PANEL DATA MODELS WITH TIME-VARYING
LATENT GROUP STRUCTURES

By

Yiren Wang, Peter C. B. Phillips and Liangjun Su

COWLES FOUNDATION PAPER NO. 1864

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Journal of Econometrics 240 (2024) 105685

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Panel data models with time-varying latent group structures[☆]

Yiren Wang [a], Peter C.B. Phillips [b,c,d], Liangjun Su [e,*]

[a] College of Finance and Statistics, Hunan University, China
[b] Yale University, United States of America
[c] University of Auckland, New Zealand
[d] Singapore Management University, Singapore
[e] School of Economics and Management, Tsinghua University, China

A R T I C L E I N F O

JEL classification:
C23
C33
C38
C51

Keywords:
Interactive fixed effects
Latent group structure
Structural break
Nuclear norm regularization
Sequential testing K-means algorithm

A B S T R A C T

This paper considers a linear panel model with interactive fixed effects and unobserved individual and time heterogeneities that are captured by some latent group structures and an unknown structural break, respectively. To enhance realism, the model may have different numbers of groups and/or different group memberships before and after the break. With preliminary nuclear norm regularized estimation followed by row- and column-wise linear regressions, we estimate the break point based on the idea of binary segmentation and the latent group structures together with the number of groups before and after the break by sequential testing K-means algorithm simultaneously. It is shown that the break point, the number of groups and the group memberships can each be estimated correctly with probability approaching one. Asymptotic distributions of the estimators of the slope coefficients are established. Monte Carlo simulations demonstrate excellent finite sample performance for the proposed estimation algorithm. An empirical application to real house price data across 377 Metropolitan Statistical Areas in the US from 1975 to 2014 suggests the presence both of structural breaks and of changes in group membership.

1. Introduction

Heterogeneous panel data models have been widely used in empirical research in economics because they can capture a rich degree of unobserved heterogeneity. But models with complete heterogeneity along either the cross-section or time dimension tend to possess too many parameters to be identified, which results in slow convergence and inefficient estimates. For this reason, researchers now frequently advocate the use of panel data models with certain structures imposed along either the cross-section or time dimension. On the one hand, the recent burgeoning of panels with latent group structures can be motivated from the observation that different groups of individuals respond differently to exogenous shocks. For instance, Durlauf and Johnson (1995), Berthelemy and Varoudakis (1996), and Ben-David (1998) show economies in different groups of income per capita and/or education level may converge to different steady state equilibria. Klapper and Love (2011), Chu (2012), and Zhang and Cheng (2019) show an exogenous shock like policy implementation has different impacts on different individuals, and Long et al. (2012) argue that the

☆ The authors thank the co-editor Michael Jansson, one associate editor and two anonymous referees for their many valuable and insightful comments on the paper. They also thank Jia Li, Jun Yu and Yichong Zhang for discussions on the subject matter. Phillips acknowledges support from a Lee Kong Chian Fellowship at SMU, the Kelly Fund at the University of Auckland, and the National Science Foundation under Grant No. SES 18-50860. Su gratefully acknowledges the support from the National Natural Science Foundation of China under Grant No. 72133002.

* Correspondence to: School of Economics and Management, Tsinghua University, Beijing, 100084, China.
E-mail address: sulj@sem.tsinghua.edu.cn (L. Su).

https://doi.org/10.1016/j.jeconom.2024.105685
Received 11 June 2023; Received in revised form 20 January 2024; Accepted 21 January 2024
Available online 28 January 2024

influence of the 2008 financial crisis on economic growth differs for emergent and developed economies. On the other hand, the recent popularity of panels that evidence structural change can be motivated by events such as financial crises, the economic impact of technological progress, and more general economic transitions that occur during the time periods covered by the data. See Qian and Su (2016) for a survey of panel data models and research that consider estimation and inference concerning structural change.

In spite of the large literatures that now separately study individual heterogeneity or time heterogeneity in the slope coefficients of panel models, few works consider both types of heterogeneity simultaneously. Exceptions include Keane and Neal (2020) and Lu and Su (2023) who consider linear panel data models with two-dimensional unobserved heterogeneity in the slope coefficients that are modeled via the usual additive structure, and Chernozhukov et al. (2020) and Wang et al. (2022) who model the slope coefficients via the use of low-rank matrices for conditional mean and quantile regressions, respectively. In addition, Okui and Wang (2021) and Lumsdaine et al. (2023) consider both individual heterogeneity and time heterogeneity by modeling them as a grouped pattern and as structural breaks, respectively. Specifically, Okui and Wang (2021) develop a new panel data model with latent groups where the number of groups and the group memberships do not change over time but the coefficients within each group can change over time and they may have different break-dates; Lumsdaine et al. (2023) consider the panels with a grouped pattern of heterogeneity when the latent group membership structure and/or the values of slope coefficients change at a break point. Both papers provide algorithms to recover the latent group structure based on linear panel models with or without individual fixed effects, but cannot allow for the presence of more complicated fixed effects such as interactive fixed effects (IFEs) to capture strong cross-sectional dependence in the data.

This paper proposes a linear panel data model with IFEs that enable the slope coefficients to exhibit two-way heterogeneity. Following the lead of Okui and Wang (2021) and Lumsdaine et al. (2023) and to encourage the parameter parsimony, we use a latent group structure to capture individual heterogeneity and an unknown structural break to capture time heterogeneity. The latent group structure of the model accommodates different group numbers and different group memberships before and after the break. Given this complicated structure, the approach proposed is to estimate the break point, the number of groups before and after the break, the group membership before and after the break, and the group-specific parameters in multiple steps. The key insight that permits this degree of complication is that the slope coefficients of each of the p regressors in the model are permitted to vary across both cross-section and time dimensions by means of a factor structure with a fixed number of factors so that they may be conveniently stacked into a low-rank matrix.

In the first step, the low-rank nature of the slope matrices is explored and initial estimates are obtained by nuclear norm regularization (NNR), a machine learning technique popular in computer science that is increasingly used in econometrics. Such initial matrix estimates are consistent in terms of the Frobenius norm but do not have pointwise or uniform convergence for their elements. Despite this, by applying singular value decomposition (SVD) to these estimates, we can obtain estimates of the associated factors and loadings that are also consistent in terms of the Frobenius norm. In the second step, we use the first-step initial estimates of the factors and loadings to run the row- and column-wise linear regressions to update the estimates of the factors and loadings which now possess pointwise and uniform consistency and can be used for subsequent analyses. In the third step, we estimate the break point by using the celebrated idea of binary segmentation, as commonly used for break point estimation in the time series literature. Once the break point is estimated, the full sample is naturally split into two subsamples. In the fourth step, we follow the lead of Lin and Ng (2012) and Jin et al. (2022) to focus on each subsample before and after the estimated break point and propose a sequential testing K-means algorithm to recover the latent group structure and obtain the number of groups simultaneously. In the last step, we use the estimated group structure to estimate the group-specific parameters. Asymptotic analyses show that the break point, the number of groups and the group memberships can be consistently estimated in Steps 3–4, so that the final-step estimates for the group-specific coefficients can enjoy the oracle property. This means they have the same asymptotic distributions as the ones obtained by knowing the break point and the latent group structures before and after the break point.

The present paper makes contributions that relate to two branches of literature. First, it contributes to the panel data literature on one-way heterogeneity, especially with either latent group structures or structural breaks. With respect to latent group structures, there are several popular ways to recover the latent groups. The first approach is the K-means algorithm. Lin and Ng (2012) apply the K-means algorithm to linear panel data models with grouped slope coefficients and propose an information criterion and a sequential testing approach to estimate the true number of groups. Sarafidis and Weber (2015) analyze the unknown grouped slopes in the large N and fixed T framework, and Zhang et al. (2019) provide an iterative algorithm based on K-means clustering for a panel quantile regression model. Bonhomme and Manresa (2015) and Ando and Bai (2016) consider panels with grouped fixed effects. The second approach is the Classifier-Lasso (C-Lasso) that has become a popular clustering method since (Su et al., 2016). This method is extended by Lu and Su (2017), Su and Ju (2018), Su et al. (2019), Wang et al. (2019), and Huang et al. (2020) to various contexts. In addition, both the clustering algorithm in regression via a data-driven segmentation (CARDS) approach and binary segmentation are also considered in Ke et al. (2015), Wang et al. (2018), Ke et al. (2016) and Wang and Su (2021), among others. As for panel models with structural breaks, binary segmentation has become a common approach to estimate the break point. See Bai (2010), Lin and Hsu (2011), Kim (2011, 2014) and Baltagi et al. (2017), among others. These papers focus on the case of a single break point in the model. In contrast, Qian and Su (2016) and Li et al. (2016) allow for multiple breaks in linear panel models with either classical fixed effects or IFEs, and propose an adaptive grouped fused lasso (AGFL) approach to estimate the break points. Compared to the existing panel literature on one-way heterogeneity, our model allows for two-way heterogeneity. In particular, not only are different membership structures in different time blocks permitted but also changes in the number of groups over time. As a result, our model is more flexible than all existing models that allow only for latent group structures or structural breaks, but not both.

Second, this paper contributes to the recent burgeoning literature that models two-way heterogeneity in the slope coefficients of a panel model. As mentioned above, there are two approaches to model two-way heterogeneity in the slope coefficients. One approach models them in an additive structure so that both individual and time effects enter the slope coefficients additively, as in Keane and Neal (2020) and Lu and Su (2023). The other approach imposes certain low-rank structures on the slope coefficient matrices in which case one models each slope coefficient via the use of IFEs to capture strong cross-sectional dependence in the panel. In view of the low-rank structures, we can resort to NNR estimation which has attracted increasing attention recently in panel data analyses. NNR has been used in recent econometric research — see Moon and Weidner (2018), Bai and Ng (2019), Chernozhukov et al. (2020), Belloni et al. (2023), Miao et al. (2023), Feng (2023), and Hong et al. (2023), among others. But none of these papers imposes any latent group structures on the slope coefficients. With latent group structures and structural breaks imposed, Okui and Wang (2021) allow the slope coefficients within each group to have common breaks and the break points to vary across different groups, and they propose to estimate the latent group structures, the structural breaks, and the group-specific regression parameters by the grouped adaptive group fused lasso (GAGFL). But neither the number of groups nor the group memberships is allowed to change over time in Okui and Wang (2021). In a companion paper, Lumsdaine et al. (2023) allow the latent group membership structure and/or the values of slope coefficients to change at a break point, and propose an estimation algorithm similar to the K-means of Bonhomme and Manresa (2015). Both Okui and Wang (2021) and Lumsdaine et al. (2023) allow for at most one-way heterogeneity (individual fixed effects) in the intercept and neither allows for IFEs to capture strong cross-section dependence. In contrast, this paper proposes an algorithm to detect the unknown break point and to recover the group structure based on linear panel model with IFEs, which involves a more general model. In addition, Lumsdaine et al. (2023) first assume the number of groups is known in the estimation algorithm and then estimate the number of groups via an information criterion but they do not establish consistency for such an estimate. Instead, we estimate the number of groups and group membership simultaneously by the sequential testing K-means algorithm and establish the consistency of the number of groups estimator.

The rest of the paper is organized as follows. We first introduce the linear panel model with time-varying latent group structures in Section 2 and provide the estimation algorithm in Section 3. The asymptotic properties are given in Section 4. In Section 5, we propose an alternative approach to detect the break point and discuss the potential extension. Sections Section 6 provides simulation evidence on the finite sample performance of our methods and Section 7 reports an empirical application to housing price data in the US that explores group membership and structural breaks. Section 8 concludes. All proofs are provided in the online supplement.

Notation. Let $\|\cdot\|_{\max}$, $\|\cdot\|_{op}$, $\|\cdot\|$, and $\|\cdot\|_*$ be the (elementwise) maximum norm, operator norm, Frobenius norm, and nuclear norm, respectively. Let $\odot$ denote the element-wise Hadamard product. $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor and ceiling functions, respectively. Let $a \vee b = \max(a,b)$ and $a \wedge b = \min(a,b)$. $a_n \lesssim b_n$ means $a_n/b_n = O_p(1)$ and $a_n \gg b_n$ means $b_n a_n^{-1} = o(1)$. Let $A = \{A_{it}\}$ be a matrix with its (i,t)-th entry denoted as A_{it}. Let $\{A_j\}_{j\in[p]\cup\{0\}}$ be the collection of matrices A_j, $j \in \{0,1,\ldots,p\}$. For a specific $A \in \mathbb{R}^{m\times n}$ with rank n, let $P_A = A(A'A)^{-1}A'$ and $M_A = I_m - P_A$. When A is symmetric, $\lambda_{\max}(A)$, $\lambda_{\min}(A)$ and $\lambda_n(A)$ denote its largest, smallest and nth largest eigenvalues, respectively. The operators $\rightsquigarrow$ and $\xrightarrow{p}$ denote convergence in distribution and in probability, respectively. Denote $[n] = \{1,\ldots,n\}$ for any positive integer n, and let $\mathbf{1}\{\cdot\}$ be the usual indicator function, w.p.a.1 and a.s. abbreviate "with probability approaching 1" and "almost surely", respectively.

2. Model setup

In this paper we consider the following linear panel model with IFEs:[1]

$$Y_{it} = \lambda_i^{0\prime} f_t^0 + X_{it}'\Theta_{it}^0 + e_{it}, \tag{1}$$

where $i \in [N]$, $t \in [T]$, Y_{it} is the dependent variable, $X_{it} = (X_{1,it},\ldots,X_{p,it})'$ is a $p\times 1$ vector of regressors, $\Theta_{it}^0 = (\Theta_{1,it}^0,\ldots,\Theta_{p,it}^0)'$ is a $p\times 1$ vector of slope coefficients, λ_i^0 and f_t^0 are individual and time fixed effects, and e_{it} is the error term. Let $\Theta_{0,it}^0 := \lambda_i^{0\prime} f_t^0$ denote the intercept term that exhibits a factor structure with r_0 factors. Here, we assume r_0 is a fixed integer that does not change as $(N,T)\to\infty$. Let $\Lambda^0 = \left(\lambda_1^0,\ldots,\lambda_N^0\right)'$ and $F^0 = \left(f_1^0,\ldots,f_T^0\right)'$. Moreover, let $Y = \{Y_{it}\}$, $X_j = \{X_{j,it}\}$, $\Theta_j^0 = \{\Theta_{j,it}^0\}$ and $E = \{e_{it}\}$, all of which are $N\times T$ matrices. Then we can rewrite (1) in matrix form as

$$Y = \Lambda^0 F^{0\prime} + \sum_{j=1}^{p} X_j \odot \Theta_j^0 + E = \Theta_0^0 + \sum_{j=1}^{p} X_j \odot \Theta_j^0 + E. \tag{2}$$

We assume that the slope coefficients follow time-varying latent group structures, viz.,

$$\Theta_{it}^0 = \sum_{k\in[K_t]} \alpha_{kt}\mathbf{1}\{i \in G_{kt}\},$$

where $\{G_{kt}\}_{k\in[K_t]}$ forms a partition of $[N]$ for each specific time t with K_t being the number of groups at time t. Moreover, we assume that the group-specific slope coefficients α_{kt} or the memberships change at an unknown time point T_1, i.e.,

$$\alpha_{kt} = \begin{cases} \alpha_k^{(1)}, & \text{for} \quad t = 1,\ldots,T_1,\ k = 1,\ldots,K^{(1)}, \\ \alpha_k^{(2)}, & \text{for} \quad t = T_1+1,\ldots,T,\ k = 1,\ldots,K^{(2)}, \end{cases}$$

[1] A referee expressed interest in specifying the intercept in terms of other forms of generalized fixed effects such as $g\left(\lambda_i^0, f_t^0\right)$ and $g_i\left(f_t^0\right)$, where the functional form of either $g(\cdot,\cdot)$ or the $g_i(\cdot)$ is unknown, $\left(\lambda_i^0, f_t^0\right)$ and the f_t^0 are unobserved in the first and second specifications, respectively. Such specifications would require different identification and estimation methods and are left for future study.

$$G_{kt} = \begin{cases} G_k^{(1)}, & \text{for} \quad t = 1, \dots, T_1,\ k = 1, \dots, K^{(1)}, \\ G_k^{(2)}, & \text{for} \quad t = T_1 + 1, \dots, T,\ k = 1, \dots, K^{(2)}, \end{cases}$$

with $K^{(1)}$ and $K^{(2)}$ being the number of latent groups before and after the break point, respectively. Let $g_i^{(1)}$ and $g_i^{(2)}$ respectively denote the individual group indices before and after the break:

$$g_i^{(1)} = \sum_{k\in[K^{(1)}]} k\mathbf{1}\{i \in G_k^{(1)}\} \quad \text{and} \quad g_i^{(2)} = \sum_{k\in[K^{(2)}]} k\mathbf{1}\{i \in G_k^{(2)}\}.$$

Let r_j be the rank of Θ_j^0 for $j \in [p] \cup \{0\}$. By the SVD for $\Theta_j^0/\sqrt{NT}$, we have

$$\Theta_j^0 = \sqrt{NT}\mathcal{U}_j^0 \Sigma_j^0 \mathcal{V}_j^{0\prime} := U_j^0 V_j^{0\prime}, \quad j \in [p] \cup \{0\}, \tag{3}$$

where $\mathcal{U}_j^0 \in \mathbb{R}^{N\times r_j}$, $\mathcal{V}_j^0 \in \mathbb{R}^{T\times r_j}$, $\Sigma_j^0 = \operatorname{diag}(\sigma_{1,j}, \dots, \sigma_{r_j,j})$, $U_j^0 = \sqrt{N}\mathcal{U}_j^0\Sigma_j^0$ with each row being $u_{i,j}^{0\prime}$, and $V_j^0 = \sqrt{T}\mathcal{V}_j^0$ with each row being $v_{t,j}^{0\prime}$. Later, we will show that Θ_j^0 exhibits a low-rank structure for all j, i.e., r_j is some fixed number.

Note that we allow $\{\Theta_{it}^0\}_{i=1}^N$ to exhibit latent group structures before and after the break. For a particular $j \in [p]$, the $N \times T$ matrix Θ_j^0 may have no group structure before or after the break, or no break, or more or fewer groups after the break. Let $K_j^{(1)}$ and $K_j^{(2)}$ denote the number of groups before and after the break, respectively, for $\{\Theta_{j,it}^0\}_{i=1}^N$. Let $\mathcal{G}_j^{(\ell)} = \{G_{1,j}^{(\ell)}, \dots, G_{K_j^{(\ell)},j}^{(\ell)}\}$, $\ell = 1, 2$, denote the associated latent group structures. Define $N_{k,j}^{(\ell)} = |G_{k,j}^{(\ell)}|$ and $\pi_{k,j}^{(\ell)} = \frac{N_{k,j}^{(\ell)}}{N}$ for $\ell = 1, 2$, where $|A|$ denotes the cardinality of set A. Further define $\tau_T := \frac{T_1}{T}$. We show that Θ_j^0 for all $j \in [p]$ has a low-rank structure in all of the following cases:

Case 1: For some $j \in [p]$, Θ_j^0 exhibits neither structural break nor group structure.

In this case, $K_j^{(1)} = K_j^{(2)} = 1$, and $\Theta_{j,it}^0 = \alpha_j\ \forall (i,t) \in [N] \times [T]$. Without loss of generality, assume that $\alpha_j > 0$. Then by the SVD for Θ_j^0 as in (3), we have

$$\begin{aligned} &\mathcal{U}_j^0 = \frac{1}{\sqrt{N}}\iota_N \in \mathbb{R}^{N\times 1}, \quad \Sigma_j^0 = \alpha_j, \quad \mathcal{V}_j^0 = \frac{1}{\sqrt{T}}\iota_T \in \mathbb{R}^{T\times 1}, \\ &U_j^0 = \alpha_j \iota_N \in \mathbb{R}^{N\times 1}, \quad V_j^0 = \iota_T \in \mathbb{R}^{T\times 1}, \end{aligned}$$

where $\iota_d = (1, \dots, 1)' \in \mathbb{R}^{d\times 1}$ for any natural number d. Obviously, $r_j = 1$ in Case 1.

Case 2: For some $j \in [p]$, Θ_j^0 exhibits no structural break but a group structure.

In this case, $K_j^{(1)} = K_j^{(2)} = K_j$, $G_{k,j}^{(1)} = G_{k,j}^{(2)} = G_{k,j}$, $N_{k,j}^{(1)} = N_{k,j}^{(2)} = N_{k,j}$, $\pi_{k,j}^{(1)} = \pi_{k,j}^{(2)} = \pi_{k,j}\ \forall k \in [K_j]$, and $\Theta_{j,it}^0 = \sum_{k\in[K_j]} \alpha_{k,j}\mathbf{1}\{i \in G_{k,j}\}$ for $t \in [T]$. Therefore, we have

$$\begin{aligned} &\mathcal{U}_{j,i}^0 = \frac{\sum_{k\in[K_j]} \alpha_{k,j}\mathbf{1}\{i \in G_{k,j}\}}{\sqrt{\sum_{k\in[K_j]} N_{k,j}\left(\alpha_{k,j}\right)^2}}, \quad \Sigma_j^0 = \sqrt{\sum_{k\in[K_j]} \pi_{k,j}\left(\alpha_{k,j}\right)^2}, \quad \mathcal{V}_j^0 = \frac{1}{\sqrt{T}}\iota_T, \\ &u_{i,j}^0 = \sum_{k\in[K_j]} \alpha_{k,j}\mathbf{1}\{i \in G_{k,j}\}, \quad V_j^0 = \iota_T, \end{aligned}$$

where $\mathcal{U}_{j,i}^0$ is the ith element in $\mathcal{U}_j^0$. Obviously, $r_j = 1$ in this case.

Case 3: For some $j \in [p]$, Θ_j^0 exhibits both a structural break and a group structure.

- **(i)** $K_j^{(1)} \neq K_j^{(2)}$, where we have different numbers of groups before and after the break;
- **(ii)** $K_j^{(1)} = K_j^{(2)} = K_j$ and $G_{k,j}^{(1)} \neq G_{k,j}^{(2)}$ for some $k \in [K_j]$, where we have the same number of groups before and after the break, but the group memberships change after the break point;
- **(iii)** $K_j^{(1)} = K_j^{(2)} = K_j$, $G_{k,j}^{(1)} = G_{k,j}^{(2)} = G_{k,j}$ for $\forall k \in [K_j]$, and $\alpha_{k,j}^{(1)} \neq \alpha_{k,j}^{(2)}$ for at least one $k \in [K_j]$, where even though neither the number of groups nor group membership changes after the break, there exists at least one group whose slope coefficients change.

Case 3 is of most interest to us. It suggests that from the perspective of the entire sample, it is possible that either the group-specific coefficients, or the group memberships of certain cross-section units, or both may change in Case 3. Specifically, the number of groups must change over time in Case 3(i), where we must observe the changes of both group memberships and group-specific slope coefficients. For example, consider the case where there are two groups before the break, with one group (say, group 2) split into two groups (say, groups 2 and 3) after the break. In this case, we have three groups after the break and some individuals in the pre-break group 2 must change their slope coefficients. Of course, it is also possible all the slope coefficients in the post-break three groups differ from the pre-break ones in this subcase. Similarly, we observe that in Case 3(ii), there is no change in the number of groups after the break, while the group memberships change for some individuals. This subcase includes a special case where the group-specific slope coefficients do not change while only the memberships of some individuals change over time. Apart from this special case, we may have both change in the group-specific slope coefficients and group memberships. In Case 3(iii), there is no change in the group memberships and only the group-specific slope coefficients change over time.

For any positive integer d, we use $\mathbf{0}_d$ to denote a $d \times 1$ vector of zeros. The following lemma lays down the foundation for break point detection in our model.

Lemma 2.1. *For any $j \in [p]$ such that Θ_j^0 lies in Case 3 above, we have $\operatorname{rank}(\Theta_j^0) \leq 2$. When $\operatorname{rank}(\Theta_j^0) = 2$, for the SVD for $\Theta_j^0/\sqrt{NT}$, i.e., $\Theta_j^0/\sqrt{NT} := \mathcal{U}_j^0 \Sigma_j^0 \mathcal{V}_j^{0\prime}$, we have*

(i) $\Theta_j^0 = U_j^0 V_j^{0\prime}$ *where* $U_j^0 = \sqrt{N}\mathcal{U}_j^0 \Sigma_j^0$, $V_j^0 = \sqrt{T}\mathcal{V}_j^0 = D_j R_j$, $D_j = \begin{bmatrix} \frac{1}{\sqrt{\tau_T}} \iota_{T_1} & \mathbf{0}_{T_1} \\ \mathbf{0}_{T-T_1} & \frac{1}{\sqrt{1-\tau_T}} \iota_{T-T_1} \end{bmatrix}$ *and* $R_j' R_j = I_2$;

(ii) $\left\| \frac{v_{t,j}^0}{\left\| v_{t,j}^0 \right\|} - \frac{v_{t^*,j}^0}{\left\| v_{t^*,j}^0 \right\|} \right\| = \sqrt{2}$ *for any $t \leq T_1$ and $t^* > T_1$.*

By Lemma 2.1 for Case 3 and the above analyses for Cases 1 and 2, we conclude that Θ_j^0 is a low-rank matrix with rank equal to or less than 2. In view of the low-rank structure of the slope matrices, we propose to adopt the NNR to obtain the preliminary estimates below. Moreover, under Case 3, Lemma 2.1(ii) indicates that singular vectors of the slope matrix with rank 2 contain the structural break information.

3. Estimation

This section develops the estimation algorithm. We first assume that the ranks r_j for $j \in [p] \cup \{0\}$ are known, and then propose a singular value thresholding (SVT) procedure to estimate them. After we recover the break point and the latent group structures, we propose consistent estimates of the group-specific parameters.

3.1. Estimation algorithm

Given r_j, $\forall j \in [p] \cup \{0\}$, we propose the following four-step procedure to estimate the break point and to recover the latent group structures before and after the break.

Step 1: Nuclear Norm Regularization (NNR). We run the nuclear norm regularized regression and obtain the preliminary estimates as follows:

$$\{\tilde{\Theta}_j\}_{j\in[p]\cup\{0\}} = \underset{\{\Theta_j\}_{j=0}^p}{\operatorname{argmin}} \frac{1}{NT} \left\| Y - \sum_{j=1}^p X_j \odot \Theta_j - \Theta_0 \right\|^2 + \sum_{j=0}^p \nu_j \left\| \Theta_j \right\|_*, \tag{4}$$

where ν_j is the tuning parameter. For each j, conduct the SVD: $\frac{1}{\sqrt{NT}} \tilde{\Theta}_j = \hat{\tilde{\mathcal{U}}}_j \hat{\tilde{\Sigma}}_j \hat{\tilde{\mathcal{V}}}_j'$, where $\hat{\tilde{\Sigma}}_j$ is a diagonal matrix that contains the singular values of $\tilde{\Theta}_j$ ordered in descending order along its diagonal line. Let $\tilde{\mathcal{V}}_j$ consist of the first r_j columns of $\hat{\tilde{\mathcal{V}}}_j$, and $\tilde{V}_j = \sqrt{T}\tilde{\mathcal{V}}_j$. Let $\tilde{v}_{t,j}'$ denote the tth row of $\tilde{V}_j$ for $t \in [T]$.

Step 2: Row- and Column-Wise Regressions. First run the row-wise regressions of Y_{it} on $(\tilde{v}_{t,0}, \{\tilde{v}_{t,j} X_{j,it}\}_{j\in[p]})$ to obtain $\{\dot{u}_{i,j}\}_{j\in[p]\cup\{0\}}$ for $i \in [N]$. Then run the column-wise regressions of Y_{it} on $(\dot{u}_{i,0}, \{\dot{u}_{i,j} X_{j,it}\}_{j\in[p]})$ to obtain $\{\dot{v}_{t,j}\}_{j\in[p]\cup\{0\}}$ for $t \in [T]$. Let $\dot{\Theta}_{j,it} = \dot{u}_{i,j}' \dot{v}_{t,j}$ for $(i,t) \in [N] \times [T]$ and $j \in [p] \cup \{0\}$. Specifically, the row- and column-wise regressions are given by

$$\left\{\dot{u}_{i,j}\right\}_{j\in[p]\cup\{0\}} = \underset{\{u_{i,j}\}_{j\in[p]\cup\{0\}}}{\operatorname{argmin}} \frac{1}{T} \sum_{t\in[T]} \left(Y_{it} - u_{i,0}' \tilde{v}_{t,0} - \sum_{j=1}^p u_{i,j}' \tilde{v}_{t,j} X_{j,it} \right)^2, \quad i \in [N], \tag{5}$$

$$\left\{\dot{v}_{t,j}\right\}_{j\in[p]\cup\{0\}} = \underset{\{v_{t,j}\}_{j\in[p]\cup\{0\}}}{\operatorname{argmin}} \frac{1}{N} \sum_{i\in[N]} \left(Y_{it} - v_{t,0}' \dot{u}_{i,0} - \sum_{j=1}^p v_{t,j}' \dot{u}_{i,j} X_{j,it} \right)^2, \quad t \in [T]. \tag{6}$$

Step 3: Break Point Estimation. We estimate the break point as follows:

$$\hat{T}_1 = \underset{s\in\{2,\ldots,T-1\}}{\operatorname{argmin}} \frac{1}{pNT} \sum_{j\in[p]} \sum_{i\in[N]} \left\{ \sum_{t=1}^s \left(\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}_{j,i}^{(1s)} \right)^2 + \sum_{t=s+1}^T \left(\dot{\Theta}_{j,it} - \bar{\dot{\Theta}}_{j,i}^{(2s)} \right)^2 \right\}, \tag{7}$$

where $\bar{\dot{\Theta}}_{j,i}^{(1s)} = \frac{1}{s} \sum_{t=1}^s \dot{\Theta}_{j,it}$ and $\bar{\dot{\Theta}}_{j,i}^{(2s)} = \frac{1}{T-s} \sum_{t=s+1}^T \dot{\Theta}_{j,it}$.

Step 4: Sequential Testing K-means (STK). In this step, we estimate the number of groups and the group membership before and after the break by using the STK algorithm. For each $j \in [p]$, define $\dot{\Theta}_{j,i}^{(1)} = (\dot{\Theta}_{j,i1}, \ldots, \dot{\Theta}_{j,i\hat{T}_1})'$, $\dot{\Theta}_{j,i}^{(2)} = (\dot{\Theta}_{j,i,\hat{T}_1+1}, \ldots, \dot{\Theta}_{j,iT})'$, $\dot{\beta}_i^{(1)} = \frac{1}{\sqrt{\hat{T}_1}} (\dot{\Theta}_{1,i}^{(1)\prime}, \ldots, \dot{\Theta}_{p,i}^{(1)\prime})'$, and $\dot{\beta}_i^{(2)} = \frac{1}{\sqrt{\hat{T}_2}} (\dot{\Theta}_{1,i}^{(2)\prime}, \ldots, \dot{\Theta}_{p,i}^{(2)\prime})'$. Let z_ς be some predetermined value which will be specified in the

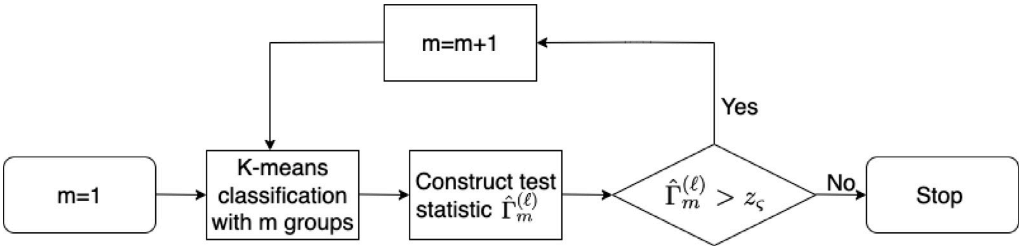

Fig. 1. The flow chart of STK algorithm.

next subsection. Given the subsample before and after the estimated break point, initialize $m = 1$ and classify each subsample into m groups by the K-means algorithm with group membership obtained as $\hat{\mathcal{G}}_m^{(\ell)} := \{\hat{G}_{k,m}^{(\ell)}\}_{k\in[m]}$. Next, we construct a suitable test statistic $\hat{\Gamma}_m^{(\ell)}$, defined by (11) in the next subsection, and compare it to its critical value z_ς at significance level ς under the null hypothesis of m subgroups, setting $m = m + 1$ and moving to the next iteration if $\hat{\Gamma}_m^{(\ell)} > z_\varsigma$ and stopping the STK algorithm otherwise. Lastly, define $\hat{K}^{(\ell)} = m$ and $\hat{\mathcal{G}}^{(\ell)} = \hat{\mathcal{G}}_m^{(\ell)}$. See the next subsection for specific details of the STK algorithm.

Several remarks are in order. First, the ranks of the intercept and slope matrices are assumed to be known in Step 1; otherwise they can be consistently estimated via SVT (see Section 3.3 below). Second, we obtain preliminary estimates by NNR based on the low-rank structure of the intercept and slope matrices in the model. These estimates are consistent in terms of the Frobenius norm but pointwise or uniform convergence for their elements is not established. Nonetheless, SVD can be employed to obtain preliminary estimates of the factors and loadings to be used subsequently. Third, row- and column-wise linear regressions are conducted to obtain updated estimates of the factors and loadings for which we can establish pointwise and uniform convergence rates. As one referee kindly pointed out, one can iterate this step to improve the finite sample performance in practice. Fourth, using the consistent estimates obtained in the second step, we can estimate the break point in Step 3 consistently by using a binary segmentation process. Fifth, the STK algorithm in Step 4 then yields the estimated number of groups and the group memberships together.

In the latent group literature, it is standard and popular to assume the number of groups in the K-means algorithm is known and then to estimate the number of groups by using certain information criteria. In this case, one needs to consider not only under- and just-fitting cases, but also over-fitting cases. It is well known that the major difficulty with this approach is to show that the over-fitting case occurs with probability approaching zero. The STK algorithm ensures a focus on the under-and just-fitting cases, which helps to avoid the difficulty caused by K-means classification with a larger than true number of groups. We will prove that the STK algorithm helps to eliminate the under-fitting case in which parameters cannot be consistently estimated in general. Although we adopt a sequential testing algorithm, we control the false discovery rate (FDR) by specifying a significance level that shrinks to zero (see Theorem 4.3 below), which also helps to deliver a consistent estimator of the number of groups. In addition, other approaches can also be used to estimate the latent group structures. For example, it is possible to apply the sequential binary segmentation algorithm (SBSA) of Wang and Su (2021) given the uniform consistent estimates $\{\dot{\Theta}_{it}^{(1)}, \dot{\Theta}_{it}^{(2)}\}$. It is also possible to extend the C-Lasso of Su et al. (2016) to allow for both latent group structures and structural changes despite the technical difficulty. But to save space we do not formally study these alternative approaches in our setup.

3.2. The STK algorithm

This subsection describes the K-means algorithm and the construction of the test statistics $\hat{\Gamma}_m^{(\ell)}$ that are used in the STK algorithm for $\ell \in \{1, 2\}$.

First, we define the objective function for the K-means algorithm with m clusters at each iteration. Let $a_{k,m}^{(\ell)}$ be a $p\hat{T}_1 \times 1$ and $p(T - \hat{T}_1) \times 1$ vector for $\ell = 1, 2$, respectively. We obtain the group membership with m groups by solving the following minimization problem

$$\left\{\dot{a}_{k,m}^{(\ell)}\right\}_{k\in[m]} = \underset{\left\{a_{k,m}^{(\ell)}\right\}_{k\in[m]}}{\operatorname{argmin}} \frac{1}{N} \sum_{i\in[N]} \min_{k\in[m]} \left\|\dot{\beta}_i^{(\ell)} - a_{k,m}^{(\ell)}\right\|^2, \tag{8}$$

which yields the membership estimates for each individual at the mth iteration as

$$\hat{g}_{i,m}^{(\ell)} = \underset{k\in[m]}{\operatorname{argmin}} \left\|\dot{\beta}_i^{(\ell)} - \dot{a}_{k,m}^{(\ell)}\right\| \quad \forall i \in [N]. \tag{9}$$

Let $\hat{G}_{k,m}^{(\ell)} := \{i \in [N] : \hat{g}_{i,m}^{(\ell)} = k\}$.

Second, we discuss the construction of the test statistic based on the idea of homogeneity test for several subsamples. At iteration m, we have m potential subgroups $(\hat{G}_{1,m}^{(\ell)}, \ldots, \hat{G}_{m,m}^{(\ell)})$ after the K-means classification for $\ell = 1$ and 2. Let $\hat{\mathcal{T}}_1 = [\hat{T}_1]$, $\hat{\mathcal{T}}_2 = [T]\backslash[\hat{T}_1]$,

$\hat{\mathcal{T}}_{1,-1} = \hat{\mathcal{T}}_1 \backslash \{\hat{T}_1\}$, $\hat{\mathcal{T}}_{2,-1} = \hat{\mathcal{T}}_2 \backslash \{T\}$, $\hat{\mathcal{T}}_{1,j} = \{1+j, \ldots, \hat{T}_1\}$, and $\hat{\mathcal{T}}_{2,j} = \{\hat{T}_1 + 1 + j, \ldots, T\}$ for some specific $j \in \hat{\mathcal{T}}_{\ell,-1}$. Based on these estimated subgroups, we can obtain the estimates of the coefficients, factors and loadings for each subgroup in regime ℓ as follows:

$$\left(\left\{\hat{\theta}^{(\ell)}_{i,k,m}\right\}_{i\in\hat{G}^{(\ell)}_{k,m}}, \hat{\Lambda}^{(\ell)}_{k,m}, \hat{F}^{(\ell)}_{k,m}\right) = \underset{\{\theta_i,\lambda_i,f_t\}_{i\in\hat{G}^{(\ell)}_{k,m}, t\in\hat{\mathcal{T}}_\ell}}{\operatorname{argmin}} \sum_{i\in\hat{G}^{(\ell)}_{k,m}} \sum_{t\in\hat{\mathcal{T}}_\ell} \left(Y_{it} - X_{it}'\theta_i - \lambda_i' f_t\right)^2,$$

where $\hat{\Lambda}^{(\ell)}_{k,m} = \{\hat{\lambda}^{(\ell)}_{i,k,m}\}_{i\in\hat{G}^{(\ell)}_{k,m}} \in \mathbb{R}^{\left|\hat{G}^{(\ell)}_{k,m}\right|\times r_0}$ and $\hat{F}^{(\ell)}_{k,m} = \{\hat{f}^{(\ell)}_{t,k,m}\}_{t\in\hat{\mathcal{T}}_\ell} \in \mathbb{R}^{\left|\hat{\mathcal{T}}_\ell\right|\times r_0}$ are the estimated individual effects matrix and time effects matrix for the estimated group $\hat{G}^{(\ell)}_{k,m}$ over the time span $\hat{\mathcal{T}}_\ell$. For all $i \in [N]$ and $t \in [T]$, define the residuals

$$\hat{e}_{it} = \sum_{\ell=1}^{2} \left(Y_{it} - \hat{f}^{(\ell)\prime}_{t,k,m}\hat{\lambda}^{(\ell)}_{i,k,m} - X_{it}'\hat{\theta}^{(\ell)}_{i,k,m}\right) \mathbf{1}\{t \in \hat{\mathcal{T}}_\ell\}.$$

Let $\hat{X}^{(1)}_i = (X_{i1}, \ldots, X_{i\hat{T}_1})'$, $\hat{X}^{(2)}_i = (X_{i,\hat{T}_1+1}, \ldots, X_{iT})'$, $\hat{T}_2 = T - \hat{T}_1$,

$$\hat{\bar{\theta}}^{(\ell)}_{k,m} = \frac{1}{|\hat{G}^{(\ell)}_{k,m}|} \sum_{i\in\hat{G}^{(\ell)}_{k,m}} \hat{\theta}^{(\ell)}_{i,k,m}, \quad \hat{S}^{(\ell)}_{ii,k,m} = \frac{1}{\hat{T}_\ell} \hat{X}^{(\ell)\prime}_i M_{\hat{F}^{(\ell)}_{k,m}} \hat{X}^{(\ell)}_i, \quad \hat{a}^{(\ell)}_{ii,k} = \hat{\lambda}^{(\ell)\prime}_{i,k,m} \left(|\hat{G}^{(\ell)}_{k,m}|^{-1} \hat{\Lambda}^{(\ell)\prime}_{k,m}\hat{\Lambda}^{(\ell)}_{k,m}\right)^{-1} \hat{\lambda}^{(\ell)}_{i,k,m},$$

and $\hat{z}^{(\ell)\prime}_{it}$ be the tth row of $M_{\hat{F}^{(\ell)}_{k,m}}\hat{X}^{(\ell)}_i$. For each subgroup $\hat{G}^{(\ell)}_{k,m}$ with $k \in [m]$, we follow the lead of Pesaran and Yamagata (2008) and Ando and Bai (2015) and define the following test statistic components

$$\hat{\Gamma}^{(\ell)}_{k,m} = \sqrt{|\hat{G}^{(\ell)}_{k,m}|} \cdot \frac{\frac{1}{|\hat{G}^{(\ell)}_{k,m}|}\sum_{i\in\hat{G}^{(\ell)}_{k,m}} \hat{\mathbb{S}}^{(\ell)}_{i,k,m} - p}{\sqrt{2p}}, \tag{10}$$

where

$$\hat{\mathbb{S}}^{(\ell)}_{i,k,m} = \hat{T}_\ell(\hat{\theta}^{(\ell)}_{i,k,m} - \hat{\bar{\theta}}^{(\ell)}_{k,m})'\hat{S}^{(\ell)}_{ii,k,m}(\hat{\Omega}^{(\ell)}_{i,k,m})^{-1}\hat{S}^{(\ell)}_{ii,k,m}(\hat{\theta}^{(\ell)}_{i,k} - \hat{\bar{\theta}}^{(\ell)}_{k})\left(1 - \frac{\hat{a}^{(\ell)}_{ii,k}}{|\hat{G}^{(\ell)}_{k,m}|}\right)^2,$$

$$\hat{\Omega}^{(\ell)}_{i,k,m} = \frac{1}{\hat{T}_\ell}\sum_{t\in\hat{\mathcal{T}}_\ell} \hat{z}^{(\ell)}_{it}\hat{z}^{(\ell)\prime}_{it}\hat{e}^2_{it} + \frac{1}{\hat{T}_\ell}\sum_{j\in\hat{\mathcal{T}}_{\ell,-1}} k(j/S_T) \sum_{t\in\hat{\mathcal{T}}_{\ell,j}} [\hat{z}^{(\ell)}_{it}\hat{z}^{(\ell)\prime}_{i,t-j}\hat{e}_{it}\hat{e}_{i,t-j} + \hat{z}^{(\ell)}_{i,t-j}\hat{z}^{(\ell)\prime}_{it}\hat{e}_{i,t-j}\hat{e}_{i,t}],$$

$k(\cdot)$ is a kernel function, S_T is a bandwidth/truncation parameter, and $\hat{\Omega}^{(\ell)}_{i,k,m}$ is the traditional HAC estimator. Using the components (10), we now define the test statistic

$$\hat{\Gamma}^{(\ell)}_m = \max_{k\in[m]} (\hat{\Gamma}^{(\ell)}_{k,m})^2. \tag{11}$$

We will show that $\hat{\Gamma}^{(\ell)}_m$ is asymptotically distributed as the maximum of m independent $\chi^2(1)$ random variables under the null hypothesis that the slope coefficients in each of the m subsamples are homogeneous, whereas it diverges to infinity under the alternative. Let z_ς denote the critical value at significance level ς, which is calculated from the maximum of m independent $\chi^2(1)$ random variables. We reject the null of m subgroups in favor of more groups at level ς if $\hat{\Gamma}^{(\ell)}_m > z_\varsigma$.

3.3. Rank estimation

Given the NNR estimators from (4), we propose to estimate the rank r_j of Θ_j via SVT:

$$\hat{r}_j = \sum_{i=1}^{N\wedge T} \mathbf{1}\left\{\sigma_i\left(\tilde{\Theta}_j\right) \geq 0.5\left(\nu_j \left\|\tilde{\Theta}_j\right\|_{op}\right)^{1/2}\right\} \quad \forall j \in \{0\} \cup [p],$$

where $\sigma_i(A)$ denotes the ith largest singular value of A and $N \wedge T = \min(N, T)$. By arguments as used in the proof of Proposition D.1 in Chernozhukov et al. (2020) and that of Theorem 3.2 in Hong et al. (2023), we can show that $\mathbb{P}(\hat{r}_j = r_j) \to 1$ for each j as $(N, T) \to \infty$.

3.4. Parameter estimation

Once we obtain the estimated break point, the number of groups and the group membership before and after the estimated break point, we can estimate the group-specific slope coefficients $\{\alpha^{(\ell)}_k\}_{k\in[\hat{K}^{(\ell)}]}$ along with the factors and loadings as follows

$$\left(\hat{\Lambda}^{(\ell)}, \hat{F}^{(\ell)}, \left\{\hat{\alpha}^{(\ell)}_k\right\}_{k\in[\hat{K}^{(\ell)}]}\right) = \operatorname{argmin} \mathbb{L}\left(\Lambda, F, \left\{a^{(\ell)}_k\right\}_{k\in[\hat{K}^{(\ell)}]}\right), \tag{12}$$

where $\mathbb{L}\left(\Lambda, F, \left\{a^{(\ell)}_k\right\}_{k\in[\hat{K}^{(\ell)}]}\right) = \frac{1}{N\hat{T}_\ell}\sum_{k=1}^{\hat{K}^{(\ell)}}\sum_{i\in\hat{G}^{(\ell)}_k}\sum_{t\in\hat{\mathcal{T}}_\ell}\left(Y_{it} - \lambda_i' f_t - X_{it}' a^{(\ell)}_k\right)^2$. Here, we ignore the fact that the prior- and post-break regimes share the same set of factor loadings and estimate the group-specific parameters separately for the two regimes at the cost of sacrificing some efficiency for the factor loading estimates. Alternatively, we can pool the observations before and after the break to estimate the parameters as

$$\left(\hat{\Lambda}, \hat{F}, \left\{\hat{\alpha}_k^{(1)}\right\}_{k\in[\hat{K}^{(1)}]}, \left\{\hat{\alpha}_k^{(2)}\right\}_{k\in[\hat{K}^{(2)}]}\right) = \operatorname{argmin} \mathbb{L}\left(\Lambda, F, \left\{a_k^{(1)}\right\}_{k\in[\hat{K}^{(1)}]}, \left\{a_k^{(2)}\right\}_{k\in[\hat{K}^{(2)}]}\right)$$

where

$$\mathbb{L}\left(\Lambda, F, \left\{a_k^{(1)}\right\}_{k\in[\hat{K}^{(1)}]}, \left\{a_k^{(2)}\right\}_{k\in[\hat{K}^{(2)}]}\right) = \mathbb{L}\left(\Lambda, F, \left\{a_k^{(1)}\right\}_{k\in[\hat{K}^{(1)}]}\right) + \mathbb{L}\left(\Lambda, F, \left\{a_k^{(2)}\right\}_{k\in[\hat{K}^{(2)}]}\right). \tag{13}$$

In either case, as should be clear because of the presence of the group structures, establishment of the asymptotic properties of the post-classification estimators of the group-specific slope coefficients becomes much more involved than in Bai (2009) and Moon and Weidner (2017). For this reason, we will focus on the estimates defined in (12).[2]

4. Asymptotic theory

This section develops the asymptotic properties of the estimators introduced above.

4.1. Basic assumptions

Define $e_i = \left(e_{i1}, \ldots, e_{iT}\right)'$ and $X_{j,i} = \left(X_{j,i1}, \ldots, X_{j,iT}\right)'$. Let V_j^0 be a $T \times r_j$ matrix with its tth row being $v_{t,j}^{0\prime}$, and U_j^0 be the $N \times r_j$ matrix with its ith row being $u_{i,j}^{0\prime}$. Throughout the paper, we treat the factors $\{V_j^0\}_{j\in[p]\cup\{0\}}$ as random and their loadings $\{U_j^0\}_{j\in[p]\cup\{0\}}$ as deterministic. Let $\mathscr{D} := \sigma(\{V_j^0\}_{j\in[p]\cup\{0\}})$ denote the minimum σ-field generated by $\{V_j^0\}_{j\in[p]\cup\{0\}}$. Similarly, let $\mathscr{G}_t := \sigma(\mathscr{D}, \left\{X_{is}\right\}_{i\in[N],s\le t+1}, \left\{e_{is}\right\}_{i\in[N],s\le t})$. Let $\max_i = \max_{i\in[N]}$, $\max_t = \max_{t\in[T]}$ and $\max_{i,t} = \max_{i\in[N],t\in[T]}$. Let M and C be generic bounded positive constants which may vary across lines.

Assumption 1.

(i) $\left\{e_{it}, X_{it}\right\}_{t\in[T]}$ are conditionally independent across i given $\mathscr{D}$.
(ii) $\mathbb{E}(e_{it}|X_{it}, \mathscr{D}) = 0$.
(iii) For each i, $\left\{\left(e_{it}, X_{it}\right), t \ge 1\right\}$ is strong mixing conditional on $\mathscr{D}$ with the mixing coefficient $\alpha_i(\cdot)$ satisfying $\max_i \alpha_i(z) \le M\vartheta^z$ for some constant $\vartheta \in (0, 1)$.
(iv) There exists a constant $C > 0$ such that $\max_i \frac{1}{T}\sum_{t\in[T]} \|\xi_{it}\|^2 \le C$ *a.s.* and $\max_t \frac{1}{N}\sum_{i\in[N]} \|\xi_{it}\|^2 \le C$ *a.s.* for $\xi_{it} = e_{it}$, X_{it} and $X_{it}e_{it}$.
(v) $\max_{i,t} \mathbb{E}[\|\xi_{it}\|^q |\mathscr{D}] \le M$ *a.s.* and $\max_{i,i^*,t} \mathbb{E}[\|X_{it}e_{i^*t}\|^q |\mathscr{D}] \le M$ *a.s.* for some $q > 8$ and $\xi_{it} = e_{it}$, X_{it} and $X_{it}e_{it}$.
(vi) As $(N,T) \to \infty$, $\sqrt{N}(\log N)^2 T^{-1} \to 0$ and $T(\log N)^2 N^{-3/2} \to 0$.

Assumption 1*. (i), (iv), (v), and (vi) are same as Assumption 1(i), (iv), (v) and (vi). In addition:

(ii) $\mathbb{E}(e_{it}|\mathscr{G}_{t-1}) = 0\ \forall (i,t) \in [N]\times[T]$, and $\max_{i,t} \mathbb{E}(e_{it}^2|\mathscr{G}_{t-1}) \le M$ *a.s.*.
(iii) $\left\{e_{it}\right\}_{i\in[N]}$ is conditionally independent across t given $\mathscr{D}$.

Assumption 1(i) imposes conditional independence on $\left\{e_{it}, X_{it}\right\}_{t\in[T]}$ across the cross-sectional units. Assumption 1(ii) is the conditional moment condition. Assumption 1(iii) imposes conditional strong mixing conditions along the time dimension. See Prakasa Rao (2009) for the definition of conditional strong mixing and Su and Chen (2013) for an application in the panel setup. Assumptions 1(iv) and (v) impose conditions that restrict the tail behavior of ξ_{it}. Note that neither the regressors nor the errors are constrained to be bounded. Assumption 1(vi) imposes restrictions on N and T but does not require N and T to diverge at the same rate. It is possible to allow N to diverge to infinity faster but not too much faster than T, and vice versa. Note that N and T enter Assumption 1(vi) asymmetrically because the estimation is based on a sequential approach, which does not treat the cross-section and time dimensions in a symmetric way.

Assumption 1* is used for the study of dynamic panel data models. To be specific, Assumption 1*(ii) requires that the error sequence $\left\{e_{it}, t \ge 1\right\}$ be a martingale difference sequence (m.d.s.) with respect to the filtration $\mathscr{G}_t$, which allows for lagged dependent variables in X_{it}. Assumption 1*(iii) imposes conditional independence of the errors over t. The presence of serially correlated errors in dynamic panels typically induces endogeneity, which invalidates least-squares-based PCA estimation.

Assumption 2. $\operatorname{rank}(\Theta_j^0) = r_j \le \bar{r}$ for $j \in [p]\cup\{0\}$ and some fixed $\bar{r}$, and $\max_{j\in[p]\cup\{0\}} \|\Theta_j^0\|_{\max} \le M$.

---

[2] If there is certain prior knowledge that both the slope coefficients and the memberships of some cross-section units in the sample are not affected by the break, we can still run the above estimation algorithm perhaps at the cost of certain efficiency loss. To incorporate such prior knowledge, one can impose some constraints in the above estimation procedure. For example, if we know individuals $1, \ldots, N_1$ belong to group 1 before and after the break and the group 1 slope coefficient vector remains constant over time, we can explore such information in the estimation procedure by grouping units 1 to N_1 into group 1 and use all T time periods to estimate the constant slope coefficient vector for group 1.

Assumption 2 imposes low-rank conditions on the coefficient matrices, which facilitate the use of NNR in obtaining preliminary estimates in the first step. As discussed in the previous section, we see that the low-rank assumption for the slope matrices is satisfied for the model introduced in Section 2. Moreover, we follow Ma et al. (2022) and assume the elements of the coefficient matrices are uniformly bounded to simplify the proofs. The boundedness of the slope coefficients is reasonable given that their cardinality does not grow with the sample size. The boundedness assumption for the intercept coefficient can be relaxed at the cost of more lengthy arguments.

Assumption 3. Let $\sigma_{l,j}$ denote the lth largest singular values of Θ_j^0 for $j \in [p] \cup \{0\}$. There exist some constants C_σ and c_σ such that

$$\infty > C_\sigma \geq \lim \sup_{(N,T)\to\infty} \max_{j\in[p]} \sigma_{1,j} \geq \lim \inf_{(N,T)\to\infty} \min_{j\in[p]} \sigma_{r_j,j} \geq c_\sigma > 0.$$

Assumption 3 imposes some conditions on the singular values of the coefficient matrices. These ensure that only pervasive factors are allowed when the matrices are written as a factor structure. The assumption can be readily verified given the latent group structures of the slope coefficients.

Consider the SVD: $\Theta_j^0/\sqrt{NT} = \mathcal{U}_j^0 \Sigma_j^0 \mathcal{V}_j^{0\prime}$ for all $j \in [p] \cup \{0\}$. Decompose $\mathcal{U}_j^0 = \left(\mathcal{U}_{j,r}, \mathcal{U}_{j,0}\right)$ and $\mathcal{V}_j^0 = \left(\mathcal{V}_{j,r}, \mathcal{V}_{j,0}\right)$ with $\left(\mathcal{U}_{j,r}, \mathcal{V}_{j,r}\right)$ being the singular vectors corresponding to nonzero singular values and $\left(\mathcal{U}_{j,0}, \mathcal{V}_{j,0}\right)$ being the singular vectors corresponding to zero singular values. Hence, for any matrix $W \in \mathbb{R}^{N\times T}$, we define

$$\mathcal{P}_j^\perp(W) = NT\mathcal{U}_{j,0}\mathcal{U}_{j,0}^\prime W \mathcal{V}_{j,0}\mathcal{V}_{j,0}^\prime, \quad \mathcal{P}_j(W) = W - \mathcal{P}_j^\perp(W),$$

where $\mathcal{P}_j(W)$ can be seen as the linear projection of matrix W into the low-rank space with $\mathcal{P}_j^\perp(W)$ being its orthogonal space. Let $\Delta_{\Theta_j} = \Theta_j - \Theta_j^0$ for any Θ_j. Based on the spaces constructed above, with some positive constants C_1 and C_2, we define the restricted set for full sample parameters as follows:

$$\mathcal{R}(C_1, C_2) := \Bigg\{ (\{\Delta_{\Theta_j}\}_{j\in[p]\cup\{0\}}) : \sum_{j\in[p]\cup\{0\}} \left\|\mathcal{P}_j^\perp(\Delta_{\Theta_j})\right\|_* \leq C_1 \sum_{j\in[p]\cup\{0\}} \left\|\mathcal{P}_j(\Delta_{\Theta_j})\right\|_*,$$
$$\sum_{j\in[p]\cup\{0\}} \left\|\Theta_j\right\|^2 \geq C_2\sqrt{NT} \Bigg\}. \tag{14}$$

Lemma B.4 in the online supplement shows that our nuclear norm estimators are in a restricted set larger than (14), which derives from the restriction on the Frobenius norm in the definition of $\mathcal{R}\left(C_1, C_2\right)$. Intuitively, the first restriction in (14) means the projection onto the orthogonal low-rank space of the estimator error can be controlled by its projection onto the low-rank space. Theorem 4.1 largely hinges on this property.

Assumption 4. For any $C_2 > 0$, there are constants C_3 and C_4 such that for any $(\{\Delta_{\Theta_j}\}_{j\in[p]\cup\{0\}}) \in \mathcal{R}(3, C_2)$, we have

$$\left\|\Delta_{\Theta_0} + \sum_{j=1}^p \Delta_{\Theta_j} \odot X_j\right\|^2 \geq C_3 \sum_{j\in[p]\cup\{0\}} \left\|\Delta_{\Theta_j}\right\|^2 - C_4(N+T) \quad \text{w.p.a.1.}$$

Assumption 4 imposes the restricted strong convexity (RSC) condition, which is similar to Assumption 3.1 in Chernozhukov et al. (2020). The latter authors also provide some sufficient conditions to justify such an assumption.

Let $r = \sum_{j\in[p]\cup\{0\}} r_j$. Define the following $r \times r$ matrices:

$$\Phi_i = \frac{1}{T}\sum_{t=1}^T \phi_{it}^0\phi_{it}^{0\prime} \quad \forall i \in [N] \text{ and } \Psi_t = \frac{1}{N}\sum_{i\in[N]} \psi_{it}^0\psi_{it}^{0\prime} \quad \forall t \in [T],$$

where $\phi_{it}^0 = (v_{t,0}^{0\prime}, v_{t,1}^{0\prime}X_{1,it}, \ldots, v_{t,p}^{0\prime}X_{p,it})'$, and $\psi_{it}^0 = (u_{i,0}^{0\prime}, u_{i,1}^{0\prime}X_{1,it}, \ldots, u_{i,p}^{0\prime}X_{p,it})'$.

Assumption 5. There exist constants C_ϕ and c_ϕ such that

$$\infty > C_\phi \geq \lim\sup_T \max_{t\in[T]} \lambda_{\max}(\Psi_t) \geq \lim\inf_T \min_{t\in[T]} \lambda_{\min}(\Psi_t) \geq c_\phi > 0,$$
$$\infty > C_\phi \geq \lim\sup_N \max_{i\in[N]} \lambda_{\max}(\Phi_i) \geq \lim\inf_N \min_{i\in[N]} \lambda_{\min}(\Phi_i) \geq c_\phi > 0.$$

Assumption 5 is similar to Assumption 8 in Ma et al. (2022), and it imposes some rank conditions.

4.2. Asymptotics of NNR estimators and singular vector estimators

Let $\eta_{N,1} = \frac{\sqrt{\log T}}{\sqrt{N\wedge T}}$ and $\eta_{N,2} = \frac{\sqrt{\log(N\vee T)}}{\sqrt{N\wedge T}}(NT)^{1/q}$. Let $\tilde{\sigma}_{k,j}$ denote the kth largest singular value of $\tilde{\Theta}_j$ for $j \in [p] \cup \{0\}$. Our first main result is about the consistency of the first-stage NNR estimators and the second-stage singular vector estimators.

Theorem 4.1. *Suppose that Assumptions 1 and 4 hold. Then $\forall j \in [p] \cup \{0\}$, we have* *(i)* $\frac{1}{\sqrt{NT}}\|\tilde{\Theta}_j - \Theta_j^0\| = O_p(\eta_{N,1})$, $\max_{k\in[r_j]}|\tilde{\sigma}_{k,j} - \sigma_{k,j}| = O_p(\eta_{N,1})$, *and* $\|V_j^0 - \tilde{V}_j O_j\| = O_p(\sqrt{T}\eta_{N,1})$ *where* O_j *is some orthogonal matrix. If in addition Assumption 5 is also satisfied, then we have*

(ii) $\max_{i\in[N]}\|\dot{u}_{i,j} - O_j u_{i,j}^0\| = O_p(\eta_{N,2})$, $\max_{t\in[T]}\left\|\dot{v}_{t,j} - O_j v_{t,j}^0\right\|_2 = O_p(\eta_{N,2})$,

(iii) $\max_{i\in[N],t\in[T]}|\dot{\Theta}_{j,it} - \Theta_{j,it}^0| = O_p(\eta_{N,2})$.

Theorem 4.1(i) reports the error bounds for $\tilde{\Theta}_j$, $\tilde{\sigma}_{k,j}$, and $\tilde{V}_j$. The $\log T$ term in the numerator of $\eta_{N,1}$ is due to the use of some exponential inequality for (conditional) strong mixing processes. Theorem 4.1(ii) reports the uniform convergence rate of the factor and loading estimators. Theorem 4.1(iii) reports the uniform convergence rate of the intercept and slope estimators. The extra $(NT)^{1/q}$ term in the definition of $\eta_{N,2}$ is due to the nonboundedness of $X_{j,it}$ in Assumption 1(v), and it disappears when $X_{j,it}$ is assumed to be uniformly bounded.

4.3. Consistency of the break point estimate

Recall that $g_i^{(1)}$ and $g_i^{(2)}$ denote the true group individual i belongs to before and after the break, respectively. To estimate the break point consistently, we add the following condition.

Assumption 6.

(i) $\sqrt{\frac{1}{N}\sum_{i\in[N]}\|\alpha_{g_i^{(1)}} - \alpha_{g_i^{(2)}}\|^2} = C_5\zeta_{NT}$, where C_5 is a positive constant and $\zeta_{NT} \gg \eta_{N,2}$.

(ii) $\tau_T := \frac{T_1}{T} \to \tau \in (0,1)$ as $T \to \infty$.

Assumption 6(i) imposes conditions on the break size in order to identify the break point. Note that we allow the average break size to shrink to zero at a rate slower than $\eta_{N,2} = \sqrt{\frac{\log(N\vee T)}{N\wedge T}}(NT)^{1/q}$. The reason why we need the average break size ζ_{NT} to be larger in order than $\eta_{N,2}$ is that breaking point detection is based on the singular vector estimators $\dot{u}_{i,j}$ and $\dot{v}_{t,j}$, which converge to their true values uniformly at rate $\eta_{N,2}$. This rate is of much bigger magnitude than the optimal $(NT)^{-1/2}$-rate that can be detected in the panel threshold regressions (PTRs) for several reasons. First, in PTRs, the slope coefficients are usually assumed to be homogeneous so that each individual is subject to the same change in the slope coefficients and one can use the cross-sectional information effectively. In contrast, we allow for heterogeneous slope coefficients here and the change can occur only for a subset of cross-section units but not all. In addition, in the presence of latent group structures, we not only allow the slope coefficients of some specific groups to change with group membership fixed, but also allow the slope coefficient to remain the same for some groups while the group memberships change after the break. Second, our break point estimation relies on the binary segmentation idea borrowed from the time series literature where one can allow break sizes of bigger magnitude than $T^{-1/2}$ in order to identify the break ratio consistently but not the break point consistently. As is apparent, even though we require bigger break sizes, we can estimate the break date consistently by using information from both the cross-section and time dimensions. Third, as mentioned above, the additional term $\log(N\vee T)$ in the above rate is mainly due to the use of an exponential inequality and the factor $(NT)^{1/q}$ is due to the fact that we only assume the existence of qth order moments for some random variables.

The following theorem indicates that we can estimate the break date T_1 consistently.

Theorem 4.2. *Suppose Assumptions 1 and 6 hold, with the true break point being T_1 and the estimator defined in (7). Then $\mathbb{P}(\hat{T}_1 = T_1) \to 1$ as $(N,T) \to \infty$.*

Theorem 4.2 shows that we can estimate the true break date consistently w.p.a.1 despite the fact that we allow the break size to shrink to zero at a certain rate.

Remarks. Following a suggestion from a referee, an alternative approach is to minimize the sum of squared residuals (SSR)

$$\underset{s\in\{2,\ldots,T-1\}}{\operatorname{argmin}} \frac{1}{NT}\sum_{i\in[N]}\left[\sum_{t=1}^{s}\left(Y_{it} - \dot{\Theta}_{0,it} - \sum_{j\in[p]} X_{j,it}\bar{\dot{\Theta}}_{j,i}^{(1s)}\right)^2 + \sum_{t=s+1}^{T}\left(Y_{it} - \dot{\Theta}_{0,it} - \sum_{j\in[p]} X_{j,it}\bar{\dot{\Theta}}_{j,i}^{(2s)}\right)^2\right],$$

where $\bar{\dot{\Theta}}_{j,i}^{(1s)} = \frac{1}{s}\sum_{l=1}^{s}\dot{\Theta}_{j,il}$ and $\bar{\dot{\Theta}}_{j,i}^{(2s)} = \frac{1}{s}\sum_{l=s+1}^{T}\dot{\Theta}_{j,il}$. The analysis is then similar to that of Bai (1997). The major complication comes from the fact that $\dot{\Theta}_{0,it}$, $\bar{\dot{\Theta}}_{j,i}^{(1s)}$, and $\bar{\dot{\Theta}}_{j,i}^{(2s)}$, are all estimated from the data in Step 2. Let $\tilde{T}_1$ denote the solution to the above minimization problem. We conjecture that one can establish $\mathbb{P}\left(\tilde{T}_1 = T_1\right) \to 1$ as $(N,T) \to \infty$ under some conditions. In Step 3, we propose to minimize the sum of (scaled) sample variances before and after the break points. We can do this because we have variations in the regression coefficients along both the cross section and time dimensions. There are two major advantages in favor of this approach. First, our objective function in Step 3 is based on a location model for $\dot{\Theta}_{it} = (\dot{\Theta}_{0,it}, \dot{\Theta}_{1,it}, \ldots, \dot{\Theta}_{p,it})'$ instead of a regression model with covariates, so the asymptotic analysis can be greatly simplified. Second, this approach allows us to impose conditions on the average break size as characterized by $\sqrt{\frac{1}{N}\sum_{i\in[N]}\|\alpha_{g_i^{(1)}} - \alpha_{g_i^{(2)}}\|^2}$ in Assumption 6(i) directly without the need to consider the behavior of the covariates before and after the break. In contrast, the SSR-based approach requires certain assumption on the covariates and their interactions with Θ_{it}^0 before and after the break. In addition, our simulations show that the two approaches deliver comparable accuracy for the estimation of the break points. So we focus on the approach stated in Step 3.

4.4. Consistency of the estimates of the number of groups and the latent group structures

To study the asymptotic properties of the estimates of the number of groups and the recovery of the latent group structures, we add the following assumptions.

Assumption 7.

(i) Let k and k^* be different group indices. Assume that $\min_{1\le k<k^*\le K^{(\ell)}}\left\|\alpha_k^{(\ell)}-\alpha_{k^*}^{(\ell)}\right\|_2 \ge C_5$ for $\ell \in \{1,2\}$.

(ii) Let $N_k^{(\ell)}$ be the number of individuals in group k for $k \in [K^{(\ell)}]$. Define $\pi_k^{(\ell)} = \frac{N_k^{(\ell)}}{N}$ for $\ell = 1,2$. Assume $K^{(\ell)}$ is fixed and $\liminf_N \inf_{k\in[K^{(\ell)}]} \pi_k^{(\ell)} \ge \underline{c} > 0$ for $\ell = 1,2$.

(iii) For any permutation of n true groups with $n \in \{2,\dots,K^{(\ell)}\}$, we have

$$\frac{T_\ell}{\sqrt{N}}\sum_{k=1}^{n} N_k^{(\ell)}\left\|\sum_{k^*\in[n],k^*\neq k}(\alpha_{k^*}^{(\ell)}-\alpha_k^{(\ell)})\right\|^2 /(\log N)^{1/2} \to \infty,\ \ \ell = 1,2.$$

Assumption 7(i)–(ii) are the standard assumptions for K-means algorithm, which are comparable to Assumption 4 in Su et al. (2020) and greatly facilitate the subsequent analyses. Assumption 7(i) assumes that the minimum distance of two distinct groups is bounded away from 0, and Assumption 7(ii) imposes that each group has asymptotically non-negligible number of units. Assumption 7(iii) is needed in the proof of the last part of Theorem 4.3(ii) and we now show that it holds under some mild conditions. When $n = 2$, it is clear that

$$\begin{aligned}\frac{T_\ell}{\sqrt{N}}\sum_{k=1}^{n} N_k^{(\ell)}\left\|\sum_{k^*\in[n],k^*\neq k}(\alpha_{k^*}^{(\ell)}-\alpha_k^{(\ell)})\right\|^2 &= \frac{T_\ell}{\sqrt{N}}\left(N_1^{(\ell)}\left\|\alpha_2^{(\ell)}-\alpha_1^{(\ell)}\right\|^2 + N_2^{(\ell)}\left\|\alpha_1^{(\ell)}-\alpha_2^{(\ell)}\right\|^2\right)\\ &\ge \frac{C_5^2 T_\ell (N_1^{(\ell)}+N_2^{(\ell)})}{\sqrt{N}} = \Omega(T\sqrt{N})\end{aligned}$$

by Assumptions 6(ii) and 7(i)–(ii). Here, $\Omega(T\sqrt{N})$ signifies that the order is of exact order $T\sqrt{N}$. When $n > 2$, we consider a special case such that $S_k =: \left\|\sum_{k^*\in[n],k^*\neq k}(\alpha_{k^*}^{(\ell)}-\alpha_k^{(\ell)})\right\| = 0$ for some specific $k = k_0 \in [n]$. Then it is easy to see S_s is non-zero for all $s \in [n]\backslash\{k_0\}$ under Assumption 7(i). Hence, if we assume S_s is lower bounded by a constant c for any $s \in [n]\backslash\{k_0\}$, Assumption 7(iii) holds naturally. Similar arguments hold for the other general cases.

Assumption 8. Let $\mathcal{T}_1 = \lceil T_1\rceil$ and $\mathcal{T}_2 = [T]\backslash\lceil T_1\rceil$. $\frac{1}{T_\ell}\sum_{t\in\mathcal{T}_\ell} f_t^0 f_t^{0\prime} \xrightarrow{p} \Sigma_F^{(\ell)} > 0$ as $T\to\infty$. $\frac{1}{N_k^{(\ell)}}\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)} \xrightarrow{p} \Sigma_{\Lambda,k}^{(\ell)} > 0$ as $N\to\infty$, where $\Lambda_k^{0,(\ell)}$ is a stack of λ_i^0 for all individuals in group k and $k\in[K^{(\ell)}]$.

Assumption 8 imposes some standard assumptions on the factors and loadings. The next result details the asymptotic properties of the STK estimators.

Theorem 4.3. *Let $\varsigma = \varsigma_N \to 0$ at rate N^{-c} for some $c > 0$ as $N\to\infty$. Suppose that Assumption 1* and Assumptions 2–8 hold. Then for $\ell \in \{1,2\}$, we have*

(i) if $m = K^{(\ell)}$,

(a) $\max_{i\in[N]} \mathbf{1}\{\hat{g}_{i,K^{(\ell)}}^{(\ell)} \neq g_i^{(\ell)}\} = 0$ w.p.a.1,

(b) $\hat{\Gamma}_{K^{(\ell)}}^{(\ell)}$ is asymptotically distributed as the maximum of $K^{(\ell)}$ independent $\chi^2(1)$ random variables,

(c) $\mathbb{P}(\hat{K}^{(\ell)} \le K^{(\ell)}) \ge 1-\varsigma+o(1)$,

(ii) if $m < K^{(\ell)}$, $\hat{\Gamma}_m^{(\ell)}/\log N \to \infty$ w.p.a.1. Thus $\mathbb{P}(\hat{K}^{(\ell)} \neq K^{(\ell)}) \le \varsigma + o(1)$.

In Theorem 4.3 we allow $\varsigma = \varsigma_N$ to shrink to zero at rate N^{-c}, so that the critical value z_ς diverges to infinity at rate $\log N$ as $N\to\infty$ by virtue of the tail properties of $\chi^2(1)$ random variables. At iteration m such that $m < K^{(\ell)}$, w.p.a.1, the test statistic $\hat{\Gamma}_m^{(\ell)}$ diverges to infinity at a rate faster than $\log N$, which means the iteration will continue at the $(m+1)$-th iteration. At iteration m with $m = K^{(\ell)}$, however, we can easily find that $z_\varsigma \to \infty$ while the test statistic $\hat{\Gamma}_m^{(\ell)}$ is stochastically bounded. As a result, the iteration stops w.p.a.1 and we have $\mathbb{P}(\hat{K}^{(\ell)} = K^{(\ell)}) \to 1$. As aforementioned, Theorem 4.3 ensures the application of K-means algorithm only for the under-fitting and just-fitting cases, and it avoids the theoretical challenge of handling the over-fitting case in the classification.

For dynamic panels, we can focus on Assumption 1*, where the error term is an m.d.s. Under this assumption, the HAC estimator $\hat{\Omega}_{i,k,m}^{(\ell)}$ degenerates to $\frac{1}{\hat{T}_\ell}\sum_{t\in\hat{\mathcal{T}}_\ell}\hat{z}_{it}^{(\ell)}\hat{z}_{it}^{(\ell)\prime}\hat{e}_{it}^2$. For static panels, we typically allow for serially correlated errors and employ the HAC estimator, and the results in Theorem 4.3 continue to hold with Assumption 1* replaced by Assumption 1. For the kernel function and bandwidth, we can follow Andrews (1991) and let $k(\cdot)$ belong to the following class of kernels

$$\mathcal{K} = \left\{k(\cdot) : \mathbb{R}\mapsto[-1,1] \mid k(0)=1,\ k(u)=k(-u),\ \int |k(u)|\,du < \infty,\right.$$

$k(\cdot)$ is continuous at 0 and at all but a finite number of other points $\Big\}$.

See, e.g., Andrews (1991) and White (2014) for details.

4.5. Distribution theory for the group-specific slope estimators

For $\ell \in \{1,2\}$, let $\{\hat{\alpha}_k^{*(\ell)}\}_{k\in K^{(\ell)}}$ be the oracle estimators of the group-specific slope coefficients before and after the break point by using the true break and membership information for all individuals. To study the asymptotic distribution theory for $\{\hat{\alpha}_k^{(\ell)}\}_{k\in K^{(\ell)}}$, $\ell \in \{1,2\}$, we only need to show that for the oracle estimators $\{\hat{\alpha}_k^{*(\ell)}\}_{k\in K^{(\ell)}}$ based on Theorems 4.2 and 4.3 by extending the result of Bai (2009) and Moon and Weidner (2017).

To proceed, we add some notation. For $\ell \in \{1,2\}$, we first define the matrix notation for each subgroup. For $j \in [p]$, let $X_{j,i}^{(1)} = \left(X_{j,i1},\dots,X_{j,iT_1}\right)'$, $X_{j,i}^{(2)} = \left(X_{j,i(T_1+1)},\dots,X_{j,iT}\right)'$, $e_i^{(1)} = \left(e_{i1},\dots,e_{iT_1}\right)'$ and $e_i^{(2)} = \left(e_{i(T_1+1)},\dots,e_{iT}\right)'$. Then we use $\mathbb{X}_{j,k}^{(\ell)} \in \mathbb{R}^{N_k^{(\ell)}\times T_\ell}$ and $E_k^{(\ell)} \in \mathbb{R}^{N_k^{(\ell)}\times T_\ell}$ to denote the regressor matrix and error matrix for subgroup $k \in [K^{(\ell)}]$ with each row being $X_{j,i}^{(\ell)}$ and $e_i^{(\ell)}$ for $i \in G_k^{(\ell)}$, respectively. To state the main result in this subsection, we add the following assumption.

Assumption 9.

(i) As $(N,T)\to\infty$, $T(\log T)N^{-4/3}\to 0$.

(ii) $\operatorname{plim}_{(N,T)\to\infty} \frac{1}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathcal{T}_\ell} X_{it}X_{it}' > 0$ for $\ell\in\{1,2\}$ and $k\in[K^{(\ell)}]$.

(iii) For $\ell \in \{1,2\}$ and $k \in [K^{(\ell)}]$, separate the p regressors of each subgroups into p_1 "low-rank regressors" $\mathbb{X}_{j,k}^{(\ell)}$ such that $\operatorname{rank}(\mathbb{X}_{j,k}^{(\ell)}) = 1$, for any $j \in \{1,\dots,p_1\}$, and "high-rank regressors" $\mathbb{X}_{j,k}^{(\ell)}$ such that $\operatorname{rank}(\mathbb{X}_{j,k}^{(\ell)}) > 1$, for any $j \in \{p_1+1,\dots,p\}$. Let $p_2 := p - p_1$. These two types of regressors satisfy:

(iii.a) Consider the linear combinations $b\cdot\mathbb{X}_{high,k}^{(\ell)} := \sum_{j=p_1+1}^{p} b_j\mathbb{X}_{j,k}^{(\ell)}$ for high-rank regressors with p_2-vectors b such that $\|b\|_2 = 1$ and $b = \left(b_{p_1+1},\dots,b_p\right)'$. There exists a positive constant C_b such that

$$\min_{\{\|b\|_2=1\}} \sum_{n=2r_0+p_1+1}^{N_k^{(\ell)}} \lambda_n\left[\frac{1}{N_k^{(\ell)}T_\ell}\left(b\cdot\mathbb{X}_{high,k}^{(\ell)}\right)\left(b\cdot\mathbb{X}_{high,k}^{(\ell)}\right)'\right] \geq C_b \quad w.p.a.1.$$

(iii.b) For $j \in [p_1]$, write $\mathbb{X}_{j,k}^{(\ell)} = w_{j,k}^{(\ell)}v_{j,k}^{(\ell)\prime}$ with $N_k^{(\ell)}$-vectors $w_{j,k}^{(\ell)}$ and T_ℓ-vectors $v_{j,k}^{(\ell)}$. Let $w_k^{(\ell)} = (w_{1,k}^{(\ell)},\dots,w_{p_1,k}^{(\ell)}) \in \mathbb{R}^{N\times p_1}$ and $v_k^{(\ell)} = (v_{1,k}^{(\ell)},\dots,v_{p_1,k}^{(\ell)}) \in \mathbb{R}^{T_\ell\times p_1}$. There exists a positive constant C_B such that $(N_k^{(\ell)})^{-1}\Lambda_k^{0,(\ell)\prime}M_{w_k^{(\ell)}}\Lambda_k^{0,(\ell)} > C_B I_{r_0}$ and $T_\ell^{-1}F^{0,(\ell)\prime}M_{v_k^{(\ell)}}F^{0,(\ell)} > C_B I_{r_0}$ w.p.a.1.

(iv) For $\forall j\in[p]$, $\ell\in\{1,2\}$, and $k\in K^{(\ell)}$,

$$\frac{1}{N_k^{(\ell)}T_\ell^2}\sum_{i\in G_k^{(\ell)}}\sum_{t_1\in\mathcal{T}_\ell}\sum_{t_2\in\mathcal{T}_\ell}\sum_{s_1\in\mathcal{T}_\ell}\sum_{s_2\in\mathcal{T}_\ell}\left|\operatorname{Cov}\left(e_{it_1}\tilde{X}_{j,it_2}, e_{is_1}\tilde{X}_{j,is_2}\right)\right| = O(1),$$

where $\tilde{X}_{j,it} = X_{j,it} - \mathbb{E}\left(X_{j,it}|\mathscr{D}\right)$.

Assumption 9 imposes some conditions to help derive the asymptotic distribution theory for the panel model with IFEs which allows for dynamics. Assumption 9(i) slightly strengthens Assumption 1(vi). Assumption 9(ii) is the standard non-collinearity condition for regressors, which is analogous to Assumption 4(i) in Moon and Weidner (2017). Assumption 9(iii) is the identification assumption, which is comparable to Assumption 4 in Moon and Weidner (2017). Assumption 9(iv) requires $\{\tilde{X}_{j,it}e_{it}\}$ to be weakly dependent over the time, which parallels Assumption 5(v) in Moon and Weidner (2017) and can be verified under some weak dependence and moment conditions.

To show the distribution of the slope estimator, we further define $\mathcal{X}_{j,k}^{(\ell)} = M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}M_{F^{0,(\ell)}} \in \mathbb{R}^{N_k^{(\ell)}\times T_\ell}$ with the (i,t)-th entry given by $\mathcal{X}_{j,k,it}^{(\ell)}$, $\mathcal{X}_{k,it}^{(\ell)} = (\mathcal{X}_{1,k,it}^{(\ell)},\dots,\mathcal{X}_{p,k,it}^{(\ell)})'$, and

$$\mathbb{B}_{NT,1,j,k}^{(\ell)} = \frac{1}{N_k^{(\ell)}}tr\left[P_{F^{0,(\ell)}}\mathbb{E}\left(E_k^{(\ell)\prime}\mathbb{X}_{j,k}^{(\ell)}\middle|\mathscr{D}\right)\right],$$

$$\mathbb{B}_{NT,2,j,k}^{(\ell)} = \frac{1}{T_\ell}tr\left[\mathbb{E}\left(E_k^{(\ell)}E_k^{(\ell)\prime}\middle|\mathscr{D}\right)M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)}F^{0,(\ell)}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\Lambda_k^{0,(\ell)\prime}\right],$$

$$\mathbb{B}_{NT,3,j,k}^{(\ell)} = \frac{1}{N_k^{(\ell)}}tr\left[\mathbb{E}\left(E_k^{(\ell)\prime}E_k^{(\ell)}\middle|\mathscr{D}\right)M_{F^{0,(\ell)}}\mathbb{X}_{j,k}^{(\ell)\prime}\Lambda_k^{0,(\ell)}\left(\Lambda_k^{0,(\ell)\prime}\Lambda_k^{0,(\ell)}\right)^{-1}\left(F^{0,(\ell)\prime}F^{0,(\ell)}\right)^{-1}F^{0,(\ell)\prime}\right],$$

$$\mathbb{B}_{NT,m,k}^{(\ell)} = \left(\mathbb{B}_{NT,m,1,k}^{(\ell)},\dots,\mathbb{B}_{NT,m,p,k}^{(\ell)}\right)', \quad \forall m\in\{1,2,3\}.$$

Then the overall bias term for each subgroup is

$$\mathbb{B}_{NT,k}^{(\ell)} = -\rho_k^{(\ell)}\mathbb{B}_{NT,1,k}^{(\ell)} - (\rho_k^{(\ell)})^{-1}\mathbb{B}_{NT,2,k}^{(\ell)} - \rho_k^{(\ell)}\mathbb{B}_{NT,3,k}^{(\ell)}, \tag{15}$$

where $\rho_k^{(\ell)} = \sqrt{\frac{N_k^{(\ell)}}{T_\ell}}$. Define the two $p \times p$ matrices $\mathbb{W}_{NT,k}^{(\ell)} = \left\{[\mathbb{W}_{NT,k}^{(\ell)}]_{j_1 j_2}\right\}$ and $\Omega_k^{(\ell)}$ respectively as

$$[\mathbb{W}_{NT,k}^{(\ell)}]_{j_1 j_2} = \frac{1}{N_k^{(\ell)}T_\ell} tr\left(M_{F^{0,(\ell)}}\mathbb{X}_{j_1,k}^{(\ell)\prime}M_{\Lambda_k^{0,(\ell)}}\mathbb{X}_{j_2,k}^{(\ell)}\right) \text{ for } j_1, j_2 \in [p] \tag{16}$$

$$\Omega_k^{(\ell)} = \frac{1}{N_k^{(\ell)}T_\ell}\sum_{i\in G_k^{(\ell)}}\sum_{t\in\mathcal{T}_\ell} e_{it}^2 \mathcal{X}_{k,it}^{(\ell)}\mathcal{X}_{k,it}^{(\ell)\prime} \text{ for } k \in \left[K^{(\ell)}\right]. \tag{17}$$

The following theorem establishes the asymptotic distribution of $\{\hat{\alpha}_k^{(\ell)}\}_{k\in K^{(\ell)}}$.

Theorem 4.4. *Suppose that Assumption 1 or Assumption 1* and Assumptions 2 and 9 hold. For $\ell \in \{1,2\}$, the estimators $\{\hat{\alpha}_k^{(\ell)}\}_{k\in K^{(\ell)}}$ are asymptotically equivalent to the oracle estimators $\{\hat{\alpha}_k^{*(\ell)}\}_{k\in K^{(\ell)}}$, and we have*

$$\mathbb{W}_{NT}^{(\ell)}\mathbb{D}_{NT}^{(\ell)}\begin{pmatrix}\hat{\alpha}_1^{(\ell)} - \alpha_1^{(\ell)} \\ \vdots \\ \hat{\alpha}_{K^{(\ell)}}^{(\ell)} - \alpha_{K^{(\ell)}}^{(\ell)}\end{pmatrix} - \mathbb{B}_{NT}^{(\ell)} \rightsquigarrow \mathbb{N}\left(0, \Omega^{(\ell)}\right),$$

such that $\mathbb{D}_{NT}^{(\ell)} = \operatorname{diag}\left(\sqrt{N_1^{(\ell)}T_\ell}I_p, \ldots, \sqrt{N_{K^{(\ell)}}^{(\ell)}T_\ell}I_p\right)$, $\mathbb{W}_{NT}^{(\ell)} = \operatorname{diag}\left(\mathbb{W}_{NT,1}^{(\ell)}, \ldots, \mathbb{W}_{NT,K^{(\ell)}}^{(\ell)}\right)$, $\mathbb{B}_{NT}^{(\ell)} = \left(\mathbb{B}_{NT,1}^{(\ell)\prime}, \ldots, \mathbb{B}_{NT,K^{(\ell)}}^{(\ell)\prime}\right)'$ *and* $\Omega^{(\ell)} = \operatorname{diag}\left(\Omega_1^{(\ell)}, \ldots, \Omega_{K^{(\ell)}}^{(\ell)}\right)$, *where* $\mathbb{B}_{NT,1}^{(\ell)}$, $\mathbb{W}_{NT,k}^{(\ell)}$ *and* $\Omega_k^{(\ell)}$ *are defined in (15), (16) and (17), respectively.*

Theorem 4.4 establishes the asymptotic distribution for the estimators of the group-specific slope coefficients before and after the break. It shows that the parameter estimators from our algorithm enjoy the oracle property given the results in Theorems 4.2 and 4.3.

Remarks. The approach in this paper seeks to extend the work of Su et al. (2016) to allow for both IFEs (to model the strong cross-section dependence) and time-varying latent group structures (to take into account possible breaks over time). Our goal is modest in that we only consider discrete heterogeneity along both the cross-section and time dimensions. Over time, we currently focus on the case of one-time break; cross-sectionally we focus on the latent group structures with a fixed number of groups both before and after the break. By doing so, our panel data model can be regarded as a special case of the low-rank panel data regression model considered by Chernozhukov et al. (2020). As in that work, we model the intercept term as an IFE but we consider a special case of low-rank structure for the slope coefficients where the rank of Θ_j^0 is typically either 1 or 2 depending on whether we have structural changes in the model. The advantage of this device is two-fold. First, we maintain a balance between parameter heterogeneity and parsimony so that the number of parameters in the slope coefficients is significantly reduced from the order $O(N+T)$ in Chernozhukov et al. (2020) to $\sum_{\ell=1}^2 K^{(\ell)}p$, which is a fixed integer. Second, we can obtain the usual $\sqrt{NT}$-rate of convergence for the estimators of the slope parameters perhaps after bias correction and derive the asymptotic normality as usual without calling upon the sample splitting or de-biasing procedure used by Chernozhukov et al. (2020). From the technical perspective, there are two major differences between our estimators and the low-rank estimators in Chernozhukov et al. (2020). One is that Chernozhukov et al. (2020) rely on sample splitting and serial independence of the error terms in both the y- and x-equations whereas we do not need either sample splitting or serial dependence. The second is that Chernozhukov et al. (2020) rely on a partialling-out de-biasing procedure to remove the strong-cross-sectional dependence in the regressors whereas we do not need that procedure.

5. Alternatives and extensions

This section first considers an alternative method to estimate the break point and then discusses a possible extension.

5.1. Alternative for break point detection

The algorithm proposed in Section 3 uses low-rank estimates of Θ_j^0 to find the break point estimates. However, by Lemma 2.1(ii), we observe that the right singular vector matrix of Θ_j^0, i.e., V_j^0, contains the structural break information when $r_j = 2$. For this reason, we can propose an alternative way to estimate the break point under the case where the maximum rank of the slope matrix in the model is 2. Let $\dot{v}_{t,j}^* := \frac{\dot{v}_{t,j}}{\left\|\dot{v}_{t,j}\right\|}$ and $\dot{v}_t^* := \left(\dot{v}_{t,1}^{*\prime}, \ldots, \dot{v}_{t,p}^{*\prime}\right)'$, with the true values being $v_{t,j}^* := \frac{O_j v_{t,j}^0}{\left\|O_j v_{t,j}^0\right\|}$ and $v_t^* := \left(v_{t,1}^{*\prime}, \ldots, v_{t,p}^{*\prime}\right)'$, respectively. Then Step 3 can be replaced by Step 3* below:

Step 3*: Break Point Estimation by Singular Vectors. We estimate the break point as follows:

$$\tilde{T}_1 = \underset{s\in\{2,\ldots,T-1\}}{\operatorname{argmin}} \frac{1}{T}\left\{\sum_{t=1}^{s}\left\|\dot{v}_t^* - \bar{\dot{v}}^{*(1)s}\right\|^2 + \sum_{t=s+1}^{T}\left\|\dot{v}_t^* - \bar{\dot{v}}^{*(2)s}\right\|^2\right\}, \tag{18}$$

where $\bar{\dot{v}}^{*(1)s} = \frac{1}{s}\sum_{t=1}^{s}\dot{v}_t^*$ and $\bar{\dot{v}}^{*(2)s} = \frac{1}{T-s}\sum_{t=s+1}^{T}\dot{v}_t^*$.

The following two theorems establish the consistency of $\dot{\upsilon}_t^*$ and $\tilde{T}_1$, respectively.

Theorem 5.1. *Suppose that Assumptions 1–5 hold. Then* $\max_t \left\|\dot{\upsilon}_t^* - \upsilon_t^*\right\| = O_p(\eta_{N,2})$.

Theorem 5.2. *Suppose that Assumptions 1–6 hold. Then* $\mathbb{P}(\tilde{T}_1 = T_1) \to 1$ *as* $(N,T) \to \infty$.

Since the singular vectors of the slope matrices contain the structural change information, Theorem 5.2 indicates that we can consistently estimate the break point by using the factor estimates instead of the slope matrix estimates in (7). Given Theorem 5.1 and Lemma 2.1(iii), we can prove Theorem 5.2 with arguments analogous to those used in the proof of Theorem 4.2.

5.2. The case of multi-dimensional clustering

In this paper, we follow the lead of Su et al. (2016) and extend their work to allow for both IFEs (to model the strong cross-section dependence) and structural changes along the time dimension (to model time-varying latent group structures). For this reason, we consider the case where all the elements in Θ_{it}^0 as a whole share a latent group structure before and after the break point. Let $\beta_i^{(1)}$ and $\beta_i^{(2)}$ denote Θ_{it}^0 before and after the break point T_1, respectively. It is also possible to consider the case where different elements of $\beta_i^{(\ell)}$ to exhibit different latent group structures.

Cheng et al. (2021) study a multi-dimensional approach for unobserved heterogeneity in panel data models where different parameters may exhibit different group structures. Similarly, Leng et al. (2023) consider a panel quantile regression model with additive individual and time fixed effects and multi-dimensional latent group structures where the individual coefficients and slope coefficients may exhibit different group structures. It is possible to consider the clustering case as studied in these two papers. For this purpose, we write $\beta_i^{(\ell)} = \left(\beta_{1,i}^{(\ell)}, \ldots, \beta_{p,i}^{(\ell)}\right)'$ and allow $\left\{\beta_{j,i}^{(\ell)}\right\}_{i=1}^N$ to be classified into $K_j^{(\ell)}$ groups for $j \in [p]$ and $\ell \in [2]$. In principle, we can apply the STK algorithm to obtain the estimates $\left\{\dot{\beta}_{j,i}^{(\ell)}\right\}_{i=1}^N$ of $\left\{\beta_{j,i}^{(\ell)}\right\}_{i=1}^N$ to recover the different group structures for $j \in [p]$ and $\ell \in [2]$. Let $\hat{K}_j^{(\ell)}$ be the estimates of $K_j^{(\ell)}$ via the STK algorithm. Let $\left\{\hat{G}_{j,k}^{(\ell)}\right\}_{k\in[\hat{K}_j^{(\ell)}]}$ denote the corresponding estimates of the latent group structure for the jth slope coefficient at regime $\ell \in [2]$. Then we can estimate the group-specific slope coefficients $\{\alpha_{j,k}^{(\ell)}\}_{j\in[p],k_j\in[\hat{K}_j^{(\ell)}]}$ for the jth regressor along with the factors and loadings as follows

$$\left(\hat{\Lambda}^{(\ell)}, \hat{F}^{(\ell)}, \left\{\hat{\alpha}_{j,k}^{(\ell)}\right\}_{j\in[p],k\in[\hat{K}_j^{(\ell)}]}\right) = \operatorname{argmin} \mathbb{L}\left(\Lambda, F, \left\{a_k^{(\ell)}\right\}_{k\in[\hat{K}^{(\ell)}]}\right),$$

where $\mathbb{L}\left(\Lambda, F, \{\alpha_{j,k}^{(\ell)}\}_{j\in[p],k\in[\hat{K}_j^{(\ell)}]}\right) = \frac{1}{N\hat{T}_\ell}\sum_{j=1}^p \sum_{k=1}^{\hat{K}_l^{(\ell)}} \sum_{i\in\hat{G}_{j,k}^{(\ell)}} \sum_{t\in\hat{\mathcal{T}}_\ell}\left(Y_{it} - \lambda_i' f_t - \sum_{j=1}^p X_{j,it} a_{j,k}^{(\ell)}\right)^2$. As in the previous section, we can establish the consistency of $\hat{K}_j^{(\ell)}$ and $\left\{\hat{G}_{j,k}^{(\ell)}\right\}_{k\in[\hat{K}_j^{(\ell)}]}$ with obvious modifications. Difficulty arises only when we want to establish the estimators for the group-specific coefficients for each regressor as we need to introduce more complicated notations and further technical assumptions.

It is worth mentioning that the above multi-clustering case can be regarded as a special case of the one-dimensional clustering studied in Sections 3–4. To see this point clearly, focus on the case where $p = 2$ so that $\Theta_{it} = \left(\Theta_{1,it}, \Theta_{2,it}\right)'$. Suppose that

$$\Theta_{j,it}^0 = \left[\sum_{k\in[K_j^{(1)}]} \alpha_{j,k}^{(1)} \mathbf{1}\{i \in G_{j,k}^{(1)}\}\right] \mathbf{1}\{t \le T_1\} + \left[\sum_{k\in[K_j^{(2)}]} \alpha_{j,k}^{(2)} \mathbf{1}\{i \in G_{j,k}^{(2)}\}\right] \mathbf{1}\{t > T_1\},$$

where $j \in [2]$, $K_j^{(1)}$ and $K_j^{(2)}$ denote the numbers of latent groups in $\{\Theta_{j,it}^0\}_{i\in[N]}$ before and after the break point T_1, respectively, $\{G_{j,k}^{(1)}\}_{k\in K_j^{(1)}}$ (resp. $\{G_{j,k}^{(2)}\}_{k\in K_j^{(2)}}$) denotes the latent group structures for $\{\Theta_{j,it}^0\}_{i\in[N]}$ before (resp. after) the break point T_1, and $\alpha_{j,k}^{(1)}$ (resp. $\alpha_{j,k}^{(2)}$) denotes the group-specific slope coefficient before (resp. after) the break point T_1. It is easy to see that for $\ell \in [2]$, $\{G_{j,k}^{(\ell)}\}_{j\in[2],k\in K_j^{(\ell)}}$ partitions $[N]$ into $K^{(\ell)}$ groups, say $\left\{\{G_k^{(\ell)}\}_{k\in K^{(\ell)}}\right\}$, where $K^{(\ell)} \le K_1^{(\ell)} + K_2^{(\ell)} - 1$. Consequently, we can still apply the estimation procedure in Section 3 to estimate the latent group structures $\left\{\{G_k^{(\ell)}\}_{k\in K^{(\ell)}}\right\}_{\ell\in[2]}$ and the associated group-specific parameters. The consistency and asymptotic normality results obtained in Section 4 continue to apply at the cost of sacrificing some efficiency for the estimators of some group-specific parameters where the group structures are j-dependent.

6. Monte Carlo simulations

In this section, we report simulation results for the low-rank estimates, break point estimates, group membership estimates and the group number estimates based on 1000 replications, and the tuning parameter ν_j is chosen by a procedure similar to that described in Chernozhukov et al. (2020). We focus on the linear panel data model $Y_{it} = \lambda_i' f_t + X_{it}'\Theta_{it} + e_{it}$, where $X_{it} = (X_{1,it}, X_{2,it})'$ and $\Theta_{it} = (\Theta_{1,it}, \Theta_{2,it})'$.

6.1. Data generating processes (DGPs)

The following four main DGPs are employed.

DGP 1: [**Static panel with homoskedasticity**] $X_{1,it} \sim i.i.d.\ U(-2,2)$, $X_{2,it} \sim i.i.d.\ U(-2,2)$, and errors $e_{it} \sim i.i.d.\ \mathbb{N}(0,1)$. For Θ_1, we randomly choose the break point T_1 from $0.4T$ to $0.6T$.

DGP 2: [**Static panel with heteroscedasticity**] Compared to DGP 1, the errors $e_{it} \sim i.i.d.\ \mathbb{N}(0,\sigma_{it}^2)$ with $\sigma_{it}^2 \sim i.i.d.\ U(0.5,1)$. The settings for the regressors and break point are the same as those in DGP 1.

DGP 3: [**Serially correlated error**] Compared to DGP 2, the errors $e_{it} = 0.2e_{i,t-1} + \eta_{it}$, where $\eta_{it} \sim i.i.d.\ \mathbb{N}(0,1)$ and all other settings are the same as in DGP 2.

DGP 4: [**Dynamic panel**] In this case, $X_{1,it} = Y_{i,t-1}$ with $Y_{i,0} \sim i.i.d.\ \mathbb{N}(0,1)$. $X_{2,it} \sim i.i.d.\ U(-2,2)$, and $e_{it} \sim i.i.d.\ \mathbb{N}(0,0.5)$.

For each DGP above, we set $r_0 = 1$ and draw λ_i and f_t from $\mathbb{N}(0,1)$ independently. We define the slope coefficients based on three subcases below.

DGP X.1: In this case, the group membership and the number of groups do not change after the break point and only the values of the slope coefficients change. We set the number of groups to be 2, the ratio of individuals among the two groups is $N_1 : N_2 = 0.5 : 0.5$, and the group membership G_1 is obtained by a random draw from $[N]$ without replacement. For DGPs 1.1, 2.1, and 3.1,

$$\Theta_{1,it} = \Theta_{2,it} = \begin{cases} 0.1, & i \in G_1,\ t \in \{1,\dots,T_1\}, \\ 0.9, & i \in G_2,\ t \in \{1,\dots,T_1\}, \\ 0.05, & i \in G_1,\ t \in \{T_1+1,\dots,T\}, \\ 0.45, & i \in G_2,\ t \in \{T_1+1,\dots,T\}. \end{cases}$$

For DGP 4.1, $\Theta_{2,it}$ is same as other DGPs X.1 for X$\in \{1,2,3\}$, and

$$\Theta_{1,it} = \begin{cases} 0.1, & i \in G_1,\ t \in \{1,\dots,T_1\}, \\ 0.7, & i \in G_2,\ t \in \{1,\dots,T_1\}, \\ 0.05, & i \in G_1,\ t \in \{T_1+1,\dots,T\}, \\ 0.35, & i \in G_2,\ t \in \{T_1+1,\dots,T\}. \end{cases}$$

DGP X.2: Compared to DGP X.1, the values of the slope coefficients for different groups do not change after the break point, but the group membership changes. The number of groups is 2, the ratio of individuals among the groups is still $N_1 : N_2 = 0.5 : 0.5$. Nevertheless, $\{G_1^{(1)}, G_2^{(1)}\}$ is different from $\{G_1^{(2)}, G_2^{(2)}\}$ so that elements in both $G_1^{(1)}$ and $G_1^{(2)}$ are independent draws from $[N]$ without replacement. In addition, for DGPs 1.2, 2.2, and 3.2,

$$\Theta_{1,it} = \Theta_{2,it} = \begin{cases} 0.1, & i \in G_1^{(1)},\ t \in \{1,\dots,T_1\}, \\ 0.9, & i \in G_2^{(1)},\ t \in \{1,\dots,T_1\}, \\ 0.1, & i \in G_1^{(2)},\ t \in \{T_1+1,\dots,T\}, \\ 0.9, & i \in G_2^{(2)},\ t \in \{T_1+1,\dots,T\}. \end{cases}$$

For DGP 4.2, $\Theta_{2,it} = 0.5$ and

$$\Theta_{1,it} = \begin{cases} 0.1, & i \in G_1^{(1)},\ t \in \{1,\dots,T_1\}, \\ 0.7, & i \in G_2^{(1)},\ t \in \{1,\dots,T_1\}, \\ 0.1, & i \in G_1^{(2)},\ t \in \{T_1+1,\dots,T\}, \\ 0.7, & i \in G_2^{(2)},\ t \in \{T_1+1,\dots,T\}. \end{cases}$$

DGP X.3: Under this scenario, the number of groups changes after the break. We set $N_1^{(1)} : N_2^{(1)} = 0.5 : 0.5$ and $N_1^{(2)} : N_2^{(2)} : N_3^{(2)} = 0.4 : 0.3 : 0.3$ before and after the break, respectively. Specifically, for DGPs 1.3, 2.3, and 3.3, we have

$$\Theta_{1,it} = \Theta_{2,it} = \begin{cases} 0.1, & i \in G_1^{(1)},\ t \in \{1,\dots,T_1\}, \\ 0.9, & i \in G_2^{(1)},\ t \in \{1,\dots,T_1\}, \\ 0.1, & i \in G_1^{(2)},\ t \in \{T_1+1,\dots,T\}, \\ 0.5, & i \in G_2^{(2)},\ t \in \{T_1+1,\dots,T\}, \\ 0.9, & i \in G_3^{(2)},\ t \in \{T_1+1,\dots,T\}. \end{cases}$$

Table 1
Frequency of correct rank estimation.

N		100		200		N		100		200	
T		100	200	100	200	T		100	200	100	200
DGP 1.1	$r_0 = 1$	1.000	1.000	1.000	1.000	DGP 3.1	$r_0 = 1$	1.000	1.000	1.000	1.000
	$r_1 = 1$	1.000	1.000	1.000	1.000		$r_1 = 1$	1.000	1.000	1.000	1.000
	$r_2 = 1$	1.000	1.000	1.000	1.000		$r_2 = 1$	1.000	1.000	1.000	1.000
DGP 1.2	$r_0 = 1$	1.000	1.000	1.000	1.000	DGP 3.2	$r_0 = 1$	1.000	1.000	1.000	1.000
	$r_1 = 2$	1.000	1.000	1.000	1.000		$r_1 = 2$	1.000	1.000	1.000	1.000
	$r_2 = 2$	1.000	1.000	1.000	1.000		$r_2 = 2$	1.000	1.000	1.000	1.000
DGP 1.3	$r_0 = 1$	1.000	1.000	1.000	1.000	DGP 3.3	$r_0 = 1$	1.000	1.000	1.000	1.000
	$r_1 = 2$	1.000	1.000	1.000	1.000		$r_1 = 2$	1.000	1.000	1.000	1.000
	$r_2 = 2$	1.000	1.000	1.000	1.000		$r_2 = 2$	0.998	1.000	1.000	1.000
DGP 2.1	$r_0 = 1$	1.000	1.000	1.000	1.000	DGP 4.1	$r_0 = 1$	1.000	1.000	1.000	1.000
	$r_1 = 1$	1.000	1.000	1.000	1.000		$r_1 = 1$	1.000	1.000	1.000	1.000
	$r_2 = 1$	1.000	1.000	1.000	1.000		$r_2 = 1$	1.000	1.000	1.000	1.000
DGP 2.2	$r_0 = 1$	1.000	1.000	1.000	1.000	DGP 4.2	$r_0 = 1$	1.000	1.000	1.000	1.000
	$r_1 = 2$	1.000	1.000	1.000	1.000		$r_1 = 2$	1.000	1.000	1.000	1.000
	$r_2 = 2$	1.000	1.000	1.000	1.000		$r_2 = 1$	1.000	1.000	1.000	1.000
DGP 2.3	$r_0 = 1$	1.000	1.000	1.000	1.000	DGP 4.3	$r_0 = 1$	1.000	1.000	1.000	1.000
	$r_1 = 2$	1.000	1.000	1.000	1.000		$r_1 = 2$	1.000	1.000	1.000	1.000
	$r_2 = 2$	1.000	1.000	1.000	1.000		$r_2 = 1$	1.000	1.000	1.000	1.000

Table 2
Frequency of correct break point estimation.

N	100		200		N	100		200	
T	100	200	100	200	T	100	200	100	200
DGP 1.1	0.980	0.993	1.000	1.000	DGP 3.1	0.985	0.972	1.000	0.999
DGP 1.2	0.999	1.000	1.000	1.000	DGP 3.2	1.000	1.000	1.000	1.000
DGP 1.3	1.000	1.000	1.000	1.000	DGP 3.3	1.000	1.000	1.000	1.000
DGP 2.1	0.998	0.999	1.000	1.000	DGP 4.1	1.000	1.000	1.000	1.000
DGP 2.2	1.000	1.000	1.000	1.000	DGP 4.2	1.000	1.000	1.000	1.000
DGP 2.3	1.000	1.000	1.000	1.000	DGP 4.3	1.000	1.000	1.000	1.000

For DGP 4.3, $\Theta_{2,it} = 0.5$ and

$$\Theta_{1,it} = \begin{cases} 0.1, & i \in G_1^{(1)},\ t \in \{1, \dots, T_1\}, \\ 0.7, & i \in G_2^{(1)},\ t \in \{1, \dots, T_1\}, \\ 0.1, & i \in G_1^{(2)},\ t \in \{T_1+1, \dots, T\}, \\ 0.4, & i \in G_2^{(2)},\ t \in \{T_1+1, \dots, T\}, \\ 0.7, & i \in G_3^{(2)},\ t \in \{T_1+1, \dots, T\}. \end{cases}$$

6.2. Results

Table 1 reports the proportion of correct rank estimation for the intercept (IFE) and slope matrices based on the SVT in Section 3.3. Note that r_0 denotes the true rank of the intercept matrix and r_1 and r_2 denote those of the two slope matrices. From Table 1, we notice that the true ranks of both the intercept and slope matrices can be almost perfectly estimated for the sample sizes under investigation.

Table 2 reports the results for break point estimation in Step 3 based on some different (N, T) combinations. We summarize some important findings from Table 2. First, when the group membership and the number of groups do not change as in DGP X.1 for X$\in$ [4], the frequency of correct break point estimation may not be 1 especially if N is not large. This suggests that the binary segmentation does not work perfectly in such a scenario. Second, the change of group membership or the number of groups help to identify the break point as reflected in the simulation results for DGP X.2 and X.3 for X$\in$ [4]. In general, the binary segmentation works well in our setting.

Table 3 reports the results for the group membership estimation when the number of groups are either known (infeasible in practice) or estimated from the data (feasible). Define the frequency of correct group membership estimation in a replication as

$$\frac{1}{N} \sum_{i \in [N]} \mathbf{1}\left\{ \hat{g}_i^{(\ell)} = g_i^{(\ell)} \right\} \text{ for } \ell \in \{1, 2\},$$

where $\hat{g}_i^{(\ell)}$ and $g_i^{(\ell)}$ are the estimated group label and true group label for individual i. The above values are then averaged over 1000 replications to obtain the frequency of correct group membership estimation in 1000 replications. With known number of

Table 3
Frequency of correct group membership estimation.

	N		100		200			N		100		200	
	T		100	200	100	200		T		100	200	100	200
Infeasible	DGP 1.1	G_B	1.000	1.000	1.000	1.000	Feasible	DGP 1.1	G_B	1.000	1.000	1.000	1.000
		G_A	1.000	1.000	1.000	1.000			G_A	1.000	1.000	1.000	1.000
	DGP 1.2	G_B	1.000	1.000	1.000	1.000		DGP 1.2	G_B	1.000	1.000	1.000	1.000
		G_A	1.000	1.000	1.000	1.000			G_A	1.000	1.000	1.000	1.000
	DGP 1.3	G_B	1.000	1.000	1.000	1.000		DGP 1.3	G_B	1.000	1.000	1.000	1.000
		G_A	0.989	0.999	0.978	0.999			G_A	0.989	0.999	0.978	0.999
	DGP 2.1	G_B	1.000	1.000	1.000	1.000		DGP 2.1	G_B	1.000	1.000	1.000	1.000
		G_A	1.000	1.000	1.000	1.000			G_A	1.000	1.000	1.000	1.000
	DGP 2.2	G_B	1.000	1.000	1.000	1.000		DGP 2.2	G_B	0.989	0.999	0.992	0.999
		G_A	1.000	1.000	1.000	1.000			G_A	0.992	0.999	0.977	0.998
	DGP 2.3	G_B	1.000	1.000	1.000	1.000		DGP 2.3	G_B	0.992	0.999	0.961	0.999
		G_A	0.998	1.000	0.999	1.000			G_A	0.989	0.999	0.992	0.999
	DGP 3.1	G_B	1.000	1.000	1.000	1.000		DGP 3.1	G_B	0.981	0.999	0.949	0.999
		G_A	1.000	1.000	1.000	1.000			G_A	0.981	0.993	0.979	0.996
	DGP 3.2	G_B	1.000	1.000	1.000	1.000		DGP 3.2	G_B	0.985	0.996	0.962	0.993
		G_A	1.000	1.000	1.000	1.000			G_A	0.985	0.994	0.973	0.998
	DGP 3.3	G_B	1.000	1.000	1.000	1.000		DGP 3.3	G_B	0.985	0.998	0.973	0.995
		G_A	0.981	0.997	0.982	0.999			G_A	0.971	0.994	0.968	0.998
	DGP 4.1	G_B	1.000	1.000	1.000	1.000		DGP 4.1	G_B	0.975	0.999	0.984	0.999
		G_A	1.000	1.000	1.000	1.000			G_A	0.985	0.998	0.949	0.997
	DGP 4.2	G_B	1.000	1.000	1.000	1.000		DGP 4.2	G_B	0.972	0.996	0.964	0.996
		G_A	1.000	1.000	1.000	1.000			G_A	0.989	0.998	0.967	0.998
	DGP 4.3	G_B	1.000	1.000	1.000	1.000		DGP 4.3	G_B	0.981	0.994	0.942	0.995
		G_A	0.981	1.000	0.988	0.999			G_A	0.958	0.997	0.964	0.997

Table 4
Frequency of correct estimation of the number of groups.

N		100		200		N		100		200	
T		100	200	100	200	T		100	200	100	200
DGP 1.1	$K^{(1)}=2$	1.000	1.000	0.999	1.000	DGP 3.1	$K^{(1)}=2$	0.880	0.993	0.675	0.985
	$K^{(2)}=2$	1.000	1.000	1.000	1.000		$K^{(2)}=2$	0.890	0.960	0.873	0.971
DGP 1.2	$K^{(1)}=2$	1.000	1.000	1.000	1.000	DGP 3.2	$K^{(1)}=2$	0.868	0.985	0.759	0.940
	$K^{(2)}=2$	1.000	1.000	1.000	0.999		$K^{(2)}=2$	0.897	0.971	0.829	0.987
DGP 1.3	$K^{(1)}=2$	0.999	1.000	1.000	1.000	DGP 3.3	$K^{(1)}=2$	0.889	0.988	0.802	0.965
	$K^{(2)}=3$	1.000	0.999	1.000	1.000		$K^{(2)}=3$	0.932	0.977	0.907	0.988
DGP 2.1	$K^{(1)}=2$	1.000	1.000	1.000	1.000	DGP 4.1	$K^{(1)}=2$	0.807	0.981	0.825	0.982
	$K^{(2)}=2$	1.000	1.000	1.000	1.000		$K^{(2)}=2$	0.919	0.988	0.714	0.980
DGP 2.2	$K^{(1)}=2$	0.919	0.995	0.940	0.994	DGP 4.2	$K^{(1)}=2$	0.794	0.962	0.754	0.974
	$K^{(2)}=2$	0.930	0.993	0.809	0.982		$K^{(2)}=2$	0.918	0.988	0.779	0.988
DGP 2.3	$K^{(1)}=2$	0.940	0.989	0.724	0.990	DGP 4.3	$K^{(1)}=2$	0.922	0.978	0.784	0.980
	$K^{(2)}=3$	0.946	0.995	0.952	0.992		$K^{(2)}=3$	0.860	0.980	0.845	0.983

groups, the STK algorithm degenerates to the traditional K-means algorithm. The "Infeasible" part of Table 3 reports the frequency of correct group membership estimation before and after the estimated break point, G_B and G_A, based on the known true number of groups and K-means algorithm. Evidently, the K-means classification exhibits excellent performance in this case. Nevertheless, without prior information on the true number of groups, the STK algorithm is able to estimate the group membership and the number of groups simultaneously. In this case, the frequency of correct estimation of the group membership and that of the number of groups are shown in the "Feasible" part in Table 3 and in Table 4, respectively. To implement the STK algorithm with unknown number of groups, we set $\varsigma_N = N^{-2}$ to ensure the consistency of the group number estimators. As expected, the performance of the STK algorithm is slightly worse than that of the K-means algorithm with knowledge of the true number of groups. But the performance improves when both N and T increase. Table 4 suggests that the number of groups can be nearly perfectly estimated in DGPs 1.1, 1.2, 1.3 and 2.1. For the more complicated DGPs (e.g., the dynamic case in DGPs 4.1, 4.2, and 4.3 or the static panel with serially correlated errors in DGPs 3.1, 3.2, and 3.3), the performance is not as good as that in the simple DGPs.

Table 5 presents more detailed results for the estimation of the number of groups. For DGPs 1.X and DGP 2.X where we have static panels with independent errors, the results show that the group membership and the number of groups can be well estimated with nearly 100% accuracy under different (N,T) combinations. For DGPs 3.X and 4.X where we have static panels with serially correlated errors and dynamic panels, respectively, the frequency of correct estimation of both the group membership and the

Table 5
Determination of the number of groups.

DGP	N	T	$\hat{K}^{(1)}$					$\hat{K}^{(2)}$				
			1	2	3	4	≥5	1	2	3	4	≥5
DGP 1.1	100	100	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
		200	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
	200	100	0.000	**0.999**	0.001	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
		200	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
DGP 1.2	100	100	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
		200	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
	200	100	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
		200	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
DGP 1.3	100	100	0.000	**0.999**	0.001	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000
		200	0.000	**1.000**	0.000	0.000	0.000	0.000	0.000	**0.999**	0.001	0.000
	200	100	0.000	**1.000**	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000
		200	0.000	**1.000**	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000
DGP 2.1	100	100	0.000	**0.933**	0.058	0.009	0.000	0.000	**0.936**	0.060	0.003	0.001
		200	0.000	**0.990**	0.010	0.000	0.000	0.000	**0.987**	0.013	0.000	0.000
	200	100	0.000	**0.864**	0.126	0.010	0.000	0.000	**0.901**	0.090	0.009	0.000
		200	0.000	**0.989**	0.011	0.000	0.000	0.000	**0.990**	0.010	0.000	0.000
DGP 2.2	100	100	0.000	**0.919**	0.074	0.007	0.000	0.000	**0.930**	0.067	0.003	0.000
		200	0.000	**0.995**	0.003	0.000	0.002	0.000	**0.993**	0.006	0.000	0.001
	200	100	0.000	**0.940**	0.056	0.004	0.000	0.000	**0.809**	0.164	0.027	0.000
		200	0.000	**0.994**	0.006	0.000	0.000	0.000	**0.982**	0.018	0.000	0.000
DGP 2.3	100	100	0.000	**0.940**	0.055	0.005	0.000	0.000	0.000	**0.946**	0.039	0.015
		200	0.000	**0.989**	0.011	0.000	0.000	0.000	0.000	**0.995**	0.002	0.003
	200	100	0.000	**0.724**	0.230	0.046	0.000	0.000	0.000	**0.952**	0.031	0.017
		200	0.000	**0.990**	0.010	0.000	0.000	0.000	0.000	**0.992**	0.006	0.002
DGP 3.1	100	100	0.000	**0.880**	0.097	0.022	0.001	0.000	**0.890**	0.062	0.031	0.017
		200	0.000	**0.993**	0.007	0.000	0.000	0.000	**0.960**	0.019	0.012	0.009
	200	100	0.000	**0.675**	0.224	0.099	0.002	0.000	**0.873**	0.081	0.041	0.005
		200	0.000	**0.985**	0.015	0.000	0.000	0.000	**0.971**	0.023	0.005	0.001
DGP 3.2	100	100	0.000	**0.868**	0.109	0.023	0.000	0.000	**0.897**	0.099	0.004	0.000
		200	0.000	**0.985**	0.008	0.003	0.004	0.000	**0.971**	0.021	0.006	0.002
	200	100	0.000	**0.759**	0.198	0.042	0.001	0.000	**0.829**	0.147	0.024	0.000
		200	0.000	**0.940**	0.055	0.005	0.000	0.000	**0.987**	0.013	0.000	0.000
DGP 3.3	100	100	0.000	**0.889**	0.100	0.011	0.000	0.000	0.000	**0.932**	0.055	0.013
		200	0.000	**0.988**	0.009	0.003	0.000	0.000	0.000	**0.977**	0.013	0.010
	200	100	0.000	**0.802**	0.175	0.023	0.000	0.000	0.000	**0.907**	0.073	0.020
		200	0.000	**0.965**	0.035	0.000	0.000	0.000	0.000	**0.988**	0.010	0.002
DGP 4.1	100	100	0.000	**0.807**	0.084	0.089	0.020	0.000	**0.919**	0.041	0.019	0.021
		200	0.000	**0.981**	0.013	0.004	0.002	0.000	**0.988**	0.004	0.005	0.003
	200	100	0.000	**0.825**	0.107	0.061	0.007	0.000	**0.714**	0.118	0.084	0.084
		200	0.000	**0.982**	0.011	0.006	0.001	0.000	**0.980**	0.010	0.004	0.006
DGP 4.2	100	100	0.000	**0.794**	0.152	0.047	0.007	0.000	**0.918**	0.063	0.015	0.004
		200	0.000	**0.962**	0.034	0.004	0.000	0.000	**0.988**	0.009	0.001	0.002
	200	100	0.000	**0.754**	0.195	0.048	0.003	0.000	**0.779**	0.175	0.046	0.000
		200	0.000	**0.974**	0.022	0.004	0.000	0.000	**0.988**	0.011	0.001	0.000
DGP 4.3	100	100	0.000	**0.922**	0.055	0.018	0.005	0.000	0.000	**0.860**	0.087	0.053
		200	0.000	**0.978**	0.015	0.007	0.000	0.000	0.000	**0.980**	0.016	0.004
	200	100	0.000	**0.784**	0.133	0.080	0.003	0.000	0.000	**0.845**	0.116	0.039
		200	0.000	**0.980**	0.015	0.005	0.000	0.000	0.000	**0.983**	0.011	0.006

number of groups is not great when T is small, but gradually approaches unity as T increases. One reason for this is the larger bias of the slope estimator caused by the presence of lagged dependent variables or serially correlated errors. Though the biases can be removed asymptotically, their estimation still has some impact on the finite sample performance. The other reason is that we need to use HAC estimates of certain long-run variance objects in the STK algorithm and it is well known that a relatively large value of T is required in order for the HAC estimates to be reasonably well behaved in finite samples.

Tables 6 and 7 show results for the post-classification estimates for the first and second slope coefficients, respectively. We follow (Su et al., 2016) to define the evaluation criteria as bias and coverage. Specifically, we define the bias to be the weighted versions of bias for slope estimator from all estimated groups, i.e. Bias $= \sum_{k=1}^{K^{(\ell)}} \text{Bias}(\alpha_{k,1}^{(\ell)})$ for $\ell \in \{1, 2\}$. Similarly, we define the weighted version of the coverage ratio of the 95% confidence interval estimators. The "Infeasible" panel shows the result assuming the number of groups is known, and the "Feasible" panel shows the result without knowing the number of groups. From Tables 6 and 7, we notice that the coverage ratio for DGPs 1 and 2 is close to 95% under different combinations of N and T for both the "Infeasible" and "Feasible" panels, which is due to the higher correct classification ratio. For DGPs 3 and 4, by using the STK algorithm, the coverage ratio is a bit lower for $T = 100$, which is due to the inaccuracy of the group number and membership estimators; but the coverage probabilities approaches 0.95 quickly when T doubles.

Table 6
Point estimation of $\alpha^{(1)}_{\cdot,1}$ and $\alpha^{(2)}_{\cdot,1}$.

DGP	N	T	Infeasible				Feasible			
			Before the break		After the break		Before the break		After the break	
			Bias ($\times 10^6$)	Coverage	Bias ($\times 10^6$)	Coverage	Bias ($\times 10^6$)	Coverage	Bias ($\times 10^6$)	Coverage
1.1	100	100	2.585	0.951	−2.869	0.946	2.585	0.951	−2.869	0.946
		200	−1.944	0.944	−8.920	0.945	−1.958	0.944	−8.920	0.945
	200	100	−1.096	0.943	1.407	0.947	−1.096	0.943	1.407	0.947
		200	−1.910	0.945	0.960	0.947	−1.910	0.945	0.960	0.947
1.2	100	100	−1.050	0.949	−27.398	0.941	−1.050	0.949	−27.398	0.941
		200	−5.449	0.930	7.616	0.953	−5.449	0.930	7.655	0.953
	200	100	4.770	0.949	1.866	0.951	4.770	0.949	1.866	0.951
		200	1.317	0.941	1.874	0.945	1.317	0.941	1.874	0.945
1.3	100	100	−0.961	0.943	11.417	0.944	−1.050	0.949	−27.398	0.941
		200	−4.213	0.951	−5.002	0.941	−5.449	0.930	7.655	0.953
	200	100	−1.571	0.938	−3.756	0.938	4.770	0.949	1.866	0.951
		200	0.403	0.941	−4.159	0.945	1.317	0.941	1.874	0.945
2.1	100	100	14.840	0.944	9.410	0.950	14.816	0.943	9.406	0.950
		200	−7.222	0.951	1.795	0.951	−7.222	0.951	1.795	0.951
	200	100	0.916	0.940	3.575	0.948	0.916	0.940	3.575	0.948
		200	0.452	0.948	−0.797	0.947	0.452	0.948	−0.797	0.947
2.2	100	100	−21.379	0.946	0.234	0.937	−21.379	0.946	0.234	0.937
		200	0.264	0.942	−15.542	0.953	0.264	0.942	−15.542	0.953
	200	100	−1.379	0.945	−1.489	0.951	−1.379	0.944	−1.489	0.951
		200	−1.101	0.950	1.127	0.949	−1.101	0.950	1.127	0.949
2.3	100	100	−8.610	0.945	5.254	0.952	−8.610	0.945	5.261	0.952
		200	0.927	0.949	5.840	0.949	0.927	0.949	5.840	0.949
	200	100	−1.560	0.943	−2.569	0.941	−1.560	0.943	−2.569	0.941
		200	−0.775	0.947	4.408	0.947	−0.775	0.947	4.386	0.947
3.1	100	100	−20.928	0.955	−73.947	0.945	−26.250	0.927	−77.613	0.920
		200	3.066	0.949	−12.443	0.937	2.884	0.940	−13.116	0.934
	200	100	−2.663	0.951	−8.742	0.944	−3.517	0.857	−7.730	0.888
		200	−3.747	0.949	−2.107	0.945	−3.642	0.939	−1.971	0.938
3.2	100	100	−55.980	0.952	−10.846	0.943	−58.714	0.926	−15.109	0.863
		200	−2.774	0.950	4.690	0.946	−3.218	0.945	4.913	0.942
	200	100	6.979	0.951	8.879	0.945	6.287	0.858	6.894	0.848
		200	−1.704	0.947	0.438	0.945	−2.122	0.928	0.381	0.940
3.3	100	100	−25.340	0.950	37.639	0.907	−29.905	0.924	37.016	0.890
		200	2.042	0.947	−6.431	0.960	1.667	0.940	−6.245	0.960
	200	100	−2.391	0.946	14.364	0.892	−2.735	0.891	13.680	0.840
		200	4.339	0.943	4.890	0.942	4.493	0.932	5.113	0.938
4.1	100	100	800.620	0.930	−466.590	0.929	777.650	0.928	−454.980	0.924
		200	126.760	0.931	550.210	0.942	126.220	0.942	548.160	0.943
	200	100	−224.960	0.931	−339.020	0.939	−214.900	0.904	−313.430	0.876
		200	417.320	0.938	412.500	0.947	415.110	0.941	410.700	0.944
4.2	100	100	−3.211	0.944	6.014	0.951	−4.703	0.825	8.954	0.905
		200	6.454	0.942	0.517	0.946	6.842	0.923	0.522	0.942
	200	100	1.971	0.940	4.962	0.936	1.368	0.801	4.866	0.808
		200	0.255	0.947	−0.288	0.957	0.267	0.932	−0.300	0.951
4.3	100	100	−4.827	0.949	6.005	0.916	−4.505	0.903	7.458	0.855
		200	−3.008	0.941	−4.816	0.943	−3.223	0.927	−5.046	0.934
	200	100	3.672	0.943	0.956	0.891	2.999	0.794	0.904	0.822
		200	−1.348	0.944	0.737	0.943	−1.340	0.931	0.801	0.935

7. Empirical study

In this section we apply the proposed model to analyze the time-varying latent group structure of real house price changes in Metropolitan Statistical Areas (MSAs) in the United States.

7.1. Model

Studies of U.S. house price changes are plentiful in the literature. Malpezzi (1999), Capozza et al. (2002), Gallin (2006), and Ortalo-Magne and Rady (2006) all show that the house price changes are closely correlated with real income in the long run. Su et al. (2023) consider a heterogeneous spatial panel and show that real income growth affects the U.S. house prices in different ways

Table 7
Point estimation of $\alpha^{(1)}_{\cdot,2}$ and $\alpha^{(2)}_{\cdot,2}$.

DGP	N	T	Infeasible				Feasible			
			Before the break		After the break		Before the break		After the break	
			Bias ($\times 10^6$)	Coverage	Bias ($\times 10^6$)	Coverage	Bias ($\times 10^6$)	Coverage	Bias ($\times 10^6$)	Coverage
1.1	100	100	−4.719	0.940	−6.319	0.946	−4.719	0.940	−6.319	0.946
		200	5.155	0.946	14.429	0.945	5.142	0.945	14.429	0.945
	200	100	−2.820	0.950	1.496	0.948	−2.820	0.950	1.496	0.948
		200	3.204	0.945	0.981	0.954	3.204	0.945	0.981	0.954
1.2	100	100	8.533	0.943	−11.589	0.946	8.533	0.943	−11.589	0.946
		200	8.833	0.944	9.115	0.934	8.833	0.944	9.129	0.934
	200	100	−1.538	0.937	0.576	0.936	−1.538	0.937	0.576	0.936
		200	0.995	0.948	−0.492	0.954	0.995	0.948	−0.492	0.954
1.3	100	100	−8.825	0.942	−32.224	0.945	8.533	0.943	−11.589	0.946
		200	−8.965	0.948	−2.501	0.945	8.833	0.944	9.129	0.934
	200	100	−2.725	0.949	0.993	0.938	−1.538	0.937	0.576	0.936
		200	0.106	0.951	2.084	0.942	0.995	0.948	−0.492	0.954
2.1	100	100	8.126	0.942	9.711	0.947	8.189	0.941	9.746	0.946
		200	−2.655	0.940	9.249	0.950	−2.655	0.940	9.249	0.950
	200	100	−1.002	0.950	1.024	0.953	−1.002	0.950	1.024	0.953
		200	−1.151	0.944	1.497	0.948	−1.151	0.944	1.497	0.948
2.2	100	100	9.206	0.942	−4.963	0.952	9.206	0.942	−4.963	0.952
		200	−0.573	0.946	1.023	0.936	−0.573	0.946	1.023	0.936
	200	100	−0.595	0.938	−0.130	0.952	−0.587	0.938	−0.130	0.952
		200	−0.744	0.950	0.621	0.953	−0.744	0.950	0.621	0.953
2.3	100	100	3.676	0.940	1.835	0.940	3.676	0.940	1.839	0.940
		200	1.101	0.945	−2.676	0.944	1.101	0.945	−2.676	0.944
	200	100	1.829	0.949	−2.656	0.944	1.829	0.949	−2.656	0.944
		200	−1.445	0.947	−5.561	0.943	−1.445	0.947	−5.576	0.943
3.1	100	100	−11.637	0.955	1.353	0.945	−14.410	0.927	4.101	0.920
		200	16.489	0.949	−17.001	0.937	17.287	0.940	−18.482	0.934
	200	100	17.505	0.951	−3.263	0.944	15.237	0.857	−3.898	0.888
		200	−2.454	0.949	1.127	0.945	−2.526	0.939	1.170	0.938
3.2	100	100	2.483	0.952	37.786	0.943	3.202	0.926	40.094	0.863
		200	−10.538	0.950	−3.229	0.946	−11.557	0.945	−3.690	0.942
	200	100	−1.127	0.951	5.599	0.945	0.515	0.858	5.876	0.848
		200	−3.999	0.947	1.501	0.945	−4.038	0.928	1.419	0.940
3.3	100	100	4.883	0.950	36.994	0.907	6.623	0.924	37.547	0.890
		200	−2.816	0.947	−36.774	0.960	−2.898	0.940	−37.607	0.960
	200	100	0.862	0.946	1.760	0.892	0.893	0.891	0.970	0.840
		200	−1.075	0.943	−0.074	0.942	−1.272	0.932	0.098	0.938
4.1	100	100	−0.701	0.954	−13.215	0.942	−1.035	0.928	−11.774	0.924
		200	1.478	0.945	−1.950	0.947	1.918	0.942	−2.076	0.943
	200	100	−5.307	0.949	7.131	0.942	−4.162	0.904	8.046	0.876
		200	−1.571	0.947	0.887	0.946	−1.572	0.941	0.825	0.944
4.2	100	100	−103.180	0.942	−38.453	0.935	−96.421	0.944	−33.687	0.939
		200	6.837	0.942	1.376	0.941	6.480	0.946	0.894	0.945
	200	100	−9.551	0.924	−7.917	0.911	−8.665	0.941	−9.627	0.932
		200	21.592	0.945	−14.624	0.945	21.663	0.951	−14.520	0.950
4.3	100	100	−62.162	0.906	59.367	0.901	−59.560	0.931	54.449	0.928
		200	−17.481	0.942	−13.186	0.942	−17.536	0.939	−13.243	0.939
	200	100	5.517	0.925	83.088	0.915	5.671	0.944	79.796	0.940
		200	−15.681	0.947	6.905	0.945	−15.653	0.945	6.860	0.944

for different MSAs. In this application, we examine whether there exist latent group structures for the real income growth elasticity of house price changes and whether these structures change over the time dimension. More specifically, we consider the following panel data model with IFEs and two-way slope heterogeneity

$$\pi_{it} = \lambda_i' f_t + \Theta_{1,it} ginc_{it} + \Theta_{2,it} ginc_{i,t-1} + e_{it}, \tag{19}$$

where the dependent variable π_{it} measures the percentage of real house price growth for MSA i at time period t. The λ_i and f_t are the individual fixed effects and time fixed effects, the covariate $ginc_{it}$ denotes the percentage of income growth for MSA i at time period t, and $ginc_{i,t-1}$ is the lagged value of $ginc_{it}$. Unlike Aquaro et al. (2021) and Su et al. (2023) who consider individual fixed effects and additive two-way fixed effects, respectively, we allow the model to have IFEs. In the above model, we allow the slope

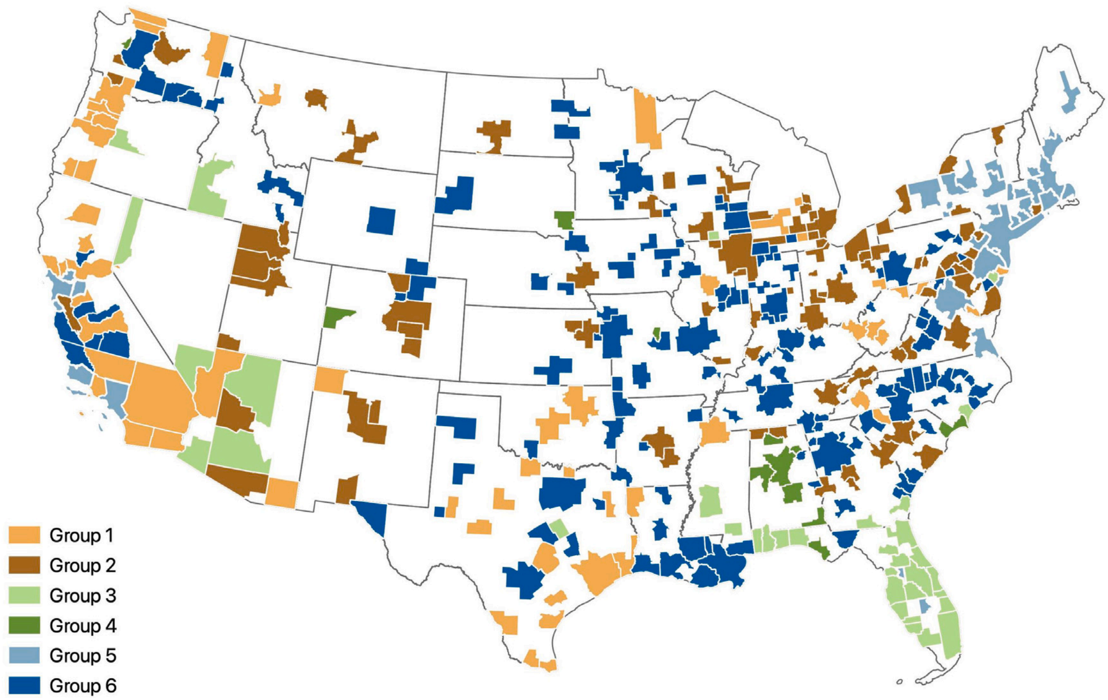

Fig. 2. Group classification result 1975Q3-1987Q4.

parameters $\left(\Theta_{1,it}, \Theta_{2,it}\right)$ to exhibit latent group structures along the cross-sectional dimension and an unknown break along the time dimension.

7.2. Data

The dataset is obtained from Aquaro et al. (2021), which contains the quarterly data for 377 MSAs over 1975 to 2014. To construct the growth rate and the lagged term, we lose two periods of observations, giving $T = 158$. Similar to Su et al. (2023), we deseasonalize the growth rate of real house price and real income. We do not de-factor the variables since our model contains IFEs to control the common shocks.

7.3. Empirical results

We first apply the SVT to estimate the ranks of $\Theta_0 = \{\lambda_i' f_t\}$, $\Theta_1 = \{\Theta_{1,it}\}$ and $\Theta_2 = \{\Theta_{2,it}\}$. The estimates are: $\hat{r}_0 = 1$, $\hat{r}_1 = 2$, and $\hat{r}_2 = 2$. Before applying the proposed estimation algorithm in Section 3, we first test the presence of a structural break as in Section F.1 of the online supplement. We first construct the sup-F test statistic $\sup_{T_1 \in \mathcal{T}_\epsilon} F_i(T_1)$ for each MSA, and then obtain the final test statistic as $F_{NT}(1|0) = \max_{i \in [N]} \sup_{T_1 \in \mathcal{T}_\epsilon} F_i(T_1) = 2161.65$. As given in Bai and Perron (1998), the 1% critical value for the sup-F test statistic for an individual time series regression is 19.25. Even though we have not formally established the asymptotic distribution of $F_{NT}(1|0)$, its value for our data is much larger than the 1% upper-percentile of the supremum of 377 independent variables that share the null distribution of $\sup_{T_1 \in \mathcal{T}_\epsilon} F_i(T_1)$ as formally given by Bai and Perron (1998). This provides strong evidence for the rejection of the null that there is no structural break in the slope coefficients $\left\{\Theta_{1,it}, \Theta_{2,it}\right\}$.

Assuming the presence of a structural break, we apply the proposed multi-stage estimation result in Section 3 to estimate the break date and numbers of groups before and after the break.[3] The estimated break date is given by $\hat{T}_1 = 51$, which suggests that the structural break happens at the first quarter in 1988. We conjecture that this break may be related to the catastrophic stock market crash that occurred on October 1987, which is considered to be the first contemporary global financial crisis event.

By setting $\varsigma_N = N^{-2}$ for the STK algorithm as in the simulations, we obtain the estimated prior- and post-break numbers of groups given by $\hat{K}^{(1)} = 6$ and $\hat{K}^{(2)} = 2$, respectively.[4] As for the group structure, Figs. 2 and 3 use six and two different types of

[3] In the presence of multiple breaks, we conjecture we can either estimate the multiple breaks simultaneously or estimate one break at a time via a sequential approach.

[4] Fig. 1 indicates that the number of groups is estimated via the homogeneity test. So our algorithm does not require one to set the maximum number of groups per se, which is quite different from existing approaches based on either K-means or C-Lasso. Nevertheless, in empirical applications, one can set the

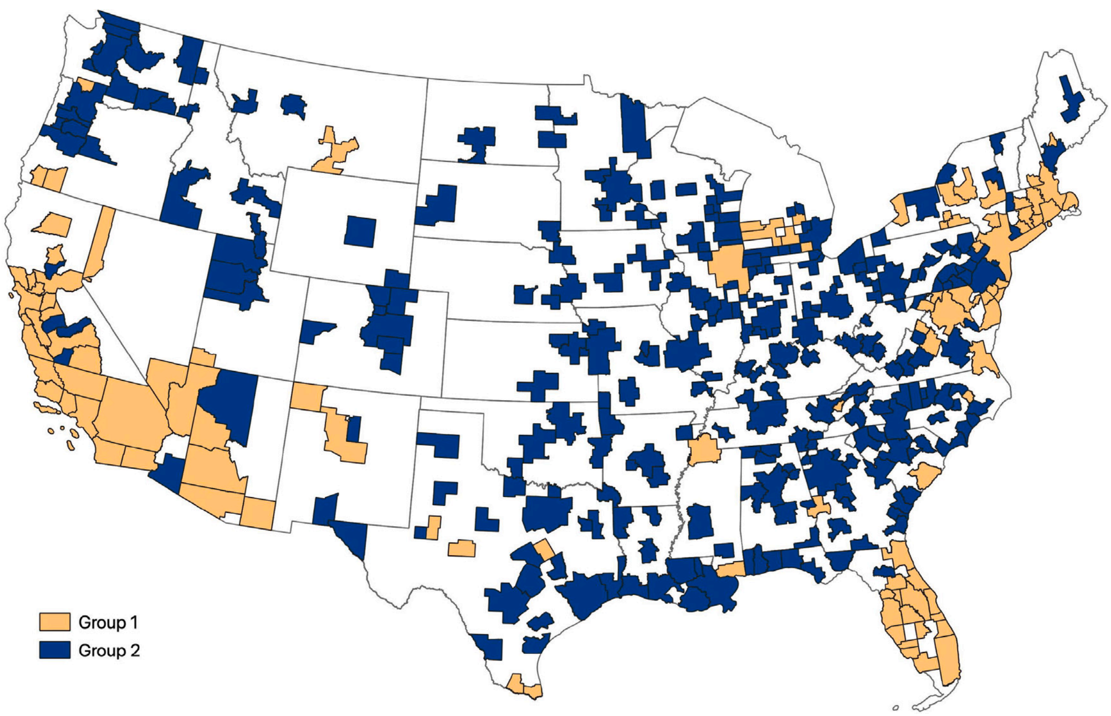

Fig. 3. Group classification result 1988Q1-2014Q4.

Table 8
Results for the pooled regressions.

	Pooled (full sample) (1)	Pooled $(1975Q3-1987Q4)$ (2)	Pooled $(1988Q1-2014Q4)$ (3)
$ginc_{it}$	0.1021***	0.0904***	0.0702***
	(0.0067)	(0.0119)	(0.0065)
$ginc_{i,t-1}$	0.0590***	0.0392***	0.0401***
	(0.0067)	(0.0117)	(0.0066)
#individuals	377	377	377

Note: Column (1) reports the pooled regression results for the full sample. Columns (2) and (3) report the pooled regression results for the subsamples before and after the estimated break point, respectively. Slope estimates are all bias-corrected with the analytical form shown in Theorem 4.4. Values in parentheses are standard errors and *** indicates significance at 1% level.

color to show the classification results for the 377 MSAs during 1975Q3 to 1987Q4 and 1988Q1 to 2014Q4, respectively. Table 8 reports the pooled regression results for the full sample in column (1), the subsample before the estimated break point in column (2), and the subsample after the estimated break point in column (3). All the slope estimators are bias-corrected following the analytical form shown in Theorem 4.4. The pooled regression results in Table 8 show that real income growth has positive and significant effect on house prices. Comparing the two subsamples before and after the estimated break, we observe that, with a 1 percentage increase in real income growth, the real house price growth rate will increase 0.09 percentage before 1988, which is 0.02 percentage points higher than that after 1988. The slope estimates for the lagged term are similar for the two subsamples.

To examine the difference for each of the 6 estimated groups before the break, Table 9 reports the post-classification regression results for each estimated group before the estimated break. Even though the effects of real income growth are positive for all estimated groups, they differ vastly across groups. The effect of real income growth for Group 2 is the highest, followed by Groups 5 and 3, and the effects of real income growth on real house prices in all of these three groups are higher than 0.15 percent. In contrast, the effects of real income growth for the remaining three groups, viz., Groups 1, 4, and 6, are less than 0.07 percent. Similarly, Table 10 reports the post-classification regression results for each estimated group after the estimated break. The estimated

maximum number of groups that serves as an upper bound for the algorithm to stop. Here we set the maximum number of groups to be 10, which is not reached in the iteration.

Table 9
Results for the post-classification regressions before the break.

	Group 1 (1)	Group 2 (2)	Group 3 (3)	Group 4 (4)	Group 5 (5)	Group 6 (6)
$ginc_{it}$	0.0301	0.3169***	0.1522***	0.0168	0.1877**	0.0661***
	(0.0345)	(0.0292)	(0.0408)	(0.0561)	(0.0775)	(0.0153)
$ginc_{i,t-1}$	0.1217***	−0.0191	−0.0089	−0.0298	−0.1269*	0.0331***
	(0.0348)	(0.0288)	(0.0407)	(0.0560)	(0.0754)	(0.0151)
#individuals	60	92	35	12	36	142

Note: Each column reports the regression results for each estimated group during 1977Q3-1987Q4. Slope estimates are all bias-corrected with the analytical form shown in Theorem 4.4. Values in parentheses are standard errors. ***, ** , and * indicate significance at 1% level, 5% level, 10% level, respectively.

Table 10
Results for the post-classification regressions after the break.

	Group 1 (1)	Group 2 (2)
$ginc_{it}$	0.0670***	0.0714***
	(0.0130)	(0.0079)
$ginc_{i,t-1}$	0.0870***	0.0275***
	(0.0135)	(0.0079)
#individuals	103	274

Note: Each column reports the regression results for each estimated group during 1988Q1-2014Q4. Slope estimators are all bias-corrected. Values in parentheses are standard errors. *** indicates significance at 1% level.

slope coefficients for both groups are statistically significant. Especially, during 1988Q1-2014Q1, the slope estimate for the lagged term in the first group is much higher than that for the second group.

We also applied the C-Lasso algorithm of Su et al. (2016) to estimate the group structure before and after the estimated break point. The C-Lasso approach in conjunction with IC detects 2 groups both before and after the break. In view of the six groups detected by our present algorithm, we conjecture that the difference may due to the smaller time periods before the break.

8. Conclusion

This paper considers a linear panel data model with interactive fixed effects and two-way heterogeneity such that the heterogeneity across individuals is captured by latent group structures and the heterogeneity across time is captured by an unknown structural break. We allow the model to have different group numbers, or different group memberships, or just changes in the slope coefficients for some specific groups before and after the break. To estimate the unknown structural break, the number of groups and group memberships before and after the break point, we propose an estimation algorithm with initial nuclear norm regularized estimates, followed by row- and column-wise linear regressions. Then, the break point estimator is obtained by binary segmentation and the group structure together with the number of groups are estimated simultaneously using a sequential testing K-means algorithm. We show that the structural break estimator, the group number estimators, and the group membership estimators before and after the break point are all consistent, and the final post-classification slope coefficient estimators enjoy the oracle property.

There are several interesting topics for further research. First, even though we discuss a possible test for the existence of a break in the panel data models with latent group structures in the online supplement, we have not fully worked out the asymptotic theory, a challenge that deserves separate treatment. Second, we assume the presence of a single break in the data and it is interesting to extend our theory to allow for multiple breaks. Third, the present treatment rules out both unit-root-type nonstationarity and nonstochastic trending nonstationarity and it is interesting to extend our theory to allow for nonstationarity in the data. We will pursue these topics in future research.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Liangjun Su reports financial support was provided by National Natural Science Foundation of China. Peter C.B. Phillips reports financial support was provided by National Science Foundation. Peter C.B. Phillips reports financial support was provided by Kelly Fund at the University of Auckland. Peter C.B. Phillips reports financial support was provided by Lee Kong Chian Fellowship at Singapore Management University.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2024.105685.

References

Ando, T., Bai, J., 2015. A simple new test for slope homogeneity in panel data models with interactive effects. Econom. Lett. 136, 112–117.
Ando, T., Bai, J., 2016. Panel data models with grouped factor structure under unknown group membership. J. Appl. Econometrics 31 (1), 163–191.
Andrews, D.W., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.
Aquaro, M., Bailey, N., Pesaran, M.H., 2021. Estimation and inference for spatial models with heterogeneous coefficients: An application to US house prices. J. Appl. Econometrics 36 (1), 18–44.
Bai, J., 1997. Estimation of a change point in multiple regression models. Rev. Econ. Stat. 79 (4), 551–563.
Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77 (4), 1229–1279.
Bai, J., 2010. Common breaks in means and variances for panel data. J. Econometrics 157 (1), 78–92.
Bai, J., Ng, S., 2019. Rank regularized estimation of approximate factor models. J. Econometrics 212 (1), 78–96.
Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66 (1), 47–78.
Baltagi, B.H., Kao, C., Liu, L., 2017. Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term. Econometric Rev. 36 (1–3), 85–102.
Belloni, A., Chen, M., Madrid Padilla, O.H., Wang, Z., 2023. High-dimensional latent panel quantile regression with an application to asset pricing. Ann. Statist. 51 (1), 96–121.
Ben-David, D., 1998. Convergence clubs and subsistence economies. J. Dev. Econ. 55 (1), 155–171.
Berthelemy, J.-C., Varoudakis, A., 1996. Economic growth, convergence clubs, and the role of financial development. Oxf. Econ. Pap. 48 (2), 300–328.
Bonhomme, S., Manresa, E., 2015. Grouped patterns of heterogeneity in panel data. Econometrica 83 (3), 1147–1184.
Capozza, D.R., Hendershott, P.H., Mack, C., Mayer, C.J., 2002. Determinants of Real House Price Dynamics. National Bureau of Economic Research Cambridge, Mass., USA.
Cheng, X., Schorfheide, F., Shao, P., 2021. Clustering for multi-dimensional heterogeneity. Working Paper, Department of Economics, University of Pennsylvania.
Chernozhukov, V., Hansen, C.B., Liao, Y., Zhu, Y., 2020. Inference for heterogeneous effects using low-rank estimations. Technical Report, CEMMAP working paper, arXiv preprint arXiv:1812.08089.
Chu, Z., 2012. Logistics and economic growth: a panel data approach. Ann. Reg. Sci. 49 (1), 87–102.
Durlauf, S.N., Johnson, P.A., 1995. Multiple regimes and cross-country growth behaviour. J. Appl. Econometrics 10 (4), 365–384.
Feng, J., 2023. Nuclear norm regularized quantile regression with interactive fixed effects. Econom. Theory 1–31.
Gallin, J., 2006. The long-run relationship between house prices and income: evidence from local housing markets. Real Estate Econ. 34 (3), 417–438.
Hong, S., Su, L., Jiang, T., 2023. Profile GMM estimation of panel data models with interactive fixed effects. J. Econometrics 235 (2), 927–948.
Huang, W., Jin, S., Su, L., 2020. Identifying latent grouped patterns in cointegrated panels. Econom. Theory 36 (3), 410–456.
Jin, J., Ke, Z.T., Luo, S., Wang, M., 2022. Optimal estimation of the number of network communities. J. Amer. Statist. Assoc. 1–16.
Ke, Z.T., Fan, J., Wu, Y., 2015. Homogeneity pursuit. J. Amer. Statist. Assoc. 110 (509), 175–194.
Ke, Y., Li, J., Zhang, W., 2016. Structure identification in panel data analysis. Ann. Statist. 44 (3), 1193–1233.
Keane, M., Neal, T., 2020. Climate change and US agriculture: Accounting for multidimensional slope heterogeneity in panel data. Quant. Econ. 11 (4), 1391–1429.
Kim, D., 2011. Estimating a common deterministic time trend break in large panels with cross sectional dependence. J. Econometrics 164 (2), 310–330.
Kim, D., 2014. Common breaks in time trends for large panel data with a factor structure. Econom. J. 17 (3), 301–337.
Klapper, L., Love, I., 2011. The impact of the financial crisis on new firm registration. Econom. Lett. 113 (1), 1–4.
Leng, X., Chen, H., Wang, W., 2023. Multi-dimensional latent group structures with heterogeneous distributions. J. Econometrics 233 (1), 1–21.
Li, D., Qian, J., Su, L., 2016. Panel data models with interactive fixed effects and multiple structural breaks. J. Amer. Statist. Assoc. 111 (516), 1804–1819.
Lin, C.-C., Hsu, C.-C., 2011. Change-Point Estimation for Nonstationary Panel Data. Technical Report, Working Paper, Department of Economics, National Central University.
Lin, C.-C., Ng, S., 2012. Estimation of panel data models with parameter heterogeneity when group membership is unknown. J. Econom. Methods 1 (1), 42–55.
Long, W., Li, N., Wang, H., Cheng, S., 2012. Impact of US financial crisis on different countries: based on the method of functional analysis of variance. Procedia Comput. Sci. 9, 1292–1298.
Lu, X., Su, L., 2017. Determining the number of groups in latent panel structures with an application to income and democracy. Quant. Econ. 8 (3), 729–760.
Lu, X., Su, L., 2023. Uniform inference in linear panel data models with two-dimensional heterogeneity. J. Econometrics 235 (2), 694–719.
Lumsdaine, R.L., Okui, R., Wang, W., 2023. Estimation of panel group structure models with structural breaks in group memberships and coefficients. J. Econometrics 233 (1), 45–65.
Ma, S., Su, L., Zhang, Y., 2022. Detecting latent communities in network formation models. J. Mach. Learn. Res. 23 (1), 13971–14031.
Malpezzi, S., 1999. A simple error correction model of house prices. J. Hous. Econ. 8 (1), 27–62.
Miao, K., Phillips, P.C.B., Su, L., 2023. High-dimensional VARs with common factors. J. Econometrics 233 (1), 155–183.
Moon, H.R., Weidner, M., 2017. Dynamic linear panel regression models with interactive fixed effects. Econom. Theory 33 (1), 158–195.
Moon, H.R., Weidner, M., 2018. Nuclear norm regularized estimation of panel regression models. arXiv preprint arXiv:1810.10987.
Okui, R., Wang, W., 2021. Heterogeneous structural breaks in panel data models. J. Econometrics 220 (2), 447–473.
Ortalo-Magne, F., Rady, S., 2006. Housing market dynamics: On the contribution of income shocks and credit constraints. Rev. Econom. Stud. 73 (2), 459–485.
Pesaran, M.H., Yamagata, T., 2008. Testing slope homogeneity in large panels. J. Econometrics 142 (1), 50–93.
Prakasa Rao, B.L.S., 2009. Conditional independence, conditional mixing and conditional association. Ann. Inst. Statist. Math. 61 (2), 441–460.
Qian, J., Su, L., 2016. Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. J. Econometrics 191 (1), 86–109.
Sarafidis, V., Weber, N., 2015. A partially heterogeneous framework for analyzing panel data. Oxf. Bull. Econ. Stat. 77 (2), 274–296.
Su, L., Chen, Q., 2013. Testing homogeneity in panel data models with interactive fixed effects. Econom. Theory 29 (6), 1079–1135.
Su, L., Ju, G., 2018. Identifying latent grouped patterns in panel data models with interactive fixed effects. J. Econometrics 206 (2), 554–573.
Su, L., Shi, Z., Phillips, P.C.B., 2016. Identifying latent structures in panel data. Econometrica 84 (6), 2215–2264.
Su, L., Wang, X., Jin, S., 2019. Sieve estimation of time-varying panel data models with latent structures. J. Bus. Econom. Statist. 37 (2), 334–349.
Su, L., Wang, W., Xu, X., 2023. Identifying latent group structures in spatial dynamic panels. J. Econometrics 235 (2), 1955–1980.
Su, L., Wang, W., Zhang, Y., 2020. Strong consistency of spectral clustering for stochastic block models. IEEE Trans. Inform. Theory 66 (1), 324–338.
Wang, W., Phillips, P.C.B., Su, L., 2018. Homogeneity pursuit in panel data models: Theory and application. J. Appl. Econometrics 33 (6), 797–815.
Wang, W., Phillips, P.C.B., Su, L., 2019. The heterogeneous effects of the minimum wage on employment across states. Econom. Lett. 174, 179–185.
Wang, W., Su, L., 2021. Identifying latent group structures in nonlinear panels. J. Econometrics 220 (2), 272–295.
Wang, Y., Su, L., Zhang, Y., 2022. Low-rank panel quantile regression: Estimation and inference. https://arxiv.org/abs/2210.11062.
White, H., 2014. Asymptotic Theory for Econometricians. Academic Press.
Zhang, J., Cheng, L., 2019. Threshold effect of tourism development on economic growth following a disaster shock: Evidence from the Wenchuan earthquake, PR China. Sustainability 11 (2), 371.
Zhang, Y., Wang, H.J., Zhu, Z., 2019. Quantile-regression-based clustering for panel data. J. Econometrics 213 (1), 54–67.