{1,i}(K,\lambda),\hat{\beta}_{2,i}(K,\lambda)\}$ of $\{\alpha_k,\beta_{1,i},\beta_{2,i},\}$ where we make the dependence of these estimators on (K,λ) explicit. We classify individual i into group $\hat{G}_k(K,\lambda)$ if and only if $\hat{\beta}_{1,i}(K,\lambda)=\hat{\alpha}_k(K,\lambda)$, i.e.,

$$\hat{G}_k(K,\lambda)=\{i=\{1,2,...,N\}:\hat{\beta}_{1,i}(K,\lambda)=\hat{\alpha}_k(K,\lambda)\}\text{ for }k=1,...K. \tag{C.1}$$

Let $\hat{G}(K,\lambda)=\{\hat{G}_1(K,\lambda),\cdots,\hat{G}_K(K,\lambda)\}$. We can define the post-Lasso estimators of α_k as

$$\hat{\alpha}_{\hat{G}_k(K,\lambda)}^{\text{post}}=\left(\sum_{i\in\hat{G}_k(K,\lambda)}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i}\right)^{+}\sum_{i\in\hat{G}_k(K,\lambda)}\tilde{x}_{1,i}'M_{2,i}\tilde{y}_i.$$

$\hat{\beta}_{2,i}^{\text{post}}(\hat{G}_k(K,\lambda))$ is defined as before but now we also make its dependence on $\hat{G}_k(K,\lambda)$ explicit. Let $\hat{\sigma}_{\hat{G}_k(K,\lambda)}^2=\frac{1}{NT}\sum_{k=1}^K\sum_{i\in\hat{G}_k(K,\lambda)}\sum_{t=1}^T[\hat{u}_{it}(k)]^2$, where $\hat{u}_{it}(k)=\tilde{y}_{it}-\tilde{x}_{1,it}'\hat{\alpha}_{\hat{G}_k(K,\lambda)}^{\text{post}}-\tilde{x}_{2,it}'\hat{\beta}_{2,i}^{\text{post}}(\hat{G}_k(K,\lambda))$

for $i \in \hat{G}_k(K, \lambda)$. We choose $\hat{K} = \hat{K}(\lambda)$ to minimize the following information criterion:

$$IC(K, \lambda) = \ln[\hat{\sigma}^2_{\hat{G}_k(K,\lambda)}] + \rho_{NT} p_1 K, \tag{C.2}$$

where ρ_{NT} is a tuning parameter.

Let $G^{(K)} = (G_{K,1}, ..., G_{K,K})$ be any K-partition of the set of $\{1, 2, ..., N\}$ and $\mathcal{G}_K$ is a collection of such partitions. Let $\hat{\sigma}^2_{G^{(K)}} = \frac{1}{NT} \sum_{k=1}^{K} \sum_{i \in G_{K,k}} \sum_{t=1}^{T} [\tilde{y}_{it} - \tilde{x}'_{1,it} \hat{\alpha}_{G_{K,k}} - \tilde{x}'_{2,it} \hat{\beta}_{2,i}(G_{K,k})]^2$, where $\hat{\alpha}_{G_{K,k}} = (\sum_{i \in G_{K,k}} \tilde{x}'_{1,i} M_{2,i} \tilde{x}_{1,i})^{+} \sum_{i \in G_{K,k}} \tilde{x}'_{1,i} M_{2,i} \tilde{y}_i$ and $\hat{\beta}_{2,i}(G_{K,k}) = \left(\tilde{x}'_{2,i} \tilde{x}_{2,i}\right)^{-1} \tilde{x}'_{2,i} \left(\tilde{y}_i - \tilde{x}_{1,i} \hat{\alpha}_{G_{K,k}}\right)$ for any $i \in G_{K,k}$. Define

$$\nu_{NT} = \begin{cases} N^{1/2} T^{1/2} & \text{in Case 1 where } x_{2,it} \text{ is absent in (2.1) and there is no endogeneity in } x_{1,it}, \\ T^{1/2} & \text{in Case 2 where } x_{2,it} \text{ is absent in (2.1) and there is endogeneity in } x_{1,it}, \\ p_2^{-1/2} T^{1/2} & \text{in Case 3 where } x_{2,it} \text{ is present in (2.1).} \end{cases} \tag{C.3}$$

Let $\sigma^2_{0,NT} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{i=1}^{T} \tilde{u}^2_{it}$ in Cases 1-2 and $= \frac{1}{NT} \sum_{i=1}^{N} \sum_{i=1}^{T} \left(\tilde{u}^*_{it}\right)^2$ in the Case 3. We can show that $\hat{\sigma}^2_{\hat{G}_k(K_0,\lambda)} - \sigma^2_{0,NT} = O_P\left(N^{-1} T^{-1}\right), O_P\left(T^{-1}\right)$, and $O_P\left(p_2 T^{-1}\right)$ corresponding to the above three cases, respectively.

We add the following assumption.

Assumption A.5 (i) As $(N, T) \to \infty$, $\min_{1 \leq K < K_0} \inf_{G^{(K)} \in \mathcal{G}_K} \hat{\sigma}^2_{G^{(K)}} \to \underline{\sigma}^2 > \sigma^2_0$, where $\sigma^2_0 = \text{plim}_{(N,T) \to \infty} \sigma^2_{0,NT}$.

(ii) As $(N, T) \to \infty$, $\rho_{NT} \to 0$ and $\rho_{NT} \nu^2_{NT} \to \infty$ where ν_{NT} is as defined in (C.3).

Assumption A.5(i) guarantees that all under-grouped models yield asymptotic mean square errors that are larger than σ^2_0, which can be obtained from the true model. Assumption A.5(ii) imposes the usual type of conditions for the consistency of model selection: the penalty coefficient ρ_{NT} cannot shrink to zero either too fast or too slowly.

The following theorem suggests that in large samples we can determine the correct number of groups by minimizing the information criterion defined in (C.2).

Theorem C.1 *Suppose that Assumptions A.1, A.3 and A.5 hold. Suppose that there exists a constant* $\underline{c}_{00}$ *such that* $\min_{1 \leq i \leq N} \Sigma_{00,i} \geq \underline{c}_{00} > 0$. *Then* $P(\hat{K} = K_0) \to 1$ *as* $(N, T) \to \infty$.

Theorem C.1 indicates that w.p.a.1 the use of IC(K, λ) in (C.2) determines the correct number of groups. A natural question is how to choose the tuning parameter ρ_{NT} empirically.

In simulations and applications, we recommend the use of DPLS estimation so that Case 3 applies. We will choose $p_2 = [T^{1/4}]$ and set $\rho_{NT} = \frac{1}{3}(NT)^{-1/3}$. Note that this rate converges to zero much slower than the usual $(NT)^{-1} \ln(NT)$-rate that works in Case 1. One can verify that the conditions in A.5(ii) are satisfied in this case when N and T diverge to infinity at roughly the same rate. Our simulations suggest that the choice of ρ_{NT} has little effect on the results.

Proof of Theorem C.1. Let $\mathcal{K} = \{1, 2, ..., K_{\max}\}$ where $K_{\max} \geq K_0$. We divide K into three subsets K_0, K_- and K_+ : $\mathcal{K}_0 = \{K_0\}$, $\mathcal{K}_- = \{K \in \mathcal{K} : K < K_0\}$, and $\mathcal{K}_+ = \{K \in \mathcal{K} : K > K_0\}$. First, using arguments as used in the proof of Lemma A.5(iii) we can show that

$$\hat{\sigma}^2_{\hat{G}(K_0,\lambda)} = \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in \hat{G}_k(K_0,\lambda)} \sum_{t=1}^{T} \left[\tilde{y}_{it} - \tilde{x}'_{1,it} \hat{\alpha}^{D,\text{post}}_k(K_0, \lambda) - \tilde{x}'_{2,it} \hat{\beta}^{D,\text{ post}}_{2,i}\left(\hat{G}_k(K_0, \lambda)\right)\right]^2 = \sigma^2_0 + o_P(1).$$

It follows that $IC(K_0,\lambda)=\ln[\hat{\sigma}^2_{\hat{G}(K_0,\lambda)}]+\rho_{NT}p_1K_0=\ln[\hat{\sigma}^2_{\hat{G}(K_0,\lambda)}]+o(1)\xrightarrow{P}\ln(\sigma_0^2)$. We consider the cases of under- and over-fitted models separately.

Case 1: Under-fitted model ($K\in\mathcal{K}_-$). Noting that

$$\begin{aligned}\hat{\sigma}^2_{\hat{G}(K,\lambda)} =& \frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in\ \hat{G}_k(K,\lambda)}\sum_{t=1}^{T}\left[\tilde{y}_{it}-\tilde{x}_{1,it}'\hat{\alpha}_k^{D,\text{post}}(K,\lambda)-\tilde{x}_{2,it}'\hat{\beta}_{2,i}^{D,\text{ post}}\left(\hat{G}_k(K,\lambda)\right)\right]^2\\ &\geq\min_{1\leq K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\frac{1}{NT}\sum_{k=1}^{K}\sum_{i\in G_{K,k}}\sum_{t=1}^{T}\left[\tilde{y}_{it}-\tilde{x}_{1,it}'\hat{\alpha}^D_{\hat{G}_{K,k}}-\tilde{x}_{2,it}'\hat{\beta}^D_{2,i}(\hat{G}_{K,k})\right]^2=\min_{1\leq K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\hat{\sigma}^2_{G^{(K)}}.\end{aligned}$$

By Assumption 4.2, we demonstrate

$$\min_{1\leq K<K_0}IC(K,\lambda)\geq\min_{1\leq K<K_0}\inf_{G^{(K)}\in G_K}\ln(\hat{\sigma}^2_{G^{(K)}})+\rho_{NT}pK\xrightarrow{P}\ln(\underline{\sigma}^2)>\ln(\sigma_0^2).$$

It follows that $P(\min_{K\in\mathcal{K}_-}IC(K,\lambda)>IC(K_0,\lambda))\to 1$.

Case 2: Over-fitted model ($K\in\mathcal{K}_+$).

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

where $\min_{K\in K^+}\nu_{NT}^2(\hat{\sigma}^2_{\hat{G}(K,\lambda)}-\hat{\sigma}^2_{\hat{G}(K_0,\lambda)})=O_p(1)$ by Lemma C.2 below and $\nu_{NT}^2\rho_{NT}\to\infty$ by Assumption A.5. ■

Lemma C.2 *Let* $\bar{\sigma}^2_{0,NT}=\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\tilde{u}_{it}^2$. *Let the conditions in Theorem C.1 hold. Then* $\max_{K_0<K\leq K_{\max}}|\hat{\sigma}^2_{\hat{G}(K,\lambda)}-\bar{\sigma}^2_{0,NT}|=O_p(\nu_{NT}^{-2})$.

Proof of Lemma C.2. When $K>K_0$, following the proof of Theorem 4.1, we can show that $\|\hat{\beta}_{1,i}-\beta^0_{1,i}\|=O_P(T^{-1}+\lambda)$, $\|\hat{\beta}_{2,i}-\beta^*_{2,i}\|=O_P(p_2^{1/2}(T^{-1/2}+\lambda))$, and $\frac{1}{N}\sum_{i=1}^N\prod_{k=1}^K\|\beta^0_{1,i}-\hat{\alpha}_k\|=O_P(b_TT^{-1})$. Noting that $\beta^0_{1,i}$, $i=1,...,N$, only take K_0 distinct values, the latter implies that the collection $\mathcal{C}\equiv\{\hat{\alpha}_1,...,\hat{\alpha}_K\}$ contains at least K_0 distinct vectors, say, $\hat{\alpha}_1,...,\hat{\alpha}_{K_0}$, such that $\hat{\alpha}_k-\alpha_k^0=O_P(b_TT^{-1})$ for $k=1,...,K_0$. For notational simplicity, we rename the other vectors in the above collection as $\hat{\alpha}_{K_0+1},...,\hat{\alpha}_K$. By the pointwise convergence of $\hat{\beta}^D_{1,i}$, $\hat{\alpha}_{K_0+1},...,\hat{\alpha}_K$ must converge in probability to one of the true values in $\{\alpha_1^0,...,\alpha^0_{K_0}\}$.

We classify $i\in\hat{G}_k(K,\lambda)$ if $||\hat{\beta}_{1,i}-\hat{\alpha}_k||=0$ for $k=1,...,K$, and $i\in\hat{G}_0(K,\lambda)$ otherwise. Using arguments like those used in the proof of Theorem 4.3 and that of Lemma S1.14 in SSPb, we can show that

$$\sum_{i\in G_k^0}P(\hat{E}_{kNT,i})=o(1)\text{ and }\sum_{i\in\hat{G}_k(K,\lambda)}P(\hat{F}_{kNT,i})=o(1)\text{ for }k=1,...,K_0. \quad \text{(C.4)}$$

This implies that

$$\sum_{i=1}^{N} P(i \in \hat{G}_0(K,\lambda) \cup \hat{G}_{K_0+1}(K,\lambda) \cup ... \cup \hat{G}_K(K,\lambda)) = o(1). \tag{C.5}$$

That is, the 'redundant' last $K - K_0$ groups containing empty elements asymptotically. Using the fact that $\mathbf{1}\{i \in \hat{G}_k\} = \mathbf{1}\{i \in G_k^0\} + \mathbf{1}\{i \in \hat{G}_k \backslash G_k^0\} - \mathbf{1}\{i \in G_k^0 \backslash \hat{G}_k\}$, we have

$$\hat{\sigma}^2_{\hat{G}(K,\lambda)} = \frac{1}{NT} \sum_{k=1}^{K} \sum_{i \in \hat{G}_k(K,\lambda)} \sum_{t=1}^{T} [\hat{u}_{it}(k)]^2 = D_{1NT} + D_{2NT} - D_{3NT} + D_{4NT},$$

where $D_{1NT} = \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k^0} \sum_{t=1}^{T} [\hat{u}_{it}(k)]^2$, $D_{2NT} = \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in \hat{G}_k(K,\lambda) \backslash G_k^0} \sum_{t=1}^{T} [\hat{u}_{it}(k)]^2$, $D_{3NT} = \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k^0 \backslash \hat{G}_k(K,\lambda)} \sum_{t=1}^{T} [\hat{u}_{it}(k)]^2$, and $D_{4NT} = \frac{1}{NT} \sum_{k=K_0+1}^{K} \sum_{i \in \hat{G}_k} \sum_{t=1}^{T} [\hat{u}_{it}(k)]^2$. By (C.4)-(C.5), we can readily show that $D_{\ell NT} = o_P\left((NT)^{-1}\right)$ for $\ell = 2, 3, 4$. For D_{1NT}, we discuss several cases: (1) When $x_{2,it}$ is absent in the cointegrating regression and there is no endogeneity in $x_{1,it}$, we can apply the fact $\hat{\alpha}_k^{\text{post}}$, $k = 1, ..., K_0$, converge to their true values at $N^{-1/2}T^{-1}$ and show that $D_{1NT} - \bar{\sigma}^2_{0NT} = O_P\left(N^{-1}T^{-1}\right)$; (2) When $x_{2,it}$ is absent in the cointegrating regression and there is endogeneity in $x_{1,it}$, we can apply the fact $\hat{\alpha}_k^{\text{post}}$, $k = 1, ..., K_0$, converge to their true values at rate T^{-1} and show that $D_{1NT} - \bar{\sigma}^2_{0NT} = O_P\left(T^{-1}\right)$; (3) When both $x_{1,it}$ and $x_{2,it}$ are present, we observe that $\hat{\beta}_{2,i}$ converge to their (pseudo) true values at rate $p_2^{1/2}T^{-1/2}$ and show that $D_{1NT} - \sigma_0^2 = O_P\left(pT^{-1}\right)$. As a result, we have $\hat{\sigma}^2_{\hat{G}(K,\lambda)} = \sigma_0^2 + O_P(\nu_{NT}^{-2})$ where $\nu_{NT} = N^{1/2}T^{1/2}$, $T^{1/2}$, and $p_2^{-1/2}T^{1/2}$ in the above three cases, respectively. This completes the proof of the lemma. ■

D Practical Implementation of the C-Lasso Procedure

In this section, we provide more details on the practical implementation of the C-Lasso procedure in the followings steps.

1. **Initial estimates based on the heterogenous nonstationary panels.** Obtain the initial estimates $\tilde{\beta}_{1,i}$ and $\tilde{\beta}_{2,i}$ from the LS time-series regression of $\tilde{y}_{it}$ on $\left(\tilde{x}'_{1,it}, \tilde{x}'_{2,it}\right)$. Let $Q_{NT}(\boldsymbol{\beta_1}, \boldsymbol{\beta_2}) = \frac{1}{NT^2} \sum_{i=1}^{N} \left\| \tilde{y}_i - \tilde{x}_{1,i}\beta_{1,i} - \tilde{x}_{2,i}\beta_{2,i} \right\|^2$, $\tilde{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^{T} (\tilde{y}_{it} - \tilde{\beta}'_i \tilde{x}_{it})^2$, and $\hat{Q}_{1i} = \frac{1}{T^2} \sum_{t=1}^{T} \tilde{x}_{1,it}\tilde{x}'_{1,it}$.

2. **Determining the number (K) of groups along with the tuning parameter** λ. Let

   $$\Lambda \equiv \left\{ \lambda = c_j T^{-3/4},\ c_j = c_0 \gamma^j \text{ for } j = 0, ..., J \right\} \text{ for some } c_0 > 0 \text{ and } \gamma > 1.$$

   Given any $K \in \{1, 2, ..., K_{\max}\}$ and $\lambda \in \Lambda$, compute $\text{IC}(K, \lambda)$ and $\text{IC}(\hat{K}(\lambda), \lambda)$ where $\hat{K}(\lambda) = \arg\min_{1 \le K \le K_{\max}} \text{IC}(K, \lambda)$. Choose $\hat{\lambda} \in \Lambda$ such that $\text{IC}(\hat{K}(\lambda), \lambda)$ is minimized. The estimated number of groups is then given by

   $$\hat{K} = \min_{\lambda \in \Lambda} \hat{K}(\lambda).$$

   Note that the above procedure fine-tunes the tuning parameter λ for the determination of the number of groups and is recommended by Su, Shi, and Phillips (2016a, SSPa hereafter). We find in simulations $c_0 = 0.025$, $\gamma = 2$, and $J = 3$ work fairly well for all DGPs under our investigation. If $\hat{K} = 1$, stop here and estimate a homogenous nonstationary panel as usual. Otherwise, move to the next step.

3. **C-Lasso estimation.** Given $\hat{\lambda}$ and $\hat{K} > 1$, solve the PLS problem

$$Q_{NT,\lambda}^{K}(\boldsymbol{\beta_1}, \boldsymbol{\beta}_2, \boldsymbol{\alpha}) = Q_{NT}(\boldsymbol{\beta_1}, \boldsymbol{\beta}_2) + \frac{\hat{\lambda}}{N}\sum_{i=1}^{N}(\tilde{\sigma}_i)^{2-\hat{K}}\prod_{k=1}^{\hat{K}}\left\|\hat{Q}_{1i}(\beta_{1,i} - \alpha_k)\right\|.$$

Obtain the C-Lasso estimates $\{\hat{\alpha}_k\}$ for the group-specific parameters and $\{\hat{G}_k, k = 1, ..., \hat{K}\}$ for the estimated group membership.

4. **Post-Lasso estimator with bias correction:** Given the estimated groups, $\{\hat{G}_k, k = 1, ..., \hat{K}\}$, we can obtain the post-Lasso estimators of α_k and $\beta_{2,i}$ as

$$\begin{aligned}\hat{\alpha}_k^{\text{post}} &= \left(\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i}\right)^{-1}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{y}_i \text{ for } k = 1, ..., \hat{K},\\ \hat{\beta}_{2,i}^{\text{post}} &= \left(\tilde{x}_{2,i}'\tilde{x}_{2,i}\right)^{-1}\tilde{x}_{2,i}'(\tilde{y}_i - \tilde{x}_{1,i}\hat{\alpha}_k^{\text{post}}) \text{ for } i \in \hat{G}_k,\end{aligned}$$

where to remove the bias we apply the dynamic OLS method in the post-Lasso estimation by including the lags and leads of $\Delta x_{1,it}$ into $x_{2,it}$ as in Section 4.4. If $x_{2,it}$ only contains the lags and leads of $\Delta x_{1,it}$ but no other stationary regressors, we compute the standard errors for the elements of $\hat{\alpha}_k^{\text{post}}$ as the square roots of the diagonal elements of $\frac{1}{\hat{N}_k T^2}\hat{\mathbb{Q}}_{(k)}^{-1}\hat{\mathbb{V}}_{(k)}^{\dagger}\hat{\mathbb{Q}}_{(k)}^{-1}$ where

$$\hat{\mathbb{Q}}_{(k)} = \frac{1}{\hat{N}_k T^2}\sum_{i\in\hat{G}_k}\tilde{x}_{1,i}'M_{2,i}\tilde{x}_{1,i} \text{ and } \hat{\mathbb{V}}_{(k)}^{\dagger} \equiv \frac{1}{\hat{N}_k}\sum_{i\in\hat{G}_k}\frac{1}{6}\hat{\Omega}_{00,i}^{\dagger}\hat{\Omega}_{11,i} \text{ for } k = 1, ..., \hat{K},$$

and $\hat{\Omega}_{00,i}^{\dagger}$ and $\hat{\Omega}_{11,i}$ are as defined in Section 4.4. If $x_{2,it}$ also contains other stationary covariates, then we can compute the standard errors for the elements of $\hat{\alpha}_k^{\text{post}}$ as the square roots of the diagonal elements of $\frac{1}{\hat{N}_k T^2}\hat{\mathbb{Q}}_{(k)}^{-1}\hat{\mathbb{V}}_{(k)}\hat{\mathbb{Q}}_{(k)}^{-1}$ where

$$\hat{\mathbb{V}}_{(k)} = \frac{1}{\hat{N}_k}\sum_{i\in\hat{G}_k}\left[\frac{1}{6}\hat{s}_i'\hat{\Omega}_i\hat{s}_iS_1\hat{\Omega}_iS_1' - \frac{1}{12}\left(\hat{s}_i'\hat{\Omega}_iS_1' \otimes S_1\hat{\Omega}_i\hat{s}_i\right)K_{p_1,1}\right],$$

$\hat{s}_i = S_0' - S_2'\hat{\Sigma}_{22,i}^{-1}\hat{\Sigma}_{20,i}$, $\hat{\Omega}_i$ denotes the HAC estimator of the long-run variance-covariance in Ω_i, and $\hat{\Sigma}_{22,i}$ and $\hat{\Sigma}_{20,i}$ denote the plug-in estimators of the short-run variance covariance submatrices $\Sigma_{22,i}$ and $\Sigma_{20,i}$ of Σ_i.

E Additional Simulation Results

In this appendix, we assess the performance of the information criterion (IC) proposed in Section C. We set $\rho_{NT} = \frac{1}{3}(NT)^{-1/3}$ and $\lambda = c_\lambda T^{-3/4}$ where $c_\lambda = 0.025$, 0.05, 0.1, or 0.2. We find that the results are not sensitive to the choice of c_λ and will only report the simulation results for the case where $c_\lambda = 0.1$ to save space. Table A.1 displays the empirical probability with which a particular group number from 1 to 6 is selected according to IC based on 500 replications for each DGP. Note that the true number of groups is 3 for DGPs 1, 2, 3, and 5 and 2 for DGP 4. When $T = 40$, the probabilities of correct choices are higher than 95 % in all cases and they reach the unity when $T = 80$. The simulation results show that our information criterion works fairly well.

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

	N	T	1	2	3	4	5	6
DGP1	50	40	0	0	0.992	0.008	0	0
	50	80	0	0	1	0	0	0
	100	40	0	0	1	0	0	0
	100	80	0	0	1	0	0	0
DGP2	50	40	0	0	0.966	0.034	0	0
	50	80	0	0	0.998	0.002	0	0
	100	40	0	0	0.982	0.018	0	0
	100	80	0	0	1	0	0	0
DGP3	50	40	0	0	0.988	0.012	0	0
	50	80	0	0	1	0	0	0
	100	40	0	0	1	0	0	0
	100	80	0	0	1	0	0	0
DGP4	50	40	0	0.976	0.024	0	0	0
	50	80	0	1	0	0	0	0
	100	40	0	0.956	0.044	0	0	0
	100	80	0	1	0	0	0	0
DGP5	50	40	0	0	0.990	0.010	0	0
	50	80	0	0	1	0	0	0
	100	40	0	0	0.986	0.014	0	0
	100	80	0	0	1	0	0	0

F Additional application results

In this section, we report some additional results for the empirical application

F.1 Information criterion for the quarterly data

Table A.2 reports the information criterion (IC) for the quarterly data with different tuning parameter values: $\lambda = c_\lambda \times T^{-3/4}$ where $c_\lambda = 0.025$, 0.05, 0.1, and 0.2. Following the majority rule, we decide to select $K = 2$ groups for the period 1975.Q1-1998.Q4 and $K = 3$ groups for the period 1999.Q1-2014.Q2. Note that the IC is minimized at $c_\lambda = 0.1$ and 0.05 for the first and second subsamples, respectively. For this reason, we choose $c_\lambda = 0.1$ and 0.05 for these two subsamples, in the paper.

F.2 Results for the monthly data

In this section we provide the application results for the monthly data.

Table A.3 reports the information criterion (IC) for the monthly data with different tuning parameter values: $\lambda = c_\lambda \times T^{-3/4}$ where $c_\lambda = 0.025$, 0.05, 0.1, and 0.2. As is evident from Table A.3, for the monthly data our information criterion tends to choose 2 groups for the first subsample and 3 groups for the second subsample, too. We set $c_\lambda = 0.05$ to report the estimation results in Table A.4 and classification results in Table A.5.

Comparing the estimation results in Table 4 for the quarterly data with those in Table A.4 for the monthly data, we find that the estimates for either group in either subsample period of the monthly

Table A.2: The information criterion for different numbers of groups (quarterly data)

	From 1975.Q1-1998.Q4				From 1999.Q1-2014.Q2			
K/c_λ	0.025	0.05	0.10	0.20	0.025	0.05	0.10	0.20
1	-0.7503	-0.7503	-0.7503	-0.7503	-0.2074	-0.2074	-0.2074	-0.2074
2	-1.1262	**-1.1262**	**-1.1262**	**-1.0716**	-0.4719	-0.4730	-0.4902	**-0.4836**
3	**-1.1622**	-0.7961	-1.0956	-0.7135	**-0.5230**	**-0.5319**	**-0.5268**	-0.4418
4	-0.7719	-0.7507	-0.7507	-1.0596	-0.5037	-0.4994	-0.4958	-0.3815
5	-0.7233	-0.7203	-0.6750	-0.6750	-0.4789	-0.4749	-0.3499	-0.2093
6	-0.6946	-0.6405	-0.6005	-0.6844	-0.4454	-0.4358	-0.3566	-0.1720

Table A.3: The information criterion for different numbers of groups (monthly data)

	From 1975-1998				From 1999-2014			
$K\backslash c_\lambda$	0.025	0.05	0.10	0.20	0.025	0.05	0.10	0.20
1	-0.8042	-0.8042	-0.8042	-0.8042	-0.1953	-0.1953	-0.1953	-0.1953
2	-1.0907	**-1.0907**	**-1.0907**	**-1.0140**	-0.4686	-0.4753	-0.4837	**-0.4748**
3	**-1.1460**	-0.8480	-0.8404	-0.8365	**-0.5312**	**-0.5311**	**-0.5230**	-0.3940
4	-1.0966	-1.0897	-0.8292	-0.9159	-0.5161	-0.5132	-0.5086	-0.3139
5	-0.9044	-1.0646	-0.9047	-0.7949	-0.5032	-0.4987	-0.3630	-0.2711
6	-0.8782	-1.0379	-0.7875	-0.7678	-0.4768	-0.4753	-0.3016	-0.2466

data are reasonably close to the corresponding estimates based on the quarterly data. This suggests the robustness of our results. The countries in bold in Table A.5 suggest good coincidences of the classification results based on the monthly and quarterly datasets.

References

Corradi, V. (1999) Deciding between I(0) and I(1) via FLIL-based bounds. *Econometric Theory* 15(5), 643-663.

Eberlein, E. (1986) On strong invariance principles under dependence assumptions. *Annals of Probability* 14, 260-270.

Freedman, D.A. (1975) On tail probabilities for martingales. *Annals of Probability* 3, 100-118.

Table A.4: Estimation results for the monthly data

	Panel A: From 1975.M1-1998.M12						
	Pool	Group 1		Group 2			
	DOLS	C-Lasso	post-Lasso	C-Lasso	post-Lasso		
β_i	0.7553	0.8350	0.8350	-0.7175	-0.7175		
	(0.0116)	(0.0097)	(0.0097)	(0.0554)	(0.0554)		
	Panel B: From 1999.M1-2014.M7						
	Pool	Group 1		Group 2		Group 3	
	DOLS	C-Lasso	post-Lasso	C-Lasso	post-Lasso	C-Lasso	post-Lasso
β_i	0.3886	0.8197	0.8197	-0.5097	-0.5096	0.1577	0.1577
	(0.0069)	(0.0086)	(0.0086)	(0.0154)	(0.0154)	(0.0179)	(0.0179)

Table A.5: Classification results for the monthly data

Panel A: From 1975.M1-1998.M12				
Group 1 ($N_1 = 53$)				
Algeria	**Austria**	**Bahrain**	**Belgium**	**Bolivia**
Botswana	**Canada**	**Colombia**	**Costa Rica**	**Cyprus**
Denmark	**Egypt**	**Finland**	**France**	**Ghana**
Greece	**Honduras**	**Hungary**	**India**	**Indonesia**
Israel	**Italy**	**Ivory Coast**	**Jamaica**	**Japan**
Jordan	**Kenya**	**South Korea**	**Luxembourg**	**Malta**
Mauritius	**Mexico**	**Morocco**	**Nepal**	**Netherlands**
Nigeria	**Norway**	**Pakistan**	**Paraguay**	**Philippines**
Portugal	**Singapore**	**South Africa**	**Spain**	**Sri Lanka**
Sudan	**Sweden**	**Switzerland**	**Thailand**	**Trinidad and Tobago**
Turkey	**Uruguay**	**Venezuela**		
Group 2 ($N_2 = 3$)				
Ecuador	**Kuwait**	**Myanmar**		
Panel B: From 1999.M1-2014.M7				
Group 1 ($N_1 = 53$)				
Angola	**Argentina**	**Austria**	**Bangladesh**	**Belgium**
Botswana	Cambodia	**Canada**	**Costa Rica**	**Denmark**
Dominican	Egypt	**Europe**	**Finland**	**France**
Germany	**Ghana**	**Honduras**	**Iceland**	India
Iran	**Italy**	**Jamaica**	**Japan**	**Jordan**
Luxembourg	**Malawi**	Mauritius	**Mexico**	**Mongolia**
Morocco	**Mozambique**	Nepal	**Netherlands**	**Nigeria**
Norway	**Pakistan**	**Romania**	**Saudi Arabia**	**Sri Lanka**
Sudan	**Sweden**	**Switzerland**	**Tanzania**	**Trinidad and Tobago**
Tunisia	**Turkey**	**Uganda**	**United Kingdom**	**Ukraine**
Uruguay	**Venezuela**	**Viet Nam**		
Group 2 ($N_2 = 20$)				
Albania	**Armenia**	**Brazil**	**Bulgaria**	**Colombia**
Congo	**Croatia**	**Georgia**	**Hungary**	**Ireland**
Ivory Coast	**Kuwait**	**Latvia**	Lithuania	**Macau**
Moldova	**Peru**	**Philippines**	**Spain**	**Thailand**
Group 3 ($N_3 = 21$)				
Algeria	Bolivia	**Czech Republic**	**Guatemala**	**Hong Kong**
Indonesia	**Israel**	**Kazakhstan**	**Kenya**	**South Korea**
Kyrgyzstan	**Laos**	**Macedonia**	**Malaysia**	**Myanmar**
Paraguay	**Poland**	**Portugal**	**Russia**	**Singapore**
South Africa				

Note: Countries in bold denote coincidences of the classification results based on the monthly and quarterly datasets.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application.* Academic Press, New York.

Park, J.Y. & P.C.B. Phillips (1988) Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4(3), 468-497.

Park, J.Y. & P.C.B. Phillips (1989) Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory* 5(1), 95-131.

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

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

Strassen, V. (1964) An invariance principle for the law of the iterated logarithm. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 3, 211-226.

Su, L., Z. Shi, & P.C.B. Phillips (2016a) Identifying latent structures in panel data. *Econometrica* 84(6), 2215-2264.

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