Determining the Number of Groups in Latent Panel Structures*

Xun Lu[a] and Liangjun Su[b]

[a]Department of Economics, Hong Kong University of Science & Technology

[b]School of Economics, Singapore Management University, Singapore

December 4, 2013

Abstract

We consider a latent group panel structure as recently studied by Su, Shi, and Phillips (2013), where the number of groups is unknown and has to be determined empirically. We propose a testing procedure to determine the number of groups. Our test is a residual-based LM-type test. We show that after being appropriately standardized, our test is asymptotically normally distributed under the null hypothesis of a given number of groups and has power to detect deviations from the null. Monte Carlo simulations show that our test performs remarkably well in finite samples. We apply our method to study the effect of income on democracy and find strong evidence of heterogeneity in the slope coefficients. Our testing procedure determines three latent groups among eighty-two countries.

Key words: Classification Lasso; Dynamic panel; Latent structure; Penalized least square; Number of groups; Test

JEL Classification: C12, C23, C33, C38, C52.

*The authors express their sincere appreciation to Stéphane Bonhomme, Xiaohong Chen, Cheng Hsiao, and Peter C. B. Phillips for discussions on the subject matter and valuable comments on the paper. Su gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. All errors are the authors' sole responsibilities. Address correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903; Phone: +65 6828 0386; e-mail: ljsu@smu.edu.sg.

1 Introduction

Recently latent group structures have received much attention in the panel data literature; see, e.g., Sun (2005), Sarafidis and Weber (2011), Bonhomme and Manresa (2012), Lin and Ng (2012), Bester and Hansen (2013), Deb and Trivedi (2013), and Su, Shi, and Phillips (2013). In comparison with some other popular approaches to model unobserved heterogeneity in panel data models such as random coefficient models (see, e.g., Hsiao (2003, chapter 6)), one important advantage of the latent group structure is that it allows flexible forms of unobservable heterogeneity while remaining parsimonious at the same time. In addition, the group structure has sound theoretical foundations from game theory or macroeconomic models where multiplicity of Nash equilibria is unavoidable (c.f. Hahn and Moon (2010)). The key question in latent group structures is how to identify each individual's group membership. Bester and Hansen (2013) assume that membership is known and determined by external information, say, external classification or geographic location, while others assume that it is unrestricted and unknown and propose statistical methods to achieve classification. Sun (2005) uses a parametric multinomial logit regression to model membership. Sarafidis and Weber (2011), Bonhomme and Manresa (2012), and Lin and Ng (2012) extend K-means classification algorithms to the panel regression framework. Deb and Trivedi (2013) propose EM algorithms to estimate finite mixture panel data models with fixed effects. Motivated by the sparse feature of the individual regression coefficients under latent group structures, Su, Shi, and Phillips (2013, SSP hereafter) propose a novel variant of the Lasso procedure, i.e., classifier Lasso (C-Lasso), to achieve classification. While these methods make important contributions by empirically grouping individuals, to implement these methods, we often need to determine the number of groups first. Some information criteria have been proposed to achieve this goal (see, e.g., Bonhomme and Manresa (2012) and SSP), which often rely on some tuning parameters. This paper provides a hypothesis-testing-based solution to determine the number of groups.

Specifically, in the same framework as in SSP, we consider the panel data structure:

$$y_{it} = \beta_i^{0\prime} X_{it} + \mu_i + u_{it}, \;\; i = 1, ..., N \text{ and } t = 1, ..., T, \tag{1.1}$$

where X_{it}, μ_i and u_{it} are the regressors, individual fixed effects, and idiosyncratic error term, respectively and β_i^0 is the slope coefficient that can depend on individual i. We assume that the N individuals belong to K groups and all individuals in the same group share the same slope coefficients. That is, β_i^0's are homogeneous within each of the K groups but heterogeneous across the K groups. For a given K, we can apply the C-Lasso method proposed in SSP to determine the group membership and estimate β_i^0's. However, in practice, K is unknown and has to be determined from data. This motivates us to test the following hypothesis:

$$\mathbb{H}_0\left(K_0\right) : K = K_0 \text{ versus } \mathbb{H}_1\left(K_0\right) : K_0 < K \leq K_{\max},$$

where K_0 and $K_{\max}$ are pre-specified by researchers. We can sequentially test the hypotheses $\mathbb{H}_0\left(1\right)$, $\mathbb{H}_0\left(2\right), ...,$ until we fail to reject $\mathbb{H}_0(K^*)$ for some $K^* \leq K_{\max}$ and conclude that the number of groups is K^*. Onatski (2009) applies a similar procedure to determine the number of latent factors in panel factor structures.

In addition to helping to determine the number of groups, testing $\mathbb{H}_0(K_0)$ itself is also useful for empirical research. When $K_0 = 1$, the test becomes a test for homogeneity in the slope coefficients, which is often assumed in empirical applications. When K_0 is some integer greater than 1, we test whether the group structure is correctly specified. Although the group structure is flexible in terms of modeling unobserved slope heterogeneity, it could still be misspecified. Inferences based on misspecified models are often misleading. Thus conducting a formal specification test is highly desirable.

Our test is a residual-based LM-type test. We estimate the model under the null hypothesis $\mathbb{H}_0(K_0)$ to obtain the restricted residuals, and the test statistic is based on whether the regressors have predictive powers on the restricted residuals. Under the null of correct number of latent groups, the regressors should not contain any useful information about the restricted residuals. We show that after being appropriately standardized, our test statistic is asymptotically normal under the null. The p-values can be obtained based on the standard normal approximation, and thus the test is easy to implement. Our test is related to the literature on testing slope homogeneity and poolability for panel data models in which case $K_0 = 1$. See, e.g., Pesaran, Smith, and Im (1996), Phillips and Sul (2003), Pesaran and Yamagata (2008), and Su and Chen (2013), among others. Nevertheless, none of the existing tests can be applied to test $K = K_0$, where $K_0 > 1$.

We conduct Monte Carlo simulations to show the excellent finite sample performance of our test. With a high probability, our method can determine the number of groups correctly. We apply our method to study the relationship between income and democracy. We find that indeed the slope coefficients (the marginal effects of income and lagged democracy on democracy) are heterogeneous with p-values being less than 0.001. Further, we determine that the number of heterogeneous groups is three and find that the slope coefficients of the three groups are substantially different from each other. Though our classification of groups is completely data-driven, we further investigate the determinants of the group pattern and find that the initial education level and long-run progress in democracy are important determinants.

The remainder of the paper is organized as follows. In Section 2, we introduce the hypotheses and the test statistic. In Section 3 we derive the asymptotic distributions of our test statistic under the null and study the global power of our test. We conduct Monte Carlo experiments to evaluate the finite sample performance of our test in Section 4 and apply it to the income-democracy dataset in Section 5. Section 6 concludes. All proofs are relegated to the Appendix.

To proceed, we adopt the following notation. For an $m \times n$ real matrix A, we denote its transpose as A' and its Frobenius norm as $\|A\|$ ($\equiv [\text{tr}(AA')]^{1/2}$) where $\equiv$ means "is defined as". Let $P_A \equiv A(A'A)^{-1} A'$ and $M_A \equiv I_m - P_A$, where I_m denotes an $m \times m$ identity matrix. When $A = \{a_{ij}\}$ is symmetric, we use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote its maximum and minimum eigenvalues, respectively, and denote $\text{diag}(A)$ as a diagonal matrix whose (i,i)th diagonal element is given by a_{ii}. Let $P_0 \equiv T^{-1}\mathbf{i}_T\mathbf{i}_T'$ and

$M_0 \equiv I_T - T^{-1}\mathbf{i}_T\mathbf{i}_T'$, where $\mathbf{i}_T$ is a $T \times 1$ vector of ones. Moreover, the operator $\xrightarrow{P}$ denotes convergence in probability, and $\xrightarrow{D}$ convergence in distribution. We use $(N, T) \to \infty$ to denote the joint convergence of N and T when N and T pass to infinity simultaneously. We abbreviate "positive semidefinite", "with probability approaching 1", and "without loss of generality" to "p.s.d.", "w.p.a.1", and "wlog", respectively.

2 Hypotheses and test statistic

In this section we introduce the hypotheses and test statistic.

2.1 Hypotheses

We consider the panel structure model

$$y_{it} = \beta_i^{0\prime} X_{it} + \mu_i + u_{it}, \; i = 1, ..., N \text{ and } t = 1, ..., T, \tag{2.1}$$

where X_{it} is a $p \times 1$ vector of exogenous or predetermined regressors, μ_i an individual fixed effect, and u_{it} the idiosyncratic error term. We assume that β_i^0 has the following group structure:

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

where K is an integer such that $1 \leq K < N$, and $\left\{G_1^0, ..., G_K^0\right\}$ forms a partition of $\{1, ..., N\}$ such that $\cup_{k=1}^K G_k^0 = \{1, ..., N\}$ and $G_k^0 \cap G_j^0 = \varnothing$ for any $j \neq k$. Further, $\alpha_k^0 \neq a_j^0$ for any $j \neq k$. Let $N_k = |G_k^0|$ be the number of members in G_k^0, i.e., the cardinality of the set G_k^0. We assume that K, $\mathcal{G}^0 \equiv \left\{G_1^0, ..., G_K^0\right\}$, $\boldsymbol{\alpha}^0 \equiv (\alpha_1^0, ..., \alpha_K^0)$, and $\boldsymbol{\beta}^0 \equiv (\beta_1^0, ..., \beta_N^0)$ are all unknown. One key step in estimating all these parameters is to first determine K, as once K is determined, we can readily apply the C-Lasso estimation method developed in SSP. This motivates us to test the following hypothesis:

$$\mathbb{H}_0\left(K_0\right) : K = K_0 \text{ versus } \mathbb{H}_1\left(K_0\right) : K_0 < K \leq K_{\max}. \tag{2.2}$$

The testing procedure developed below can be used to determine K. Suppose that we have a priori information such that $K_{\min} \leq K \leq K_{\max}$, where $K_{\min}$ is typically 1. Then we can first test: $\mathbb{H}_0(K_{\min})$ against $\mathbb{H}_1(K_{\min})$. If we fail to reject the null, then we conclude that $K = K_{\min}$. Otherwise, we continue to test $\mathbb{H}_0(K_{\min} + 1)$ against $\mathbb{H}_1(K_{\min} + 1)$. We repeat this procedure until we fail to reject the null $\mathbb{H}_0(K^*)$ and conclude that $K = K^*$. This procedure is similar to that in Onatski (2009) for determining the number of latent factors in panel factor structures.

2.2 Estimation under the null and test statistic

Our test is a residual-based test and so we only need to estimate the model under the null $\mathbb{H}_0(K_0)$: $K = K_0$. In the special case where $K_0 = 1$, the panel structure model reduces to a homogeneous panel data model so that $\beta_i^0 = \beta^0$ for all $i = 1, ..., N$, and we can estimate the homogeneous slope coefficient using the usual within-group estimator $\hat{\beta}$. In the general case where $K_0 > 0$, we consider the C-Lasso estimation proposed by SSP. Let $\tilde{y}_{it} = y_{it} - T^{-1}\sum_{t=1}^{T} y_{it}$, and $\tilde{X}_{it} = X_{it} - T^{-1}\sum_{t=1}^{T} X_{it}$. Let $\tilde{\boldsymbol{\beta}} \equiv (\tilde{\beta}_1, ..., \tilde{\beta}_N)$ and $\hat{\boldsymbol{\alpha}}_{K_0} = (\hat{\alpha}_1, ..., \hat{\alpha}_{K_0})$ be the C-Lasso estimators proposed in SSP, which are defined as the minimizer of the following criterion function:

$$Q_{1NT,\lambda}^{(K_0)}(\boldsymbol{\beta}, \boldsymbol{\alpha}_{K_0}) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda}{N}\Pi_{k=1}^{K_0}||\beta_i - \alpha_k||,$$

where $\lambda \equiv \lambda_{NT}$ is a tuning parameter and

$$Q_{1,NT}(\boldsymbol{\beta}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{y}_{it} - \beta_i'\tilde{X}_{it}\right)^2.$$

Let $\hat{G}_k = \left\{i \in \{1, 2, ..., N\} : \tilde{\beta}_i = \hat{\alpha}_k\right\}$ for $k = 1, ..., K_0$. Let $\hat{G}_0 = \{1, 2, ..., N\} \backslash (\cup_{k=1}^{K_0}\hat{G}_k)$. Although SSP demonstrate that the number of elements in $\hat{G}_0$ shrinks to zero as $T \to \infty$, in finite samples, $\hat{G}_0$ may not be empty. To fully impose the null hypothesis $\mathbb{H}_0(K_0)$, we force all the estimates of the slope coefficients to be grouped into K_0 groups and define the final estimators of β_i^0's as $\hat{\boldsymbol{\beta}} \equiv (\hat{\beta}_1, ..., \hat{\beta}_N)$ where

$$\hat{\beta}_i = \begin{cases} \tilde{\beta}_i & \text{if } i \in \hat{G}_k \text{ for some } k = 1, ..., K_0 \\ \hat{\alpha}_{k^*} & \text{otherwise} \end{cases},$$

where $k^* \equiv \arg\min_k \left\{||\tilde{\beta}_i - \hat{\alpha}_k||, k = 1, ..., K_0\right\}$. Note that we have suppressed the dependence of $\hat{\boldsymbol{\beta}}$, $\hat{\beta}_i$'s, and $\hat{G}_k$'s on K_0 to maintain notational simplicity.

Given $\{\hat{\beta}_i\}$, we can estimate individual fixed effects using $\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{\beta}_i' X_{it})$.[1] The residuals are obtained by

$$\hat{u}_{it} \equiv y_{it} - \hat{\beta}_i' X_{it} - \hat{\mu}_i. \tag{2.3}$$

It is easy to show that

$$\begin{aligned} \hat{u}_{it} &= (y_{it} - \bar{y}_i) - \left(X_{it} - \bar{X}_i\right)'\hat{\beta}_i \\ &= u_{it} - \bar{u}_i + \left(X_{it} - \bar{X}_i\right)'\left(\beta_i^0 - \hat{\beta}_i\right), \end{aligned} \tag{2.4}$$

where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, $\bar{X}_i = T^{-1}\sum_{t=1}^{T} X_{it}$, and $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$. Under the null hypothesis, $\hat{\beta}_i$ is a consistent estimator of β_i^0.[2] Hence, $\hat{u}_{it}$ should be close to u_{it}. By the assumption, x_{it} should not have

[1] If $K_0 = 1$, we set $\hat{\beta}_i = \hat{\beta}$, the within-group estimator of the homogeneous slope coefficient. Note that we also suppress the dependence of $\hat{\mu}_i$ on K_0.

[2] Strictly speaking, $\tilde{\beta}_i$ is a consistent estimator of β_i^0 under the null. But because the cardinality of the set $\hat{G}_0$ shrinks to zero under the null as $T \to \infty$, the difference between $\tilde{\beta}_i$ and $\hat{\beta}_i$ is asymptotically negligible.

any predictive power for u_{it}. This motivates us to run the following auxiliary regression model

$$\hat{u}_{it} = \upsilon_i + \phi_i' X_{it} + \eta_{it},\ i = 1, ..., N,\ t = 1, ..., T, \tag{2.5}$$

and test the null hypothesis

$$\mathbb{H}_0^* : \phi_i = 0 \text{ for all } i = 1, ..., N.$$

We construct an LM-type test statistic by concentrating the intercept υ_i out in (2.5). Consider the Gaussian quasi-likelihood function for $\hat{u}_{it}$:

$$\ell(\boldsymbol{\phi}) = \sum_{i=1}^{N} (\hat{u}_i - M_0 X_i \phi_i)' (\hat{u}_i - M_0 X_i \phi_i),$$

where $\boldsymbol{\phi} \equiv (\phi_1, ..., \phi_N)'$, $\hat{u}_i \equiv (\hat{u}_{i1}, ..., \hat{u}_{iT})'$, and $X_i \equiv (X_{i1}, ..., X_{iT})'$. Define the LM statistic:

$$LM_{NT}(K_0) = \left(T^{-1/2} \frac{\partial \ell(0)}{\partial \boldsymbol{\phi}} \right)' \left(-T^{-1} \frac{\partial^2 \ell(0)}{\partial \boldsymbol{\phi}\, \partial \boldsymbol{\phi}'} \right) \left(T^{-1/2} \frac{\partial \ell(0)}{\partial \boldsymbol{\phi}} \right), \tag{2.6}$$

where we make the dependence of $LM_{NT}(K_0)$ on K_0 explicit. We can verify that

$$LM_{NT}(K_0) = \sum_{i=1}^{N} \hat{u}_i' M_0 X_i (X_i' M_0 X_i)^{-1} X_i' M_0 \hat{u}_i, \tag{2.7}$$

where the dependence of $LM_{NT}(K_0)$ on K_0 is through that of $\hat{u}_i$ on K_0. We will show that after being appropriately scaled and centered, $LM_{NT}(K_0)$ is asymptotically normally distributed under $\mathbb{H}_0(K_0)$ and diverges to infinity under $\mathbb{H}_1(K_0)$.

Remark. Similar statistics are proposed by Su and Chen (2013) to test for slope homogeneity in panel data models with interactive fixed effects. Note that we have included a constant term in the regression in (2.5). Under the assumption that $E(u_{it}) = 0$ and N and T pass to infinity jointly, one can also omit the constant term and obtain the following LM test statistic:

$$\overline{LM}_{NT}(K_0) = \sum_{i=1}^{N} \hat{u}_i' X_i (X_i' X_i)^{-1} X_i' \hat{u}_i. \tag{2.8}$$

The asymptotic distribution of $\overline{LM}_{NT}(K_0)$ can be similarly studied with little modification. In case T is not very large as in our empirical applications, we recommend including a constant term in the auxiliary regression in (2.5) and thus only focus on the study of $LM_{NT}(K_0)$ below.

3 Asymptotic properties

In this section we first present a set of assumptions that are necessary for asymptotic analyses, and then study the asymptotic distributions of $LM_{NT}(K_0)$ under both $\mathbb{H}_0(K_0)$ and $\mathbb{H}_1(K_0)$.

3.1 Assumptions

Let $\|A\|_q \equiv [E(\|A\|^q)]^{1/q}$ for $q \geq 1$. Let $\hat{\Omega}_i \equiv T^{-1}X_i'M_0X_i$ and $\Omega_i \equiv E(\hat{\Omega}_i)$. Define $\mathcal{F}_{NT,t} \equiv \sigma(\{X_{i,t+1}, X_{it}, u_{it}, X_{i,t-1}, u_{i,t-1}, ...\}_{i=1}^N)$. Let $C < \infty$ be a generic constant that may vary across lines. We make the following assumptions.

Assumption A.1. (i) $\max_{1\leq i\leq N}\max_{1\leq t\leq N}\|\zeta_{it}\|_{8+4\sigma} \leq C$ for some $\sigma > 0$ for $\zeta_{it} = X_{it}$, u_{it}, and $X_{it}u_{it}$.

(ii) There exist positive constants $\underline{c}_\Omega$ and $\bar{c}_\Omega$ such that $\underline{c}_\Omega \leq \min_{1\leq i\leq N}\lambda_{\min}(\Omega_i) \leq \max_{1\leq i\leq N}\lambda_{\max}(\Omega_i) \leq \bar{c}_\Omega$.

(iii) For each $i = 1, ..., N$, $\{(X_{it}, u_{it}) : t = 1, 2, ...\}$ is a strong mixing process with mixing coefficients $\{\alpha_{NT,i}(\cdot)\}$. $\alpha(\cdot) \equiv \alpha_{NT}(\cdot) \equiv \max_{1\leq i\leq N}\alpha_{NT,i}(\cdot)$ satisfies $\alpha(s) = O_{a.s.}(s^{-\rho})$ where $\rho = 3(2+\sigma)/\sigma + \epsilon$ for some arbitrarily small $\epsilon > 0$. In addition, there exist integers τ_0, $\tau_* \in (1, T)$ such that $NT\alpha(\tau_0) = o(1)$, $T(T + N^{1/2})\alpha(\tau_*)^{(1+\sigma)/(2+\sigma)} = o(1)$, and $N^{1/2}T^{-1}\tau_*^2 = o(1)$.

(iv) Let $u_i \equiv (u_{i1}, ..., u_{iT})'$. (X_i, u_i), $i = 1, ..., N$, are mutually independent of each other.

(v) For each $i = 1, ..., N$, $E(u_{it}|\mathcal{F}_{NT,t-1}) = 0$ a.s.

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

(ii) As $(N, T) \to \infty$, $T\lambda \to \infty$ and $T\lambda^4 \to c_0 \in [0, \infty)$.

(iii) For any $c > 0$, $N\max_{1\leq i\leq N} P\left(\left\|T^{-1}\sum_{t=1}^T \tilde{X}_{it}u_{it}\right\| \geq c\sqrt{\lambda}\right) \to 0$ as $(N, T) \to \infty$.

(iv) For each $k = 1, ..., K_0$, $\bar{\Phi}_k \equiv \frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^T \tilde{X}_{it}\tilde{X}_{it}' \xrightarrow{P} \Phi_k > 0$ as $(N, T) \to \infty$.

(v) For each $k = 1, ..., K_0$, $\frac{1}{\sqrt{N_kT}}\sum_{i\in G_k^0}\sum_{t=1}^T \tilde{X}_{it}u_{it} - B_{kNT} \xrightarrow{D} N(0, \Psi_k)$ as $(N, T) \to \infty$ where $B_{kNT} = \frac{1}{\sqrt{N_kT}}\sum_{i\in G_k^0}\sum_{t=1}^T E(\tilde{X}_{it}u_{it})$ is either 0 or $O(\sqrt{N_k/T})$ depending on whether X_{it} is strictly exogenous.

Assumption A.3. There exist finite nonnegative numbers c_1 and c_2 such that $\limsup_{(N,T)\to\infty} N\log(NT)/T^2 = c_1$ and $\limsup_{(N,T)\to\infty}\log(NT)N^{(3+\sigma)/(4+2\sigma)}T^{-(5+3\sigma)/(4+2\sigma)} = c_2$.

A.1(i) imposes moment conditions on X_{it} and u_{it}. A.1(ii) requires that Ω_i be positive definite uniformly in i. A.1(iii) requires that each individual time series $\{(X_{it}, u_{it}) : t = 1, 2, ...\}$ be strong mixing. This condition can be verified if X_{it} does not contain lagged dependent variables no matter whether one treats the fixed effects μ_i's as random or fixed. In the case of dynamic panel data models, Hahn and Kuersteiner (2011) assume that μ_i's are nonrandom and uniformly bounded, in which case the strong mixing condition can also be verified. In the case of random fixed effects, they suggest adopting the concept of *conditional strong mixing* where the mixing coefficient is defined by conditioning on the fixed effects. Su and Chen (2013) also consider conditional strong mixing processes where the conditioning set is given by the common factors and factor loadings in their panel factor model. The dependence of the mixing rate on σ defined in A.1(i) reflects the trade-off between the degree of dependence and the moment bounds of the process $\{(X_{it}, u_{it}), t \geq 1\}$. The last set of conditions in A.1(iii) can easily be met. In particular, if the process is strong mixing with a geometric mixing rate, the conditions on

$\alpha(\cdot)$ can be met simply by specifying $\tau_0 = \tau_* = \lfloor C_\tau \log T \rfloor$ for some sufficiently large C_τ, where $\lfloor a \rfloor$ denotes the integer part of a. A.1(iv) rules out cross sectional dependence among (X_i, u_i) and greatly facilitates our asymptotic analysis. A.1(v) requires that the error term u_{it} be a martingale difference sequence (m.d.s.) with respect to the filter $\mathcal{F}_{NT,t}$ which allows for lagged dependent variables in X_{it}, and conditional heteroskedasticity, skewness, or kurtosis of unknown form in u_{it}.

The conditions in A.2 are borrowed from SSP. A.2(i) is identical to Assumption A1(iv) in SSP and implies that each group has asymptotically non-negligible members as $N \to \infty$. A.2(ii)-(iii) and A.2(iv)-(v) parallel Assumptions A2(i)-(ii) and A3(i)-(ii) in SSP, respectively. The conditions in Assumption A1(i)-(iii) in SSP are implied by our Assumption A.1. According to Theorem 2.3 in SSP, under our A.1-A.2 and $\mathbb{H}_0(K_0)$, the C-Lasso estimates $\{\hat{\alpha}_k\}$ have the following asymptotic property:

$$\hat{\alpha}_k - \alpha_k^0 = O_P\left((NT)^{-1/2}\right) \text{ if } B_{kNT} = 0 \text{ in A.2(v) and } O_P\left((NT)^{-1/2} + T^{-1}\right) \text{ otherwise.} \tag{3.1}$$

A.3 imposes conditions on the rates at which N and T pass to infinity, and the interaction between (N, T) and σ. Note that we allow N and T to pass to infinity at either identical or suitably restricted different rates. The appearance of the logarithm terms is due to the use of a Bernstein inequality for strong mixing processes. If the mixing process $\{(X_{it}, u_{it}), t \geq 1\}$ has a geometric decay rate, one can take an arbitrarily small σ in A.1(i) . In this case, A.3 puts the most stringent restrictions on (N, T) by passing $\sigma \to 0 : N^{3/5}/T \to 0$ as $(N, T) \to \infty$, ignoring the logarithm term. On the other hand, if $\sigma \geq 1$ in A.1(i), then the second condition in A.3 becomes redundant given the first condition. In the case of conventional panel data models with strictly exogenous regressors only, Pesaran and Yamagata (2008) require that either $\sqrt{N}/T \to 0$ or $\sqrt{N}/T^2 \to 0$ for two of their tests; but for stationary dynamic panel data models, they prove the asymptotic validity of their test only under the condition that $N/T \to \kappa \in [0, \infty)$.

3.2 Asymptotic null distribution

Let $h_{i,ts}$ denote the (t, s)'th element of $H_i \equiv M_0 X_i (X_i' M_0 X_i)^{-1} X_i' M_0$. Let $X_{it}^\dagger \equiv X_{it} - T^{-1} \sum_{s=1}^T E(X_{is})$ and $\bar{b}_{it} \equiv \Omega_i^{-1/2} X_{it}^\dagger$. Define

$$B_{NT} \equiv N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T u_{it}^2 h_{i,tt} \text{ and } V_{NT} \equiv 4T^{-2} N^{-1} \sum_{i=1}^N \sum_{t=2}^T E\left[u_{it} \bar{b}_{it}' \sum_{s=1}^{t-1} \bar{b}_{is} u_{is}\right]^2. \tag{3.2}$$

The following theorem states the asymptotic null distribution of the infeasible statistic LM_{NT}.

Theorem 3.1 *Suppose Assumptions A.1-A.3 hold. Then under* $\mathbb{H}_0(K_0)$,

$$J_{NT}(K_0) \equiv \left(N^{-1/2} LM_{NT}(K_0) - B_{NT}\right) / \sqrt{V_{NT}} \xrightarrow{D} N(0, 1).$$

The proof of the above theorem is tedious and relegated to the appendix. The key step in the proof is to show that under $\mathbb{H}_0(K_0)$, $\sqrt{V_{NT}} J_{NT}(K_0) = A_{NT} + o_P(1)$, where $A_{NT} \equiv \sum_{t=2}^T Z_{NT,t}$ and

$Z_{NT,t} \equiv 2T^{-1}N^{-1/2}\sum_{i=1}^{N}\sum_{s=1}^{t-1} u_{it}u_{is}\bar{b}_{it}'\bar{b}_{is}$. By construction, $\{Z_{NT,t}, \mathcal{F}_{NT,t}\}$ is an m.d.s. so that we can apply the martingale central limit theorem (e.g., Pollard (1984, p. 171)) to show that $A_{NT} \xrightarrow{D} N(0, V_0)$ under Assumptions A.1-A.3, where $V_0 = \lim_{(N,T)\to\infty} V_{NT}$.

To implement the test, we need consistent estimates of both B_{NT} and V_{NT}. We propose to estimate them respectively by

$$\hat{B}_{NT}(K_0) = N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2 h_{i,tt} \text{ and } \hat{V}_{NT}(K_0) = 4T^{-2}N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[\hat{u}_{it}\hat{b}_{it}'\sum_{s=1}^{t-1}\hat{b}_{is}\hat{u}_{is}\right]^2 \tag{3.3}$$

where $\hat{b}_{it} = \hat{\Omega}_i^{-1/2}(X_{it} - T^{-1}\sum_{s=1}^{T} X_{is})$.[3] Then we can define a feasible test statistic:

$$\hat{J}_{NT}(K_0) \equiv \left(N^{-1/2}LM_{NT}(K_0) - \hat{B}_{NT}(K_0)\right)/\sqrt{\hat{V}_{NT}(K_0)}. \tag{3.4}$$

The following theorem establishes the consistency of $\hat{B}_{NT}(K_0)$ and $\hat{V}_{NT}(K_0)$ and the asymptotic distribution of $\hat{J}_{NT}(K_0)$ under $\mathbb{H}_0(K_0)$.

Theorem 3.2 *Suppose Assumptions A.1-A.3 hold. Then under* $\mathbb{H}_0(K_0)$, $\hat{B}_{NT}(K_0) = B_{NT} + o_P(1)$, $\hat{V}_{NT}(K_0) = V_{NT} + o_P(1)$, *and* $\hat{J}_{NT}(K_0) \xrightarrow{D} N(0,1)$.

Theorem 3.2 implies that the test statistic $\hat{J}_{NT}(K_0)$ is asymptotically pivotal under $\mathbb{H}_0(K_0)$. We can compare $\hat{J}_{NT}(K_0)$ with the one-sided critical value z_α, i.e., the upper αth percentile from the standard normal distribution, and reject the null when $\hat{J}_{NT}(K_0) > z_\alpha$ at the asymptotic α significance level.

We obtain the above distributional results despite the fact that the individual effects μ_i's can only be estimated at the slower rate $T^{-1/2}$ than the rate $(NT)^{-1/2}$ or $(NT)^{-1/2} + T^{-1}$ at which the group-specific parameter estimates $\{\hat{\alpha}_k, k = 1, ..., K_0\}$ converge to their true values under $\mathbb{H}_0(K_0)$. The slow convergence rate of these individual effect estimates does not have adverse asymptotic effects on the estimation of the bias term B_{NT}, the variance term V_{NT}, and the asymptotic distribution of $\hat{J}_{NT}(K_0)$. Nevertheless, they can play an important role in finite samples, which we verify through Monte Carlo simulations.

3.3 Consistency

Let $\mathcal{G}_K = \{(G_1, ..., G_K) : \cup_{k=1}^{K} G_k = \{1, ..., N\} \text{ and } G_k \cap G_j = \varnothing \text{ for any } j \neq k\}$. That is, $\mathcal{G}_K$ denotes the class of all possible K-group partitions of $\{1, ..., N\}$. To study the consistency of our test, we add the following assumption.

Assumption A.4. (i) $N^{-1}\sum_{i=1}^{N}\left\|\beta_i^0\right\|^2 = O_P(1)$.

(ii) $\inf_{(G_1,...,G_{K_0})\in\mathcal{G}_{K_0}} \min_{(\alpha_1,...,\alpha_{K_0})} N^{-1}\sum_{k=1}^{K_0}\sum_{i\in G_k}\left\|\beta_i^0 - \alpha_k\right\|^2 \xrightarrow{P} \underline{c}_{K_0} > 0$ as $N \to \infty$.

[3] Let $\hat{b}_i \equiv (\hat{b}_{i1}, ..., \hat{b}_{iT})'$. Note that $\hat{b}_i = M_0 X_i \hat{\Omega}_i^{-1/2}$.

A.4(i) is trivially satisfied if β_i^0's are uniformly bounded or random with finite second moments. A.4(ii) essentially says that one cannot group the N parameter vectors $\{\beta_i^0, 1 \le i \le N\}$ into K_0 groups by leaving out an insignificant number of unclassified individuals. It is satisfied for a variety of global alternatives:

1. The number of groups is $K = K_0 + r$ for some positive integer r such that $N_k/N \to \tau_k \in (0,1)$ for each $k = 1, ..., K_0 + r$.

2. There is no grouped pattern among $\{\beta_i^0, 1 \le i \le N\}$ such that we have a completely heterogeneous population of individuals.

3. The regression model is actually a random coefficient model: $\beta_i^0 = \beta^0 + v_i$, where β^0 is a fixed parameter vector, and v_i's are independent and identical draws from a continuous distribution with zero mean and finite variance.

4. The regression model is a hierarchical random coefficient model:

$$\beta_i^0 = \begin{cases} \alpha_1^0 + v_{1i} & \text{if } i \in G_1^0 \\ \vdots & \vdots \\ \alpha_K^0 + v_{Ki} & \text{if } i \in G_K^0 \end{cases},$$

   where $\alpha_1^0, ..., \alpha_K^0$ are defined as before, v_{ji}'s (for $j = 1, ..., K$) are independent and identical draws from a continuous distribution with zero mean and finite variance, and K may be different from K_0.

The following theorem establishes the consistency of $\hat{J}_{NT}$.

Theorem 3.3 *Suppose Assumptions A.1 and A.3-A.4 hold. Then under* $\mathbb{H}_1(K_0)$ *with possible diverging* $K_{\max}$ *and random coefficients,* $P(\hat{J}_{NT}(K_0) \ge c_{NT}) \to 1$ *as* $(N, T) \to \infty$ *for any non-stochastic sequence* $c_{NT} = o\left(N^{1/2}T\right)$.

The above theorem indicates that our test statistic $\hat{J}_{NT}(K_0)$ is divergent at $N^{1/2}T$-rate under $\mathbb{H}_1(K_0)$ and thus has power to detect any alternatives such that A.4 is satisfied.

3.4 Issues related to sequential multiple testing

It is well known that we need to take into account the multiplicity of individual tests when controlling sizes in multiple testing procedures. In this subsection, we briefly discuss the related issues in our context. Suppose that the null hypothesis $\mathbb{H}_0(K_0)$ is true. We consider the sequential testing procedure described above to determine the number of groups, in which case we reject $\mathbb{H}_0(K_0)$ in either of the following two cases:

Case (i) : We fail to reject one of the individual hypotheses: $\mathbb{H}_0(1), \mathbb{H}_0(2), ..., \mathbb{H}_0(K_0-1)$, say $\mathbb{H}_0(K^*)$, then we conclude that $K = K^*$, thus $\mathbb{H}_0(K_0)$ is rejected;

Case (ii) : We reject all the individual hypotheses: $\mathbb{H}_0(1), \mathbb{H}_0(2), ..., \mathbb{H}_0(K_0-1)$, and $\mathbb{H}_0(K_0)$.

Note that when $\mathbb{H}_0(K_0)$ is true, $\mathbb{H}_0(1), \mathbb{H}_0(2), ..., \mathbb{H}_0(K_0-1)$ are all false. Let $\alpha(K_0)$ be the *asymptotic* type I error of testing the single hypothesis $\mathbb{H}_0(K_0)$. Further let $\alpha^*(K_0)$ be the *asymptotic* type I error of the sequential testing procedure described above. The next lemma shows that $\alpha(K_0)$ and $\alpha^*(K_0)$ are equal.

Lemma 3.4 *Suppose that Assumptions A.1-A.3 hold. Then* $\alpha^*(K_0) = \alpha(K_0)$.

The key condition to ensure $\alpha^*(K_0) = \alpha(K_0)$ is that our test is consistent for the individual hypotheses: $\mathbb{H}_0(1)$, $\mathbb{H}_0(2)$, ..., and $\mathbb{H}_0(K_0-1)$.

Suppose that our testing procedure starts with testing $\mathbb{H}_0(1)$ and ends with testing $\mathbb{H}_0(K^*)$ where $K^* \geq 2$. This means that we have conducted K^* tests, rejected $\mathbb{H}_0(1), ..., \mathbb{H}_0(K^*-1)$ but failed to reject $\mathbb{H}_0(K^*)$ at some prescribed nominal significance level α. We consider the family-wise error rate (FWER) for our sequential tests of $\mathbb{H}_0(k)$, $1 \leq k \leq K^*$, which is defined as

$$\text{FWER} = P\left(\text{Reject at least one hypothesis } \mathbb{H}_0(k) \text{ that is true}, 1 \leq k \leq K^* \mid \mathbb{H}_0(K_0)\right).$$

To control the FWER, one may consider the use of the Holm-Bonferroni method (see, e.g., Holm (1979) and Hochberg (1988)).[4] We order the individual p-values from the smallest to the largest as $p_{(1)} \leq p_{(2)} \leq ... \leq p_{(K^*)}$ with their corresponding null hypotheses labelled accordingly as $\mathbb{H}_{0(1)}$, $\mathbb{H}_{0(2)}$, ..., $\mathbb{H}_{0(K^*)}$. Then, we reject $\mathbb{H}_{0(k)}$ when for all $j = 1, ..., k$, $p_{(j)} \leq \alpha/(K^* - j + 1)$. Thus, to control the FWER at asymptotic level α for the K^* tests we end up with, we can use the step-down Holm adjusted p-values for testing $\mathbb{H}_{0(k)}$:

$$\text{adjusted-}p_{(k)} = \min\left((K^* - k + 1)\, p_{(k)}, 1\right)$$

and compare it with α.

The Holm adjusted p-values do not account for the dependence of the multiple tests. More sophisticated methods, such as those based on resampling, can also be used to improve test powers (see, e.g., Romano and Wolf (2005a, 2005b)).

[4] When $K^* < K_0$, none of the hypotheses $\mathbb{H}_0(k)$, $k = 1, ..., K^*$, is true. Thus, the FWER $= 0$. When $K^* = K_0$, only one of the hypotheses $(\mathbb{H}_0(K_0))$ is true. Then, the FWER is the same as the probability of falsely rejecting $\mathbb{H}_0(K_0)$, which is well controlled as shown in Lemma 3.4. Nevertheless, in the case $K^* > K_0$, $\mathbb{H}_0(K_0), ..., \mathbb{H}_0(K^*)$ can all be regarded as true in the sense that our test does not have power against $\mathbb{H}_0(k)$ for $k = K_0 + 1, ..., K^*$ when $\mathbb{H}_0(K_0)$ holds. In this case, the FWER is different from the size of individual tests. In practice, we do not know whether $K^* < K_0$, $K^* = K_0$, or $K^* > K_0$, thus we recommend the use of conservative bounds for p-values.

4 Monte Carlo simulations

In this section, we conduct Monte Carlo simulations to examine the finite sample performance of our proposed testing method.

4.1 Data generating processes and implementation

We consider the following four data generating processes (DGPs):
DGP 1: $y_{it} = \beta_{1i}^0 X_{1it} + \beta_{2i}^0 X_{2it} + \mu_i + u_{it}$,
DGPs 2-4: $y_{it} = \beta_{1i}^0 X_{1it} + \beta_{2i}^0 y_{i,t-1} + \mu_i + u_{it}$,
where $X_{it} = \xi_{it} + \mu_i$, and μ_i, ξ_{it}, u_{it} are IID $N(0,1)$ variables, and mutually independent of each other. DGP 1 is a static panel structure, while DGPs 2-4 are dynamic panel structures. In DGPs 1 and 2, $\left(\beta_{1i}^0, \beta_{2i}^0\right)$ has a group structure:

$$\left(\beta_{1i}^0, \beta_{2i}^0\right) = \begin{cases} (0.5, -0.5) & \text{with probability } 0.3 \\ (-0.5, 0.5) & \text{with probability } 0.3 \\ (0,0) & \text{with probability } 0.4 \end{cases} .$$

Therefore in DGPs 1 and 2, the true number of groups is 3. In DGP 3, we consider a completely heterogeneous (random coefficient) panel structure where β_{1i}^0 and β_{2i}^0 follow $N(0.5, 1)$ and $\mathrm{U}(-0.5, 0.5)$, respectively. In principle, the true number of groups is the cross-section dimension N in this case. In DGP 4, $(\beta_{1i}^0, \beta_{2i}^0)$ is similar to that in DGPs 1 and 2 except that it has some additional small disturbance. Specifically,

$$\left(\beta_{1i}^0, \beta_{2i}^0\right) = \begin{cases} (0.5 + 0.1\nu_{1i}, -0.5 + 0.1\nu_{2i}) & \text{with probability } 0.3 \\ (-0.5 + 0.1\nu_{1i}, 0.5 + 0.1\nu_{2i}) & \text{with probability } 0.3 \\ (0.1\nu_{1i}, 0.1\nu_{2i}) & \text{with probability } 0.4 \end{cases} ,$$

where ν_{1i} and ν_{2i} are each IID $N(0,1)$, mutually independent, and independent of μ_i, ξ_{it}, and u_{it}. DGP 4 can be thought of as a small deviation from a group structure.

For each DGP, we first test the null hypotheses: $\mathbb{H}_0(1)$, $\mathbb{H}_0(2)$, and $\mathbb{H}_0(3)$ to examine the levels and powers of our test.

We then use our tests to determine the number of groups as described in Section 2.1. We set $K_{\max} = 8$ and let the nominal size decrease with the time series dimension T to ensure that the type I error decreases with T. Specifically, we let the nominal size be $1/T$, which equals to 0.10, 0.05 and 0.025 for $T = 10$, 20 and 40, respectively.[5] If all eight hypotheses, $\mathbb{H}_0(1)$, ..., and $\mathbb{H}_0(8)$ are rejected, then we stop and conclude that the number of groups is greater than 8.

For the combination of N and T, we consider the typical case in empirical applications that T is smaller than or comparable to N and let $(N,T) = (40,10)$, $(40,20)$, $(40,40)$, $(80,10)$, $(80,20)$ and $(80,40)$. The number of replications in the simulations is 1000.

[5] We also try fixing the nominal level at 0.05. The results are similar and available upon request.

One important step in implementing our testing procedure is to choose the tuning parameter λ. Following the theory in SSP, we let $\lambda = c \cdot s_Y \cdot T^{-3/4}$, where s_Y is the sample standard deviation of Y_{it} and c is some constant. We use three different values of c (0.25, 0.5 and 1) to examine the sensitivity of our results to c (thus λ).

4.2 Simulation results

Table 1 shows the level and power behavior of our test statistics for testing the three null hypotheses: $\mathbb{H}_0(1)$, $\mathbb{H}_0(2)$, and $\mathbb{H}_0(3)$. We choose three conventional nominal levels: 0.01, 0.05 and 0.10. For DGPs 1 and 2, the true number of groups is 3. For $\mathbb{H}_0(1)$, the rejection frequencies are all almost 1 for all combinations of N and T at all three nominal levels. For $\mathbb{H}_0(2)$, the powers of the test increase rapidly with both N and T. For example, when $T = 40$, the rejection frequencies are all 1 at all three nominal levels. For $\mathbb{H}_0(3)$, we examine the level of our test and find that the rejection frequencies are fairly close to the nominal levels, especially when T is large.

For the heterogeneous DGP 3, our test rejects all the three hypotheses with the frequencies being 1 or nearly 1 at all three nominal levels. This reflects the power of our test against global alternatives. For DGP 4 which represents a small deviation from the group structure, our test shows reasonable power for large T, though it rejects $K = 3$ with a small frequency when T is small, as expected. Also note that all the testing results are quite robust to the values of c (thus λ).

Table 2 shows the proportions of the replications in which the number of groups determined by our method is equal to a certain number. For DGPs 1 and 2, our method determines the correct number of groups (3) with a large probability. For example, when $T = 40$, the number of groups determined by our testing procedure equals the true number 3 with probabilities ranging from 0.969 to 0.988. This is consistent with the recommended nominal level 0.025 ($1/T$) for $T = 40$. For DGP 3 where the true number of groups is N, our method determines a large number of groups (greater than 8) with probabilities higher than 0.96 even when $T = 20$. DGP 4 represents a small deviation from a 3-group structure. With a high probability, the number of groups determined by our method is 3 when N and T are small, and is equal to or larger than 5 when N and T are large. These results are reasonable. Intuitively, if N and T are small, the data can only provide limited information on the underlying DGP, and it is reasonable to apply a 3-group structure to serve as a good approximation to the true model. As N and T become large, more information on the underlying DGP is revealed, and it is sensible to adopt a larger number of groups to approximate the true model more accurately.

5 Empirical application: income and democracy

The relationship between income and democracy has attracted much attention in empirical research; see, e.g., Lipset (1959), Barro (1999), Acemoglu, Johnson, Robinson, and Yared (2008, AJRY hereafter), and Bonhomme, and Manresa (2012, BM hereafter). To the best of our knowledge, none of the existing studies allows for heterogeneity in the slope coefficients in their model specifications. As discussed in AJRY, "*societies may embark on divergent political-economic development paths*". Thus, ignoring the heterogeneity in the slope coefficients may result in model misspecification and invalidates subsequent inferences. Hence, it is important to know whether the data support the assumption of homogeneous slope coefficients. If not, then we need to determine the number of heterogeneous groups and classify the countries using statistic methods. We apply our new method to study this important question.

5.1 Data and implementation

We let y_{it} be a measure of democracy for country i in period t and X_{it} be the logarithm of its real GDP per capita. The measure of democracy and real GDP per capita are from the Freedom House and Penn World Tables, respectively.[6] Note that the Freedom House measures of democracy (y_{it}) are normalized to be between 0 and 1.

We consider the fixed effect specification,

$$y_{it} = \beta_{1i} X_{i,t-1} + \beta_{2i} y_{i,t-1} + \mu_i + u_{it},$$

and assume that (β_{1i}, β_{2i}) has a group structure to account for possible heterogeneity.[7] In a closely related paper, BM consider a group structure in the interactive fixed effects and assume (β_{1i}, β_{2i}) is constant for all i.

We use a balanced panel dataset similar to that in BM. The number of countries (N) is 82. The time index is $t = 1, ..., 8$, where each period corresponds to a five-year interval over the period 1961-2000. For example, $t = 1$ refers to years 1961-1965. Because the lagged y_{it} is used as a regressor, the number of time periods (T) is 7. The choice of the countries is determined by data availability. In addition, we exclude the countries whose measures of democracy remain constant over all eight periods. The list of the 82 countries can be found in Table 7. Table 3 presents the summary statistics. The details of implementation of our method are the same as in the simulations.

5.2 Testing and estimation results

We first test the hypothesis $\mathbb{H}_0(1)$, i.e., we test whether (β_{1i}, β_{2i}) is constant for almost all i. For all three values of the tuning parameter $\lambda = c \cdot s_Y \cdot T^{-3/4}$ ($c = 0.25$, 0.5, and 1), we soundly reject this hypothesis. All the p-values are less than 0.001. This provides strong evidence that the slope coefficients are not homogeneous. We then test the hypothesis $\mathbb{H}_0(2)$ and reject the hypothesis again at 5% level with p-values being 0.006, 0.024 and 0.010 for $c = 0.25$, 0.5 and 1, respectively. We continue to test $\mathbb{H}_0(3)$

[6] All the data are directly from AJRY: http://economics.mit.edu/faculty/acemoglu/data/ajry2008.

[7] We do not include time fixed effects, as our econometric theory so far has not yet been developed to allow time fixed effect. We leave this important question for future research.

and find that the p-values are 0.269, 0.173 and 0.045 for $c = 0.25$, 0.5 and 1, respectively. Considering that the p-values are above or close to 5%, we stop the testing procedure and conclude that the number of groups is 3. Table 4 presents all the testing results and Table 7 shows the country membership of the three groups. Note that we also report Holm adjusted p-values in the last row in Table 4, which also lend strong support to the conclusion of three groups in the data.

Table 5 presents the estimation results. We report both C-Lasso estimates and post C-Lasso estimates. C-Lasso estimates are defined in Section 2.2. The post C-Lasso is implemented on the data of each group based on the results of our classification. Both estimates are bias-corrected and the standard errors are obtained using the asymptotic theory developed in SSP. The C-Lasso and post C-Lasso estimation results are similar for different values of $c = 0.25$, 0.5 and 1. In the following discussion, we focus on the post C-Lasso estimates with $c = 0.5$. It is clear that the estimated slope coefficients exhibit substantial heterogeneity. For β_{1i}, the estimates for the three groups are -0.427, 0.078, and 0.341. All of them are significant at the 5% level. It is interesting to note that not only the magnitudes but also the signs of estimates are different among the three groups. For group 1, income has a negative effect on democracy, while for groups 2 and 3, the effects are positive with different magnitudes. For β_{2i}, the three group estimates are 0.115, -0.107 and 0.380. The first two estimates are not significant at the 5% level, while the third estimate is significant at the 1% level. We also present the point estimates of cumulative income effect (CIE), which is defined as $\beta_{1i}/(1 - \beta_{2i})$. The estimates of CIE for the three groups are -0.482, 0.070 and 0.550, which imply that a 10% increase in income per capita is associated with increases of -0.0482, 0.007 and 0.055 in the measures of democracy, respectively.

Note that if we assume that β_{1i} and β_{2i} are homogeneous across i, then the common estimates are 0.130 and 0.290, respectively. The common CIE estimate is 0.183. All the common estimates fall in the middle of their corresponding group estimates.

To understand the heterogeneity in the data intuitively, we select a country from each of the three groups (Malaysia, Indonesia, and Nepal) and show their time-series data in Panel A of Table 6. We simply calculate the correlations between the dependent variable Y_{it} and the key explanatory variable $X_{i,t-1}$. Even the simple correlations exhibit substantial heterogeneity with the values being -0.863, 0.069 and 0.658. This suggests that it is implausible that the slope coefficients are the same for all countries even without doing any sophisticated analysis.

This application shows that ignoring the heterogeneity in the slope coefficients can mask the true underlying relationship among economic variables.

5.3 Explaining the group pattern

According to the estimates of β_{1i}, we refer to groups 1, 2 and 3 as the "negative effect", "small effect" and "large effect" groups, respectively. Apparently, the group membership listed in Table 7 does not match the countries' geographic locations, though most of the countries in the "negative effect" group are in Africa or Central America. One natural question is how we explain the group membership. For example, we may wonder why China and the United States, two very different countries, are both classified into the same "small effect" group. This is actually not difficult to understand. We list the original time-series data for the two countries in Panel B of Table 6. China's measures of democracy over this period do not show much progress, though it has a fast economic growth. Hence, intuitively, China's income effect should be small. On the other hand, the United States' measures of democracy remain constant at the highest level (1) over the period $t = 2, ...8$, which explains why United States is also in the "small effect" group.

So far, our classification of the groups is completely statistical and does not use any a priori information. We further investigate the group pattern by using a cross-section multinomial logit model. We let the dependent variable be group membership, which takes one of three values: 1, 2 or 3.[8] The explanatory variables include (*i*) initial education level in 1965, (*ii*) initial income level in 1965, (*iii*) initial democracy level in 1965, (*iv*) a measure of constraints on the executive at independence, (*v*) independence year/100, (*vi*) 500-year change in income per capita over 1500-2000 and (*vii*) 500-year change in democracy over 1500-2000. Among them, (*i*) , (*ii*) and (*iii*) are the initial key economic variables. Acemoglu, Johnson and Robinson (2005) suggest that (*iv*) is an important determinant of democracy. (*v*) measures how recent a country became independent. (*vi*) and (*vii*) present long-run changes in income and democracy levels, respectively. All the data are taken directly from AJRY.

Table 8 provides summary statistics for each of the three groups. The initial education levels are clearly substantially different among the three groups. The average initial education level of the "negative effect" group is only half of that of the "small effect" and "large effect" groups. The 500-year changes of income per capita and of democracy also differ noticeably among the three groups. The "negative effect" group has smaller values of these two variables than the other two groups, which suggests that "negative effect" group has achieved relatively slow long-run progress in economic growth and democracy.

Table 9 presents the multinomial logit regression results for various model specifications. We choose group 3 (the "large effect" group) as the base group. Compared with the "large effect" group, at the 5% level, a higher education, a later independence year, or a larger 500-year change in democracy leads to a reduced likelihood of being in the "negative effect" group, while a larger constraint on the executive at independence leads to an increased likelihood. On the other hand, a higher education leads to a higher likelihood of being in the "small effect" group and a larger 500-year change in democracy leads to a lower likelihood. In summary, we find that the initial education level and the long-run progress in democracy are important determinants of our group pattern.

[8] We only report the results for $c = 0.5$. The results for $c = 0.25$ and 1 are similar and available upon request.

6 Conclusion

We develop a data-driven procedure to determine the number of groups in a latent group panel structure proposed in Su, Shi, and Phillips (2013). The procedure is based on conducting hypothesis testing on the model specifications. The test is a residual-based LM type test and is asymptotically normally distributed under the null. We apply our new method to study the relationship between income and democracy and find strong evidence that the slope coefficients are heterogeneous and form three distinct groups. Further, we find that the initial education level and the long-run progress in democracy are important determinants of the group pattern.

There are several interesting topics for further research. Here we apply our testing procedure to determine the number of groups for slope coefficients. The same idea can be applied to other group structures, such as those considered in Bonhomme and Manresa (2012) where fixed effects have a grouped pattern. We may also extend our methods to non-linear panel data models such as discrete choice models.

Table 1: Empirical rejection frequency

		N	T	$c = 0.25$ 0.01	0.05	0.10	$c = 0.5$ 0.01	0.05	0.10	$c = 1$ 0.01	0.05	0.10
DGP 1	$K=1$	40	10	0.998	0.999	1	0.998	0.999	1	0.998	0.999	1
		40	20	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
		80	10	1	1	1	1	1	1	1	1	1
		80	20	1	1	1	1	1	1	1	1	1
		80	40	1	1	1	1	1	1	1	1	1
	$K=2$	40	10	0.214	0.459	0.568	0.176	0.400	0.525	0.144	0.358	0.484
		40	20	0.708	0.88	0.939	0.683	0.865	0.935	0.665	0.855	0.932
		40	40	1	1	1	1	1	1	1	1	1
		80	10	0.471	0.702	0.792	0.385	0.63	0.749	0.323	0.581	0.715
		80	20	0.970	0.997	0.999	0.964	0.996	0.999	0.966	0.993	0.999
		80	40	1	1	1	1	1	1	1	1	1
	$K=3$	40	10	0.022	0.062	0.110	0.016	0.061	0.104	0.012	0.049	0.094
		40	20	0.019	0.043	0.074	0.014	0.039	0.077	0.009	0.037	0.075
		40	40	0.008	0.036	0.063	0.007	0.034	0.061	0.012	0.036	0.065
		80	10	0.022	0.107	0.168	0.025	0.080	0.152	0.021	0.078	0.138
		80	20	0.015	0.035	0.057	0.020	0.039	0.063	0.019	0.043	0.078
		80	40	0.008	0.045	0.078	0.009	0.042	0.073	0.007	0.041	0.078
DGP 2	$K=1$	40	10	1	1	1	1	1	1	1	1	1
		40	20	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
		80	10	1	1	1	1	1	1	1	1	1
		80	20	1	1	1	1	1	1	1	1	1
		80	40	1	1	1	1	1	1	1	1	1
	$K=2$	40	10	0.206	0.433	0.547	0.152	0.361	0.482	0.118	0.289	0.421
		40	20	0.792	0.937	0.966	0.766	0.924	0.959	0.739	0.907	0.957
		40	40	1	1	1	1	1	1	1	1	1
		80	10	0.468	0.715	0.837	0.376	0.623	0.769	0.266	0.503	0.659
		80	20	0.990	0.998	0.999	0.986	0.998	0.999	0.980	0.997	0.999
		80	40	1	1	1	1	1	1	1	1	1
	$K=3$	40	10	0.008	0.053	0.102	0.008	0.036	0.062	0.004	0.026	0.054
		40	20	0.011	0.038	0.069	0.014	0.039	0.075	0.009	0.034	0.073
		40	40	0.007	0.021	0.054	0.006	0.021	0.052	0.008	0.023	0.050
		80	10	0.028	0.088	0.127	0.014	0.043	0.087	0.008	0.034	0.068
		80	20	0.017	0.042	0.087	0.022	0.044	0.086	0.019	0.045	0.080
		80	40	0.016	0.049	0.095	0.013	0.048	0.088	0.014	0.049	0.086

Table 1: Empirical rejection frequency (cont'd)

		N	T	c = 0.25 0.01	0.05	0.10	c = 0.5 0.01	0.05	0.10	c = 1 0.01	0.05	0.10
DGP 3	$K=1$	40	10	1	1	1	1	1	1	1	1	1
		40	20	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
		80	10	1	1	1	1	1	1	1	1	1
		80	20	1	1	1	1	1	1	1	1	1
		80	40	1	1	1	1	1	1	1	1	1
	$K=2$	40	10	0.998	1	1	0.999	1	1	0.997	1	1
		40	20	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
		80	10	1	1	1	1	1	1	1	1	1
		80	20	1	1	1	1	1	1	1	1	1
		80	40	1	1	1	1	1	1	1	1	1
	$K=3$	40	10	0.970	0.994	0.997	0.974	0.995	0.997	0.981	0.996	0.999
		40	20	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
		80	10	1	1	1	1	1	1	1	1	1
		80	20	1	1	1	1	1	1	1	1	1
		80	40	1	1	1	1	1	1	1	1	1
DGP 4	$K=1$	40	10	1	1	1	1	1	1	1	1	1
		40	20	1	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1	1
		80	10	1	1	1	1	1	1	1	1	1
		80	20	1	1	1	1	1	1	1	1	1
		80	40	1	1	1	1	1	1	1	1	1
	$K=2$	40	10	0.308	0.568	0.696	0.243	0.501	0.633	0.181	0.404	0.552
		40	20	0.921	0.974	0.994	0.907	0.97	0.991	0.891	0.958	0.985
		40	40	1	1	1	1	1	1	1	1	1
		80	10	0.670	0.864	0.931	0.590	0.805	0.893	0.476	0.705	0.839
		80	20	0.999	1	1	0.999	1	1	0.999	1	1
		80	40	1	1	1	1	1	1	1	1	1
	$K=3$	40	10	0.031	0.113	0.172	0.034	0.089	0.145	0.019	0.079	0.124
		40	20	0.115	0.229	0.312	0.104	0.208	0.290	0.098	0.202	0.282
		40	40	0.388	0.623	0.737	0.380	0.617	0.733	0.370	0.611	0.725
		80	10	0.091	0.201	0.300	0.066	0.157	0.260	0.052	0.142	0.222
		80	20	0.224	0.408	0.525	0.209	0.385	0.512	0.210	0.379	0.501
		80	40	0.773	0.915	0.957	0.769	0.913	0.954	0.759	0.909	0.947

Table 2: Frequency of number of groups determined

		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 1	$c=0.25$	40	10	0	0.432	0.460	0.081	0.025	0.002
		40	20	0	0.118	0.838	0.032	0.011	0.001
		40	40	0	0	0.981	0.015	0.004	0
		80	10	0	0.209	0.623	0.132	0.035	0.001
		80	20	0	0.003	0.961	0.027	0.007	0.002
		80	40	0	0	0.973	0.018	0.009	0
	$c=0.5$	40	10	0	0.475	0.421	0.075	0.026	0.003
		40	20	0	0.133	0.828	0.025	0.014	0
		40	40	0	0	0.981	0.013	0.006	0
		80	10	0	0.251	0.599	0.127	0.023	0
		80	20	0	0.004	0.957	0.031	0.007	0.001
		80	40	0	0	0.976	0.014	0.010	0
	$c=1$	40	10	0	0.516	0.390	0.065	0.026	0.003
		40	20	0	0.144	0.818	0.027	0.010	0.001
		40	40	0	0	0.977	0.016	0.006	0.001
		80	10	0	0.286	0.577	0.115	0.020	0.002
		80	20	0	0.007	0.949	0.029	0.015	0
		80	40	0	0	0.977	0.011	0.012	0
		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 2	$c=0.25$	40	10	0	0.453	0.448	0.077	0.020	0.002
		40	20	0	0.063	0.898	0.033	0.006	0.000
		40	40	0	0	0.988	0.011	0.000	0.001
		80	10	0	0.163	0.710	0.097	0.028	0.002
		80	20	0	0.002	0.956	0.036	0.006	0
		80	40	0	0	0.969	0.027	0.004	0
	$c=0.5$	40	10	0	0.518	0.421	0.048	0.013	0
		40	20	0	0.075	0.886	0.033	0.006	0
		40	40	0	0	0.987	0.012	0.000	0.001
		80	10	0	0.232	0.682	0.063	0.020	0.003
		80	20	0	0.002	0.954	0.034	0.010	0
		80	40	0	0	0.971	0.024	0.005	0
	$c=1$	40	10	0	0.580	0.367	0.037	0.015	0.001
		40	20	0	0.091	0.875	0.028	0.006	0
		40	40	0	0	0.986	0.012	0.001	0.001
		80	10	0	0.342	0.590	0.055	0.012	0.001
		80	20	0	0.003	0.952	0.032	0.013	0
		80	40	0	0	0.974	0.022	0.004	0

Note: the numbers in the main entries are the proportions of the replications in which the number of groups determined by our method is equal to, less than (<5 in DGP 3), or greater than (>5 in DGPs 1, 2, and 4, >8 in DGP 3) a number.

Table 2: Frequency of number of groups determined (cont'd)

		N	T	$K<5$	$K=5$	$K=6$	$K=7$	$K=8$	$K>8$
DGP 3	$c=0.25$	40	10	0.030	0.058	0.047	0.100	0.028	0.737
		40	20	0	0.002	0.008	0.013	0.011	0.966
		40	40	0	0	0	0	0	1
		80	10	0	0.003	0	0.014	0.002	0.981
		80	20	0	0	0	0	0	1
		80	40	0	0	0	0	0	1
	$c=0.5$	40	10	0.016	0.042	0.023	0.067	0.016	0.836
		40	20	0	0	0.004	0.013	0.012	0.971
		40	40	0	0	0	0	0	1
		80	10	0	0.001	0	0.005	0	0.994
		80	20	0	0	0	0	0	1
		80	40	0	0	0	0	0	1
	$c=1$	40	10	0.008	0.022	0.012	0.027	0.014	0.917
		40	20	0	0	0	0.004	0.004	0.992
		40	40	0	0	0	0	0	1
		80	10	0	0	0	0.002	0	0.998
		80	20	0	0	0	0	0	1
		80	40	0	0	0	0	0	1
		N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
DGP 4	$c=0.25$	40	10	0	0.305	0.524	0.110	0.050	0.011
		40	20	0	0.023	0.747	0.167	0.050	0.013
		40	40	0	0	0.484	0.269	0.181	0.066
		80	10	0	0.069	0.632	0.169	0.116	0.014
		80	20	0	0	0.592	0.257	0.122	0.029
		80	40	0	0	0.131	0.334	0.297	0.238
	$c=0.5$	40	10	0	0.367	0.490	0.094	0.042	0.007
		40	20	0	0.030	0.761	0.140	0.056	0.013
		40	40	0	0	0.490	0.262	0.179	0.069
		80	10	0	0.107	0.634	0.152	0.097	0.010
		80	20	0	0	0.615	0.232	0.122	0.031
		80	40	0	0	0.139	0.301	0.314	0.246
	$c=1$	40	10	0	0.450	0.426	0.080	0.038	0.006
		40	20	0	0.041	0.756	0.124	0.071	0.008
		40	40	0	0	0.506	0.231	0.195	0.068
		80	10	0	0.162	0.616	0.136	0.081	0.005
		80	20	0	0	0.618	0.205	0.139	0.038
		80	40	0	0	0.145	0.251	0.356	0.248

Table 3: Summary statistics ($N = 82$)

Time period t	Years	Y_{it} : democracy			X_{it}: logarithm of real GDP per capita		
		mean	median	s.d.	mean	median	s.d.
1	1961 - 1965	0.590	0.580	0.272	7.805	7.762	0.864
2	1966 - 1970	0.419	0.333	0.364	7.948	7.837	0.905
3	1971 - 1975	0.394	0.333	0.350	8.045	8.122	0.936
4	1976 - 1980	0.465	0.333	0.344	8.152	8.224	0.978
5	1981 - 1985	0.498	0.500	0.367	8.177	8.188	0.997
6	1986 - 1990	0.543	0.500	0.354	8.227	8.186	1.068
7	1991 - 1995	0.577	0.667	0.343	8.273	8.270	1.151
8	1996 - 2000	0.632	0.667	0.332	-	-	-

Table 4: Test statistics

	$c = 0.25$			$c = 0.5$			$c = 1$		
Null hypothesis	$K = 1$	$K = 2$	$K = 3$	$K = 1$	$K = 2$	$K = 3$	$K = 1$	$K = 2$	$K = 3$
Statistics	3.706	2.518	0.619	3.706	1.975	0.944	3.706	2.323	1.699
p-values	0.0001	0.006	0.269	0.0001	0.024	0.173	0.0001	0.010	0.045
Holm adjusted p-values	0.0003	0.012	0.269	0.0003	0.048	0.173	0.0003	0.020	0.045

Table 5: Estimation results

			β_{1i}			β_{2i}			CIE
			estimates	s.e.	t-stat	estimates	s.e.	t-stat	estimates
common estimation			0.130	0.031	4.160	0.290	0.043	6.770	0.183
$c = 0.25$	C-Lasso	Group 1	-0.449	0.054	-8.360	0.181	0.080	2.250	-0.548
		Group 2	0.121	0.032	3.804	-0.144	0.074	-1.948	0.106
		Group 3	0.334	0.080	4.174	0.417	0.069	6.022	0.573
	Post C-Lasso	Group 1	-0.394	0.054	-7.333	0.129	0.077	1.611	-0.452
		Group 2	0.103	0.034	3.245	-0.084	0.068	-1.135	0.095
		Group 3	0.393	0.064	4.907	0.394	0.069	5.691	0.649
$c = 0.5$	C-Lasso	Group 1	-0.512	0.058	-8.778	0.163	0.088	1.863	-0.612
		Group 2	0.121	0.036	3.332	-0.156	0.079	-1.973	0.105
		Group 3	0.301	0.057	5.280	0.413	0.064	6.503	0.513
	Post C-Lasso	Group 1	-0.427	0.058	-7.316	0.115	0.084	1.307	-0.482
		Group 2	0.078	0.035	2.138	-0.107	0.074	-1.351	0.070
		Group 3	0.341	0.051	5.976	0.380	0.063	5.983	0.550
$c = 1$	C-Lasso	Group 1	-0.496	0.077	-6.480	0.223	0.099	2.251	-0.638
		Group 2	0.053	0.025	2.100	-0.365	0.102	-3.576	0.039
		Group 3	0.264	0.043	6.155	0.357	0.057	6.276	0.411
	Post C-Lasso	Group 1	-0.363	0.057	-4.740	0.139	0.089	1.405	-0.422
		Group 2	0.021	0.022	0.828	-0.240	0.074	-2.352	0.017
		Group 3	0.286	0.041	6.677	0.350	0.057	6.157	0.440

Note: CIE stands for cumulative income effect, which is defined as $(\beta_{1i}/(1-\beta_{2i}))$.

Table 6: Correlation between Y_{it} and $X_{i,t-1}$ for selected countries

Time period t	Panel A: Representative countries						Panel B: China v.s. U.S.			
	Malaysia		Indonesia		Nepal		China		U.S.	
	Y_{it}	X_{it}	Y_{it}	X_{it}	Y_{it}	X_{it}	Y_{it}	X_{it}	Y_{it}	X_{it}
1	0.800	7.823	0.100	6.798	0.290	6.652	0.160	6.643	0.920	9.595
2	0.833	7.967	0.333	6.992	0.167	6.704	0	6.703	1	9.702
3	0.667	8.186	0.333	7.257	0.167	6.785	0	6.811	1	9.800
4	0.667	8.492	0.333	7.547	0.667	6.756	0.167	6.974	1	9.968
5	0.667	8.603	0.333	7.731	0.667	6.908	0.167	7.296	1	10.070
6	0.333	8.783	0.167	7.955	0.500	6.991	0	7.488	1	10.183
7	0.500	9.072	0.000	8.201	0.667	7.125	0	7.944	1	10.255
8	0.333	9.202	0.667	8.200	0.667	7.286	0	8.229	1	10.413
Correlation between Y_{it} and $X_{i,t-1}$ $(t = 2, ..., 8)$	-0.863		0.069		0.658		-0.330		N.A.	

Table 7: Classification of countries

Group 1 ("negative effect" group) ($N_1 = 19$)				
Burkina Faso	Central African Rep.	Colombia	Guatemala	Iran
Kenya	Sri Lanka	Madagascar	Mauritania	Malaysia
Niger	Nicaragua	Sierra Leone	El Salvador	Syrian Arab Rep.
Chad	Togo	Turkey	South Africa	
Group 2 ("small effect" group) ($N_2 = 30$)				
Argentina	Austria	Burundi	China	Cote d'Ivoire
Cameroon	Congo Rep.	Costa Rica	Dominican Rep.	Egypt Arab Rep.
France	Gabon	United Kingdom	Ghana	Indonesia
Ireland	Italy	Japan	Luxembourg	Mexico
Nigeria	Rwanda	Singapore	Sweden	Thailand
Tunisia	Uganda	United States	Congo Dem. Rep.	Zambia
Group 3 ("large effect" group) ($N_3 = 33$)				
Benin	Bolivia	Brazil	Chile	Cyprus
Algeria	Ecuador	Spain	Finland	Guinea
Greece	Guyana	Honduras	India	Israel
Jamaica	Jordan	Korea Rep.	Morocco	Mali
Malawi	Nepal	Panama	Peru	Philippines
Portugal	Paraguay	Romania	Trinidad and Tobago	Taiwan
Tanzania	Uruguay	Venezuela RB		

Table 8: Summary statistics by groups

		Group 1 "negative effect"		Group2 "small effect"		Group 3 "large effect"	
variables	variable description	mean	s.d.	mean	s.d.	mean	s.d.
edu65	education level in 1965	1.678	1.160	3.967	2.713	3.232	1.634
inc65	logarithm of real GDP per capita in 1965	7.568	0.582	7.903	1.073	7.852	0.783
dem65	measure of democracy in 1965	0.542	0.233	0.625	0.290	0.585	0.278
constraint	constraints on the executive at independence	0.353	0.343	0.295	0.338	0.335	0.367
indcent	year of independence/100	19.094	0.690	18.889	0.735	18.951	0.685
democ	500 year democracy change	0.616	0.274	0.661	0.303	0.826	0.211
growth	500 year income per capita change	1.288	0.931	2.157	1.237	2.091	1.014

Table 9: Determinants of the group pattern

Group 1 ("negative effect" group)							
edu65	-0.566***	-0.847**	-1.013***	-1.294***	-1.491***	-1.249***	-0.990**
	(0.199)	(0.339)	(0.347)	(0.388)	(0.418)	(0.415)	(0.423)
inc65	-	0.861	0.869	1.363 *	1.249	1.144	1.817*
	-	(0.718)	(0.727)	(0.778)	(0.845)	(0.873)	(0.962)
dem65	-	-	1.740	0.781	0.549	0.068	-0.356
	-	-	(1.606)	(1.718)	(1.717)	(1.720)	(1.904)
constraints	-	-	-	2.116	3.303**	4.282**	4.838***
	-	-	-	(1.313)	(1.487)	(1.704)	(1.716)
indcent	-	-	-	-	-0.906	-1.723*	-1.963**
	-	-	-	-	(0.694)	(0.978)	(0.971)
demco	-	-	-	-	-	-5.099*	-5.627**
	-	-	-	-	-	(2.766)	(2.778)
growth	-	-	-	-	-	-	-1.074
	-	-	-	-	-	-	(0.684)
Group 2 ("small effect" group)							
edu65	0.167	0.260	0.249	0.299	0.301	0.795**	0.780**
	(0.141)	(0.225)	(0.260)	(0.259)	(0.262)	(0.334)	(0.364)
inc65	-	-0.295	-0.298	-0.427	-0.380	-0.714	-0.906
	-	(0.606)	(0.601)	(0.618)	(0.674)	(0.730)	(0.788)
dem65	-	-	0.142	0.535	0.495	-0.157	0.145
	-	-	(1.401)	(1.580)	(1.583)	(1.760)	(1.784)
constraints	-	-	-	-0.862	-0.939	-0.340	-0.245
	-	-	-	(0.839)	(1.099)	(1.211)	(1.216)
indcent	-	-	-	-	0.084	-0.983	-1.123
	-	-	-	-	(0.653)	(0.753)	(0.788)
democ	-	-	-	-	-	-7.445***	-7.529***
	-	-	-	-	-	(2.572)	(2.532)
growth	-	-	-	-	-	-	0.105
	-	-	-	-	-	-	(0.540)

Note: *, ** and *** denote significance at the 10%, 5% and 1% levels, respectively. The results are based on a multinomial logit regression where Group 3 ("large effect" group) is taken as the reference group. The standard errors are calculated without taking into account the fact that the dependent variables are estimated.

APPENDIX

In this appendix we prove the main results in the paper. The proof relies on some technical lemmas given in Appendix B.

A Proof of the main results

Proof of Theorem 3.1. Let $\hat{u}_i = (\hat{u}_{i1}, ..., \hat{u}_{iT})'$ and $\bar{P}_{X_i} = X_i(X_i' M_0 X_i)^{-1} X_i'$. Then by (2.4),

$$\hat{u}_i = M_0 u_i + M_0 X_i (\beta_i^0 - \hat{\beta}_i), \tag{A.1}$$

and

$$\begin{aligned}
\sqrt{V_{NT}} J_{NT}(K_0) &= N^{-1/2} \sum_{i=1}^{N} \left[u_i + X_i\left(\beta_i^0 - \hat{\beta}_i\right)\right]' M_0 \bar{P}_{X_i} M_0 \left[u_i + X_i\left(\beta_i^0 - \hat{\beta}_i\right)\right] - B_{NT} \\
&= \left(N^{-1/2} \sum_{i=1}^{N} u_i' M_0 \bar{P}_{X_i} M_0 u_i - B_{NT}\right) + N^{-1/2} \sum_{i=1}^{N} \left(\beta_i^0 - \hat{\beta}_i\right)' X_i' M_0 X_i \left(\beta_i^0 - \hat{\beta}_i\right) \\
&\quad + 2N^{-1/2} \sum_{i=1}^{N} u_i' M_0 X_i \left(\beta_i^0 - \hat{\beta}_i\right) \\
&\equiv (A_{1NT} - B_{NT}) + A_{2NT} + 2A_{3NT}, \text{ say,}
\end{aligned} \tag{A.2}$$

by the fact that $\bar{P}_{X_i} M_0 X_i = X_i$. We complete the proof by showing that under $\mathbb{H}_0(K_0)$, (i) $A_{1NT} - B_{NT} \xrightarrow{D} N(0, V_0)$, (ii) $A_{2NT} = o_P(1)$, and (iii) $A_{3NT} = o_P(1)$. We prove (i)-(iii) in Propositions A.1-A.3 below.

Proposition A.1 *$A_{1NT} - B_{NT} \xrightarrow{D} N(0, V_0)$ under $\mathbb{H}_0(K_0)$.*

Proof. Recall that $H_i = M_0 \bar{P}_{X_i} M_0$ and that $h_{i,ts}$ denotes the (t,s)'th element of H_i : $h_{i,ts} = T^{-1} \sum_{r=1}^{T} \sum_{q=1}^{T} \eta_{tr}\ X_{ir}' \left(T^{-1} X_i' M_0 X_i\right)^{-1} X_{iq} \eta_{qs}$, where $\eta_{tr} = \mathbf{1}_{tr} - T^{-1}$ and $\mathbf{1}_{tr} = \mathbf{1}\{t = r\}$. Let $\bar{h}_{i,ts} \equiv T^{-1} \sum_{r=1}^{T} \sum_{q=1}^{T} \eta_{tr}\ X_{ir}' \Omega_i^{-1} X_{iq} \eta_{qs}$. Observe that

$$\begin{aligned}
A_{1NT} - B_{NT} &= 2N^{-1/2} \sum_{i=1}^{N} \sum_{1 \le s < t \le T} u_{it} u_{is} \bar{h}_{i,ts} + 2N^{-1/2} \sum_{i=1}^{N} \sum_{1 \le s < t \le T} u_{it} u_{is} \left(h_{i,ts} - \bar{h}_{i,ts}\right) \\
&\equiv A_{1NT,1} + A_{1NT,2}, \text{ say.}
\end{aligned}$$

It suffices to show that: (i) $A_{1NT,1} \xrightarrow{D} N(0, V_0)$ and (ii) $A_{1NT,2} = o_P(1)$.

First, we show (i). Using $\eta_{tr} = \mathbf{1}_{tr} - T^{-1}$, we have

$$\begin{aligned}
A_{1NT,1} &= \frac{2}{T\sqrt{N}} \sum_{i=1}^{N} \sum_{1 \le s < t \le T} \sum_{r=1}^{T} \sum_{q=1}^{T} u_{it} u_{is} \eta_{tr} X_{ir}' \Omega_i^{-1} X_{iq} \eta_{qs} \\
&= \frac{2}{T\sqrt{N}} \sum_{i=1}^{N} \sum_{1 \le s < t \le T} u_{it} u_{is} X_{it}' \Omega_i^{-1} X_{is} - \frac{2}{T^2\sqrt{N}} \sum_{i=1}^{N} \sum_{1 \le s < t \le T} \sum_{r=1}^{T} u_{it} u_{is} X_{ir}' \Omega_i^{-1} X_{is}
\end{aligned}$$

$$-\frac{2}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{q=1}^{T}u_{it}u_{is}X_{it}'\Omega_i^{-1}X_{iq}+\frac{2}{T^3\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\sum_{q=1}^{T}u_{it}u_{is}X_{ir}'\Omega_i^{-1}X_{iq}$$

$$\begin{aligned}
&= \frac{2}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}u_{it}u_{is}X_{it}^{\dagger\prime}\Omega_i^{-1}X_{is}^{\dagger}\\
&\quad -\frac{2}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}u_{it}u_{is}\left[X_{ir}-E\left(X_{ir}\right)\right]'\Omega_i^{-1}X_{is}\\
&\quad -\frac{2}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{q=1}^{T}u_{it}u_{is}X_{it}'\Omega_i^{-1}\left[X_{iq}-E\left(X_{iq}\right)\right]\\
&\quad +\frac{2}{T^3\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\sum_{q=1}^{T}u_{it}u_{is}\left[X_{ir}-E\left(X_{ir}\right)\right]'\Omega_i^{-1}\left[X_{iq}-E\left(X_{iq}\right)\right]\\
&\quad +\frac{4}{T^3\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\sum_{q=1}^{T}u_{it}u_{is}\left[X_{ir}-E_{\mathcal{D}}\left(X_{ir}\right)\right]'\Omega_i^{-1}E_{\mathcal{D}}\left(X_{iq}\right)\\
&\equiv A_{1NT,11}+A_{1NT,12}+A_{1NT,13}+A_{1NT,14}+A_{1NT,15},\ \text{say.}
\end{aligned}$$

By Lemma B.4(ii)-(v), $A_{1NT,12}+A_{1NT,13}+A_{1NT,14}+A_{1NT,15}=o_P(1)$. We are left to show that $A_{1NT,11}\overset{D}{\to}N(0,V_0)$. Observe that

$$A_{1NT,11}=\frac{2}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}u_{it}u_{is}X_{it}^{\dagger\prime}\Omega_i^{-1}X_{is}^{\dagger}=\sum_{t=2}^{T}Z_{NT,t},$$

where $Z_{NT,t}\equiv 2T^{-1}N^{-1/2}\sum_{i=1}^{N}\sum_{s=1}^{t-1}u_{it}u_{is}\bar{b}_{it}'\bar{b}_{is}$ and $\bar{b}_{it}\equiv\Omega_i^{-1/2}X_{it}^{\dagger}$. By Assumption A.1(v)

$$E\left(Z_{NT,t}|\mathcal{F}_{NT,t-1}\right)\equiv 2T^{-1}N^{-1/2}\sum_{i=1}^{N}\sum_{s=1}^{t-1}u_{is}\bar{b}_{it}'\bar{b}_{is}E(u_{it}|\mathcal{F}_{NT,t-1})=0.$$

That is, $\{Z_{NT,t},\ \mathcal{F}_{NT,t}\}$ is an m.d.s. By the martingale CLT (e.g., Pollard (1984, p. 171)), it suffices to show that

$$\mathcal{Z}\equiv\sum_{t=2}^{T}E_{\mathcal{F}_{NT,t-1}}\left|Z_{NT,t}\right|^4=o_P(1),\ \text{and}\ \sum_{t=2}^{T}Z_{NT,t}^2-V_{NT}=o_P(1)\qquad\text{(A.3)}$$

where $E_{\mathcal{F}_{NT,t-1}}$ denotes expectation conditional on $\mathcal{F}_{NT,t-1}$. Observing that $\mathcal{Z}\ge 0$, it suffices to show that $\mathcal{Z}=o_P(1)$ by showing that $E(\mathcal{Z})=o_P(1)$ by Markov inequality. By Assumptions A.1(iv)-(v),

$$\begin{aligned}
E(\mathcal{Z}) &= \frac{16}{T^4N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{1\le r,s,q,v\le t-1}E\left(\bar{b}_{it}'\bar{b}_{is}\bar{b}_{jt}'\bar{b}_{jr}\bar{b}_{kt}'\bar{b}_{kq}\bar{b}_{lt}'\bar{b}_{lv}u_{is}u_{jr}u_{kq}u_{lv}u_{it}u_{jt}u_{kt}u_{lt}\right)\\
&= 48\mathcal{Z}_1+16\mathcal{Z}_2,
\end{aligned}$$

where

$$\mathcal{Z}_1\equiv\frac{1}{T^4N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{1\le r,s,q,v\le t-1}E\left(\bar{b}_{it}'\bar{b}_{is}\bar{b}_{it}'\bar{b}_{ir}u_{is}u_{ir}u_{it}^2\right)E\left(\bar{b}_{jt}'\bar{b}_{jq}\bar{b}_{jt}'\bar{b}_{jv}u_{jq}u_{jv}u_{jt}^2\right),\qquad\text{(A.4)}$$

$$\mathcal{Z}_2\equiv\frac{1}{T^4N^2}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{1\le r,s,q,v\le t-1}E\left(\bar{b}_{it}'\bar{b}_{is}\bar{b}_{it}'\bar{b}_{ir}\bar{b}_{it}'\bar{b}_{iq}\bar{b}_{it}'\bar{b}_{iv}u_{is}u_{ir}u_{iq}u_{iv}u_{it}^4\right).\qquad\text{(A.5)}$$

For the moment we assume that $p=1$ so that we can treat the $p\times 1$ vector $\bar{b}_{it}$ as a scalar. (The general case follows from Slutsky lemma and the fact that $\bar{b}_{it}'\bar{b}_{is}\bar{b}_{it}'\bar{b}_{ir}=\sum_{k=1}^{p}\sum_{l=1}^{p}\bar{b}_{it,k}\bar{b}_{is,k}\bar{b}_{it,l}\bar{b}_{ir,l}$ where $\bar{b}_{it,k}$ denotes the k'th element of $\bar{b}_{it}$.) To bound the summation in (A.4), we consider three cases for the time indices in $S\equiv\{r,s,q,v,t-1\}$: (a) $\#S=5$, (b) $\#S=4$, and (c) $\#S\leq 3$. We use EZ_{1a}, EZ_{1b} and EZ_{1c} to denote the corresponding summations when the time indices are restricted to cases (a), (b), and (c) respectively. In case (a), applying Davydov inequality (e.g., Hall and Heyde (1980, p. 278)) yields

$$\left|E\left(\bar{b}_{it}\bar{b}_{is}\bar{b}_{it}\bar{b}_{ir}u_{is}u_{ir}u_{it}^{2}\right)\right|\leq 8C_{i,tsr}\left(t,s,r\right)\alpha\left(t-1-\left(s\vee r\right)\right)^{(1+\sigma)/(2+\sigma)}, \tag{A.6}$$

where $a\vee b\equiv\max\left(a,b\right)$ and $C_{1i,tsr}\equiv\max_{i,t,s,r}\left\|\bar{b}_{is}\bar{b}_{ir}u_{is}u_{ir}\right\|_{4+2\sigma}\left\|\bar{b}_{it}^{2}u_{it}^{2}\right\|_{4+2\sigma}$. A similar inequality holds for $E(\bar{b}_{jt}\bar{b}_{jq}\bar{b}_{jt}\bar{b}_{jv}u_{jq}u_{jv}u_{jt}^{2})$. By the repeated use of Cauchy-Schwarz's and Jensen's inequalities and Assumption A.1(i),

$$\begin{aligned}\left|C_{1i,tsr}\right| &\leq \frac{1}{2}\left[\left\|\bar{b}_{is}\bar{b}_{ir}u_{is}u_{ir}\right\|_{4+2\sigma}^{2}+\left\|\bar{b}_{it}^{2}u_{it}^{2}\right\|_{4+2\sigma}^{2}\right]\\ &\leq \frac{1}{4}\left\{\left\|\bar{b}_{is}u_{is}\right\|_{8+4\sigma}^{2}+\left\|\bar{b}_{ir}u_{ir}\right\|_{8+4\sigma}^{2}+2\left\|\bar{b}_{it}^{2}u_{it}^{2}\right\|_{4+2\sigma}^{2}\right\}\leq C_{1}\end{aligned} \tag{A.7}$$

for some $C_1<\infty$. With this, we can readily show that under A.1(iii),

$$EZ_{1a}\leq\frac{64C_{1}^{2}}{T^{4}}\sum_{t=2}^{T}\left\{\sum_{1\leq s,r\leq t-1}\alpha\left(t-1-\left(s\vee r\right)\right)^{(1+\sigma)/(2+\sigma)}\right\}^{2}=O\left(T^{-1}\right).$$

In case (b), we consider two subcases: (b1) one and only one of r,s,q,v equals $t-1$, (b2) $\#\{r,s,q,v\}=3$. We use EZ_{1b1} and EZ_{1b2} to denote the corresponding summations when the individual indices are restricted to subcases (b1) and (b2), respectively. In subcase (b1), wlog we assume that $v=t-1$, and apply

$$\left|E\left(\bar{b}_{jt}\bar{b}_{jq}\bar{b}_{jt}\bar{b}_{j,t-1}u_{jq}u_{j,t-1}u_{jt}^{2}\right)\right|\leq 8C_{2j,tq}\alpha\left(t-1-q\right)^{(1+\sigma)/(2+\sigma)}$$

for $C_{2j,tq}\equiv\left\|\bar{b}_{jq}u_{jq}\right\|_{8+4\sigma,\mathcal{D}}\left\|\bar{b}_{jt}^{2}\bar{b}_{j,t-1}u_{j,t-1}u_{jt}^{2}\right\|_{(8+4\sigma)/3,\mathcal{D}}\leq C_{2}$ for some $C_2<\infty$ and (A.6)-(A.7) to obtain

$$\begin{aligned}EZ_{1b1} &\leq \frac{64C_{1}C_{2}}{T^{3}}\sum_{t=2}^{T}\left\{\frac{1}{T}\sum_{1\leq s,r\leq t-1}\alpha\left(t-1-\left(s\vee r\right)\right)^{(1+\sigma)/(2+\sigma)}\right\}\left\{\sum_{1\leq q\leq t-1}\alpha\left(t-1-q\right)^{(1+\sigma)/(2+\sigma)}\right\}\\ &= O\left(T^{-2}\right).\end{aligned}$$

In subcase (b2), wlog we assume that $q=v$ and $r<s<t-1$. We consider two subsubcases: (b21) either $t-1-s>\tau_{*}$ or $s-r>\tau_{*}$, and (b22) $t-1-s\leq\tau_{*}$ and $s-t\leq\tau_{*}$. In the first case, we have

$$\left|E\left(\bar{b}_{it}\bar{b}_{is}\bar{b}_{it}\bar{b}_{ir}u_{is}u_{ir}u_{it}^{2}\right)\right|\leq\left\{\begin{array}{ll}8C_{3i,tsr}\alpha\left(\tau_{*}\right)^{(1+\sigma)/(2+\sigma)} & \text{if } t-1-s>\tau_{*}\\ 8C_{4i,tsr}\left(t,s,r\right)\alpha\left(\tau_{*}\right)^{(1+\sigma)/(2+\sigma)} & \text{if } s-r>\tau_{*}\end{array}\right.,$$

where $C_{3i,tsr}\equiv\left\|\bar{b}_{it}\bar{b}_{it}u_{it}^{2}\right\|_{4+2\sigma}\left\|\bar{b}_{is}\bar{b}_{ir}u_{is}u_{ir}\right\|_{4+2\sigma}\leq C_{3}<\infty$ and $C_{4i,tsr}\equiv\left\|\bar{b}_{it}\bar{b}_{is}\bar{b}_{it}u_{is}u_{it}^{2}\right\|_{(8+4\sigma)/3}$ $\left\|\bar{b}_{ir}u_{ir}\right\|_{8+4\sigma}\leq C_{4}<\infty$. These results, in conjunction with the fact that the total number of terms in the summation in subcase (b22) is of order $O\left(N^2T^3\tau_*^2\right)$ and Assumption A.1(iii), imply that

$$EZ_{1b2} \leq O\left[T^2\alpha\left(\tau_*\right)^{(1+\sigma)/(2+\sigma)} + T^{-4}N^{-2}N^2T^3\tau_*^2\right] = O\left(T^2\alpha\left(\tau_*\right)^{(1+\sigma)/(2+\sigma)} + T^{-1}\tau_*^2\right) = o\left(1\right).$$

Consequently, $EZ_{1b} = o\left(1\right).$ In case (c), we have $EZ_{1c} = O\left(T^{-1}\right)$ as the number of terms in the summation is $O\left(N^2T^3\right)$ and each term in absolute value has a bounded expectation. It follows that $\mathcal{Z}_1 = o_P\left(1\right).$

To bound $\mathcal{Z}_2$, we consider two cases for the set of indices $S \equiv \{r, s, q, v, t-1\}$: (a) $\#S = 5$, and (b) all the other cases. We use EZ_{2a} and EZ_{2b} to denote the corresponding summations when the individual indices are restricted to subcases (a) and (b), respectively. In the first case, letting $c = \max(s, r, q, v)$ we have

$$\left|E\left(\bar{b}_{it}^4\bar{b}_{is}\bar{b}_{ir}\bar{b}_{iq}\bar{b}_{iv}u_{is}u_{ir}u_{iq}u_{iv}u_{it}^4\right)\right| \leq 8C_{5i,t,s,r,q,v}\left(t,s,r,q,v\right)\alpha\left(t-1-c\right)^{\sigma/(2+\sigma)},$$

where $C_{5i,t,s,r,q,v} \equiv \left\|\bar{b}_{is}\bar{b}_{ir}\bar{b}_{iq}\bar{b}_{iv}u_{is}u_{ir}u_{iq}u_{iv}\right\|_{2+\sigma}\left\|\bar{b}_{it}^4u_{it}^4\right\|_{2+\sigma} \leq C_5 < \infty$. Then $EZ_{2a} \leq 8CN^{-1}\sum_{s=1}^{T}\alpha\left(s\right)^{\sigma/(2+\sigma)} = O\left(N^{-1}\right).$ In case (b), we have $EZ_{2b} = O\left(N^{-1}\right).$ It follows that $\mathcal{Z}_2 = O\left(N^{-1}\right)$ and thus $\mathcal{Z} = o_P\left(1\right).$ Consequently the first part of (A.3) follows.

For the second part of (A.3), by Assumptions A.1(iv)-(v) we have

$$\begin{aligned}\sum_{t=2}^{T}E(Z_{NT,t}^2) &= 4T^{-2}N^{-1}\sum_{t=2}^{T}E\left[\sum_{i=1}^{N}\sum_{s=1}^{t-1}u_{it}u_{is}\bar{b}_{it}'\bar{b}_{is}\right]^2 \\ &= 4T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}E(u_{it}^2u_{is}u_{ir}\bar{b}_{it}'\bar{b}_{is}\bar{b}_{it}'\bar{b}_{ir}) = V_{NT}.\end{aligned}$$

In addition, we can show by straightforward moment calculations that $E(\sum_{t=2}^{T}Z_{NT,t}^2)^2 = V_{NT}^2 + o\left(1\right).$ Thus $\mathrm{Var}(\sum_{t=2}^{T}Z_{NT,t}^2) = o\left(1\right)$ and the second part of (A.3) follows. This completes the proof of (i).

In addition, by Lemma B.4(i), $A_{1NT,2} = o_P\left(1\right).$ ■

Proposition A.2 $A_{2NT} = o_P\left(1\right)$ *under* $\mathbb{H}_0\left(K_0\right).$

Proof. Noting that $\mathbf{1}\{i \in G_k^0\} = \mathbf{1}\{i \in \hat{G}_k\} + \mathbf{1}\{i \in G_k^0\backslash\hat{G}_k\} - \mathbf{1}\{i \in \hat{G}_k\backslash G_k^0\}$, under $\mathbb{H}_0\left(K_0\right)$ we have

$$\begin{aligned}A_{2NT} &= N^{-1/2}\sum_{k=1}^{K_0}\sum_{i\in G_k^0}\left(\alpha_k^0 - \hat{\beta}_i\right)'X_i'M_0X_i\left(\alpha_k^0 - \hat{\beta}_i\right) \\ &= N^{-1/2}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k}\left(\alpha_k^0 - \hat{\alpha}_k\right)'X_i'M_0X_i\left(\alpha_k^0 - \hat{\alpha}_k\right) \\ &\quad + N^{-1/2}\sum_{k=1}^{K_0}\sum_{i\in G_k^0\backslash\hat{G}_k}\left(\alpha_k^0 - \hat{\beta}_i\right)'X_i'M_0X_i\left(\alpha_k^0 - \hat{\beta}_i\right) \\ &\quad - N^{-1/2}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k\backslash G_k^0}\left(\alpha_k^0 - \hat{\alpha}_k\right)'X_i'M_0X_i\left(\alpha_k^0 - \hat{\alpha}_k\right) \\ &\equiv A_{2NT,1} + A_{2NT,2} - A_{2NT,3}, \text{ say.}\end{aligned}$$

Let $\|\cdot\|_{\text{sp}}$ denote the spectral norm. Note that $\|A\| \leq$rank$(A)\,\|A\|_{\text{sp}}$ and $\|A\|_{\text{sp}} \leq \|A\|$ for any matrix A. By these properties, the submultiplicative property of the spectral norm, and the fact that $\|M_0\|_{\text{sp}} = 1$,

$$\begin{aligned} A_{2NT,1} &\leq N^{-1/2}\sum_{k=1}^{K_0}\left\|\alpha_k^0-\hat{\alpha}_k\right\|^2\sum_{i\in\hat{G}_k}\|X_i'M_0X_i\| \leq pN^{-1/2}\sum_{k=1}^{K_0}\left\|\alpha_k^0-\hat{\alpha}_k\right\|^2\sum_{i\in\hat{G}_k}\|X_i\|^2 \\ &= N^{-1/2}O_P\left((NT)^{-1}+T^{-2}\right)O_P\left(NT\right) = N^{-1/2}O_P\left(1+NT^{-1}\right) = o_P\left(1\right). \end{aligned}$$

Define

$$E_{kNT,i} = \left\{i \notin \hat{G}_k \mid i \in G_k^0\right\} \text{ and } \hat{F}_{kNT,i} = \left\{i \notin G_k^0 \mid i \in \hat{G}_k\right\}, \tag{A.8}$$

where $i = 1, ..., N$ and $k = 1, ..., K_0$. Let $\hat{E}_{kNT} = \cup_{i\in G_k^0}\hat{E}_{kNT,i}$ and $\hat{F}_{kNT} = \cup_{i\in\hat{G}_k}\hat{F}_{kNT,i}$. Then by the proof of Theorem 2.2 in SSP, for any $\epsilon > 0$,

$$P\left(A_{2NT,2} \geq \epsilon\right) \leq P\left(\cup_{k=1}^{K_0}\hat{E}_{kNT}\right) \to 0, \text{ and } P\left(A_{2NT,3} \geq \epsilon\right) \leq P\left(\cup_{k=1}^{K_0}\hat{F}_{kNT}\right) \to 0.$$

It follows that $A_{2NT,2} = o_P\left(1\right)$ and $A_{2NT,3} = o_P\left(1\right)$. Consequently $A_{2NT} = o_P\left(1\right)$ under $\mathbb{H}_0\left(K_0\right)$. ■

Proposition A.3 $A_{3NT} = o_P\left(1\right)$ *under* $\mathbb{H}_0\left(K_0\right)$.

Proof. As in the proof of Proposition A.2, we make the following decomposition:

$$\begin{aligned} A_{3NT} &= N^{-1/2}\sum_{k=1}^{K_0}\sum_{i\in G_k^0}u_i'M_0X_i(\alpha_k^0-\hat{\beta}_i) \\ &= N^{-1/2}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k}u_i'M_0X_i\left(\alpha_k^0-\hat{\alpha}_k\right)+N^{-1/2}\sum_{k=1}^{K_0}\sum_{i\in G_k^0\backslash\hat{G}_k}u_i'M_0X_i(\alpha_k^0-\hat{\beta}_i) \\ &\quad -N^{-1/2}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k\backslash G_k^0}u_i'M_0X_i\left(\alpha_k^0-\hat{\alpha}_k\right) \\ &\equiv A_{3NT,1}+A_{3NT,2}-A_{3NT,3}, \text{ say.} \end{aligned}$$

Using the same arguments as those used in the study of $A_{2NT,2}$ and $A_{2NT,3}$, we can show that $A_{3NT,2} = o_P\left(1\right)$ and $A_{3NT,3} = o_P\left(1\right)$. Noting that $\alpha_k^0 - \hat{\alpha}_k = O_P\left((NT)^{-1/2}+T^{-1}\right)$ for $k = 1, ..., K_0$ under $\mathbb{H}_0\left(K_0\right)$, it suffices to prove that $A_{3NT,1} = o_P\left(1\right)$ by showing that

$$\bar{A}_{3NT,1k} \equiv N^{-1/2}\sum_{i\in\hat{G}_k}X_i'M_0u_i = o_P\left(\min((NT)^{1/2},T)\right) \text{ for } k = 1, ..., K_0.$$

By the fact that $\mathbf{1}\{i \in \hat{G}_k\} = \mathbf{1}\{i \in G_k^0\} + \mathbf{1}\{i \in \hat{G}_k\backslash G_k^0\} - \mathbf{1}\{i \in G_k^0\backslash\hat{G}_k\}$ and the arguments used in the study of $A_{2NT,2}$ and $A_{2NT,3}$, we can show that $\bar{A}_{3NT,1k} \equiv \dot{A}_{3NT,1k} + o_P\left(1\right)$, where $\dot{A}_{3NT,1k} =$

$N^{-1/2}\sum_{i\in G_k^0} X_i' M_0 u_i$. Using $X_i'M_0u_i = \sum_{t=1}^T X_{it}(u_{it}-\bar{u}_i)$, we can decompose $\dot{A}_{3NT,1k}$ as follows

$$\begin{aligned}\dot{A}_{3NT,1k} &= N^{-1/2}\sum_{i\in G_k^0}\sum_{t=1}^T X_{it}u_{it} - N^{-1/2}T^{-1}\sum_{i\in G_k^0}\sum_{t=1}^T\sum_{s=1}^T X_{it}u_{is} \\ &= \left(1-T^{-1}\right)N^{-1/2}\sum_{i\in G_k^0}\sum_{t=1}^T X_{it}u_{it} - N^{-1/2}T^{-1}\sum_{i\in G_k^0}\sum_{1\le t<s\le T} X_{it}u_{is} - N^{-1/2}T^{-1}\sum_{i\in G_k^0}\sum_{1\le s<t\le T} X_{it}u_{is} \\ &\equiv \dot{A}_{3NT,1k1} - \dot{A}_{3NT,1k2} - \dot{A}_{3NT,1k3}, \text{ say.}\end{aligned}$$

Using Chebyshev inequality, we can readily show that $\dot{A}_{3NT,1k1} = O_P\left(T^{1/2}\right)$ under Assumptions A.1(i), (iv) and (v). Let $\omega_p = (\omega_{p1},...,\omega_{pp})'$ be an arbitrary $p\times 1$ nonrandom vector with $\|\omega_p\| = 1$. By Assumptions A.1(i), (iv) and (v), and Jensen inequality,

$$\begin{aligned}E\left[\left(\omega_p'\dot{A}_{3NT,1k2}\right)^2\right] &= N^{-1}T^{-2}\sum_{i\in G_k^0}\sum_{j\in G_k^0}\sum_{1\le t<s\le T}\sum_{1\le r<q\le T} E\left(\omega_p'X_{it}u_{is}\omega_p'X_{jr}u_{jq}\right) \\ &= N^{-1}T^{-2}\sum_{i\in G_k^0}\sum_{1\le t<s\le T}\sum_{1\le r<q\le T} E\left(\omega_p'X_{it}u_{is}\omega_p'X_{ir}u_{iq}\right) \\ &= N^{-1}T^{-2}\sum_{i\in G_k^0}\sum_{1\le t,r<s\le T} E\left(\omega_p'X_{it}\omega_p'X_{ir}u_{is}^2\right) = O\left(T\right).\end{aligned}$$

Then $\dot{A}_{3NT,1k2} = O_P\left(T^{1/2}\right)$ by Chebyshev inequality. Next,

$$\begin{aligned}E\left[\left(\omega_p'\dot{A}_{3NT,1k3}\right)^2\right] &= N^{-1}T^{-2}\sum_{i\in G_k^0}\sum_{1\le s<t\le T}\sum_{1\le q<r\le T} E\left(\omega_p'X_{it}u_{is}\omega_p'X_{ir}u_{iq}\right) \\ &\quad + N^{-1}T^{-2}\sum_{i\in G_k^0}\sum_{j\in G_k^0, j\ne i}\sum_{1\le s<t\le T}\sum_{1\le q<r\le T} E\left(\omega_p'X_{it}u_{is}\right)E\left(\omega_p'X_{jr}u_{jq}\right) \\ &\equiv I + II, \text{ say.}\end{aligned}$$

Let $S\equiv\{t,s,q,r\}$. To bound I, we consider two cases: (a) $\#S=4$ and (b) $\#S\le 3$, and denote the corresponding summations as I_a and I_b such that $I = I_a + I_b$. Apparently, $I_b = O(T)$. For I_a, wlog we consider three subcases: (a1) $s<t<q<r$, (a2) $s<q<t<r$, (a3) $s<q<r<t$, and denote the corresponding summations as I_{a1}, I_{a2}, and I_{a3}, respectively. (Note $I_a = 2(I_{a1}+I_{a2}+I_{a3})$.) In subcase (a1), we apply Davydov inequality to obtain

$$|I_{a1}| \le 8N^{-1}T^{-2}\sum_{i\in G_k^0}\sum_{1\le s<t<q<r\le T} c_{i,tsrq}\alpha\left(t-s\right)^{(1+\sigma)/(2+\sigma)} \le 8CT\sum_{\tau=1}^{\infty}\alpha\left(\tau\right)^{(1+\sigma)/(2+\sigma)} = O\left(T\right),$$

where $c_{i,tsrq} = \|u_{is}\|_{8+4\sigma}\left\|\omega_p'X_{it}\omega_p'X_{ir}u_{iq}\right\|_{(8+4\sigma)/3} \le C < \infty$ by Assumption A.1(i) and Jensen inequality. Analogously, we can show that $I_{a2} = O\left(T\right)$ and $I_{a3} = O(T)$. It follows that $I = O(T)$. For II, we apply Davydov inequality to obtain

$$\begin{aligned}|II| &\le N^{-1}T^{-2}\left\{\sum_{i\in G_k^0}\sum_{1\le s<t\le T}\left|E\left(\omega_p'X_{it}u_{is}\right)\right|\right\}^2 \le N^{-1}T^{-2}\left\{\sum_{i\in G_k^0}\sum_{1\le s<t\le T} c_{i,ts}\alpha\left(t-s\right)^{(3+2\sigma)/(4+2\sigma)}\right\}^2 \\ &= N^{-1}T^{-2}O\left(N^2T^2\right) = O\left(N\right),\end{aligned}$$

where $c_{i,ts} = \left\|\omega_p' X_{it}\right\|_{8+4\sigma} \left\|\omega_p' X_{it}\right\|_{8+4\sigma} \leq C < \infty$ by Assumption A.1(i). Consequently, $E\{[\omega_p' \dot{A}_{3NT,1k1}(3)]^2]\}$ $= O(N+T)$ and $\dot{A}_{3NT,1k3} = O_P\left(N^{1/2} + T^{1/2}\right)$. In sum, we have $\dot{A}_{3NT,1k} = O_P\left(N^{1/2} + T^{1/2}\right)$. It follows that $\bar{A}_{3NT,1k} = O_P\left(N^{1/2} + T^{1/2}\right) = o_P(\min((NT)^{1/2}, T))$. ■

Proof of Theorem 3.2. By Theorem 3.1 and the Slutsky lemma, it suffices to prove the first two parts of the theorem.

Step 1. We prove (i) $\hat{B}_{NT}(K_0) = B_{NT} + o_P(1)$ **under** $\mathbb{H}_0(K_0)$. Let ι_s denote a $T \times 1$ vector with 1 in the t'th position and zeros everywhere else. Then $h_{i,ts} = \iota_t' M_0 \bar{P}_{X_i} M_0 \iota_s = \sum_{r=1}^T \sum_{q=1}^T \eta_{tr} X_{ir}' (X_i M_0 X_i)^{-1}$ $X_{iq}\eta_{qs}$. Using $\hat{u}_{it}^2 - u_{it}^2 = (\hat{u}_{it} - u_{it})^2 + 2(\hat{u}_{it} - u_{it}) u_{it}$, we decompose $\hat{B}_{NT} - B_{NT}$ as follows:

$$\hat{B}_{NT}(K_0) - B_{NT} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{t=1}^T (\hat{u}_{it} - u_{it})^2 h_{i,tt} + \frac{2}{\sqrt{N}} \sum_{i=1}^N \sum_{t=1}^T (\hat{u}_{it} - u_{it}) u_{it} h_{i,tt} \equiv \hat{B}_{NT,1} + 2\hat{B}_{NT,2}, \text{ say.}$$

Noting that diag(H_i) is p.s.d., we have by (A.1) and Cauchy-Schwarz inequality

$$\begin{aligned}
\hat{B}_{NT,1} &= N^{-1/2} \sum_{i=1}^N (\hat{u}_i - u_i)' \operatorname{diag}(H_i)(\hat{u}_i - u_i) \\
&\leq 2N^{-1/2} \sum_{i=1}^N u_i' P_0 \operatorname{diag}(H_i) P_0 u_i + 2N^{-1/2} \sum_{i=1}^N (\hat{\beta}_i - \beta_i^0)' X_i' M_0 \operatorname{diag}(H_i) M_0 X_i (\hat{\beta}_i - \beta_i^0) \\
&\equiv 2\hat{B}_{NT,11} + 2\hat{B}_{NT,12}, \text{ say.}
\end{aligned}$$

We will show that $\hat{B}_{NT,1s} = o_P(1)$ for $s = 1$ and 2. By the fact that $\sum_{t=1}^T \iota_t \iota_t' = I_T$ and M_0 is idempotent, we have

$$\begin{aligned}
\mathbf{i}_T' \operatorname{diag}(H_i) \mathbf{i}_T &= \operatorname{tr}[\mathbf{i}_T' \operatorname{diag}(H_i) \mathbf{i}_T] = \sum_{t=1}^T \operatorname{tr}\left(\iota_t' M_0 \bar{P}_{X_i} M_0 \iota_t\right) = \operatorname{tr}\left(M_0 \bar{P}_{X_i} M_0\right) \\
&= \operatorname{tr}\left(M_0 X_i (X_i' M_0 X_i)^{-1} X_i' M_0\right) = p.
\end{aligned}$$

This, in conjunction with Davydov inequality, implies that

$$\begin{aligned}
E\left|\hat{B}_{NT,11}\right| &= T^{-2} N^{-1/2} \sum_{i=1}^N E\left[u_i' \mathbf{i}_T (\mathbf{i}_T' \operatorname{diag}(H_i) \mathbf{i}_T) \mathbf{i}_T' u_i\right] = pT^{-2} N^{-1/2} \sum_{i=1}^N E(u_i' \mathbf{i}_T \mathbf{i}_T' u_i) \\
&= pT^{-2} N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T E\left(u_{it}^2\right) + 2pT^{-2} N^{-1/2} \sum_{i=1}^N \sum_{1 \leq t < s \leq T} E(u_{it} u_{is}) \\
&= O(N^{1/2} T^{-1}) + O(N^{1/2} T^{-1}) = O(N^{1/2} T^{-1}).
\end{aligned}$$

Consequently $\hat{B}_{NT,11} = O_P(N^{1/2} T^{-1}) = o_P(1)$ by Markov inequality. Using $\mathbf{1}\{i \in G_k^0\} = \mathbf{1}\{i \in \hat{G}_k\} + \mathbf{1}\{i \in G_k^0 \backslash \hat{G}_k\} - \mathbf{1}\{i \in \hat{G}_k \backslash G_k^0\}$ and following similar arguments as those used in the proof of

Proposition A.2, we can show that under $\mathbb{H}_0(K_0)$,

$$\begin{aligned}\hat{B}_{NT,12} &= N^{-1/2}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k}(\hat{\alpha}_k-\alpha_k^0)'X_i'M_0\mathrm{diag}(H_i)M_0X_i(\hat{\alpha}_k-\alpha_k^0)+o_P(1)\\ &\leq N^{-1/2}\sum_{k=1}^{K_0}\left\|\hat{\alpha}_k-\alpha_k^0\right\|^2\sum_{i\in\hat{G}_k}\|X_i'M_0\mathrm{diag}(H_i)M_0X_i\|+o_P(1)\\ &= N^{-1/2}O_P\left((NT)^{-1}+T^{-2}\right)O_P(N)+o_P(1)=o_P(1),\end{aligned}$$

based on the fact that $M_0AM_0\leq A$ for any p.s.d. matrix A, and

$$\begin{aligned}\sum_{i\in\hat{G}_k}\|X_i'M_0\mathrm{diag}(H_i)M_0X_i\| &\leq \sum_{i\in\hat{G}_k}\|X_i'\mathrm{diag}(H_i)X_i\|\leq\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\left\|X_{it}'\iota_t'M_0\bar{P}_{X_i}M_0\iota_tX_{it}\right\|\\ &\leq \sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\left\|X_{it}'\iota_t'\bar{P}_{X_i}\iota_tX_{it}\right\|=\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\left\|X_{it}'X_{it}\left(X_i'M_0X_i\right)^{-1}X_{it}'X_{it}\right\|\\ &\leq \max_{1\leq i\leq N}\left\|\hat{\Omega}_i^{-1}\right\|T^{-1}\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\|X_{it}\|^4=O_P(N).\end{aligned}$$

Consequently, we have shown that $\hat{B}_{NT,1}=o_P(1)$.

For $\hat{B}_{NT,2}$, we first apply (A.1) to decompose it as follows

$$\begin{aligned}\hat{B}_{NT,2} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{u}_i-u_i)'\,\mathrm{diag}(H_i)u_i\\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}u_i'P_0\mathrm{diag}(H_i)u_i+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{\beta}_i-\beta_i^0)'X_i'M_0\mathrm{diag}(H_i)u_i\equiv\hat{B}_{NT,21}+\hat{B}_{NT,22},\ \text{say.}\end{aligned}$$

Observe that

$$\begin{aligned}\hat{B}_{NT,21} &= \frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{it}\iota_s'M_0X_i\hat{\Omega}_i^{-1}X_i'M_0\iota_su_{is}\\ &= \frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{it}\iota_s'M_0X_i\Omega_i^{-1}X_i'M_0\iota_su_{is}\\ &\quad+\frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{it}\iota_s'M_0X_i\left[\hat{\Omega}_i^{-1}-\Omega_i^{-1}\right]X_i'M_0\iota_su_{is}\equiv\hat{B}_{NT,211}+\hat{B}_{NT,212},\ \text{say.}\end{aligned}$$

Noting that $\|\iota_s'M_0X_i\|\leq\|\iota_s'X_i\|=\|X_{is}\|$, we apply Lemma B.3 to obtain

$$\begin{aligned}\left|\hat{B}_{NT,212}\right| &\leq \max_{1\leq i\leq N}\left\|\hat{\Omega}_i^{-1}-\Omega_i^{-1}\right\|\frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\left|\sum_{t=1}^{T}u_{it}\right|\sum_{s=1}^{T}\|X_{is}\|^2\|u_{is}\|\\ &= O_P(a_{NT})\,O_P\left(N^{1/2}T^{-1/2}\right)=o_P(1).\end{aligned}$$

For $\hat{B}_{NT,211}$, we have

$$\begin{aligned}
\hat{B}_{NT,211} &= \frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}u_{it}\eta_{sr}X_{ir}'\Omega_i^{-1}X_{iq}\eta_{qs}u_{is} \\
&= \frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{it}X_{is}'\Omega_i^{-1}X_{is}u_{is} - \frac{2}{T^3\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}u_{it}X_{ir}'\Omega_i^{-1}X_{is}u_{is} \\
&\quad + \frac{1}{T^4\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}u_{it}X_{ir}'\Omega_i^{-1}X_{iq}u_{is} \\
&\equiv \hat{B}_{NT,211a} - 2\hat{B}_{NT,211b} + \hat{B}_{NT,221c}, \text{ say.}
\end{aligned}$$

We further decompose $\hat{B}_{NT,211a}$ as follows $\hat{B}_{NT,211a} = \frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}'\Omega_i^{-1}X_{it}u_{it}^2 + \frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le t<s\le T}u_{it}$ $\times X_{is}'\Omega_i^{-1}X_{is}u_{is} + \frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{1\le s<t\le T}u_{it}X_{is}'\Omega_i^{-1}X_{is}u_{is} = \hat{B}_{NT,211a}(1) + \hat{B}_{NT,211a}(2) + \hat{B}_{NT,211a}(3)$. Apparently, $\hat{B}_{NT,211a}(1) = O_P\left(N^{1/2}T^{-1}\right)$ by Markov inequality. Noting that $E[\hat{B}_{NT,211a}(2)] = 0$, by Davydov inequality we can readily show that

$$E\left[\hat{B}_{NT,211a}(2)\right]^2 = \frac{1}{T^4N}\sum_{i=1}^{N}\sum_{1\le t<s\le T}\sum_{1\le r<q\le T}E\left[u_{it}X_{is}'\Omega_i^{-1}X_{is}u_{is}\ u_{ir}X_{iq}'\Omega_i^{-1}X_{iq}u_{iq}\right] = O\left(T^{-1}\right).$$

It follows that $\hat{B}_{NT,211a}(2) = O_P\left(T^{-1/2}\right)$. Similarly, $\hat{B}_{NT,211a}(3) = O_P\left(T^{-1/2}\right)$. Then $\hat{B}_{NT,211a} = O_P\left(N^{1/2}T^{-1} + T^{-1/2}\right) = o_P(1)$. Analogously, we can show that $\hat{B}_{NT,211s} = o_P(1)$ for $s = b, c$. Then we have $\hat{B}_{NT,211} = o_P(1)$ and $\hat{B}_{NT,21} = o_P(1)$.

For $\hat{B}_{NT,22}$, using the same arguments as those used in the proof of Proposition A.2, we can show that under $\mathbb{H}_0(K_0)$,

$$\begin{aligned}
\hat{B}_{NT,22} &= \frac{1}{\sqrt{N}}\sum_{k=1}^{K_0}(\hat{\alpha}_k - \alpha_k^0)'\sum_{i\in\hat{G}_k}X_i'M_0\text{diag}(H_i)u_i + o_P(1) \\
&= \frac{1}{\sqrt{N}}\sum_{k=1}^{K_0}(\hat{\alpha}_k - \alpha_k^0)'\sum_{i\in G_k^0}X_i'M_0\text{diag}(H_i)u_i + o_P(1) \equiv \bar{B}_{NT,22} + o_P(1).
\end{aligned}$$

Let $B_k = \frac{1}{\sqrt{N}}\sum_{i\in G_k^0}X_i'M_0\text{diag}(H_i)u_i$. Then as in the proof of Proposition A.2 and analysis of $\hat{B}_{NT,211a}(2)$, we can show that

$$\begin{aligned}
B_k &= \frac{1}{T\sqrt{N}}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}X_{it}\eta_{ts}\iota_s M_0X_i\hat{\Omega}_i^{-1}X_i'M_0\iota_s u_{is} \\
&= \frac{1}{T\sqrt{N}}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}X_{it}\eta_{ts}\iota_s M_0X_i\Omega_i^{-1}X_i'M_0\iota_s u_{is} + o_P(1) \\
&= \frac{1}{T\sqrt{N}}\sum_{i\in G_k^0}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}X_{it}\eta_{ts}\eta_{sr}X_{ir}'\Omega_i^{-1}X_{iq}\eta_{qs}u_{is} + o_P(1) = O_P\left(N^{1/2} + T^{1/2}\right).
\end{aligned}$$

It follows that $\hat{B}_{NT,22} = O_P\left((NT)^{-1/2} + T^{-1}\right) O_P\left(N^{1/2} + T^{1/2}\right) = o_P(1)$. This completes the proof of (i1).

Step 2. We prove (ii) $\hat{V}_{NT}(K_0) = V_{NT} + o_P(1)$. Observe that $\hat{V}_{NT}(K_0) - V_{NT} = 4V_{NT,1} + 4V_{NT,2}$, where

$$
\begin{aligned}
V_{NT,1} &= T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left\{\left[\hat{u}_{it}\hat{b}_{it}'\sum_{s=1}^{t-1}\hat{b}_{is}\hat{u}_{is}\right]^2 - \left[u_{it}\bar{b}_{it}'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2\right\}, \text{ and} \\
V_{NT,2} &= T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left\{\left[u_{it}\bar{b}_{it}'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2 - E\left[u_{it}\bar{b}_{it}'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2\right\}.
\end{aligned}
$$

Noting that $E(V_{NT,2}) = 0$ and $\operatorname{Var}(V_{NT,2}) = o(1)$ by direct moment calculations, we have $V_{NT,2} = o_P(1)$ by Chebyshev's inequality. Thus we are left to show that $V_{NT,1} = o_P(1)$. Again, using $a^2 - b^2 = (a-b)^2 + 2(a-b)b$ yields

$$
\begin{aligned}
V_{NT,1} &= T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[\hat{u}_{it}\hat{b}_{it}'\sum_{s=1}^{t-1}\hat{b}_{is}\hat{u}_{is} - u_{it}\bar{b}_{it}'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2 \\
&\quad + 2T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[\hat{u}_{it}\hat{b}_{it}'\sum_{s=1}^{t-1}\hat{b}_{is}\hat{u}_{is} - u_{it}\bar{b}_{it}'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]u_{it}\bar{b}_{it}'\sum_{r=1}^{t-1}\bar{b}_{ir}u_{ir} \\
&\equiv V_{NT,11} + 2V_{NT,12}.
\end{aligned}
$$

Let $\bar{V}_{NT,12} \equiv T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}[u_{it}\bar{b}_{it}'\sum_{r=1}^{t-1}\bar{b}_{ir}u_{ir}]^2$. By Cauchy-Schwarz inequality $V_{NT,12} \le \{V_{NT,11}\}^{1/2}\{\bar{V}_{NT,12}\}^{1/2}$. It is straightforward to show that $\bar{V}_{NT,12} = O_P(1)$ so that we can prove that $V_{NT,1} = o_P(1)$ by showing that $V_{NT,11} = o_P(1)$. Using $\hat{u}_{it}\hat{b}_{it} = (\hat{u}_{it}\hat{b}_{it} - u_{it}\bar{b}_{it}) + u_{it}\bar{b}_{it}$ and Cauchy-Schwarz inequality,

$$
\begin{aligned}
V_{NT,11} &\le 3T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[\left(\hat{u}_{it}\hat{b}_{it} - u_{it}\bar{b}_{it}\right)'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2 + 3T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[u_{it}\bar{b}_{it}'\sum_{s=1}^{t-1}\left(\hat{b}_{is}\hat{u}_{is} - \bar{b}_{is}u_{is}\right)\right]^2 \\
&\quad + 3T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[\left(\hat{u}_{it}\hat{b}_{it} - u_{it}\bar{b}_{it}\right)'\sum_{s=1}^{t-1}\left(\hat{b}_{is}\hat{u}_{is} - \bar{b}_{is}u_{is}\right)\right]^2 \\
&\equiv 3V_{NT,111} + 3V_{NT,112} + 3V_{NT,113}.
\end{aligned}
$$

We complete the proof of (ii) by showing that (ii1) $V_{NT,111} = o_P(1)$, (ii2) $V_{NT,112} = o_P(1)$, and (ii3) $V_{NT,113} = o_P(1)$.

We first show (ii1) $V_{NT,111} = o_P(1)$. Using $\hat{u}_{it}\hat{b}_{it} - u_{it}\bar{b}_{it} = (\hat{u}_{it} - u_{it})\bar{b}_{it} + u_{it}(\hat{b}_{it} - \bar{b}_{it}) + (\hat{u}_{it} - u_{it})(\hat{b}_{it} - \bar{b}_{it})$ and Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
V_{NT,111} &\le 3T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[(\hat{u}_{it} - u_{it})\bar{b}_{it}'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2 + 3T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[u_{it}\left(\hat{b}_{it} - \bar{b}_{it}\right)'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2 \\
&\quad + 3T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[(\hat{u}_{it} - u_{it})\left(\hat{b}_{it} - \bar{b}_{it}\right)'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2 \equiv 3V_{NT,111a} + 3V_{NT,111b} + 3V_{NT,111c}.
\end{aligned}
$$

By Markov and Davydov inequalities we can show that $T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}[\bar{b}_{it}'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}]^2 = O_P(1)$. By Boole inequality, Doob inequality (e.g., Hall and Heyde (1980, pp.14-15)) for m.d.s., and then Davydov inequality, for any $\epsilon > 0$ we have

$$\begin{aligned} P\left(\max_{1\le i\le N}\max_{2\le t\le T}\left\|T^{-1/2}\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right\| > N^{1/8}\epsilon\right) &\le \sum_{i=1}^{N} P\left(\max_{2\le t\le T}\left\|T^{-1/2}\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right\| > N^{1/8}\epsilon\right) \\ &\le \frac{1}{NT^4\epsilon^8}\sum_{i=1}^{N} E\left\|\sum_{s=1}^{T-1}\bar{b}_{is}u_{is}\right\|^8 = O(1). \end{aligned}$$

It follows that

$$\max_{1\le i\le N}\max_{2\le t\le T}\left\|\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right\| = O_P\left(T^{1/2}N^{1/8}\right). \quad (A.9)$$

The same conclusion follows when one replaces $\bar{b}_{is}$ by X_{is} or 1. Let $\bar{B}_i$ =diag$(\left\|\bar{b}_{i1}\right\|^2, ..., \left\|\bar{b}_{iT}\right\|^2)$. By (A.1), we can readily show that

$$\begin{aligned} N^{-1}\sum_{i=1}^{N}\|\hat{u}_i - u_i\|^2 &\le 2N^{-1}\sum_{i=1}^{N}\|P_0u_i\|^2 + 2N^{-1}\sum_{i=1}^{N}\left\|M_0X_i(\beta_i^0 - \hat{\beta}_i)\right\|^2 \\ &= O_P(1) + o_P(1) = O_P(1), \end{aligned} \quad (A.10)$$

and

$$\begin{aligned} N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}(\hat{u}_{it} - u_{it})^2\left\|\bar{b}_{it}\right\|^2 &= N^{-1}\sum_{i=1}^{N}(\hat{u}_i - u_i)'\bar{B}_i(\hat{u}_i - u_i) \\ &\le 2N^{-1}\sum_{i=1}^{N}u_i'P_0\bar{B}_iP_0u_i + 2N^{-1}\sum_{i=1}^{N}(\beta_i^0 - \hat{\beta}_i)'X_i'M_0\bar{B}_iM_0X_i(\beta_i^0 - \hat{\beta}_i) \\ &= O_P(1) + o_P(1) = O_P(1). \end{aligned} \quad (A.11)$$

By (A.9), (A.11), and Assumption A.3,

$$\begin{aligned} V_{NT,111a} &= T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}(\hat{u}_{it} - u_{it})^2\left[\bar{b}_{it}'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2 \\ &\le T^{-2}\max_{1\le i\le N}\max_{2\le t\le T}\left\|\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right\|^2\left\{N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}(\hat{u}_{it} - u_{it})^2\left\|\bar{b}_{it}\right\|^2\right\} \\ &= T^{-2}O_P\left(TN^{1/4}\right)O_P(1) = o_P(1). \end{aligned}$$

To determine the probability order of $V_{NT,111b}$ and $V_{NT,111c}$, we use the uniform probability order of $\hat{b}_{it} - \bar{b}_{it}$. We decompose $\hat{b}_{it} - \bar{b}_{it}$ as follows:

$$\begin{aligned} \hat{b}_{it} - \bar{b}_{it} &= \hat{\Omega}_i^{-1/2}\left[X_{it} - T^{-1}\sum_{r=1}^{T}X_{ir}\right] - \Omega_i^{-1/2}\left[X_{it} - T^{-1}\sum_{r=1}^{T}E(X_{ir})\right] \\ &= e_iX_{it} - e_iT^{-1}\sum_{r=1}^{T}X_{ir} - \Omega_i^{-1/2}T^{-1}\sum_{r=1}^{T}[X_{ir} - E(X_{ir})] \equiv b_{1it} - b_{2it} - b_{3it}, \end{aligned} \quad (A.12)$$

where $e_i \equiv \hat{\Omega}_i^{-1/2} - \Omega_i^{-1/2}$. By Lemma B.3 and the fact that $\max_{1\le i\le N}\max_{1\le t\le T}\|X_{it}\| = o_P((NT)^{1/(8+4\sigma)})$ by Boole and Markov inequalities, we have $\max_{1\le i\le N}\max_{1\le t\le T}\|b_{1it}\| = o_P(a_{NT}(NT)^{1/(8+4\sigma)})$. Following the proof of Lemma B.3(v), we can show that

$$
\begin{aligned}
\max_{1\le i\le N}\left\|T^{-1}\sum_{r=1}^{T}X_{ir}\right\| &\le \max_{1\le i\le N}\left\|T^{-1}\sum_{r=1}^{T}[X_{ir}-E(X_{ir})]\right\| + \max_{1\le i\le N}\left\|T^{-1}\sum_{r=1}^{T}E(X_{ir})\right\| \\
&= O_P\left(\max\{(NT)^{1/(8+4\sigma)}\log(NT)/T, (\log(NT)/T)^{1/2}\}\right) + O(1) \\
&= O(1). \qquad \text{(A.13)}
\end{aligned}
$$

It follows that $\max_{1\le i\le N}\max_{1\le t\le T}\|b_{2it}\| = O_P(a_{NT})$. Also, $\max_{1\le i\le N}\max_{1\le t\le T}\|b_{3it}\| = O_P(a_{NT})$. Thus $\max_{1\le i\le N}\max_{1\le t\le T}\left\|\hat{b}_{it}-\bar{b}_{it}\right\| = o_P(a_{NT}(NT)^{1/(8+4\sigma)})$. In addition, using (A.12) and the above bounds, we have

$$
\begin{aligned}
& T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}u_{it}^2\left\|\hat{b}_{it}-\bar{b}_{it}\right\|^2 \\
\le\ & 3T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}u_{it}^2\|b_{1it}\|^2 + 3T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}u_{it}^2\|b_{2it}\|^2 + 3T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}u_{it}^2\|b_{3it}\|^2 \\
\le\ & 3T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}u_{it}^2\|b_{1it}\|^2 + O_P(a_{NT}^2) + O_P(a_{NT}^2) = O_P(a_{NT}^2),
\end{aligned}
$$

where the last equality follows from the fact that $T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}u_{it}^2\|b_{1it}\|^2 \le \max_{1\le i\le N}\|e_i\|^2 T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}u_{it}^2\|X_{it}\|^2 = O_P(a_{NT}^2)$. Then by Assumption A.3,

$$
\begin{aligned}
V_{NT,111b} &= T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[u_{it}\left(\hat{b}_{it}-\bar{b}_{it}\right)'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2 \\
&\le T^{-1}\max_{1\le i\le N}\max_{1\le t\le T}\left[\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2\left[T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}u_{it}^2\left\|\hat{b}_{it}-\bar{b}_{it}\right\|^2\right] \\
&= T^{-1}O_P\left(TN^{1/4}\right)O_P(a_{NT}^2) = o_P(N^{1/4}a_{NT}^2) = o_P(1),
\end{aligned}
$$

and

$$
\begin{aligned}
V_{NT,111c} &= T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[(\hat{u}_{it}-u_{it})\left(\hat{b}_{it}-\bar{b}_{it}\right)'\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right]^2 \\
&= T^{-2}\max_{1\le i\le N}\max_{1\le t\le T}\left\|\hat{b}_{it}-\bar{b}_{it}\right\|^2\max_{1\le i\le N}\max_{1\le t\le T}\left\|\sum_{s=1}^{t-1}\bar{b}_{is}u_{is}\right\|^2 N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}(\hat{u}_{it}-u_{it})^2 \\
&= T^{-2}o_P(a_{NT}^2(NT)^{1/(4+2\sigma)})\,O_P\left(TN^{1/4}\right)O_P(1) = o_P(N^{1/4}a_{NT}^2) = o_P(1).
\end{aligned}
$$

It follows that $V_{NT,111} = o_P(1)$.

To show (ii2) and (ii3), we find that it is convenient to bound $S_{it} \equiv T^{-1/2}\sum_{s=1}^{t-1}(\hat{b}_{is}\hat{u}_{is} - \bar{b}_{is}u_{is})$. Using $\hat{u}_{it}\hat{b}_{it} - u_{it}\bar{b}_{it} = u_{it}(\hat{b}_{it} - \bar{b}_{it}) + (\hat{u}_{it} - u_{it})\,\bar{b}_{it} + (\hat{u}_{it} - u_{it})\,(\hat{b}_{it} - \bar{b}_{it})$, we have

$$\begin{aligned} S_{it} &\equiv T^{-1/2}\sum_{s=1}^{t-1} u_{is}(\hat{b}_{is} - \bar{b}_{is}) + T^{-1/2}\sum_{s=1}^{t-1}(\hat{u}_{is} - u_{is})\,\bar{b}_{is} + T^{-1/2}\sum_{s=1}^{t-1}(\hat{u}_{is} - u_{is})\,(\hat{b}_{is} - \bar{b}_{is}) \\ &\equiv S_{1it} + S_{2it} + S_{3it},\ \text{say.} \end{aligned}$$

By (A.12), $S_{1it} = T^{-1/2}\sum_{s=1}^{t-1} u_{is}\,(b_{1is} - b_{2is} - b_{3is}) \equiv S_{1it,1} - S_{1it,2} - S_{1it,3}$, say. By the remark after (A.9), (A.13), and Lemma B.3,

$$\begin{aligned} \max_{1\le i\le N}\max_{1\le t\le T}\|S_{1it,1}\| &\le \max_{1\le i\le N}\|e_i\| \max_{1\le i\le N}\max_{1\le t\le T}\left\|T^{-1/2}\sum_{s=1}^{t-1} X_{is}u_{is}\right\| = O_P\left(a_{NT}\right)O_P(N^{1/4}) = o_P\left(1\right), \\ \max_{1\le i\le N}\max_{1\le t\le T}\|S_{1it,2}\| &\le \max_{1\le i\le N}\|e_i\| \max_{1\le i\le N}\left\|T^{-1}\sum_{r=1}^{T} X_{ir}\right\| \max_{1\le i\le N}\max_{1\le t\le T}\left\|T^{-1/2}\sum_{s=1}^{t-1} u_{is}\right\| \\ &= O_P\left(a_{NT}\right)O_P\left(1\right)O_P(N^{1/8}) = o_P\left(1\right), \end{aligned}$$

and

$$\begin{aligned} \max_{1\le i\le N}\max_{1\le t\le T}\|S_{1it,3}\| &\le \max_{1\le i\le N}\left\|\Omega_i^{-1/2}\right\| \max_{1\le i\le N}\left\|T^{-1}\sum_{r=1}^{T}\left[X_{ir} - E\left(X_{ir}\right)\right]\right\| \max_{1\le i\le N}\max_{1\le t\le T}\left\|T^{-1/2}\sum_{s=1}^{t-1} u_{is}\right\| \\ &= O\left(1\right)O_P\left(a_{NT}\right)O_P(N^{1/8}) = o_P\left(1\right). \end{aligned}$$

It follows that $\max_{1\le i\le N}\max_{1\le t\le T}\|S_{1it}\| = o_P\left(1\right)$. Similarly we can show that $\max_{1\le i\le N}\max_{1\le t\le T}\|S_{2it}\| = o_P\left(1\right)$ and $\max_{1\le i\le N}\max_{1\le t\le T}\|S_{3it}\| = o_P\left(1\right)$. Hence $\max_{1\le i\le N}\max_{1\le t\le T}\|S_{it}\| = o_P\left(1\right)$. It follows that

$$\begin{aligned} V_{NT,112} &= T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[u_{it}\bar{b}_{it}'\sum_{s=1}^{t-1}\left(\hat{b}_{is}\hat{u}_{is} - \bar{b}_{is}u_{is}\right)\right]^2 \\ &\le \left\{\max_{1\le i\le N}\max_{1\le t\le T}\|S_{it}\|^2\right\}\left\{T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left\|u_{it}\bar{b}_{it}'\right\|^2\right\} = o_P\left(1\right)O_P\left(1\right) = o_P\left(1\right), \end{aligned}$$

and

$$\begin{aligned} V_{NT,113} &= T^{-2}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[\left(\hat{u}_{it}\hat{b}_{it} - u_{it}\bar{b}_{it}\right)'\sum_{s=1}^{t-1}\left(\hat{b}_{is}\hat{u}_{is} - \bar{b}_{is}u_{is}\right)\right]^2 \\ &\le \left\{\max_{1\le i\le N}\max_{1\le t\le T}\|S_{it}\|^2\right\}\left\{T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left\|\hat{u}_{it}\hat{b}_{it} - u_{it}\bar{b}_{it}\right\|^2\right\} = o_P\left(1\right)o_P\left(1\right) = o_P\left(1\right), \end{aligned}$$

as one can readily show that $T^{-1}N^{-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left\|\hat{u}_{it}\hat{b}_{it} - u_{it}\bar{b}_{it}\right\|^2 = o_P\left(1\right)$. Thus $V_{NT,11} = o_P\left(1\right)$. This completes the proof of (ii). ■

Proof of Theorem 3.3. Observe that $\sqrt{\hat{V}_{NT}(K_0)}\hat{J}_{NT}(K_0) = A_{1NT} - \hat{B}_{NT}(K_0) + A_{2NT} + 2A_{3NT}$, where A_{1NT}, A_{2NT}, and A_{3NT} are as defined in (A.2). We study the probability order of each term in the last expression. Noting that $\|X_i'M_0u_i\|^2 \le 2\|X_i'u_i\|^2 + 2\|X_i'P_0u_i\|^2$, we have $N^{-1}T^{-2}\sum_{i=1}^N \|X_i'M_0u_i\|^2 \le 2a_1 + 2a_2$, where $a_1 = 2N^{-1}T^{-2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T u_{it}X_{it}'X_{is}u_{is}$ and $a_2 = N^{-1}T^{-4}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T\sum_{r=1}^T \sum_{q=1}^T u_{it}X_{is}'X_{ir}u_{iq}$. By Assumption A.1 and Markov inequality, we can readily show that $a_1 = O_P(T^{-1})$ and $a_2 = O_P(T^{-1})$. It follows that $N^{-1}T^{-2}\sum_{i=1}^N \|X_i'M_0u_i\|^2 = O_P(T^{-1})$. Then by Lemma B.3(v),

$$\begin{aligned}
N^{-1/2}T^{-1}A_{1NT} &= N^{-1}T^{-1}\sum_{i=1}^N u_i'M_0\bar{P}_{X_i}M_0u_i \le \max_{1\le i\le N}\lambda_{\max}\left(\hat{Q}_i\right)N^{-1}T^{-2}\sum_{i=1}^N u_i'M_0X_iX_i'M_0u_i \\
&= O_P(1)\,O_P\left(T^{-1}\right) = O_P\left(T^{-1}\right).
\end{aligned}$$

By (A.1) and Cauchy-Schwarz inequality

$$\begin{aligned}
N^{-1/2}T^{-1}\hat{B}_{NT} &= N^{-1}T^{-1}\sum_{i=1}^N\sum_{t=1}^T \hat{u}_{it}^2 h_{i,tt} = N^{-1}T^{-1}\sum_{i=1}^N \hat{u}_i'\text{diag}(H_i)\,\hat{u}_i \\
&\le 2N^{-1}T^{-1}\sum_{i=1}^N u_i'M_0\text{diag}(H_i)\,M_0u_i, \\
&\quad +2N^{-1}T^{-1}\sum_{i=1}^N(\beta_i^0 - \hat{\beta}_i)'X_i'M_0\text{diag}(H_i)\,M_0X_i(\beta_i^0 - \hat{\beta}_i) \\
&\equiv 2b_1 + 2b_2, \text{ say.}
\end{aligned}$$

By Cauchy-Schwarz inequality, $b_1 \le 2N^{-1}T^{-1}\sum_{i=1}^N u_i'\text{diag}(H_i)\,u_i + 2N^{-1}T^{-1}\sum_{i=1}^N u_i'P_0\text{diag}(H_i)\,P_0u_i \equiv 2b_{1,1} + 2b_{1,2}$, say. By the fact $H_i = M_0\bar{P}_{X_i}M_0 \le [\lambda_{\min}(\hat{Q}_i)]^{-1}T^{-1}M_0X_iX_i'M_0$ and Lemma B.3(v), we have

$$\begin{aligned}
b_{1,1} &\le [\lambda_{\min}(\hat{Q}_i)]^{-1}N^{-1}T^{-2}\sum_{i=1}^N u_i'\text{diag}(M_0X_iX_i'M_0)\,u_i \\
&= [\lambda_{\min}(\hat{Q}_i)]^{-1}N^{-1}T^{-2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T\sum_{r=1}^T u_{it}\eta_{ts}X_{is}'X_{ir}\eta_{tr}u_{ir} = O_P\left(T^{-1}\right),
\end{aligned}$$

as we can readily show that $N^{-1}T^{-2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T\sum_{r=1}^T u_{it}\eta_{ts}X_{is}'X_{ir}\eta_{tr}u_{ir} = O_P\left(T^{-1}\right)$ based on the fact that $\eta_{ts} = \mathbf{1}\{t=s\} - T^{-1}$ and Markov inequality. As in the analysis of $\hat{B}_{NT,11}$, we can readily apply the fact that $\mathbf{i}_T'\text{diag}(H_i)\,\mathbf{i}_T = p$ to obtain

$$\begin{aligned}
b_{1,2} &= N^{-1}T^{-3}\sum_{i=1}^N u_i'\mathbf{i}_T\mathbf{i}_T'\text{diag}(H_i)\,\mathbf{i}_T\mathbf{i}_T'u_i = pN^{-1}T^{-3}\sum_{i=1}^N u_i'\mathbf{i}_T\mathbf{i}_T'u_i \\
&= pN^{-1}T^{-3}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T u_{it}u_{is} = O_P\left(T^{-2}\right).
\end{aligned}$$

It follows that $b_1 = O_P\left(T^{-1}\right)$. By Cauchy-Schwarz inequality, $b_2 \le 2N^{-1}T^{-1}\sum_{i=1}^N(\beta_i^0 - \hat{\beta}_i)'X_i'\text{diag}(H_i)\,X_i(\beta_i^0 - \hat{\beta}_i) + 2N^{-1}T^{-1}\sum_{i=1}^N(\beta_i^0 - \hat{\beta}_i)'X_i'P_0\text{diag}(H_i)\,P_0X_i(\beta_i^0 - \hat{\beta}_i) \equiv 2b_{2,1} + 2b_{2,2}$, say. Noting that $\mathrm{diag}(M_0X_iX_i'M_0) \leq \mathrm{diag}(X_iX_i')$, we have

$$\begin{aligned}
b_{2,1} &\leq [\lambda_{\min}(\hat{Q}_i)]^{-1}N^{-1}T^{-2}\sum_{i=1}^{N}(\beta_i^0-\hat{\beta}_i)'X_i'\mathrm{diag}\left(M_0X_iX_i'M_0\right)X_i(\beta_i^0-\hat{\beta}_i) \\
&\leq [\lambda_{\min}(\hat{Q}_i)]^{-1}N^{-1}T^{-2}\sum_{i=1}^{N}(\beta_i^0-\hat{\beta}_i)'X_i'\mathrm{diag}\left(X_iX_i'\right)X_i(\beta_i^0-\hat{\beta}_i) \\
&= [\lambda_{\min}(\hat{Q}_i)]^{-1}N^{-1}T^{-2}\sum_{i=1}^{N}(\beta_i^0-\hat{\beta}_i)'X_{it}X_{it}'X_{it}X_{it}'(\beta_i^0-\hat{\beta}_i) \\
&\leq T^{-1}[\lambda_{\min}(\hat{Q}_i)]^{-1}\max_{1\leq i\leq N}T^{-1}\sum_{t=1}^{T}\|X_{it}\|^4 N^{-1}\sum_{i=1}^{N}\left\|\beta_i^0-\hat{\beta}_i\right\|^2 \\
&= T^{-1}O_P(1)\,O_P(1)\,O_P(1)=O_P\left(T^{-1}\right).
\end{aligned}$$

Using $\mathbf{i}_T'\mathrm{diag}(H_i)\,\mathbf{i}_T=p$, we have

$$\begin{aligned}
b_{2,2} &= N^{-1}T^{-3}\sum_{i=1}^{N}(\beta_i^0-\hat{\beta}_i)'X_i'\mathbf{i}_T\mathbf{i}_T'\mathrm{diag}\left(H_i\right)\mathbf{i}_T\mathbf{i}_T'X_i(\beta_i^0-\hat{\beta}_i) \\
&= pN^{-1}T^{-3}\sum_{i=1}^{N}(\beta_i^0-\hat{\beta}_i)X_i'\mathbf{i}_T\mathbf{i}_T'X_i(\beta_i^0-\hat{\beta}_i) \\
&\leq \max_{1\leq i\leq N}\left\|\bar{X}_i\right\|^2 N^{-1}T^{-1}\sum_{i=1}^{N}\left\|\beta_i^0-\hat{\beta}_i\right\|^2=O_P\left(T^{-1}\right).
\end{aligned}$$

It follows that $b_2=O_P\left(T^{-1}\right)$ and $N^{-1/2}T^{-1}\hat{B}_{NT}\left(K_0\right)=O_P\left(T^{-1}\right)$.

By Assumption A.4(ii) and Lemma B.3(v), w.p.a.1

$$\begin{aligned}
N^{-1/2}T^{-1}A_{2NT} &= N^{-1}T^{-1}\sum_{i=1}^{N}\left(\beta_i^0-\hat{\beta}_i\right)'X_i'M_0X_i\left(\beta_i^0-\hat{\beta}_i\right) \\
&\geq \lambda_{\min}\left(\hat{Q}_i\right)N^{-1}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k}\left\|\beta_i^0-\hat{\alpha}_k\right\|^2\geq\frac{1}{2}\lambda_{\min}\left(Q_i\right)\underline{c}_{K_0}.
\end{aligned}$$

Now, we decompose $N^{-1/2}T^{-1}A_{3NT}$ as follows: $N^{-1/2}T^{-1}A_{3NT}=N^{-1}T^{-1}\sum_{i=1}^{N}u_i'X_i(\beta_i^0-\hat{\beta}_i)-N^{-1}T^{-2}\sum_{i=1}^{N}u_i'\mathbf{i}_T\mathbf{i}_T'X_i(\beta_i^0-\hat{\beta}_i)\equiv A_{3NT,1}+A_{3NT,2}$, say. For the first term, we have

$$\begin{aligned}
|A_{3NT,1}| &\leq \max_{1\leq i\leq N}\left\|T^{-1}X_i'u_i\right\|N^{-1}\sum_{i=1}^{N}\left\|\beta_i^0-\hat{\beta}_i\right\| \\
&\leq \max_{1\leq i\leq N}\left\|T^{-1}X_i'u_i\right\|\left\{N^{-1}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k}\left\|\beta_i^0-\hat{\alpha}_k\right\|^2\right\}^{1/2} \\
&= O_P\left(\alpha_{NT}\right)O_P\left(1\right)=o_P\left(1\right),
\end{aligned}$$

where we use the fact that $\max_{1\le i\le N}\left\|T^{-1}X_i'u_i\right\| = O_P(\alpha_{NT})$ by using similar arguments as those used in the proof of Lemma B.3(iii). Similarly,

$$
\begin{aligned}
|A_{3NT,2}| &\le \max_{1\le i\le N}\left\|T^{-1}u_i'\mathbf{i}_T\right\|\max_{1\le i\le N}\left\|T^{-1}X_i'\mathbf{i}_T\right\| N^{-1}\sum_{i=1}^{N}\left\|\beta_i^0-\hat{\beta}_i\right\| \\
&\le \max_{1\le i\le N}\left\|T^{-1}u_i'\mathbf{i}_T\right\|\max_{1\le i\le N}\left\|T^{-1}X_i'\mathbf{i}_T\right\|\left\{N^{-1}\sum_{k=1}^{K_0}\sum_{i\in\hat{G}_k}\left\|\beta_i^0-\hat{\alpha}_k\right\|^2\right\}^{1/2} \\
&= O_P(\alpha_{NT})\,O_P(1)\,O_P(1) = o_P(1).
\end{aligned}
$$

It follows that $N^{-1/2}T^{-1}A_{3NT} = o_P(1)$.

In sum, we have $N^{-1/2}T^{-1}\sqrt{\hat{V}_{NT}(K_0)}\hat{J}_{NT}(K_0) \ge \frac{1}{2}\lambda_{\min}(Q_i)\,\underline{c}_{K_0} + o_P(1)$ w.p.a.1. In addition, we can show that $\hat{V}_{NT}(K_0)$ has a positive probability limit under $\mathbb{H}_1(K_0)$. It follows that under $\mathbb{H}_1(K_0)$, $P(\hat{J}_{NT}(K_0) \ge c_{NT}) \to 1$ as $(N,T)\to\infty$ for any $c_{NT} = o\left(N^{1/2}T\right)$. ■

Proof of Lemma 3.4. Observe that

$$
\begin{aligned}
\alpha^*(K_0) &\equiv \lim_{(N,T)\to\infty} P\left(\text{Case } (i) \text{ or Case } (ii) \text{ occurs} \mid \mathbb{H}_0(K_0)\right) \\
&\le \lim_{(N,T)\to\infty} P\left(\text{Case } (i) \text{ occurs} \mid \mathbb{H}_0(K_0)\right) + \lim_{(N,T)\to\infty} P\left(\text{Case } (ii) \text{ occurs} \mid \mathbb{H}_0(K_0)\right) \\
&= \lim_{(N,T)\to\infty} P\left(\text{Case } (ii) \text{ occurs} \mid \mathbb{H}_0(K_0)\right) \\
&\le \lim_{(N,T)\to\infty} P\left(\text{Reject } \mathbb{H}_0(K_0) \mid \mathbb{H}_0(K_0)\right) \equiv \alpha(K_0),
\end{aligned}
$$

where the equality follows from the fact that $\lim_{(N,T)\to\infty} P\left(\text{Reject } \mathbb{H}_0(k) \mid \mathbb{H}_0(K_0)\right) = 1$ for all $k < K_0$ by applying Theorem 3.3 to the case of testing $\mathbb{H}_0(k)$ with $k < K_0$. On the other hand,

$$
\begin{aligned}
\alpha^*(K_0) &\equiv \lim_{(N,T)\to\infty} P\left(\text{Case } (i) \text{ or Case } (ii) \text{ occurs} \mid \mathbb{H}_0(K_0)\right) \\
&\ge \lim_{(N,T)\to\infty} P\left(\text{Case } (ii) \text{ occurs} \mid \mathbb{H}_0(K_0)\right) \\
&\ge 1 - \lim_{(N,T)\to\infty} \sum_{k=1}^{K_0} P\left(\text{Fail to reject } \mathbb{H}_0(k) \mid \mathbb{H}_0(K_0)\right) \\
&= 1 - \lim_{(N,T)\to\infty} P\left(\text{Fail to reject } \mathbb{H}_0(K_0) \mid \mathbb{H}_0(K_0)\right) \\
&= \lim_{(N,T)\to\infty} P\left(\text{reject } \mathbb{H}_0(K_0) \mid \mathbb{H}_0(K_0)\right) \equiv \alpha(K_0),
\end{aligned}
$$

where the first equality follows again from the fact that $\lim_{(N,T)\to\infty} P\left(\text{Reject } \mathbb{H}_0(k) \mid \mathbb{H}_0(K_0)\right) = 1$ for all $k < K_0$. Combining the two results above yields $\alpha^*(K_0) = \alpha(K_0)$. ■

B Some technical lemmas

Define the mth order U-statistic $\mathcal{U}_T = \begin{pmatrix} T \\ m \end{pmatrix}^{-1} \sum_{1\le t_1<...<t_m\le T} \vartheta\left(\xi_{t_1},...,\xi_{t_m}\right)$ where ϑ is symmetric in its arguments. Let $F_t(\cdot)$ denote the distribution function of ξ_t. Let $\vartheta_{(0)} = \int\cdots\int \vartheta\left(v_{t_1},...,v_{t_m}\right) \Pi_{s=1}^m dF_{t_s}\left(v_{t_s}\right)$, and $\vartheta_{(c)}\left(v_1,...,v_c\right) = \int\cdots\int \vartheta(v_1,...,v_c,\ v_{t_{c+1}},...,v_{t_m})\ \Pi_{s=c+1}^m dF_{t_s}\left(v_{t_s}\right)$ for $c=1,...,m$. Let $h^{(1)}(v) = \vartheta_{(1)}(v) - \vartheta_{(0)}$, and $h^{(c)}\left(v_1,...,v_c\right) = \vartheta_{(c)}\left(v_1,...,v_c\right) - \sum_{j=1}^{c-1}\sum_{(c,j)} h^{(j)}\left(v_{t_1},...,v_{t_j}\right) - \vartheta_{(0)}$ for $c=2,...,m$, where the sum $\sum_{(c,j)}$ is taken over all subsets $1\le t_1<t_2<\cdots<t_j\le c$ of $\{1,2,...,c\}$. Let $\mathcal{H}_T^{(c)} = \begin{pmatrix} T \\ c \end{pmatrix}^{-1} \sum_{1\le t_1<...<t_c\le T} h^{(c)}\left(\xi_{t_1},...,\xi_{t_c}\right)$. Then by Theorem 1 in Lee (1990, p. 26), we have the following Hoeffding decomposition:

$$\mathcal{U}_T = \vartheta_{(0)} + \sum_{c=1}^m \begin{pmatrix} m \\ c \end{pmatrix} \mathcal{H}_T^{(c)}. \tag{B.1}$$

To study the second moment of $\mathcal{H}_T^{(c)}$ for $3\le c\le m$, we need the following lemma.

Lemma B.1 *Let $\{\xi_t, t\ge 1\}$ be an l-dimensional strong mixing process with mixing coefficient $\alpha(\cdot)$ and distribution function $F_t(\cdot)$. Let the integers $(t_1,...,t_m)$ be such that $1\le t_1<t_2<\cdots<t_m\le T$. Suppose that $\max\{\int\left|\vartheta\left(v_1,...,v_m\right)\right|^{1+\tilde{\sigma}} dF_{t_1,...,t_m}\left(v_1,...,v_m\right), \int\left|\vartheta\left(v_1,...,v_m\right)\right|^{1+\tilde{\sigma}} dF_{t_1,...,t_j}(v_1,...,v_j)$ $dF_{t_{j+1},...,t_m}\left(v_{j+1},...,v_m\right)\}\le C$ for some $\tilde{\sigma}>0$, where, e.g., $F_{t_1,...,t_m}\left(v_1,...,v_m\right)$ denotes the distribution function of $\left(\xi_{t_1},...,\xi_{t_m}\right)$. Then*

$$\begin{aligned} &\left|\int \vartheta\left(v_1,...,v_m\right) dF_{t_1,...,t_m}\left(v_1,...,v_m\right) - \int \vartheta\left(v_1,...,v_m\right) dF^{(1)}_{t_1,...,t_j}\left(v_1,...,v_j\right) dF_{t_{j+1},...,t_m}\left(v_{j+1},...,v_m\right)\right| \\ \le\ & 4C^{1/(1+\tilde{\sigma})}\alpha\left(t_{j+1}-t_j\right)^{\tilde{\sigma}/(1+\tilde{\sigma})}. \end{aligned}$$

Proof. See Lemma 2.1 in Sun and Chiang (1997). ■

Lemma B.2 *Let $\{\xi_t, t\ge 1\}$ be an l-dimensional strong mixing process with mixing coefficient $\alpha(\cdot)$ and distribution function $F_t(\cdot)$. Suppose that $\alpha(s) = O\left(s^{-3(2+\sigma)/\sigma-\epsilon}\right)$. If there exists $\sigma>0$ such that*

$$L_T \equiv \max\left\{\int\left|\vartheta\left(v_{t_1},\cdots,v_{t_m}\right)\right|^{2+\sigma}\ \Pi_{s=1}^m dF_{t_s}\left(v_{t_s}\right),\ E\left|\vartheta\left(\xi_{t_1},...,\xi_{t_m}\right)\right|^{2+\sigma}\right\} \le \sum_{q=1}^m C_q\left(t_q\right),$$

and $T^{-1}\sum_{q=1}^m\sum_{t_q=1}^T C_q\left(t_q\right) = O(1)$, then $E[\mathcal{H}_T^{(c)}]^2 = O_P\left(T^{-3}\right)$ for $3\le c\le m$.

Proof. The proof is analogous to that of Lemma A.6 in Su and Chen (2013) who consider conditional strong mixing processes instead. ■

Lemma B.3 *Recall $\hat{\Omega}_i \equiv X_i' M_0 X_i/T$ and $\Omega_i \equiv E(\hat{\Omega}_i)$. Let $\hat{\Omega}_{1i} \equiv X_i' X_i/T$ and $\Omega_{1i} \equiv E(\hat{\Omega}_{1i})$. Suppose Assumptions A.1-A.3 hold. Then (i) $\lambda_{\max}(\hat{\Omega}_{1i}) \le \lambda_{\max}\left(\Omega_{1i}\right) + O_P\left(T^{-1/2}\right)$, (ii) $\lambda_{\min}(\hat{\Omega}_{1i}) \ge \mu_{\min}\left(\Omega_{1i}\right) - O_P\left(T^{-1/2}\right)$, (iii) $\max_{1\le i\le N}\left\|\hat{\Omega}_{1i}-\Omega_{1i}\right\| = O_P\left(a_{NT}\right)$, (iv) $\max_{1\le i\le N}\left\|\hat{\Omega}_{1i}^{-1}-\Omega_{1i}^{-1}\right\| =$*

$O_P(a_{NT})$, *(v)* $\max_{1\le i\le N}\left\|\hat{\Omega}_i-\Omega_i\right\| = O_P(a_{NT})$ *and* $\max_{1\le i\le N}\left\|\hat{\Omega}_i^{-1}-\Omega_i^{-1}\right\| = O_P(a_{NT})$, *where* $a_{NT}\equiv\max\{(NT)^{1/(4+2\sigma)}\log(NT)/T,\ (\log(NT)/T)^{1/2}\}$.

Proof. The results in (i)-(ii) follow from Lemmas A.1(iv)-(v) in Su and Jin (2012). Su and Chen (2013, Lemma A.7) prove (iii) for the conditional strong mixing process. The result also holds for strong mixing processes with a simple application of the Bernstein-type inequality for strong mixing processes (see, e.g., Lemma 2.2 in Sun and Chiang (1997)). (iv) follows from (i)-(iii) and the submultiplicative property of the Frobenius norm.

Now we show (v). Using $M_0=I_T-T^{-1}\mathbf{i}_T\mathbf{i}_T'$, we can decompose $\hat{\Omega}_i-\Omega_i$ as follows:

$$
\begin{aligned}
\hat{\Omega}_i-\Omega_i &= T^{-1}\left[X_i'M_0X_i-E\left(X_i'M_0X_i\right)\right]\\
&= \left(\hat{\Omega}_{1i}-\Omega_{1i}\right)-\bar{X}_i\bar{X}_i'+E\left(\bar{X}_i\bar{X}_i'\right)\\
&= \left(\hat{\Omega}_{1i}-\Omega_{1i}\right)-\left[\bar{X}_i-E\left(\bar{X}_i\right)\right]\left[\bar{X}_i-E\left(\bar{X}_i\right)\right]'-\left[\bar{X}_i-E\left(\bar{X}_i\right)\right]E\left(\bar{X}_i'\right)\\
&\quad -E\left(\bar{X}_i\right)\left[\bar{X}_i-E\left(\bar{X}_i\right)\right]'+\left[E\left(\bar{X}_i\bar{X}_i'\right)-E\left(\bar{X}_i\right)E\left(\bar{X}_i'\right)\right].
\end{aligned}
$$

Following the proof of (iii), we can show that $\max_{1\le i\le N}\left\|\bar{X}_i-E\left(\bar{X}_i\right)\right\|=O_P(a_{1NT})$, where $a_{1NT}\equiv\max\{(NT)^{1/(8+4\sigma)}\log(NT)/T,(\log(NT)/T)^{1/2}\}=O(a_{NT})$. $\max_{1\le i\le N}\left\|E\left(\bar{X}_i\right)\right\|=O(1)$ by Assumption A.1(i). Let $b_{i,kl}$ denote the (k,l)th element of $E\left(\bar{X}_i\bar{X}_i'\right)-E\left(\bar{X}_i\right)E\left(\bar{X}_i'\right)$ for $k,l=1,...,p$. Then by triangle inequality, Davydov inequality, and Assumption A.2(iii),

$$
\begin{aligned}
\left|b_{i,kl}\right| &= \frac{1}{T^2}\left|\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{cov}\left(X_{it,k},X_{is,l}\right)\right|\le\frac{1}{T^2}\sum_{t=1}^{T}\left|\mathrm{cov}\left(X_{it,k},X_{it,l}\right)\right|+\frac{1}{T^2}\sum_{1\le t\ne s\le T}\left|\mathrm{cov}\left(X_{it,k},X_{is,l}\right)\right|\\
&\le O\left(T^{-1}\right)+\frac{8c_kc_l}{T}\sum_{\tau=1}^{\infty}\alpha(\tau)^{(3+2\sigma)/(4+2\sigma)}=O\left(T^{-1}\right),
\end{aligned}
$$

where $c_k\le\sup_{T,N\ge1}\max_{1\le i\le N}\max_{1\le t\le T}\left\|X_{it,k}\right\|_{8+4\sigma}$. Then by the triangle inequality, we have $\max_{1\le i\le N}\left\|\hat{\Omega}_i-\Omega_i\right\|=O_P(a_{NT})$. (vi) follows from (v) and Assumption A.1(ii). ■

Lemma B.4 *Let* $h_{i,ts}$ *and* $\bar{h}_{i,ts}$ *be as defined in the proof of Theorem 3.1. Suppose Assumptions A.1-A.3 hold. Then*

(i) $D_{1NT}\equiv N^{-1/2}\sum_{i=1}^{N}\sum_{1\le s\ne t\le T}u_{it}u_{is}\left(h_{i,ts}-\bar{h}_{i,ts}\right)=o_P(1)$,

(ii) $D_{2NT}\equiv T^{-2}N^{-1/2}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}u_{it}u_{is}\left[X_{ir}-E\left(X_{ir}\right)\right]'\Omega_i^{-1}X_{is}=o_P(1)$,

(iii) $D_{3NT}\equiv T^{-2}N^{-1/2}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{q=1}^{T}u_{it}u_{is}X_{it}'\Omega_i^{-1}\left[X_{is}-E\left(X_{iq}\right)\right]=o_P(1)$,

(iv) $D_{4NT}\equiv T^{-3}N^{-1/2}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\sum_{q=1}^{T}u_{it}u_{is}\left[X_{ir}-E\left(X_{ir}\right)\right]'\Omega_i^{-1}\left[X_{iq}-E\left(X_{iq}\right)\right]=o_P(1)$,

(v) $D_{5NT}\equiv T^{-3}N^{-1/2}\sum_{i=1}^{N}\sum_{1\le s<t\le T}\sum_{r=1}^{T}\sum_{q=1}^{T}u_{it}u_{is}\left[X_{ir}-E\left(X_{ir}\right)\right]'\Omega_i^{-1}E\left(X_{iq}\right)=o_P(1)$.

Proof. The proof of (i) is analogous to that of Lemma A.8 in Su and Chen (2013) except that we replace their Lemmas A.5-A.7 by Lemmas B.1-B.3. To show (ii), letting $c_{i,ts}\equiv\left[X_{it}-E\left(X_{it}\right)\right]'\Omega_i^{-1}X_{is}$, we can decompose D_{2NT} as follows

$$\begin{aligned} D_{2NT} &= \frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{1\leq s<t\leq T}\sum_{r=1,r\neq t,s}^{T}u_{it}u_{is}c_{i,rs}+\frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{1\leq s<t\leq T}u_{it}u_{is}c_{i,ts} \\ &\quad +\frac{1}{T^2\sqrt{N}}\sum_{i=1}^{N}\sum_{1\leq s<t\leq T}u_{it}u_{is}c_{i,ss} \\ &\equiv D_{2NT,1}+D_{2NT,2}+D_{2NT,3}, \text{ say.} \end{aligned}$$

Let $\xi_{it}=(u_{it},X_{it}')'$, $\varphi_0(\xi_{it},\xi_{is},\xi_{ir})=u_{it}u_{is}c_{i,rs}$, and $\varphi(\xi_{it},\xi_{is},\xi_{ir})=[\varphi_0(\xi_{it},\xi_{is},\xi_{ir})+\varphi_0(\xi_{it},\xi_{ir},\xi_{is})$ $+\varphi_0(\xi_{is},\xi_{it},\xi_{ir})+\varphi_0(\xi_{is},\xi_{ir},\xi_{it})+\varphi_0(\xi_{ir},\xi_{it},\xi_{is})+\varphi_0(\xi_{ir},\xi_{is},\xi_{it})]/6$. Let $d_{iNT}\equiv\binom{T}{3}^{-1}\sum_{1\leq r<s<t\leq T}$ $\varphi(\xi_{it},\xi_{is},\xi_{ir})$. Then $D_{2NT,1}=\frac{a_T}{\sqrt{N}}\sum_{i=1}^{N}d_{iNT}$, where $a_T=\frac{(T-1)(T-2)}{2T}$. By Assumption A.1 and Lemma B.2, $E\left(D_{2NT,1}^2\right)=\frac{a_T^2}{N}\sum_{i=1}^{N}E\left(d_{iNT}^2\right)=a_T^2O_P\left(T^{-3}\right)=O_P\left(T^{-1}\right)$. It follows that $D_{2NT,1}=O_P\left(T^{-1/2}\right)$. Noting that $E_{\mathcal{D}}\left(D_{2NT,2}\right)=0$ and $E\left(D_{2NT,2}^2\right)=\frac{1}{T^4N}\sum_{i=1}^{N}\sum_{1\leq s,r<t\leq T}E(u_{it}^2u_{is}u_{ir}c_{i,ts}c_{i,tr})=O_P\left(T^{-1}\right)$, we have $D_{2NT,2}=O_P\left(T^{-1/2}\right)$. Similarly, $D_{2NT,3}=O_P\left(T^{-1/2}\right)$. Then (ii) follows. The proof of (iii)-(v) is analogous to that of (ii) and thus omitted. ∎

REFERENCES

Acemoglu, D., Johnson, S., and Robinson, J. (2005), "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth," *American Economic Review* 95, 546-79.

Acemoglu, D., Johnson, S., Robinson, J., and Yared, P. (2008), "Income and Democracy," *American Economic Review* 98, 808-842.

Barro, R. J. (1999), "Determinants of Democracy," *Journal of Political Economy* 107, 158-183.

Bester, C. A., and Hansen, C. B. (2013), "Grouped Effects Estimators in Fixed Effects Models," *Journal of Econometrics*, forthcoming.

Bonhomme, S., and Manresa, E. (2012), "Grouped Patterns of Heterogeneity in Panel Data," *Working paper*, CEMFI, Madrid.

Deb, P., and Trivedi, P. K., (2013), "Finite Mixture for Panels with Fixed Effects," *Journal of Econometric Methods* 2, 35-51.

Hahn, J., and Kuersteiner, G. (2011), "Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," *Econometric Theory* 27, 1152-1191.

Hahn, J., and Moon H. R. (2010), "Panel Data Models with Finite Number of Multiple Equilibria," *Econometric Theory* 26, 863-881.

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

Hochberg, Y. (1988), "A Sharper Bonferroni Procedure for Multiple Tests of Significance," *Biometrika* 75, 800-802.

Holm, S. (1979), "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics* 6, 65-70.

Hsiao, C. (2003), *Analysis of Panel Data*. Cambridge University Press, Cambridge.

Lee, A. J. (1990), *U-statistics: Theory and Practice*. Marcel Dekker, New York.

Lin, C-C., and Ng, S. (2012), "Estimation of Panel Data Models with Parameter Heterogeneity When Group Membership Is Unknown," *Journal of Econometric Methods* 1, 42-55.

Lipset, S. M. (1959), "Some Social Requisites of Democracy: Economic Development and Political Legitimacy," *American Political Science Review* 53, 69-105.

Onatski, A. (2009), "Testing Hypotheses about the Number of Factors in Large Factor Models," *Econometrica* 77, 1447-1479.

Pesaran, M. H., Smith, R., and Im K. S., (1996), "Dynamic Linear Models for Heterogeneous Panels,". in: L. Matyas and P. Sevestre (Eds.), *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, second revised edition, pp.145-195, Kluwer Academic Publishers, Dordrecht.

Pesaran, M. H., and Yamagata, T. (2008), "Testing Slope Homogeneity in Large Panels," *Journal of Econometrics* 142, 50-93.

Phillips, P. C. B., and Sul, D. (2003), "Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence," *Econometrics Journal* 6, 217—259.

Pollard, D. (1984), *Convergence of Stochastic Processes*. Springer-Verlag, New York.

Romano, J., and Wolf, M. (2005a), "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica* 73, 1237-1282.

Romano, J., and Wolf, M. (2005b), "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing," *Journal of the American Statistical Association* 100, 94-108.

Sarafidis, V., and Weber, N. (2011), "A Partially Heterogeneous Framework for Analyzing Panel Data," *Working paper*, University of Sydney.

Su, L., and Chen, Q. (2013), "Testing Homogeneity in Panel Data Models with Interactive Fixed Effects," *Econometric Theory* 29, 1079-1135.

Su, L., and Jin, S. (2012), "Sieve Estimation of Panel Data Models with Cross Section Dependence," *Journal of Econometrics* 169, 34-4

Su, L., Shi, Z., and Phillips, P. C. B. (2013), "Identifying Latent Structures in Panel," *Working paper*, Dept. of Economics, Yale University.

Sun, S., and Chiang, C-Y. (1997), "Limiting behavior of the perturbed empirical distribution functions evaluated at U-statistics for strongly mixing sequences of random variables," *Journal of Applied Mathematics and Stochastic Analysis* 10, 3-20.

Sun, Y. (2005), "Estimation and Inference in Panel Structure Models," *Working paper*, Dept. of Economics, UCSD.