Journal of Econometrics 207 (2018) 1–29

Contents lists available at ScienceDirect

Journal of Econometrics

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

Estimation of large dimensional factor models with an unknown number of breaks[☆]

Shujie Ma [a], Liangjun Su [b,*]

[a] Department of Statistics, University of California-Riverside, Riverside, CA 92521, United States
[b] School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore

ARTICLE INFO

Article history:
Received 7 March 2016
Received in revised form 13 July 2017
Accepted 29 June 2018
Available online 17 July 2018

JEL classification:
C12
C33
C33
C38

Keywords:
Break point
Factor model
Fused Lasso
Group Lasso
Information criterion
Principal component
Structural change
Super-consistency
Time-varying parameter

ABSTRACT

In this paper we study the estimation of a large dimensional factor model when the factor loadings exhibit an unknown number of changes over time. We propose a novel three-step procedure to detect the breaks if any and then identify their locations. In the first step, we divide the whole time span into subintervals and fit a conventional factor model on each interval. In the second step, we apply the adaptive fused group Lasso to identify intervals containing a break. In the third step, we devise a grid search method to estimate the location of the break on each identified interval. We show that with probability approaching one our method can identify the correct number of changes and estimate the break locations. Simulation studies indicate superb finite sample performance of our method. We apply our method to investigate Stock and Watson's (2009) U.S. monthly macroeconomic dataset and identify five breaks in the factor loadings, spanning 1959–2006.

1. Introduction

High-dimensional time series data are frequently encountered in modern statistical and econometric studies, and they may be one of the most common types of data in the "big data" era. Examples come from many fields including economics, finance, genomics, environmental study, medical study, meteorology, chemometrics, and so forth. Hence, there is a pressing need to develop effective statistical tools for their analysis. The celebrating large-dimensional factor models which allow both the sample size and the dimension of time series to go to infinity have become a popular method in analyzing high-dimensional time series data, and therefore have received considerable attention in statistics and econometrics since Stock and Watson (1998, 2002), Bai and Ng (2002), and Forni et al. (2005). We refer to Bai and Li (2012, 2014), Fan et al. (2013,

☆ The authors gratefully thank the Co-editor Jianqing Fan, the Associate Editor, and two anonymous referees for their many constructive comments on the previous version of the paper. Ma's research was partially supported by NSF grants DMS 1306972 and DMS 1712558 and Hellman Fellowship. Su gratefully acknowledges the Singapore Ministry of Education for the Tier-2 Academic Research Fund (AcRF) under grant number MOE2012-T2-2-021 and the Lee Kong Chian Fund for Excellence.
* Corresponding author.
E-mail addresses: shujie.ma@ucr.edu (S. Ma), ljsu@smu.edu.sg (L. Su).

https://doi.org/10.1016/j.jeconom.2018.06.019
0304-4076/

2015), Lam and Yao (2012), Onatski (2012) and Wang and Fan (2017) for recent advancement in estimation methods and inference theory in large-dimensional factor modeling.

In large-dimensional factor models, it is assumed that a large number of time series are driven by low-dimensional latent factors. Most existing estimation and forecasting methods in factor models are based on the assumption of time-invariant factor loadings. However, with large-scale data spanning over a long period of time, more and more evidence shows that the factor loadings tend to exhibit structural changes over time, that is, some structural breaks may occur at some dates over a period of time in the study. Ignoring structural breaks generally leads to misleading estimation, inference, and forecasting (Hansen, 2001). Hence, it is prudent to identify structural breaks of the factor loadings before one relies on the conventional time-invariant factor models. Indeed, a growing number of researches have been devoted to studying structural changes in factor loadings recently. To the best of our knowledge, most existing works can be classified into two categories. The first category focuses on developing testing procedures to detect breaks. For example, Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), and Yamamoto and Tanaka (2015) propose various tests for a one-time structural change in the factor loadings; Corradi and Swanson (2014) propose a test to check structural stability of both factor loadings and factor-augmented forecasting regression coefficients; Su and Wang (2017) consider estimation and testing in time-varying factor models and their test allows for multiple breaks in the factor loadings. The second category considers estimation of the change points in factor models. For example, Cheng et al. (2016) consider an adaptive group-Lasso estimator for factor models with a potential one-time structural change and possible emergence of new factors; Chen (2015) proposes a consistent estimator of the break date based on the least squares loss function; Shi (2015) derives the limiting distribution of the least square estimator of a break point in factor models when the break sizes shrink to zero at an appropriate rate; Baltagi et al. (2017, 2016) consider least squares estimation of the single and multiple structural changes, respectively, in factor models based on the observations that the changes of the factor loadings can be equivalently represented by the changes in the second moments of the estimated factors. Apparently, most of these works focus on the case of a single change with two exceptions by Su and Wang (2017) and Baltagi et al. (2016). In addition, Brandom et al. (2013) consider consistent factor estimation in approximate dynamic factor models with moderate structural instability.

Frequently, one can reject the null hypothesis of constant factor loadings in empirical applications. Despite this, methods for determining the number of breaks and for identifying the locations of the break dates in factor models remained unavailable before the first version of the paper, due to great technical challenges in developing the asymptotic tools. In this paper, we propose a novel three-step structural break detection procedure, which can automatically check the existence of breaks and then identify the exact locations of breaks if any. The procedure is easy-to-implement and theoretically reliable. Specifically, in Step I, we divide the whole time span into $J+1$ subintervals and estimate a conventional factor model with time-invariant factor loadings on each interval by the means of principal component analysis (PCA) (Bai and Ng, 2002; Bai, 2003). Based on the piecewise constant PCA estimates on each subinterval, we propose a BIC-type information criterion to determine the number of common factors and show that our information criterion can identify the correct number of common factors with probability approaching one (w.p.a.1). Our method extends Bai and Ng's (2002) method to allow for an unknown number of breaks in the data and is thus robust to the presence of structural breaks in factor models. In Step II, we adopt the adaptive group fused Lasso (AGFL, Tibshirani et al., 2005; Yuan and Lin, 2006; Zou, 2006) to find intervals that contain a break point. We apply an adaptive group fusion penalty to the successive differences of the normalized factor loadings, which can identify the correct number of breaks and the subintervals that the breaks reside in w.p.a.1. In step III, we devise a grid search method to find the break locations in the identified subintervals sequentially and show that w.p.a.1 we can estimate the break points precisely.

The above three-step method provides an automatic way to detect breaks in factor models, and it is computationally fast. The major challenges in the asymptotic analysis of the proposed three-step procedure are threefold. First, some subintervals obtained in the first step may contain a break point in which case the conventional time-invariant factor model is a misspecified model. Hence, we need to develop asymptotic properties of the estimators of the factors and factor loadings in the misspecified factor models, which do not exist in the literature. We find that the properties depend on whether the break point lies in the interior or boundary region of such a time interval. Second, we consider this paper as the first work to apply the AGFL procedure to the normalized factor loadings to identify whether a subinterval contains a break point or not, where the adaptive weights behave substantially different from the weights investigated in the adaptive Lasso literature (e.g., Zou, 2006) due to the presence of misspecified factor models in the first step. In particular, the adaptive weights have distinct asymptotic behaviors when the break points occur in the interior or boundary region of a subinterval, which greatly complicates the analysis of the AGFL procedure. Third, it is technically challenging to establish the theoretical claim that the grid search in the third step identifies the true break points w.p.a.1., even after we find the subintervals that contain a break point. In fact, our grid search method appears to be the first method to estimate the *break dates* consistently in the presence of estimation errors in early stages.

We conduct a sequence of Monte Carlo simulations to evaluate the finite sample performance of our procedure. We find that our information criterion can determine the correct number of factors accurately and our three-step procedure can identify the true number of breaks and estimate the break dates precisely in large samples. We apply our method to Stock and Watson's (2009) macroeconomic dataset and detect five breaks for the period of 1959m01–2006m12.

After we finished the first version of the paper, we found that Baltagi et al. (2016, BKW hereafter) also study the estimation of large dimensional factor models with an unknown number of structural changes. Our approach differs from theirs in several aspects. First, the estimation methods are different. Following the lead of Han and Inoue (2015), BKW observe that the changes in the factor loadings can be represented as the changes in the second moment of the estimated factors, and they then apply the standard techniques in the literature on time series structural change (e.g., Bai (1997) and Bai and Perron (1998)) and consider both the joint and sequential estimation of the change points in the second moments of the estimated factors. In contrast, our method is motivated from the Lasso literature. Second, the choices of the key tuning parameters differ. Unlike our procedure which requires the division of the whole time span into $J+1$ subintervals explicitly, BKW's procedure does not need so in theory. However, in practice, it requires the choice of a tuning/trimming parameter ϵ by restricting the minimum length of a regime to be ϵT, where ϵ typically takes values from 0.05 to 0.25; see Assumption A4(ii) in Bai and Perron (1998), Section 5.1 in Bai and Perron (2003), and the discussion in Qian and Su (2016b).[1] The performance of their method highly depends on the choice of ϵ, which plays a similar role to $1/(J+1)$. Third, the asymptotic results are different. As in the study of structural changes in time series regression, BKW establish the consistency of the estimator of the *break fractions* but not that of the estimator of the *common break dates*. This is mainly because they transform the original problem of estimating structural changes in the factor loadings to the problem of determining the breaks in the second moments of the estimated factor time series process, which cannot use the common break date information across all cross-sectional units effectively. In contrast, we work on the original problem and can establish the super-consistency of our estimator of the *common break dates* by the effective use of the cross-sectional information as in Qian and Su (2016a). That being said, we notice that BKW claim in their Section 5 that through re-estimation (based on the simultaneous search of the multiple break dates after one obtains the estimated numbers of factors and breaks) they can establish the consistency of the estimators of the break dates. This step parallels to Step III in our procedure with the only difference that our method is sequential while their re-estimation is joint. In either case, the consistency of the break dates estimators are expected because both methods can rely on the large-dimensional cross-sectional information effectively.

The rest of this paper is organized as follows. In Section 2, we introduce the three-step procedure for break points detection and estimation. In Section 3, we study the asymptotic theory. In Section 4, we study the finite sample performance of our method. Section 5 provides an empirical study. Section 6 concludes. All proofs are relegated to the appendix. Further technical details are contained in the online supplementary material.

2. The factor model and estimation procedure

In this section, we consider a large-dimensional factor model with an unknown number of breaks, and then propose a three-step procedure for estimation. We first introduce some notations which will be used throughout the paper. Let $\mu_{\max}(\mathbf{B})$ and $\mu_{\min}(\mathbf{B})$ denote the largest and smallest eigenvalues of a symmetric matrix $\mathbf{B}$, respectively. We use $\mathbf{B}>0$ to denote that $\mathbf{B}$ is positive definite. For an $m\times n$ real matrix $\mathbf{A}$, we denote its transpose as $\mathbf{A}^\top$, its Moore–Penrose generalized inverse as $\mathbf{A}^+$, its rank as rank($\mathbf{A}$), its Frobenius norm as $\|\mathbf{A}\|$ ($\equiv [\mathrm{tr}(\mathbf{A}\mathbf{A}^\top)]^{1/2}$), and its spectral norm as $\|\mathbf{A}\|_{\mathrm{sp}}$ ($\equiv \sqrt{\mu_{\max}\left(\mathbf{A}^\top\mathbf{A}\right)}$). Note that the two norms are equal when $\mathbf{A}$ is a vector. We will frequently use the submultiplicative property of these norms and the fact that $\|\mathbf{A}\|_{\mathrm{sp}}\le\|\mathbf{A}\|\le\|\mathbf{A}\|_{\mathrm{sp}}\mathrm{rank}(\mathbf{A})^{1/2}$. Let $P_{\mathbf{A}}\equiv\mathbf{A}\left(\mathbf{A}^\top\mathbf{A}\right)^+\mathbf{A}^\top$ and $M_{\mathbf{A}}\equiv\mathbb{I}_m-P_{\mathbf{A}}$, where $\mathbb{I}_m$ denotes an $m\times m$ identity matrix. For any set S, we use $|S|$ to denote its cardinality. For any positive numbers a_n and b_n, let $a_n\asymp b_n$ denote $\lim_{n\to\infty}a_n/b_n=c$, for a positive constant c, and let $a_n\gg b_n$ denote $a_n^{-1}b_n=o(1)$. The operator $\xrightarrow{P}$ denotes convergence in probability and plim denotes probability limit. We use $(N,T)\to\infty$ to denote that N and T pass to infinity jointly.

2.1. The factor model

We consider the time-varying factor model:

$$X_{it}=\lambda_{it}^\intercal F_t+e_{it},\ i=1,\ldots,N,\ t=1,\ldots,T,$$

where λ_{it} is an $R\times 1$ vector of time-dependent factor loadings, F_t is an $R\times 1$ vector of unobserved common factors, e_{it} is the idiosyncratic error term, and both N and T pass to infinity. For simplicity of technical proofs, we assume that R does not depend on N and T, but it is unknown. Hence we need to estimate R from the data. Writing the above model in the vector form, we have

$$\mathbf{X}_t=\boldsymbol{\lambda}_tF_t+\mathbf{e}_t,\ t=1,\ldots,T,$$

where $\mathbf{X}_t=(X_{1t},\ldots,X_{Nt})^\intercal$, $\boldsymbol{\lambda}_t=(\lambda_{1t},\ldots,\lambda_{Nt})^\intercal$, and $\mathbf{e}_t=(e_{1t},\ldots,e_{Nt})^\intercal$.

We assume that the factor-loadings $\{\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_T\}$ exhibit certain *sparse* nature such that the total number of distinct vectors in the set is given by $m+1$, where m denotes the total number of break points in the process $\{\boldsymbol{\lambda}_t\}$ and it satisfies $T\gg m$. When $m\ge 1$, let $\{t_1,\ldots,t_m\}$ denote the m change-points satisfying

$$1\equiv t_0<t_1<\cdots<t_m<t_{m+1}\equiv T+1,$$

so that the whole time span is divided into $m+1$ regimes/segments, denoted by $I_\kappa=[t_\kappa,t_{\kappa+1})$ for $\kappa=0,1,\ldots,m-1$ and $I_m=[t_m,t_{m+1}]$. We assume that

$$\lambda_{it}=\alpha_{i\kappa}\text{ for all }t\in I_\kappa\text{ and }\kappa=0,1,\ldots,m.$$

[1] BKW choose $\epsilon=0.1$ in their simulations, which implies that the maximum number of breaks allowed is 8.

When $m=0$, we have $I_0=I_m=[t_0,t_1)=[1,T]$ and $\lambda_{it}=\alpha_{i0}$ for all $t\in[1,T]$, so that no break happens in this scenario. Let $\boldsymbol{\alpha}_\kappa=(\alpha_{1k},\ldots,\alpha_{N\kappa})^\intercal$ for $\kappa=0,1,\ldots,m$. In practice, the number of breaks, m, and the locations of the breaks are unknown if there are any breaks. Our target is to detect breaks, to find the number of breaks and identify their locations, and to estimate R, $\alpha_{i\kappa}$ and F_t. Let t_κ^0, $\alpha_{i\kappa}^0$, $\boldsymbol{\alpha}_\kappa^0$ and F_t^0 denote the true values of t_κ, $\alpha_{i\kappa}$, $\boldsymbol{\alpha}_\kappa$ and F_t, respectively.

2.2. A three-step procedure

We propose a three step procedure to automatically detect breaks, to determine the number of breaks if any, and to estimate their locations. For clarity, we assume that R is known in this section and Section 3, but we discuss how to estimate it and establish the consistency of its estimator in Section 4.1.

2.2.1. Step I: Piecewise constant estimation

Noting that $\boldsymbol{\lambda}_tF_t=\boldsymbol{\lambda}_t\left(H_t^{-1}\right)^\intercal H_t^\intercal F_t$ for any $R\times R$ nonsingular matrix H_t, $\boldsymbol{\lambda}_t$ and F_t are not separately identified, and their identification requires R^2 restrictions at each time point t. For the estimation of $\boldsymbol{\lambda}_t$ and F_t, following the lead of Bai and Ng (2002), we shall impose the following identification conditions:

$$\boldsymbol{\lambda}_t^\intercal\boldsymbol{\lambda}_t/N=\mathbb{I}_R \text{ for each } t,\ \sum_{t=1}^T F_tF_t^\intercal/T \text{ is a diagonal matrix.}$$

In this step, we propose to approximate λ_{it} by piecewise-constants, and then estimate λ_{it} and F_t accordingly. The procedure is described as follows. Let $J=J(N,T)$ be a prescribed integer that depends on (N,T), satisfying $T\gg J\gg m$. Divide $[1,T]$ into $(J+1)$ subintervals $S_j=[v_j,v_{j+1})$ for $j=0,1,\ldots,J-1$ and $S_J=[v_J,T]$, where $\{v_j\}_{j=1}^J$ is a sequence of "equally-spaced" interior knots given as $v_0\equiv 1<v_1<\cdots<v_J<T\equiv v_{J+1}$, where $v_j=\lfloor Tj/(J+1)\rfloor$ for $j=1,\ldots,J$ and $\lfloor\cdot\rfloor$ denotes the integer part of $\cdot$. Note that each interval contains $T/(J+1)$ observations, for $j=0,1,\ldots,J-1$, when $T/(J+1)$ is an integer. For any $t\in S_j$, λ_{it} is treated as a constant and can be approximated by $\lambda_{it}\approx\delta_{ij}$, so that the identification condition that $\boldsymbol{\lambda}_t^\intercal\boldsymbol{\lambda}_t/N=\mathbb{I}_R\ \forall\, t$ implies that $\sum_{i=1}^N\delta_{ij}\delta_{ij}^\intercal/N=\mathbb{I}_R$ for each $j=0,1,\ldots,J$. Denote $\boldsymbol{\Delta}_j=(\delta_{1j},\ldots,\delta_{Nj})^\intercal$. Then we need

$$\boldsymbol{\Delta}_j^\intercal\boldsymbol{\Delta}_j/N=\mathbb{I}_R \text{ for every } j=0,1,\ldots,J.$$

The estimators $\hat{\boldsymbol{\Delta}}_j$ and $\hat{F}_t$ are obtained by minimizing

$$\sum_{t\in S_j}(\mathbf{X}_t-\boldsymbol{\Delta}_jF_t)^\intercal(\mathbf{X}_t-\boldsymbol{\Delta}_jF_t)$$

subject to $\boldsymbol{\Delta}_j^\intercal\boldsymbol{\Delta}_j/N=\mathbb{I}_R$ and $\mathbf{F}_{S_j}^\intercal\mathbf{F}_{S_j}=$ diagonal, where $\mathbf{F}_{S_j}=(F_t,t\in S_j)^\intercal=(F_{v_j},\ldots,F_{v_{j+1}-1})^\intercal$. By concentrating out $F_t=(\boldsymbol{\Delta}_j^\intercal\boldsymbol{\Delta}_j/N)^{-1}(\boldsymbol{\Delta}_j^\intercal\mathbf{X}_t/N)=\boldsymbol{\Delta}_j^\intercal\mathbf{X}_t/N$, the above objective function becomes

$$\begin{aligned}&\sum_{t\in S_j}(\mathbf{X}_t-\boldsymbol{\Delta}_j\boldsymbol{\Delta}_j^\intercal\mathbf{X}_t/N)^\intercal(\mathbf{X}_t-\boldsymbol{\Delta}_j\boldsymbol{\Delta}_j^\intercal\mathbf{X}_t/N)\\ &=\sum_{t\in S_j}\mathbf{X}_t^\intercal\mathbf{X}_t-N^{-1}\mathrm{tr}(\boldsymbol{\Delta}_j^\intercal\mathbf{X}_{S_j}\mathbf{X}_{S_j}^\intercal\boldsymbol{\Delta}_j),\end{aligned}$$

where $\mathbf{X}_{S_j}=(\mathbf{X}_t,t\in S_j)$ and we have used the restriction that $\boldsymbol{\Delta}_j^\intercal\boldsymbol{\Delta}_j/N=\mathbb{I}_R$. Thus, the estimators $\hat{\boldsymbol{\Delta}}_j=(\hat\delta_{1j},\ldots,\hat\delta_{Nj})^\intercal$ can be obtained by maximizing

$$N^{-1}\mathrm{tr}(\boldsymbol{\Delta}_j^\intercal\mathbf{X}_{S_j}\mathbf{X}_{S_j}^\intercal\boldsymbol{\Delta}_j)$$

subject to $\boldsymbol{\Delta}_j^\intercal\boldsymbol{\Delta}_j/N=\mathbb{I}_R$. When $\mathrm{rank}(\mathbf{X}_{S_j}\mathbf{X}_{S_j}^\intercal)\geq R$ for every $j=0,\ldots,J$, $\hat{\boldsymbol{\Delta}}_j$ is $\sqrt{N}$ times the eigenvectors corresponding to the R largest eigenvalues of the $N\times N$ matrix $\mathbf{X}_{S_j}\mathbf{X}_{S_j}^\intercal=\sum_{t\in S_j}\mathbf{X}_t\mathbf{X}_t^\intercal$, and $\hat{F}_t=\hat{\boldsymbol{\Delta}}_j^\intercal\mathbf{X}_t/N$ for $t\in S_j$.

2.2.2. Step II: Adaptive group fused Lasso penalization for break detection

Let $\tau_j=|S_j|$ be the cardinality of the set S_j. Let $V_{N,j}$ denote the $R\times R$ diagonal matrix of the first R largest eigenvalues of $\frac{1}{N\tau_j}\mathbf{X}_{S_j}\mathbf{X}_{S_j}^\intercal$ in descending order. For those time points in the same true regime (I_κ, say), their factor loadings should be the same. By Proposition 3.1(ii) below, $\hat{\boldsymbol{\Delta}}_jV_{N,j}$ is a consistent estimator of $\boldsymbol{\alpha}_\kappa^0\Sigma_FQ_\kappa^\intercal$ for all j satisfying $S_j\subset I_\kappa$, where Σ_F is defined in Assumption A1 and Q_κ is defined in Proposition 3.1, both of which do not depend on j. Note that $\boldsymbol{\alpha}_\kappa^0\Sigma_FQ_\kappa^\intercal$ remains unchanged if two consecutive intervals, say S_j and S_{j-1}, belong to I_κ. This motivates us to consider the following objective function by imposing an AGFL penalty to detect the breaks between segments:

$$\frac{1}{2N}\sum_{j=0}^J\frac{1}{\tau_j}\sum_{t\in S_j}(\mathbf{X}_t-\boldsymbol{\Delta}_j\hat{F}_t)^\intercal(\mathbf{X}_t-\boldsymbol{\Delta}_j\hat{F}_t)+\gamma\sum_{j=1}^J w_j^0\left\|\boldsymbol{\Delta}_jV_{N,j}-\boldsymbol{\Delta}_{j-1}V_{N,j-1}\right\|,\tag{2.1}$$

where γ is a tuning parameter and w_j^0's are adaptive weights to be specified later.

Let $\breve{\Theta}_j = \hat{\boldsymbol{\Delta}}_j V_{N,j}$, $\Theta_j = N^{1/2}\boldsymbol{\Delta}_j V_{N,j}/\|\breve{\Theta}_j\|$, $\hat{Z}_{jt} = N^{-1/2}\|\breve{\Theta}_j\|V_{N,j}^{-1}\hat{F}_t$ and $\hat{\mathbf{Z}}_{S_j} = (\hat{Z}_{jt}, t \in S_j)$. We can slightly modify the objective function in (2.1) in terms of Θ_j

$$\frac{1}{2N}\sum_{j=0}^{J}\frac{1}{\tau_j}\|\mathbf{X}_{S_j} - \Theta_j\hat{\mathbf{Z}}_{S_j}\|^2 + \gamma\sum_{j=1}^{J} w_j\|\Theta_j - \Theta_{j-1}\|, \tag{2.2}$$

where $w_j = N^{-1/2}w_j^0\|\breve{\Theta}_j\|$. Note that (2.1) compares $\boldsymbol{\Delta}_j V_{N,j}$ with $\boldsymbol{\Delta}_{j-1}V_{N,j-1}$ while (2.2) contrasts their normalized versions. Let $\breve{\Theta}_{j,r}$ denote the rth column of $\breve{\Theta}_j$ for $r = 1, \ldots, R$. Let $\breve{\rho}_{j,r}$ denote the sample Pearson correlation coefficient of $\breve{\Theta}_{j,r}$ and $\breve{\Theta}_{j-1,r}$ for $j = 1, \ldots, J$. When the eigenvectors in $\hat{\boldsymbol{\Delta}}_j$ are properly normalized to ensure the sign-identification, with Proposition 3.1 below we can show that $\breve{\rho}_{j,r} \xrightarrow{P} 1$ when both S_j and S_{j-1} belong to I_κ and they may converge in probability to a value different from one otherwise. This motivates us to consider the following adaptive weights

$$w_j = \left(1 - R^{-1}\sum_{r=1}^{R}\breve{\rho}_{j,r}\right)^{-\varkappa}, \tag{2.3}$$

where $\varkappa$ is some fixed positive constant, e.g., 2. Let $\tilde{\Theta}_j$ denote the penalized estimator of Θ_j in (2.2). Then the penalized estimator of $\boldsymbol{\Delta}_j$ is given by $\tilde{\boldsymbol{\Delta}}_j = \tilde{\boldsymbol{\Delta}}_j(\gamma) = N^{-1/2}\tilde{\Theta}_j V_{N,j}^{-1}\|\breve{\Theta}_j\|$.

We apply Boyd et al.'s (2011) alternating direction method of multipliers (ADMM) algorithm to obtain the penalized estimator $\tilde{\Theta}_j$. Boyd et al. (2011) show that the ADMM algorithm has a good global convergence property. The detailed procedure is provided in Section 3 of the Supplementary Material. The tuning parameter γ is chosen by the information criterion method as given in Section 4.2.

2.2.3. *Step III: Grid search for the locations of the breaks*

Let $\tilde{\boldsymbol{\beta}}_j \equiv \tilde{\Theta}_j - \tilde{\Theta}_{j-1}$ for $j = 1, \ldots, J$. By step II, we are able to identify the subintervals containing the breaks. There are four situations that can happen for each subinterval S_j: (1) when $\tilde{\boldsymbol{\beta}}_j \neq 0$ and $\tilde{\boldsymbol{\beta}}_{j+1} \neq 0$, the break happens in the interior of the interval S_j; (2) when $\tilde{\boldsymbol{\beta}}_j \neq 0$ and $\tilde{\boldsymbol{\beta}}_{j+1} = 0$ and $\tilde{\boldsymbol{\beta}}_{j-1} = 0$, the break may happen near the left end of S_j or the right end of S_{j-1}; (3) when $\tilde{\boldsymbol{\beta}}_{j+1} \neq 0$ and $\tilde{\boldsymbol{\beta}}_j = 0$ and $\tilde{\boldsymbol{\beta}}_{j+2} = 0$, the break may happen near the right end of S_j or the left end of S_{j+1}; and (4) when $\tilde{\boldsymbol{\beta}}_j = 0$ and $\tilde{\boldsymbol{\beta}}_{j+1} = 0$, no break happens in S_j. For case (1), we can conclude that an estimated break happens in the interval S_j, and for cases (2) and (3), we have that an estimated break happens in the intervals S_{j-1}^* and S_j^*, respectively, where, e.g., $S_{j-1}^* \equiv [v_{j-1} + \lfloor\tau_{j-1}/2\rfloor + 1, v_j + \lfloor\tau_j/2\rfloor)$. Suppose that we have found $\hat{m}$ intervals that contain a break point. We denote such $\hat{m}$ intervals as $\bar{S}_{j_1}, \ldots, \bar{S}_{j_{\hat{m}}}$. Note that $\bar{S}_{j_\kappa}$ coincides with either S_{j_κ} or $S_{j_\kappa}^*$. Write $\bar{S}_{j_\kappa} = [t_{\kappa,1}^*, \ldots, t_{\kappa,\bar{\tau}_{j_\kappa}}^*]$ with $\bar{\tau}_{j_\kappa} = \left|\bar{S}_{j_\kappa}\right|$ for $\kappa = 1, \ldots, \hat{m}$. We discuss how to estimate these $\hat{m}$ break points below.

To estimate the first break point, we conduct a grid search over the interval $\bar{S}_{j_1}$ by using as many observations as possible from both pre-$\bar{S}_{j_1}$ and post-$\bar{S}_{j_1}$ intervals. If the first break point happens to be $t_{1,\ell}^*$ for some $\ell \in \left\{1, 2, \ldots, \bar{\tau}_{j_1}\right\}$, we know that observations that occur before $t_{1,\ell}^*$ belong to the first regime w.p.a.1. Similarly, the observations that occur after $t_{1,\ell}^*$ but before the first observation in $\bar{S}_{j_2}$ belong to the second regime w.p.a.1. But $t_{1,\ell}^*$ is unknown and has to be searched over all points in $\bar{S}_{j_1}$. After obtaining the first break point, we can find subsequent break points analogously.

To state the algorithm, let $S_a^b = \{t : a \le t \le b\}$ and $\mathbf{F}_{S_a^b} = (F_a, \ldots, F_b)^\intercal$ for any integers $a \le b$. Let $\boldsymbol{\alpha}_l = (\alpha_{1l}, \ldots, \alpha_{Nl})^\intercal$ for $l = 1, 2, \ldots$ The following procedure describes how we can find the locations of all $\hat{m}$ break points sequentially:

1. To search for the first break point t_1, we consider the following minimization problem:

$$\min_{\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \{F_t\}\}} Q_1(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \{F_t\}; t_1) = \sum_{t \in S_1^{t_1-1}}\|\mathbf{X}_t - \boldsymbol{\alpha}_1 F_t\|^2 + \sum_{t \in S_{t_1}^{t_{2,1}^*-1}}\|\mathbf{X}_t - \boldsymbol{\alpha}_2 F_t\|^2$$

subject to the constraints $N^{-1}\boldsymbol{\alpha}_1^\intercal\boldsymbol{\alpha}_1 = \mathbb{I}_R$, $N^{-1}\boldsymbol{\alpha}_2^\intercal\boldsymbol{\alpha}_2 = \mathbb{I}_R$, $\frac{1}{t_1-1}\mathbf{F}^\intercal_{S_1^{t_1-1}}\mathbf{F}_{S_1^{t_1-1}} =$ diagonal and $\frac{1}{t_{2,1}^*-t_1} \times \mathbf{F}^\intercal_{S_{t_1}^{t_{2,1}^*-1}}\mathbf{F}_{S_{t_1}^{t_{2,1}^*-1}} =$ diagonal. Denote the solution to the above minimization problem as $(\tilde{\boldsymbol{\alpha}}_1(t_1), \tilde{\boldsymbol{\alpha}}_2(t_1), \{\tilde{F}_t(t_1)\})$. The first break point is estimated as

$$\hat{t}_1 = \arg\min_{t_1 \in \bar{S}_{j_1}} Q_1\left(\tilde{\boldsymbol{\alpha}}_1(t_1), \tilde{\boldsymbol{\alpha}}_2(t_1), \{\tilde{F}_t(t_1)\}; t_1\right).$$

2. After obtaining the break points, $\hat{t}_1, \ldots, \hat{t}_{\kappa-1}$, we can search for the κth break point t_κ by considering the following minimization problem

$$\begin{aligned} &\min_{\{\boldsymbol{\alpha}_\kappa, \boldsymbol{\alpha}_{\kappa+1}, \{F_t\}\}} Q_\kappa(\boldsymbol{\alpha}_\kappa, \boldsymbol{\alpha}_{\kappa+1}, \{F_t\}; t_\kappa) \\ &= \sum_{t \in S_{\hat{t}_{\kappa-1}}^{t_\kappa-1}}\|\mathbf{X}_t - \boldsymbol{\alpha}_\kappa F_t\|^2 + \sum_{t \in S_{t_\kappa}^{t_{\kappa+1,1}^*-1}}\|\mathbf{X}_t - \boldsymbol{\alpha}_{\kappa+1} F_t\|^2 \end{aligned}$$

subject to the constraints $N^{-1}\boldsymbol{\alpha}_{\kappa}^{\intercal}\boldsymbol{\alpha}_{\kappa} = \mathbb{I}_R$, $N^{-1}\boldsymbol{\alpha}_{\kappa+1}^{\intercal}\boldsymbol{\alpha}_{\kappa+1} = \mathbb{I}_R$, $\frac{1}{t_\kappa - \hat{t}_{\kappa-1}}\mathbf{F}^{\intercal}_{S_{\hat{t}_{\kappa-1}}^{t_\kappa-1}}\mathbf{F}_{S_{\hat{t}_{\kappa-1}}^{t_\kappa-1}} =$ diagonal and $\frac{1}{t^*_{\kappa+1,1}-t_\kappa}\mathbf{F}^{\intercal}_{S_{t_\kappa}^{t^*_{\kappa+1,1}-1}}\mathbf{F}_{S_{t_\kappa}^{t^*_{\kappa+1,1}-1}} =$ diagonal. Denote the solution to the above minimization problem as $(\tilde{\boldsymbol{\alpha}}_\kappa(t_\kappa), \tilde{\boldsymbol{\alpha}}_{\kappa+1}(t_\kappa), \{\tilde{F}_t(t_\kappa)\})$. The κth break point is estimated as

$$\hat{t}_\kappa = \arg\min_{t_\kappa \in \bar{S}_{j_\kappa}} Q_\kappa\left(\tilde{\boldsymbol{\alpha}}_\kappa(t_\kappa), \tilde{\boldsymbol{\alpha}}_{\kappa+1}(t_\kappa), \{\tilde{F}_t(t_\kappa)\}; t_\kappa\right).$$

3. Repeat the above step until we obtain all $\hat{m}$ estimated break points.

At last, after we find the locations of the break points, $\hat{t}_1, \ldots, \hat{t}_{\hat{m}}$, the whole time span is divided into $\hat{m}+1$ regimes/segments, denoted by $\hat{I}_\kappa = [\hat{t}_{\kappa-1}, \hat{t}_\kappa)$ for $\kappa = 1, \ldots, \hat{m}+1$, where $\hat{t}_0 = 1$ and $\hat{t}_{\hat{m}+1} \equiv T+1$. On segment $\hat{I}_\kappa$, we estimate the factors and their loadings as

$$(\hat{\boldsymbol{\alpha}}_\kappa, \{\hat{F}_t\}) = \arg\min_{\boldsymbol{\alpha}_\kappa, \{F_t\}} \sum_{t \in \hat{I}_\kappa} (\mathbf{X}_t - \boldsymbol{\alpha}_\kappa F_t)^{\intercal}(\mathbf{X}_t - \boldsymbol{\alpha}_\kappa F_t)$$

subject to the constraints $N^{-1}\boldsymbol{\alpha}_{\kappa}^{\intercal}\boldsymbol{\alpha}_{\kappa} = \mathbb{I}_R$, and $\frac{1}{|\hat{I}_\kappa|}\mathbf{F}^{\intercal}_{\hat{I}_\kappa}\mathbf{F}_{\hat{I}_\kappa}$ =diagonal.

3. Asymptotic theory

In this section, we study the asymptotic properties of our estimators.

3.1. Theory for the piecewise constant estimators

For each subinterval S_j, we will establish the asymptotic property of $\hat{\boldsymbol{\Delta}}_j$ from the piecewise constant estimation in Step I. Denote $\mathbb{S} = \{0, 1, 2, \ldots, J\}$. Let $\tau_{j1} = t_\kappa^0 - v_j$ and $\tau_{j2} = \tau_j - \tau_{j1}$ when S_j contains a true break point t_κ^0 for some $\kappa = \kappa(j)$. Define

$$\begin{aligned}
\mathbb{S}_1 &= \left\{ j \in \mathbb{S} : \ S_j \subset I_\kappa \text{ for some } \kappa(j) \right\}, \\
\mathbb{S}_{2a} &= \left\{ j \in \mathbb{S} : \ S_j \text{ contains a break } t_\kappa^0 \text{ for some } \kappa(j) \text{ such that } \lim_{T\to\infty} \tau_{j1}/\tau_j = 1 \right\}, \\
\mathbb{S}_{2b} &= \left\{ j \in \mathbb{S} : \ S_j \text{ contains a break } t_\kappa^0 \text{ for some } \kappa(j) \text{ such that } \lim_{T\to\infty} \tau_{j1}/\tau_j = 0 \right\}, \\
\mathbb{S}_{2c} &= \left\{ j \in \mathbb{S} : \ S_j \text{ contains a break } t_\kappa^0 \text{ for some } \kappa(j) \text{ such that } \lim_{T\to\infty} \tau_{j1}/\tau_j \in (0,1) \right\}.
\end{aligned}$$

Let $\mathbb{S}_2 = \mathbb{S}_{2a} \cup \mathbb{S}_{2b} \cup \mathbb{S}_{2c}$. When no confusion arises, we will suppress the dependence of $\kappa = \kappa(j)$ on j. Noting that $|\mathbb{S}_2| = m \ll J$, we have $|\mathbb{S}_1|/J \to 1$.

Case 1. When no break occurs in the subinterval S_j, i.e., $S_j \subset I_\kappa$ for some segment I_κ, then we have $\lambda_{it} = \alpha_{i\kappa}^0$ for all $t \in S_j$, where $\alpha_{i\kappa}^0$ is the vector of the true factor loadings for the segment I_κ. Let F_t^0 be the vector of true factors for $t \in S_j$. Then we have

$$X_{it} = \alpha_{i\kappa}^{0\intercal} F_t^0 + e_{it}, \ i = 1, \ldots, N, \ t \in S_j.$$

Let $\mathbf{F}_{S_j}^0 = (F_t^0, t \in S_j)^{\intercal} = (F_{v_j}^0, \ldots, F_{v_{j+1}-1}^0)^{\intercal}$ and $\boldsymbol{\alpha}_\kappa^0 = (\alpha_{1\kappa}^0, \ldots, \alpha_{N\kappa}^0)^{\intercal}$. Denote $\gamma_N(s,t) = N^{-1}E\left(\mathbf{e}_s^{\intercal}\mathbf{e}_t\right)$, $\gamma_{N,F}(s,t) = N^{-1}E\left(F_s^0\mathbf{e}_s^{\intercal}\mathbf{e}_t\right)$, $\gamma_{N,FF}(s,t) = N^{-1}E(F_s^0\mathbf{e}_s^{\intercal}\mathbf{e}_t F_t^{0\intercal})$, $\zeta_{st} = N^{-1}[\mathbf{e}_s^{\intercal}\mathbf{e}_t - E(\mathbf{e}_s^{\intercal}\mathbf{e}_t)]$, $\zeta_{F,st} = N^{-1}[F_s^0\mathbf{e}_s^{\intercal}\mathbf{e}_t - E(F_s^0\mathbf{e}_s^{\intercal}\mathbf{e}_t)]$, $\zeta_{FF,st} = N^{-1}[F_s^0\mathbf{e}_s^{\intercal}\mathbf{e}_t F_t^{0\intercal} - E(F_s^0\mathbf{e}_s^{\intercal}\mathbf{e}_t F_t^{0\intercal})]$, and $\varsigma_{il}(t_1, t_2) = \frac{1}{t_2 - t_1}\sum_{s=t_1}^{t_2}[e_{is}e_{ls} - E(e_{is}e_{ls})]$. Let $\varpi_{il,ts} = E(e_{it}e_{ls})$ and $\varpi_{il,t} = \varpi_{il,tt}$. Let $C < \infty$ denote a positive constant that may vary from case to case.

We make the following assumptions.

Assumption A1. $E\|F_t^0\|^4 \le C$ and $\frac{1}{t-s}\mathbf{F}^{0\intercal}_{S_s^{t-1}}\mathbf{F}^0_{S_s^{t-1}} = \Sigma_F + O_P((t-s)^{-1/2})$ for some $R \times R$ positive definite matrix Σ_F and for any two points $t, s \in [1, T]$ satisfying $t - s \to \infty$.

Assumption A2. λ_{it}'s are nonrandom such that $\max_{1\le i\le N, 1\le t\le T}\|\lambda_{it}\| \le C$ and $\frac{1}{N}\boldsymbol{\alpha}_\kappa^{0\intercal}\boldsymbol{\alpha}_\kappa^0 = \Sigma_\kappa + O\left(N^{-1/2}\right)$ for some $R \times R$ positive definite matrix Σ_κ for $\kappa = 0, 1, \ldots, m$.

Assumption A3. (i) $E(e_{it}) = 0$ and $\max_{1\le i\le N, 1\le t\le T} E\left(e_{it}^4\right) \le C$.

(ii) $\max_{1\le t\le T}\sum_{s=1}^T \|\gamma(s,t)\| \le C$ and $\max_{1\le s\le T}\sum_{t=1}^T \|\gamma(s,t)\| \le C$ for $\gamma = \gamma_N, \gamma_{N,F}$, and $\gamma_{N,FF}$. $\max_{1\le t\le T}\left|\varpi_{il,tt}\right| \le \varpi_{il}$ for some ϖ_{il} such that $\max_{1\le l\le N}\sum_{i=1}^N \varpi_{il} \le C$.

(iv) $\left(N\tau_j\right)^{-1}\sum_{i=1}^N\sum_{l=1}^N\sum_{t\in S_j}\sum_{s\in S_j}\left|\varpi_{il,ts}\right| \le C$.

(v) $\max_{1\le i,l\le T} E\left|(t_2-t_1)^{1/2}\varsigma_{il}(t_1,t_2)\right|^4 \le C$ for all $t_1 < t_2$ such that $t_2 - t_1 \to \infty$.

(vi) $\max_{1\le s,t\le T} E\left\|N^{1/2}\zeta_{st}^{\dagger}\right\|^4 \le C$ for $\zeta_{st}^{\dagger} = \zeta_{st}, \zeta_{F,st}$ and $\zeta_{FF,st}$, and $\max_{1\le t\le T} E\left\|N^{-1/2}\boldsymbol{\alpha}_{\kappa}^{0\intercal}\mathbf{e}_t\right\|^4 \le C$ for $\kappa = 0, 1, \ldots, m$.

Assumption A4. The eigenvalues of the $R\times R$ matrices $\Sigma_{\kappa}^{1/2}\Sigma_F\Sigma_{\kappa}^{1/2}$ are distinct for $\kappa = 0, 1, \ldots, m$.

Assumptions A1–A2 parallel Assumptions A and B in Bai (2003). A1 implies that $\frac{1}{\tau_j}\mathbf{F}_{S_j}^{0\intercal}\mathbf{F}_{S_j}^0 = \Sigma_F + O_P(\tau_j^{-1/2})$ as $\tau_j \to \infty$ and A2 requires λ_{it} to be nonrandom and uniformly bounded. A3(i) imposes moment conditions on e_{it} and A3(ii)–(v) restricts the cross-sectional and serial dependence among $\{e_{it}, F_t\}$. Similar conditions are also imposed in the literature; see, Bai and Ng (2002) and Bai (2003). A4 is required to establish the convergence of certain eigenvector estimates.

Let $\eta_{N\tau_j} = \min\{\sqrt{N}, \sqrt{\tau_j}\}$ and $H_j = H_{N\tau_j,j} = (\frac{1}{\tau_j}\mathbf{F}_{S_j}^{0\intercal}\mathbf{F}_{S_j}^0)(\frac{1}{N}\boldsymbol{\alpha}_{\kappa}^{0\intercal}\hat{\boldsymbol{\Delta}}_j)V_{N,j}^{-1}$. Following Bai (2003), we can readily obtain the following results:

Proposition 3.1. *Suppose that Assumptions A1–A4 hold. Then as* $(N, \tau_j) \to \infty$,

(i) $\frac{1}{N}\|\hat{\boldsymbol{\Delta}}_j - \boldsymbol{\alpha}_{\kappa}^0 H_j\|^2 = O_P(\eta_{N\tau_j}^{-2})$ *and* $\|\hat{F}_t - H_j^{-1}F_t^0\| = O_P(\eta_{N\tau_j}^{-1})$ *for any* $t \in S_j$ *and* $j \in \mathbb{S}_1$,

(ii) $\|\frac{1}{N}\hat{\boldsymbol{\Delta}}_j^{\intercal}\boldsymbol{\alpha}_{\kappa}^0 - Q_{\kappa}\| = O_P(\eta_{N\tau_j}^{-1})$ *and* $\frac{1}{N}\|\hat{\boldsymbol{\Delta}}_j V_{N,j} - \boldsymbol{\alpha}_{\kappa}^0\Sigma_F Q_{\kappa}^{\intercal}\|^2 = O_P(\eta_{N\tau_j}^{-2})$ *for any* $j \in \mathbb{S}_1$,

where the matrix Q_{κ} *is invertible and is given by* $Q_{\kappa} = V_{\kappa}^{1/2}\Upsilon_{\kappa}^{\intercal}\Sigma_F^{-1/2}$, $V_{\kappa} = \mathrm{diag}(v_{1\kappa}, \ldots, v_{R\kappa})$, $v_{1\kappa} > v_{2\kappa} > \cdots > v_{R\kappa} > 0$ *are the eigenvalues of* $\Sigma_F^{1/2}\Sigma_{\kappa}\Sigma_F^{1/2}$, *and* Υ_{κ} *is the corresponding eigenvector matrix such that* $\Upsilon_{\kappa}^{\intercal}\Upsilon_{\kappa} = \mathbb{I}_R$.

Remark 3.1. The above result can be proved by modifying the arguments used in Bai (2003). Alternatively, they can be derived from the results in Proposition 3.2(ii) below.

Case 2. When a break point t_{κ}^0 lies in the interval $S_j = [v_j, v_{j+1})$, we have

$$\lambda_{it} = \begin{cases} \alpha_{i,\kappa-1}^0 & \text{for } t \in [v_j, t_{\kappa}^0) \\ \alpha_{i\kappa}^0 & \text{for } t \in [t_{\kappa}^0, v_{j+1}) \end{cases} \text{ for some } \kappa = \kappa(j).$$

Let $\mathbf{F}_{S_j,1}^0 = (F_t^0, t \in [v_j, t_{\kappa}^0))^{\intercal}$, $\mathbf{F}_{S_j,2}^0 = (F_t^0, t \in [t_{\kappa}^0, v_{j+1}))^{\intercal}$, and $\boldsymbol{\alpha}_{\kappa}^* = (\boldsymbol{\alpha}_{\kappa-1}^0, \boldsymbol{\alpha}_{\kappa}^0)$. Let $F_t^* = (F_t^{0\intercal}1_{jt}, F_t^{0\intercal}\bar{1}_{jt})^{\intercal}$ and $\mathbf{F}_{S_j}^* = (F_{v_j}^*, \ldots, F_{v_{j+1}}^*)^{\intercal}$, where $1_{jt} = 1\left\{v_j \le t < t_{\kappa}^0\right\}$, $\bar{1}_{jt} = 1\{t_{\kappa}^0 \le t < v_{j+1}\}$, and we suppress the dependence of F_t^* on j. Let $H_j^* = \frac{1}{\tau_j}\mathbf{F}_{S_j}^{*\intercal}\mathbf{F}_{S_j}^*\frac{1}{N}\boldsymbol{\alpha}_{\kappa}^{*\intercal}\hat{\boldsymbol{\Delta}}_j V_{N,j}^{-1}$, $H_{j,1} = (\frac{1}{\tau_j}\mathbf{F}_{S_j,1}^{0\intercal}\mathbf{F}_{S_j,1}^0)(\frac{1}{N}\boldsymbol{\alpha}_{\kappa-1}^{0\intercal}\hat{\boldsymbol{\Delta}}_j)V_{N,j}^{-1}$, and $H_{j,2} = (\frac{1}{\tau_j}\mathbf{F}_{S_j,2}^{0\intercal}\mathbf{F}_{S_j,2}^0)(\frac{1}{N}\boldsymbol{\alpha}_{\kappa}^{0\intercal}\hat{\boldsymbol{\Delta}}_j)V_{N,j}^{-1}$.

The following proposition establishes the asymptotic property of $\hat{\boldsymbol{\Delta}}_j$ in Case 2.

Proposition 3.2. *Suppose that Assumptions A1–A4 hold. Then*

(i) $\frac{1}{N}\|\hat{\boldsymbol{\Delta}}_j - \boldsymbol{\alpha}_{\kappa-1}^0 H_{j,1}\|^2 = O_P(c_{j2a}^2)$ *and* $\frac{1}{N}\|\hat{\boldsymbol{\Delta}}_j V_{N,j} - \boldsymbol{\alpha}_{\kappa-1}^0\Sigma_F Q_{\kappa-1}^{\intercal}\|^2 = O_P(c_{j2a}^2)$ *for all* $j \in \mathbb{S}_{2a}$;

(ii) $\frac{1}{N}\|\hat{\boldsymbol{\Delta}}_j - \boldsymbol{\alpha}_{\kappa}^0 H_{j,2}\|^2 = O_P(c_{j2b}^2)$ *and* $\frac{1}{N}\|\hat{\boldsymbol{\Delta}}_j V_{N,j} - \boldsymbol{\alpha}_{\kappa}^0\Sigma_F Q_{\kappa}^{\intercal}\|^2 = O_P(c_{j2b}^2)$ *for all* $j \in \mathbb{S}_{2b}$;

(iii) $\frac{1}{N}\|\hat{\boldsymbol{\Delta}}_j - \boldsymbol{\alpha}_{\kappa}^* H_j^*\|^2 = O_P(\eta_{N\tau_j}^{-2})$ *and* $\frac{1}{N}\|\hat{\boldsymbol{\Delta}}_j V_{N,j} - (N\tau_j)^{-1}(\boldsymbol{\alpha}_{\kappa-1}^0\mathbf{F}_{S_j,1}^{0\intercal}\mathbf{F}_{S_j,1}^0\boldsymbol{\alpha}_{\kappa-1}^{0\intercal} + \boldsymbol{\alpha}_{\kappa}^0\mathbf{F}_{S_j,2}^{0\intercal}\mathbf{F}_{S_j,2}^0\boldsymbol{\alpha}_{\kappa}^{0\intercal})\hat{\boldsymbol{\Delta}}_j\|^2 = O_P(\eta_{N\tau_j}^{-2})$ *for all* $j \in \mathbb{S}_{2c}$; *where* $c_{j2a} = \eta_{N\tau_j}^{-1} + \tau_{j2}/\tau_j$ *and* $c_{j2b} = \eta_{N\tau_j}^{-1} + \tau_{j1}/\tau_j$.

Remark 3.2. Proposition 3.2 indicates that the asymptotic properties of $\hat{\boldsymbol{\Delta}}_j$ and $\hat{\boldsymbol{\Delta}}_j V_{N,j}$ depend on whether j lies in $\mathbb{S}_{2a}$, $\mathbb{S}_{2b}$, or $\mathbb{S}_{2c}$. In particular, Proposition 3.2(i) (resp. (ii)) says, when the observations in S_j are mainly from regime $\kappa-1$ (resp. κ), the asymptotic property of $\hat{\boldsymbol{\Delta}}_j$ mainly depends on $\boldsymbol{\alpha}_{\kappa-1}^0$ (resp. $\boldsymbol{\alpha}_{\kappa}^0$), in which case the probability limit of $\hat{\boldsymbol{\Delta}}_j V_{N,j}$ will be different from that of $\hat{\boldsymbol{\Delta}}_{j+1}V_{N,j+1}$ (resp. $\hat{\boldsymbol{\Delta}}_{j-1}V_{N,j-1}$), given by $\boldsymbol{\alpha}_{\kappa}^0\Sigma_F Q_{\kappa}^{\intercal}$ (resp. $\boldsymbol{\alpha}_{\kappa-1}^0\Sigma_F Q_{\kappa-1}^{\intercal}$). In the case where $j \in \mathbb{S}_{2c}$, the limit of $\hat{\boldsymbol{\Delta}}_j V_{N,j}$ will be different from those of $\hat{\boldsymbol{\Delta}}_{j-1}V_{N,j-1}$ (which is given by $\boldsymbol{\alpha}_{\kappa-1}^0\Sigma_F Q_{\kappa-1}^{\intercal}$) and $\hat{\boldsymbol{\Delta}}_{j+1}V_{N,j+1}$ (which is given by $\boldsymbol{\alpha}_{\kappa}^0\Sigma_F Q_{\kappa}^{\intercal}$). This serves as the basis for the determination of the subintervals that contain a break point.

3.2. Identifying the intervals that contain a break point

Let Θ_j^* denote the elementwise probability limit of $\hat{\Theta}_j \equiv N^{1/2}\hat{\boldsymbol{\Delta}}_j V_{N,j}/\left\|\hat{\boldsymbol{\Delta}}_j V_{N,j}\right\|$, $j = 0, 1, \ldots, J$. In the absence of break points on the whole time interval $[1, T]$, we can readily show that $\Theta_j^* - \Theta_{j-1}^* = \mathbf{0}$ for $j = 1, \ldots, J$. In the general case, $\Theta_j^* - \Theta_{j-1}^*$ may be equal to or different from the zero matrix depending on whether the subinterval S_j or S_{j-1} contains a break point.

Let $|I_{\min}| = \min_{0\le\kappa\le m}|I_{\kappa}|$. To state the next result, we add the following two assumptions.

Assumption A5. For $\kappa = 1, 2, \ldots, m$, $\frac{1}{N}\left\|\boldsymbol{\alpha}_{\kappa}^0\Sigma_F Q_{\kappa}^{\intercal} - \boldsymbol{\alpha}_{\kappa-1}^0\Sigma_F Q_{\kappa-1}^{\intercal}\right\|^2 \to c_{\kappa} > 0$ as $(N, T) \to \infty$.

Assumption A6. (i) $\tau = O(N)$, $\tau\ln T = o(|I_{\min}|)$, and $m/J = o(1)$.

(ii) As $(N, T) \to \infty$, $(N\tau)^{1/2}\gamma = O(1)$ and $(N\tau)^{1/2}\gamma\eta_{N\tau}^{\varkappa}/J \to \infty$.

A5 ensures that parameters of interest in neighboring segments are distinct from each other. Note that $Q_{\kappa} = V_{\kappa}^{1/2}\Upsilon_{\kappa}^{\intercal}\Sigma_F^{-1/2}$, where V_{κ} and Υ_{κ} collect the eigenvalues and normalized eigenvectors of $\Sigma_F^{1/2}\Sigma_{\kappa}\Sigma_F^{1/2}$, and Σ_{κ} denotes the limit of $\frac{1}{N}\boldsymbol{\alpha}_\kappa^{0\top}\boldsymbol{\alpha}_\kappa^0$. If $\boldsymbol{\alpha}_\kappa^0 = \boldsymbol{\alpha}_{\kappa-1}^0$, then $\boldsymbol{\alpha}_\kappa^0 \Sigma_F Q_\kappa^\top = \boldsymbol{\alpha}_{\kappa-1}^0 \Sigma_F Q_{\kappa-1}^\top$. When $\boldsymbol{\alpha}_\kappa^0$ and $\boldsymbol{\alpha}_{\kappa-1}^0$ are distinct from each other such that $\frac{1}{N}\left\|\boldsymbol{\alpha}_\kappa^0 - \boldsymbol{\alpha}_{\kappa-1}^0\right\|^2 \to c_{\alpha\kappa}$ for some $c_{\alpha\kappa} > 0$, we generally expect A5 to be satisfied. A6(i) ensures that $\eta_{N\tau_j}^{-1} = O(\tau_j^{-1/2})$ and each interval $S_j, j = 0, 1, \ldots, J$, contains at most one break. A6(ii) requires that γ converge to zero at a suitable rate, which is required to identify all intervals that do not contain a break point.

The next proposition is crucial for identifying the intervals that contain the break points.

Proposition 3.3. *Suppose that Assumptions A1–A6 hold. Then*

(i) $N^{-1}\left\|\tilde{\Theta}_j - \Theta_j^*\right\|^2 = O_P(a_j^2)$ *for all* $j \in \mathbb{S}$,

(ii) $\frac{1}{N(J+1)}\sum_{j=0}^{J}\left\|\tilde{\Theta}_j - \Theta_j^*\right\|^2 = O_P\left(\eta_{N\tau}^{-2} + m/J\right)$,

(iii) $\Pr\left\{\left\|\tilde{\Theta}_j - \tilde{\Theta}_{j-1}\right\| = 0 \text{ for all } j,\ j-1 \in \mathbb{S}_1\right\} \to 1$ *as* $(N, T) \to \infty$,

where $a_j = \eta_{N\tau_j}^{-1}$ *if* $j \in \mathbb{S}_1 \cup \mathbb{S}_{2c}$, $a_j = c_{j2a}$ *if* $j \in \mathbb{S}_{2a}$, *and* $a_j = c_{j2b}$ *if* $j \in \mathbb{S}_{2b}$, *and* c_{j2a} *and* c_{j2b} *are defined in Proposition 3.2.*

Remark 3.3. Proposition 3.3(i) establishes the mean square convergence rates of the penalized estimators $\tilde{\Theta}_j$ which depend on whether $j \in \mathbb{S}_1, \mathbb{S}_{2a}, \mathbb{S}_{2b}$, or $\mathbb{S}_{2c}$. Proposition 3.3(ii) is the average version of (i). Proposition 3.3(iii) establishes the selection consistency of our AGFL method; it says that w.p.a.1 all the zero matrices $\{\Theta_j^* - \Theta_{j-1}^*, j, j-1 \in \mathbb{S}_1\}$ must be estimated as exactly zeros by the AGFL method. On the other hand, we notice that $\Theta_j^* - \Theta_{j-1}^* = 0$ if $j-1 \in \mathbb{S}_1$ and $j \in \mathbb{S}_{2a}$, or $j-1 \in \mathbb{S}_{2b}$ and $j \in \mathbb{S}_1$. In the latter two cases, the estimate $\tilde{\Theta}_j - \tilde{\Theta}_{j-1}$ of $\Theta_j^* - \Theta_{j-1}^*$ may be zero or nonzero, depending on whether we allow $(N\tau)^{1/2}\gamma\left(\tau_{j2}/\tau_j\right)^{-2\varkappa}$ to pass to infinity in the case where $j-1 \in \mathbb{S}_1$ and $j \in \mathbb{S}_{2a}$, and $(N\tau)^{1/2}\gamma\left(\tau_{j1}/\tau_j\right)^{-2\varkappa}$ to pass to infinity in the case where $j-1 \in \mathbb{S}_1$ and $j \in \mathbb{S}_{2b}$. If the latter two conditions are satisfied, a close examination of the proof of Proposition 3.3(iii) indicates that $\Theta_j^* - \Theta_{j-1}^*$ will also be estimated by exactly zero in large samples when $j-1 \in \mathbb{S}_1$ and $j \in \mathbb{S}_{2a}$, or $j-1 \in \mathbb{S}_{2b}$ and $j \in \mathbb{S}_1$. On the other hand, by (i), we know that the matrices $\Theta_j^* - \Theta_{j-1}^*$ can be consistently estimated by $\tilde{\Theta}_j - \tilde{\Theta}_{j-1}$. Putting these results together, Proposition 3.3 implies that the AGFL is capable of identifying the intervals among $\{S_j, j = 0, 1, \ldots, J\}$ that might contain an unknown break point. Recall that we use $\hat{m}$ to denote the estimated number of break points. A direct implication of Proposition 3.3 is that

$$\Pr\left(\hat{m} = m\right) \to 1 \text{ as } (N, T) \to \infty. \tag{3.1}$$

Remark 3.4. In order to see whether a subinterval $S_j, j = 1, \ldots, J-1$, contains a break point (say, t_κ^0) or not, we need to compare Θ_j^* with both Θ_{j-1}^* and Θ_{j+1}^* at the population level or compare $\tilde{\Theta}_j$ with both $\tilde{\Theta}_{j-1}$ and $\tilde{\Theta}_{j+1}$ at the sample level. At the population level, we have four scenarios: (1) $\Theta_{j-1}^* \neq \Theta_j^* \neq \Theta_{j+1}^*$ when $j \in \mathbb{S}_{2c}$, (2) $\Theta_{j-2}^* = \Theta_{j-1}^* \neq \Theta_j^* = \Theta_{j+1}^*$ when $j \in \mathbb{S}_{2b}$ or $j-1 \in \mathbb{S}_{2a}$, (3) $\Theta_{j-1}^* = \Theta_j^* \neq \Theta_{j+1}^* = \Theta_{j+2}^*$ when $j \in \mathbb{S}_{2a}$ or $j+1 \in \mathbb{S}_{2b}$, (d) $\Theta_{j-1}^* = \Theta_j^* = \Theta_{j+1}^*$ when $j \in \mathbb{S}_1$. In case (1), we can conclude that we have an estimated break point in the interval S_j, and for cases (2) and (3), we can conclude that a break point happens in S_{j-1}^* and S_j^*, respectively (see Section 2.2.3 for the definitions of S_{j-1}^* and S_j^*). The sample case has been discussed at the beginning of Section 2.2.3. In addition, under the condition that $|I_{\min}| \gg T/J$, any finite fixed number of consecutive intervals (e.g., S_{j-1}, S_j, and S_{j+1}) can contain at most one break, and S_0 and S_J cannot contain any break. Such information is useful to prove the result in Proposition 3.3.

3.3. Estimation of the break dates

Assumption A7. $\frac{1}{N}\left\|\left(\boldsymbol{\alpha}_\kappa^0 - \boldsymbol{\alpha}_{\kappa-1}^0\right)F_{t_\kappa^0-\ell}^0\right\|^2 \gg c_{NT}$ for $\ell = 0{,}1$ and $\kappa = 1, \ldots, m$, where $c_{NT} = |I_{\min}|^{-1/2}(\ln T)^{3/2}$ $+(\tau \ln T/|I_{\min}|)^{1/2}$.

Assumption A8. (i) $\max_{1\le i\le N}\max_{1\le r\le T-s}\|\frac{1}{s}\sum_{t=r}^{r+s}F_t^0 e_{it}\| = O_P\left((s/\ln T)^{-1/2}\right)$ for any $s \to \infty$.

(ii) $\max_{1\le s,t\le T}\frac{1}{N}\left|\mathbf{e}_t^\top \mathbf{e}_s - E\left(\mathbf{e}_t^\top \mathbf{e}_s\right)\right| = O_P\left((N/\ln T)^{-1/2}\right)$.

(iii) $\max_{0\le\kappa\le m}\max_{1\le t\le T}\frac{1}{N}\left\|\boldsymbol{\alpha}_\kappa^{0\top}\mathbf{e}_t\right\| = O_P\left((N/\ln T)^{-1/2}\right)$.

(iv) $\max_{j\in\mathbb{S}_1}\left\|\mathbf{E}_{S_j}\right\|_{\mathrm{sp}} = O_P(\max(\sqrt{N}, \sqrt{\tau_j}))$ where $\mathbf{E}_{S_j} = (\mathbf{e}_t, t \in S_j)$.

Assumption A7 is needed to consistently estimate all m break points. To understand this, we focus on the case where $\mathbf{D}_{N,\kappa} \equiv \frac{1}{N}\left(\boldsymbol{\alpha}_\kappa^0 - \boldsymbol{\alpha}_{\kappa-1}^0\right)^\top\left(\boldsymbol{\alpha}_\kappa^0 - \boldsymbol{\alpha}_{\kappa-1}^0\right) \to \mathbf{D}_\kappa > 0$. In this case,

$$\frac{1}{N}\left\|\left(\boldsymbol{\alpha}_\kappa^0 - \boldsymbol{\alpha}_{\kappa-1}^0\right)F_{t_\kappa^0-\ell}^0\right\|^2 = \operatorname{tr}\left(\mathbf{D}_{N,\kappa}F_{t_\kappa^0-\ell}^0 F_{t_\kappa^0-\ell}^{0\top}\right) \ge \mu_{\min}\left(\mathbf{D}_{N,\kappa}\right)\left\|F_{t_\kappa^0-\ell}^0\right\|^2 \gg c_{NT} \text{ almost surely.}$$

A8 is used to obtain some uniform result and can be verified under certain primitive conditions. For example, under certain strong mixing and moment conditions on the process $\{F_t^0 e_{it}, t \ge 1\}$, A8(i) can be verified by a simple use of Bernstein inequality for strong mixing processes provided that N and T diverge to infinity at comparable rates. See Moon and Weidner (2015) for primitive conditions to ensure A8(iv) to hold.

The next proposition establishes the super-consistency of the estimators of the break points.

Proposition 3.4. *Suppose that Assumptions A1–A8 hold. Then* $\Pr(\hat{t}_1 = t_1^0, \ldots, \hat{t}_m = t_m^0 | \hat{m} = m) \to 1$ *as* $(N, T) \to \infty$.

Remark 3.5. In conjunction with (3.1), the above proposition indicates that we can estimate the break dates precisely w.p.a.1. This result is much stronger that the first set of results in BKW. BKW consider both joint and sequential estimation of the break dates in large dimensional factor models with an unknown number of structural changes. Conditioning on the correct determination of the number of structural changes, they show that the distance between the estimated and true break dates are $O_P(1)$, which implies the consistency of the estimators of the break fractions $(t_\kappa^0/T, \kappa = 1, \ldots, m)$. Nevertheless, BKW also consider simultaneous search of the multiple break dates after one obtains the estimated numbers of factors and breaks; and they claim the consistency of the estimators of the break dates. This re-estimation step parallels to Step III in our procedure with the only difference that our method is sequential while their re-estimation is joint. In either case, the consistency of the break dates estimators are expected.

4. Practical issues

In this section we first discuss the determination of the number of factors and then propose an information criterion to choose the tuning parameter γ.

4.1. Determination of the number of factors

In the above analysis, we assume that the number of factors, R, is known. In practice, one has to determine R from the data. Here we assume that the true value of R, denoted as R_0, is bounded from above by a finite integer $R_{\max}$. We propose a BIC-type information criterion to determine R_0.

Now, we use $\hat{\boldsymbol{\Delta}}_j(R)$ and $\hat{F}_t(R)$ to denote the estimators of $\boldsymbol{\Delta}_j$ and F_t by using R factors defined in Section 2.2.1. Let $\breve{\boldsymbol{\Delta}}_j(R) = \left(N\tau_j\right)^{-1}\mathbf{X}_{S_j}\mathbf{X}_{S_j}^{\intercal}\hat{\boldsymbol{\Delta}}_j(R)$ for $j = 0, 1, \ldots, J$. Define

$$\begin{aligned} V(R) &= V(R, \{\breve{\boldsymbol{\Delta}}_j(R)\}) \\ &= \min_{\{F_1(R),\ldots,F_T(R)\}} (J+1)^{-1}\sum_{j=0}^{J}(N\tau_j)^{-1}\sum_{t\in S_j}(\mathbf{X}_t - \breve{\boldsymbol{\Delta}}_j(R)F_t(R))^{\intercal}(\mathbf{X}_t - \breve{\boldsymbol{\Delta}}_j(R)F_t(R)). \end{aligned}$$

Following the lead of Bai and Ng (2002), we consider the following BIC-type information criterion to determine R_0:

$$IC_1(R) = \ln V\left(R, \left\{\breve{\boldsymbol{\Delta}}_j(R)\right\}\right) + \rho_{1NT}R, \tag{4.1}$$

where ρ_{1NT} plays the role of $\ln(NT)/(NT)$ in the case of BIC. Let $\hat{R} = \arg\min_R IC_1(R)$. We add the following assumption.

Assumption A9. As $(N, T) \to \infty$, $\rho_{1NT} \to 0$ and $\rho_{1NT}/(mJ^{-1} + \eta_{N\tau}^{-2}) \to \infty$ where $\eta_{N\tau} = \min(\sqrt{N}, \sqrt{\tau})$ and $\tau = \min_{0\le j\le J}\tau_j$.

The conditions on ρ_{1NT} in A9 are typical conditions in order to estimate the number of factors consistently. The penalty coefficient ρ_{1NT} has to shrink to zero at an appropriate rate to avoid both overfitting and underfitting.

Proposition 4.1. *Suppose that Assumptions A1–A4 and A8–A9 hold. Then* $P(\hat{R} = R_0) \to 1$ *as* $(N, T) \to \infty$.

Remark 4.1. Proposition 4.1 indicates that we can minimize $IC_1(R)$ to consistently estimate R_0. To implement the information criterion, one needs to choose the penalty coefficient ρ_{1NT}. Following the lead of Bai and Ng (2002), we suggest setting $\rho_{1NT} = \frac{N+\bar{\tau}}{N\bar{\tau}}\ln\left(\frac{N\bar{\tau}}{N+\bar{\tau}}\right)$ or $\rho_{1NT} = \frac{N+\bar{\tau}}{N\bar{\tau}}\ln\eta_{N\bar{\tau}}^2$ with $\eta_{N\bar{\tau}} = \min\{\sqrt{\bar{\tau}}, \sqrt{N}\}$ and $\bar{\tau} = T/(J+1)$, and evaluate the performance of these two information criteria in our simulation studies. Define

$$\begin{aligned} IC_{1a}(R) &= \log(V(R)) + R\frac{N+\bar{\tau}}{N\bar{\tau}}\ln\left(\frac{N\bar{\tau}}{N+\bar{\tau}}\right), \\ IC_{1b}(R) &= \log(V(R)) + R\frac{N+\bar{\tau}}{N\bar{\tau}}\ln\eta_{N\bar{\tau}}^2. \end{aligned} \tag{4.2}$$

Let $\hat{R}_{1a} = \arg\min_R IC_{1a}(R)$ and $\hat{R}_{1b} = \arg\min_R IC_{1b}(R)$. When the number of breaks, m, is fixed, it appears that one can choose J such that $J \asymp \bar{\tau}$, in which case $J/m + \eta_{N\bar{\tau}}^2 \asymp \eta_{N\bar{\tau}}^2$ provided $\bar{\tau} = O(N)$.

4.2. Choice of the tuning parameter γ

We now discuss the choice of the tuning parameter γ, which is an important issue when the penalized objective function in (2.2) is used in practice. (2.2) suggests that a too large value of γ tends to under-estimate the true number of breaks, denoted as m^0 hereafter; similarly, a too small value of γ tends to over-estimate m^0. Therefore it is sensible to choose a data-driven γ such that m^0 can be identified.

To proceed, we assume the existence of a closed interval, namely, $\Gamma \equiv [\gamma_{\min}, \gamma_{\max}]$, such that when $\gamma = \gamma_{\min}$, one can identify at most $m_{\max} \geq m^0$ breaks, and when $\gamma = \gamma_{\max}$, one does not identify any break. If one believes that the number of breaks is fixed when $(N,T) \to \infty$ as in many applications, it is reasonable to conjecture a finite value for $m_{\max}$. Then $\gamma_{\min}$ and $\gamma_{\max}$ can be easily pinned down from the data.

Given $\gamma \in [\gamma_{\min}, \gamma_{\max}]$, we can apply the three-step procedure in Section 2 to obtain the break point estimates $\hat{t}_\kappa(\gamma)$, $\kappa = 1, \ldots, \hat{m}_\gamma$, where we make the dependence of $\hat{t}_\kappa(\gamma)$ and $\hat{m}_\gamma$ on γ explicit. Let $\hat{I}_\kappa(\gamma) = [\hat{t}_{\kappa-1}(\gamma), \hat{t}_\kappa(\gamma))$ for $\kappa = 1, \ldots, \hat{m}_\gamma + 1$, where $\hat{t}_0(\gamma) = 1$ and $\hat{t}_{\hat{m}_\gamma+1}(\gamma) \equiv T+1$. On segment $\hat{I}_\kappa(\gamma)$, we estimate the factors and their loadings as

$$(\hat{\boldsymbol{\alpha}}_\kappa(\gamma), \{\hat{F}_t(\gamma)\}) = \arg\min_{\boldsymbol{\alpha}_\kappa, \{F_t\}} \sum_{t \in \hat{I}_\kappa(\gamma)} (\mathbf{X}_t - \boldsymbol{\alpha}_\kappa F_t)^\intercal (\mathbf{X}_t - \boldsymbol{\alpha}_\kappa F_t)$$

subject to the constraints that $N^{-1}\boldsymbol{\alpha}_\kappa^\intercal \boldsymbol{\alpha}_\kappa = \mathbb{I}_R$ and $\mathbf{F}_{\hat{I}_\kappa(\gamma)}^\intercal \mathbf{F}_{\hat{I}_\kappa(\gamma)} = $ diagonal. Let $\hat{\mathcal{T}}_{\hat{m}_\gamma}(\gamma) = \{\hat{t}_1(\gamma), \ldots, \hat{t}_{\hat{m}_\gamma}(\gamma)\}$. Define

$$\hat{\sigma}^2\left(\hat{\mathcal{T}}_{\hat{m}_\gamma}(\gamma)\right) = \frac{1}{NT} \sum_{j=1}^{\hat{m}_\gamma+1} \sum_{t=\hat{t}_j(\gamma)}^{\hat{t}_j(\gamma)-1} \left[\mathbf{X}_t - \hat{\boldsymbol{\alpha}}_j(\gamma)\hat{F}_t(\gamma)\right]^\intercal \left[\mathbf{X}_t - \hat{\boldsymbol{\alpha}}_j(\gamma)\hat{F}_t(\gamma)\right].$$

Following the lead of Li et al. (2016), we propose to select $\gamma \in \Gamma$ to minimize the following information criterion

$$IC_2(\gamma) = \log\left[\hat{\sigma}^2\left(\hat{\mathcal{T}}_{\hat{m}_\gamma}(\gamma)\right)\right] + \rho_{2NT}\left(\hat{m}_\gamma + 1\right), \tag{4.3}$$

where ρ_{2NT} is a predetermined tuning parameter that satisfies certain conditions. Let $\hat{\gamma} = \mathrm{argmin}_{\gamma \in \Gamma} IC_2(\gamma)$. We add the following assumption.

Assumption A10. (i) For any $0 \leq m < m^0$, there exists a positive non-increasing sequence c_{1NT} and a positive constant $\underline{c}_\lambda$ such that

$$\mathrm{plim}_{(N,T)\to\infty} \min_{\mathcal{T}_m} \min_{\left\{\boldsymbol{\alpha}_j :\, N^{-1}\boldsymbol{\alpha}_j^\intercal \boldsymbol{\alpha}_j = \mathbb{I}_R\right\}} c_{1NT}^{-1} \frac{1}{NT} \sum_{j=1}^{m+1} \sum_{t=t_{j-1}}^{t_j-1} F_t^{0\top} (\boldsymbol{\alpha}_j - \boldsymbol{\lambda}_t^0)^\top \boldsymbol{M}_{\boldsymbol{\alpha}_j} (\boldsymbol{\alpha}_j - \boldsymbol{\lambda}_t^0) F_t^0 \geq \underline{c}_\lambda,$$

where $\mathcal{T}_m = \{t_1, \ldots, t_m\}$ with $1 < t_1 < \cdots < t_m < T$, and $\boldsymbol{M}_{\boldsymbol{\alpha}_j} = \mathbb{I}_N - \boldsymbol{\alpha}_j(\boldsymbol{\alpha}_j^\top \boldsymbol{\alpha}_j)^+ \boldsymbol{\alpha}_j^\top$.

(ii) As $(N,T) \to \infty$, $c_{1NT}\eta_{NI_{\min}} \to \infty$ and $c_{1NT}^{-1}\rho_{2NT} m^0 \to 0$.

(iii) As $(N,T) \to \infty$, $\rho_{2NT} c_{2NT}^{-1} \to \infty$ where $c_{2NT} = N^{-1} + I_{\min}^{-1} + m_{\max} T^{-1}$.

Assumptions A10(i) and (ii) impose conditions to avoid the selection of γ to yield fewer breaks than the true number by using $IC_2(\gamma)$ in (4.3). A10(iii) specifies conditions to avoid the selection of γ to yield more breaks than the true number. The remark after the proof of Lemma A.8 in the online Appendix B discusses cases where $c_{1NT} = 1$ when the m^0 is fixed and $I_{\min}$ is proportional to T.

Proposition 4.2. *Suppose that Assumptions A1–A8 and A10 hold. Then $P(\hat{m}_{\hat{\gamma}} = m^0) \to 1$ as $(N,T) \to \infty$.*

Remark 4.2. Proposition 4.2 indicates that by minimizing $IC_2(\gamma)$ we can obtain a data-driven $\hat{\gamma}$ that ensures the correct determination of the number of breaks asymptotically. When we minimize $IC_2(\gamma)$ in (4.3), we do not restrict γ to satisfy Assumption A6(ii). If A6(ii) is satisfied, we know from Proposition 3.3 that $\hat{m}_\gamma$ is given by the true number of breaks (m^0) w.p.a.1. But in practice, it is hard to ensure such an assumption is satisfied and Proposition 4.2 becomes handy.

To implement $IC_2(\gamma)$ in practice, it is often reasonable to assume that m^0 and $m_{\max}$ are fixed and $I_{\min} \propto T$. In this case, Assumption A10 holds with $c_{1NT} = 1$ and $c_{2NT} = N^{-1} + T^{-1}$ under some weak conditions. Then one can specify ρ_{2NT} as follows

$$\rho_{2NT} = \frac{c \log(\min(N,T))}{\min(N,T)}, \tag{4.4}$$

where c is a positive constant. Following Hallin and Liska (2007) and Li et al. (2016), one can also apply a data-driven procedure to determine c.

5. Monte Carlo simulations

In this section, we conduct simulation studies to assess the finite-sample performance of our proposed break detection procedure.

5.1. Data generating processes

We generate data under the framework of high dimensional factor models with $R = 2$ common factors:

$$X_{it} = \lambda_{it}^{\intercal} F_t + e_{it}, \ i = 1, \dots, N, \ t = 1, \dots, T,$$

where $F_t = (F_{1t}, F_{2t})^{\intercal}$, $F_{1t} = 0.6F_{1,t-1} + u_{1t}$, u_{1t} are i.i.d. $N(0, 1 - 0.6^2)$, $F_{2t} = 0.3F_{2,t-1} + u_{2t}$, u_{2t} are i.i.d. $N(0, 1 - 0.3^2)$ and independent of u_{1t}. We consider the following setups for the factor loadings λ_{it} and error terms e_{it}.

DGP1: (Single structural break)

$$\lambda_{it} = \begin{cases} \alpha_{i1} & \text{for } t = 1, 2, \dots, t_1 - 1 \\ \alpha_{i2} & \text{for } t = t_1, t_1 + 1, \dots, T \end{cases},$$

where α_{i1} are from i.i.d. $N((0.5b, 0.5b)^{\intercal}, ((1, 0)^{\intercal}, (0, 1)^{\intercal}))$ and α_{i2} are from i.i.d. $N((b, b)^{\intercal}, ((1, 0)^{\intercal}, (0, 1)^{\intercal}))$ and independent of α_{i1} . The error terms e_{it} are generated in two ways: (1) (IID) e_{it} are i.i.d. $N(0, 2)$, and (2) (CHeter) $e_{it} = \sigma_i v_{it}$, where σ_i are i.i.d. $U(0.5, 1.5)$, v_{it} are from i.i.d. $N(0, 2)$, and CHeter denotes cross-sectional heterogeneity in the error terms. Let $b = 1$,2.

DGP2: (Multiple structural breaks)

$$\lambda_{it} = \begin{cases} \alpha_{i1} & \text{for } t = 1, 2, \dots, t_1 - 1 \\ \alpha_{i2} & \text{for } t = t_1, t_1 + 1, \dots, t_2 - 1 \\ \alpha_{i3} & \text{for } t = t_2, t_2 + 1, \dots, T \end{cases},$$

where α_{i1} are from i.i.d. $N((0.5b, 0.5b)^{\intercal}, ((1, 0)^{\intercal}, (0, 1)^{\intercal}))$, α_{i2} are from i.i.d. $N((b, b)^{\intercal}, ((1, 0)^{\intercal}, (0, 1)^{\intercal}))$, α_{i3} are from i.i.d. $N((1.5b, 1.5b)^{\intercal}, ((1, 0)^{\intercal}, (0, 1)^{\intercal}))$, and they are mutually independent of each other. The error terms e_{it} are generated in two ways: (1) (IID) e_{it} are i.i.d. $N(0, 2)$, and (2) (AR(1)) $e_{it} = 0.2e_{it-1} + u_{it}$, where u_{it} are i.i.d. $N(0, 2(1 - 0.2^2))$. Let $b = 1$,2.

DGP3: (No breaks) $\lambda_{it} = \alpha_i$ and α_i are i.i.d. $N((1, 1)^{\intercal}, ((1, 0)^{\intercal}, (0, 1)^{\intercal}))$. The error terms e_{it} are i.i.d. $N(0, 2)$.

For each DGP, we simulate 1000 datasets with sample sizes $T = 250$, 500 and $N = 50$. Since the factor loadings are assumed to be nonrandom, we generate them once and fix them across the 1000 replications. We use $J+1 = 10$ subintervals for $T = 250$ and use $J + 1 = 10$, 15 subintervals for $T = 500$ in the piecewise constant estimation in Step I.

In DGP1, we consider two cases:

(Case 1) we set the break date $t_1 = T/2$, so that $t_1 = 125$ and 250 for $T = 250$ and 500, respectively;
(Case 2) we set $t_1 = T/2 + \lfloor 0.5T/(J + 1)\rfloor$, so that $t_1 = 137$ for $T = 250$ and $t_1 = 275, 266$ for $J + 1 = 10$, 15 and $T = 500$.

It is worth noting that when $T = 250$, $t_1 = 125$ is in the boundary of some subinterval and $t_1 = 137$ is located in the interior of the subinterval. When $T = 500$, $t_1 = 250$ and 266 are in the boundary of some subinterval, respectively, for $J + 1 = 10$ and 15, and $t_1 = 275$ and 250 are in the interior of some subinterval, respectively, for $J + 1 = 10$ and 15.

In DGP2, we consider two cases:

(Case 1) we set the breaks $t_1 = 0.3T$ and $t_2 = 0.7T$, so that $t_1 = 75$ and 150 for $T = 250$ and 500, and $t_2 = 175$ and 350 for $T = 250$ and 500;
(Case 2) we let $t_1 = 0.3T$ and $t_2 = 0.6T + \lfloor 0.5T/(J + 1)\rfloor$, so that $t_2 = 162$ for $T = 250$ and $t_2 = 325, 316$ for $J + 1 = 10$,15 and $T = 500$.

Similarly to DGP1, some breaks are located in the boundary of an interval and some are in the interior of an interval.

5.2. Determination of the number of factors

First, we assume that the true number of factors is unknown. We select the number of factors by the two information criteria $IC_{1a}(R)$ and $IC_{1b}(R)$ given in (4.2) of Section 4.1 . Since the information criteria also depend on J and J plays the role of the trimming parameter ϵ in Assumption A4(ii) of Bai and Perron (1998, BP hereafter), we follow Bai and Perron's (2003) recommendation and consider 5–25% of observations within each subinterval (i.e., $\epsilon \in [0.05, 0.25]$). Recall that BP requires that each regime has at least ϵT observations, and the larger value ϵ takes, the smaller number of breaks are allowed. Specifically, when $T = 250$, we set $J + 1 = 10$ which corresponds to BP's $\epsilon = 0.1$; when $T = 500$, we set $J + 1 = 10$ and 15, which correspond to BP's $\epsilon = 0.1$ and 0.0667, respectively.

Table 1 presents the average selected number of factors (AVE) and the empirical probability of correct selection (PROB) by the two information criteria for DGP1-3 with $b = 1$. We observe that the AVE is equal to or close to two, which is the true number of factors, and the PROB is equal to or close to one for all cases. The results in Table 1 demonstrate the selection consistency of the two information criteria established in Section 4.1.

To illustrate the relationship between the IC values and the number of factors, Fig. 1 shows the average value of $IC_{1a}(R)$ (thin line) and $IC_{1b}(R)$ (thick line) among 1000 replications against the number of factors for (a) DGP1-Case1 with $T = 250$ and cross-sectionally heteroskedastic error terms; (b) DGP2-Case1 with $T = 250$ and autoregressive error terms; and (c) DGP3. We observe that the average IC value reaches its minimum at $R = 2$ in these three plots. In addition, we find that $IC_{1b}(R)$ has steeper slope than $IC_{1a}(R)$ when $R > 2$ so that it helps to avoid overselecting the number of factors.

Table 1
Performance of the two information criteria in determining the number of factors: DGPs 1–3 with $b = 1$.

$(T, J+1)$	IC_{1a}			IC_{1b}		
	(250, 10)	(500, 10)	(500, 15)	(250, 10)	(500, 10)	(500, 15)
Average selected number of factors						
DGP1-IID						
Case 1	2.000	2.000	2.000	2.000	2.000	2.000
Case 2	2.000	2.000	2.000	2.000	2.000	2.000
DGP1-CHeter						
Case 1	2.000	2.000	2.000	2.000	2.000	2.000
Case 2	2.003	2.001	2.000	2.000	2.000	2.000
DGP2-IID						
Case 1	2.000	2.000	2.000	2.000	2.000	2.000
Case 2	2.001	2.000	2.000	2.000	2.000	2.000
DGP2-AR						
Case 1	2.000	2.000	2.000	1.998	2.000	2.000
Case 2	2.000	2.000	2.000	2.000	2.000	2.000
DGP3	2.000	2.000	2.000	2.000	2.000	2.000
Empirical probability of correct selection						
DGP1-IID						
Case 1	1.000	1.000	1.000	1.000	1.000	1.000
Case 2	1.000	1.000	1.000	1.000	1.000	1.000
DGP1-CHeter						
Case 1	1.000	1.000	1.000	1.000	1.000	1.000
Case 2	0.997	0.999	1.000	1.000	1.000	1.000
DGP2-IID						
Case 1	1.000	1.000	1.000	1.000	1.000	1.000
Case 2	0.999	1.000	1.000	1.000	1.000	1.000
DGP2-AR						
Case 1	1.000	1.000	1.000	0.998	1.000	1.000
Case 2	1.000	1.000	1.000	1.000	1.000	1.000
DGP3	1.000	1.000	1.000	1.000	1.000	1.000

5.3. Estimation of the break points

Following the literature on adaptive Lasso, we set $\varkappa = 2$ and 4 to determine the adaptive weight in the adaptive fused Lasso penalty given in Section 2.2.2. For a larger value of $\varkappa$, more sparsity is induced. We select the tuning parameter γ by minimizing the information criterion (4.4) given in Section 4.2. We set $c = 0.15$ as suggested in Hallin and Liska (2007). To examine the break detection performance, we calculate the percentages of correct estimation (C) of m, and conditional on the correct estimation of m, the accuracy of break date estimation, which is measured by average Hausdorff distance of the estimated and true break points divided by T (HD/T). Let $\mathcal{D}(A, B) \equiv \sup_{b\in B}\inf_{a\in A}|a - b|$ for any two sets A and B. The Hausdorff distance between A and B is defined as $\max\{\mathcal{D}(A, B), \mathcal{D}(B, A)\}$.

The results for DGP 1–2 are shown in Tables 2 and 3 for $\varkappa = 2$ and 4, respectively. All figures in the tables are in percentages (%). We observe that the percentage of correct estimation is closer to 100% for the larger signal of $b = 2$. By using the same number of subintervals with $J + 1 = 10$, the C value for $T = 500$ is larger than that for $T = 250$, and the HD/T value for $T = 500$ is smaller than that for $T = 250$ for all cases. Moreover, for the same $T = 500$, the break detection procedure performs better by using $J + 1 = 10$ subintervals than $J + 1 = 15$ subintervals by observing larger C values for most cases. Furthermore, the HD/T value for breaks located at the boundaries of the subintervals is smaller than that for breaks in the interior of the subintervals. For example, for DGP1-IID with $T = 500$, for $J + 1 = 10$, the 100×HD/T value for Case 1 (0.029) is smaller than that (0.154) for Case 2, since the break is in the boundary for Case 1 and it is in the interior of some subinterval for Case 2. However, the result is reversed for $J + 1 = 15$ by observing 0.288 and 0.025, respectively, for Case 1 and Case 2, since the break is in the boundary for Case 2 for this scenario.

To further evaluate the three-step break detection procedure for DGP1 with one break point, we calculate the frequency for all identified break points among 1000 replications. Since the percentage of correct estimation for $\varkappa = 4$ is higher than that for $\varkappa = 2$ for each case, in the following we just report the results for $\varkappa = 2$ to save spaces. Figs. 2–4 show the plots of the frequency of the identified breaks among 1000 replications for DGP1 and for $T = 250$ and $J + 1 = 10$, and $T = 500$ and $J + 1 = 10,15$, respectively. The blue shaded line with angle=135 is for $b = 1$ and the red shaded line with angle=45 is for $b = 2$. For plots (a) and (b) of Fig. 2, the true break is at $t_1 = 125$, and for plots (c) and (d) of Fig. 2, the true break is at $t_1 = 137$. We see that the height of the frequency bar around the true break is close to 1000. This indicates that the three-step procedure can identify the true break or some neighborhood point as a break with a high chance. For the stronger signal with $b = 2$, the identified breaks are more concentrated around the true break than those for the weaker signal with

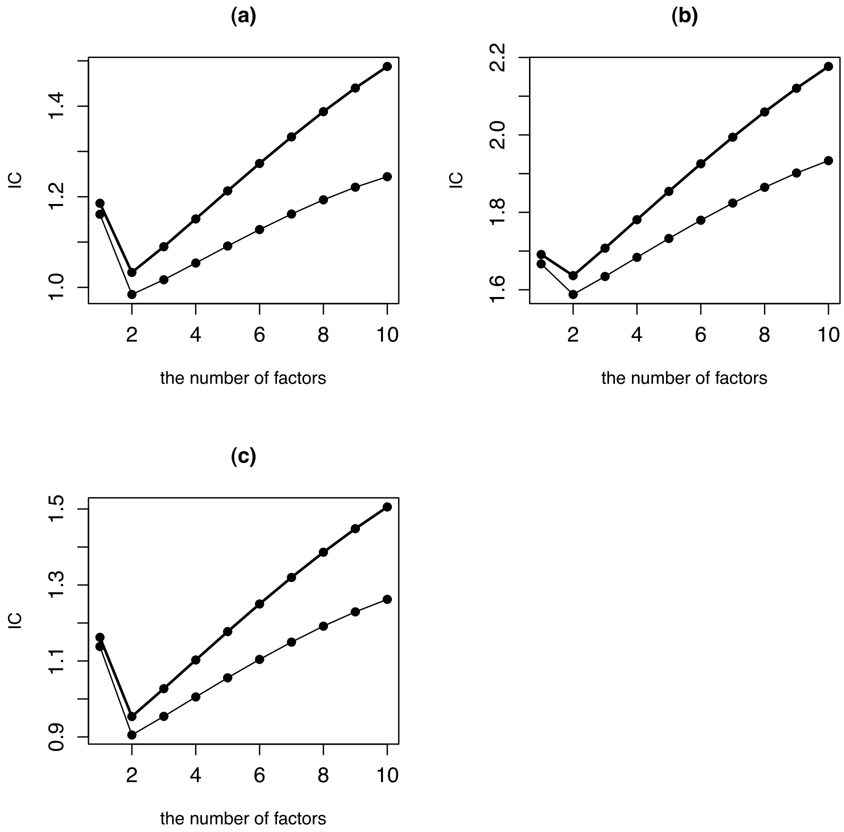

Fig. 1. Plots of the IC_{1a} (thin line) and IC_{1b} (thick line) against the number of factors with $b = 1$ for (a) DGP1-Case1 with $T = 250$ and cross-sectional heteroscedastic errors, (b) DGP2-Case1 with $T = 250$ and autoregressive errors. and (c) DGP3.

$b = 1$. Moreover, by using the same number of subintervals with $J + 1 = 10$, when we increase the T value from 250 to 500, the frequency bar around the true break is closer to 1000 as shown in Fig. 3. For $T = 500$, when we increase $J + 1$ from 10 to 15, more points are identified as breaks, especially for the weaker signal with $b = 1$ as shown in Figs. 3 and 4. Figs. 5–7 show the plots of the frequency of the identified breaks among 1000 replications for DGP2. We see that the two true breaks can be identified well. We can observe similar patterns as the frequency plots for DGP1. For example, for larger T value, the frequency bars around the true breaks have height closer to 1000.

For DGP3 with no breaks, the false detection proportion among 1000 replication by using $\varkappa = 4$ is 0.021, 0.000 and 0.008, respectively, for the three cases: $T = 250$ with $J + 1 = 10$, $T = 500$ with $J + 1 = 10$, and $T = 500$ with $J + 1 = 15$. There is no break detected for $T = 500$ and $J + 1 = 10$, while the false detection proportion is close to zero for the other two cases. This result indicates that our method works well when no break exists in the model.

6. Application

In this section, we apply our proposed method to the U.S. Macroeconomic DataSet (Stock and Watson, 2009) to detect possible structural breaks in the underlying factor model. The dataset consists of $N = 108$ monthly macroeconomic time-series variables including real economic activity measures, prices, interest rates, money and credit aggregates, stock prices, exchange rates, etc. for the United States, spanning 1959m01–2006m12. Following the literature, we transform the data by taking the first order difference, so that we obtain a total of $T = 575$ monthly observations for each macroeconomic variable. The data have been centered and standardized for the analysis. We refer to Stock and Watson (2009) for the detailed data description.

We use $J + 1 = 10$ subintervals for the piecewise constant estimation, since as demonstrated in the simulation studies that the method works well for $T = 500$ by using $J + 1 = 10$ subintervals. We let $\varkappa = 4$ in the fused penalization procedure. We first determine the appropriate number of common factors. We select the number of factors by the information criteria $\mathrm{IC}_2(R)$ given in (4.2) of Section 4.1. As a result, the number of selected factors is 6. In Fig. 8, we plot the values of $\mathrm{IC}_2(R)$ against the number of factors. We observe that the IC value reaches its minimum at $R = 6$.

Next, we apply our proposed break detection procedure with the numbers of factors of $R = 6$. The tuning parameter in the fused penalization procedure is selected by the information criterion described in Section 4.2 with $c = 0.15$. Our method

Table 2
Percentage of correct detection of the number of breaks (C) and accuracy of break-point estimation ($100\times\mathrm{HD}/T$): DGP1-2 with $\varkappa = 2$.

$(T, J+1)$	(250, 10)		(500, 10)		(500, 15)	
	C	$100\times\mathrm{HD}/T$	C	$100\times\mathrm{HD}/T$	C	$100\times\mathrm{HD}/T$
$b = 1$						
DGP1-IID						
Case 1	66.9	0.071	81.4	0.029	68.2	0.288
Case 2	65.7	0.661	77.0	0.154	73.1	0.025
DGP1-CHeter						
Case 1	66.6	0.101	79.6	0.031	67.0	0.290
Case 2	64.1	0.672	75.8	0.165	71.9	0.034
DGP2-IID						
Case 1	84.3	0.118	89.4	0.042	76.2	0.180
Case 2	81.8	0.190	88.4	0.058	74.8	0.140
DGP2-AR						
Case 1	78.6	0.187	85.2	0.078	67.5	0.232
Case 2	72.7	0.217	82.0	0.086	66.5	0.238
$b = 2$						
DGP1-IID						
Case 1	93.6	0.019	98.2	0.007	95.0	0.135
Case 2	93.5	0.173	98.0	0.019	94.4	0.006
DGP1-CHeter						
Case 1	94.6	0.025	98.0	0.008	94.0	0.148
Case 2	93.4	0.212	98.1	0.018	94.2	0.008
DGP2-IID						
Case 1	96.2	0.065	97.9	0.012	95.1	0.130
Case 2	93.5	0.145	97.2	0.023	95.0	0.085
DGP2-AR						
Case 1	84.2	0.158	95.3	0.035	88.3	0.188
Case 2	82.4	0.211	92.2	0.044	87.4	0.134

Table 3
Percentage of correct detection of the number of breaks (C) and accuracy of break-point estimation ($100\times\mathrm{HD}/T$): DGP1-2 with $\varkappa = 4$.

$(T, J+1)$	(250, 10)		(500, 10)		(500, 15)	
	C	$100\times\mathrm{HD}/T$	C	$100\times\mathrm{HD}/T$	C	$100\times\mathrm{HD}/T$
$b = 1$						
DGP1-IID						
Case 1	86.4	0.033	91.5	0.018	76.4	0.769
Case 2	82.6	0.832	89.3	0.546	77.3	0.014
DGP1-CHeter						
Case 1	85.0	0.040	91.7	0.016	76.1	0.798
Case 2	81.1	0.887	88.2	0.570	77.4	0.013
DGP2-IID						
Case 1	92.7	0.082	96.4	0.018	87.5	0.643
Case 2	90.5	0.538	96.2	0.126	86.8	0.589
DGP2-AR						
Case 1	91.6	0.102	95.8	0.030	86.5	0.742
Case 2	88.9	0.702	95.3	0.182	83.4	0.618
$b = 2$						
DGP1-IID						
Case 1	98.1	0.015	99.4	0.005	94.0	0.502
Case 2	92.8	0.577	99.0	0.420	99.6	0.006
DGP1-CHeter						
Case 1	97.9	0.016	99.3	0.007	94.1	0.585
Case 2	92.1	0.624	98.8	0.462	99.5	0.008
DGP2-IID						
Case 1	99.7	0.034	100.0	0.008	93.2	0.356
Case 2	95.5	0.325	99.8	0.121	94.5	0.318
DGP2-AR						
Case 1	99.4	0.056	100.0	0.020	98.0	0.334
Case 2	95.3	0.452	99.2	0.138	97.2	0.292

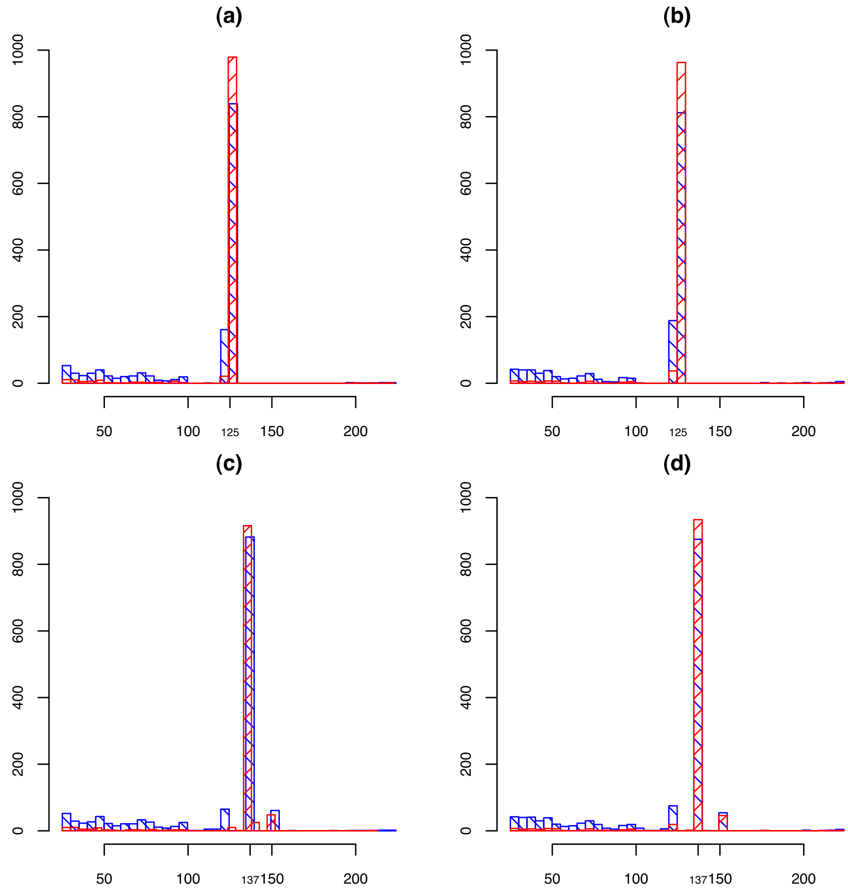

Fig. 2. Plots of the frequency of the estimated breaks among 1000 replications for DGP1 and $T = 250$ and for (a) Case 1 and IID errors, (b) Case 1 and CHeter errors, (c) Case 2 and IID errors, and (d) Case 2 and CHeter errors. The blue shaded line with angle = 135 is for $b = 1$ and the red shaded line with angle = 45 is for $b = 2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is able to identify five break dates in 1979m09, 1984m07, 1990m03, 1995m06, and 2002m01, respectively. The year of 1984 was considered as a potential break date by Stock and Watson (2009). As shown in a recent paper of Chen et al. (2014), their Sup-Wald test detected one break date around 1979–1980 (second oil price shock). This break date is also found by our proposed method. They mentioned that one possible explanation could be the impact on monetary policy in the US by the Iranian revolution in the beginning of 1979. Moreover, by using the U.S. labor productivity time-series data, Hansen (2001) plotted the sequence of Chow statistics for testing structural changes as a function of candidate break dates as shown in Figure 1 of page 120. It shows that the curve of the Chow test statistic has two peaks around the years of 1991 and 1995 which indicates that breaks may happen at these time points if any. By using our proposed method, we detected two break dates in 1990m03 and 1995m06, respectively. For the break date in the year of 2002, it may be attributed to the early 2000s recession (Kliesen, 2003).

7. Conclusion

In this paper, we propose a novel three-step procedure by utilizing nonparametric local estimation, shrinkage methods and grid search to determine the number of breaks and to estimate the break locations in large dimensional factor models. Based on the first-stage piecewise constant estimation of the factor loadings, we also propose a BIC-type information criterion to determine the number of factors. The proposed procedure is easy to implement, computationally efficient, and theoretically reliable. We show that the information criterion can consistently estimate the number of factors and our three-step procedure can consistently estimate the number of breaks and the break locations. Simulation studies demonstrate good performance of the proposed method. An application to U.S. macroeconomic dataset further illustrates the usefulness of our method.

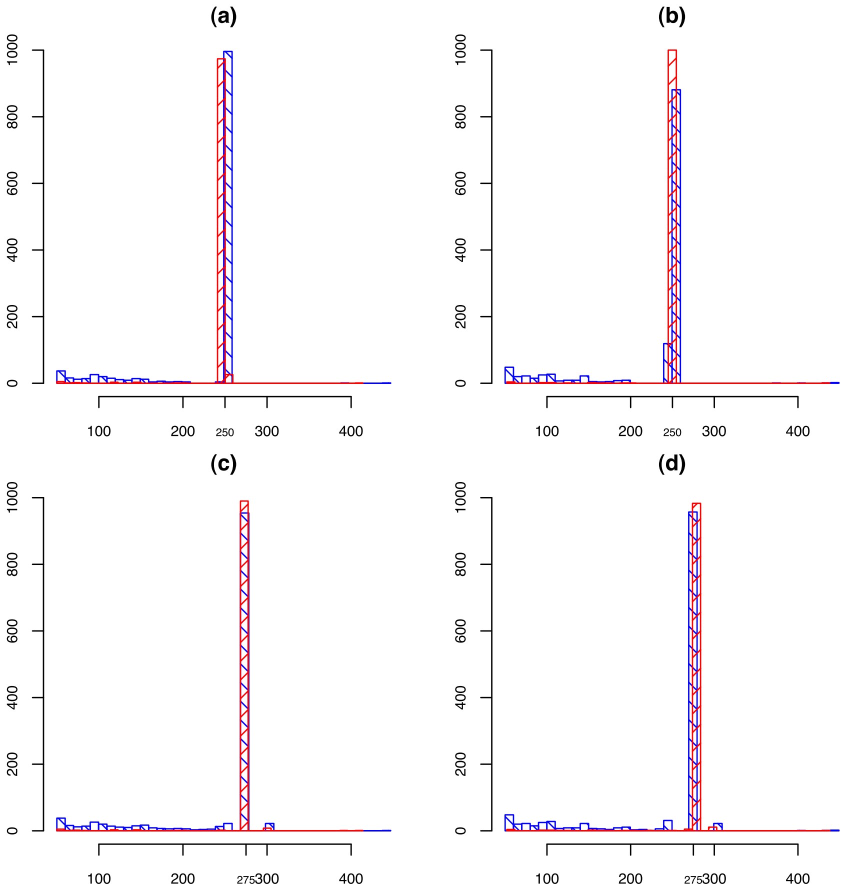

Fig. 3. Plots of the frequency of the estimated breaks among 1000 replications for DGP1 and $T = 500, J + 1 = 10$ and for (a) Case 1 and IID errors, (b) Case 1 and CHeter errors, (c) Case 2 and IID errors, and (d) Case 2 and CHeter error. The blue shaded line with angle = 135 is for $b = 1$ and the red shaded line with angle = 45 is for $b = 2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Mathematical appendix

This appendix provides the proofs of Propositions 3.2–3.3 in Section 3 and 4.1–4.2 in Section 4. The proof of Proposition 3.4 as well as that of some technical lemmas are available in the online supplementary material.

Appendix A. Proofs of the Propositions in Section 3

A.1. Proof of Proposition 3.2

Let V_κ, Υ_κ, and Q_κ be as defined in Proposition 3.1. We first state the following three lemmas that are used in proving Proposition 3.2. The proofs of these three lemmas are provided in the online Supplementary Material.

Lemma A.1. *Suppose that Assumptions A1–A4 hold. Suppose that S_j contains a break point t_κ^0 for some $\kappa = \kappa(j)$ and $j \in \mathbb{S}_{2a}$. Then*

(i) $N^{-1}\hat{\boldsymbol{\Delta}}_j^\top \left[(N\tau_j)^{-1}\mathbf{X}_{S_j}\mathbf{X}_{S_j}^\top\right]\hat{\boldsymbol{\Delta}}_j = V_{N,j} = V_{\kappa-1} + O_P(\eta_{N\tau_j}^{-1} + \tau_{j2}/\tau_j)$,

(ii) $N^{-1}\hat{\boldsymbol{\Delta}}_j^\top \boldsymbol{\alpha}_{\kappa-1}^0 = Q_{\kappa-1} + O_P(\eta_{N\tau_j}^{-1} + \tau_{j2}/\tau_j)$,

(iii) $H_{j,1} = \Sigma_F^{1/2}\Upsilon_{\kappa-1}V_{\kappa-1}^{-1/2} + O_P(\eta_{N\tau_j}^{-1} + \tau_{j2}/\tau_j)$,

(iv) $\frac{1}{N}\left\|\hat{\boldsymbol{\Delta}}_j - \boldsymbol{\alpha}_{\kappa-1}^0 H_{j,1}\right\|^2 = \frac{1}{N}\sum_{i=1}^N \left\|\hat{\Delta}_{ij} - H_{j,1}^\top \alpha_{i,\kappa-1}^0\right\|^2 = O_P(\eta_{N\tau_j}^{-2} + (\tau_{j2}/\tau_j)^2)$,

(v) $\hat{F}_t = \frac{1}{N}\hat{\boldsymbol{\Delta}}_j^\top \boldsymbol{\alpha}_\kappa^* F_t^* + O_P(N^{-1/2} + \eta_{N\tau_j}^{-1}(\eta_{N\tau_j}^{-1} + \tau_{j2}/\tau_j))$ *for each* $t \in S_j$,

(vi) $\frac{1}{\tau_j}\left\|\hat{\mathbf{F}}_{S_j} - \mathbf{F}_{S_j}^0 H_{j,1}^{\top -1}\right\|^2 = O_P(N^{-1} + \eta_{N\tau_j}^{-2}(\eta_{N\tau_j}^{-1} + \tau_{j2}/\tau_j)^2)$.

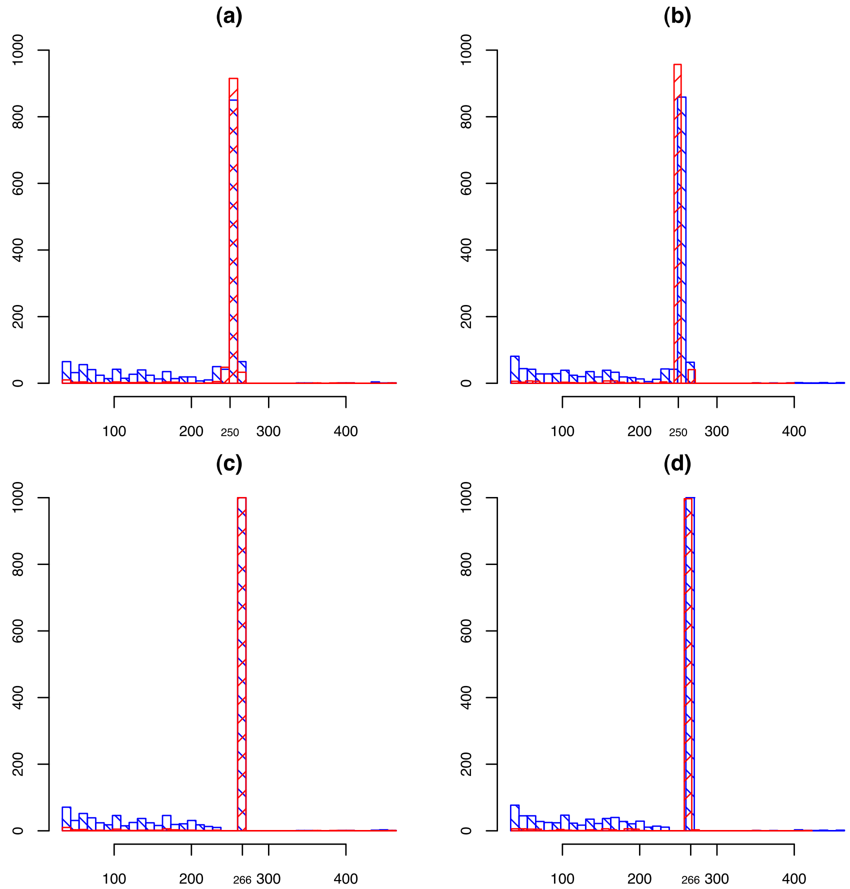

Fig. 4. Plots of the frequency of the estimated breaks among 1000 replications for DGP1 and $T = 500, J + 1 = 15$ and for (a) Case 1 and IID errors, (b) Case 1 and CHeter errors, (c) Case 2 and IID errors, and (d) Case 2 and CHeter errors. The blue shaded line with angle = 135 is for $b = 1$ and the red shaded line with angle = 45 is for $b = 2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Lemma A.2. *Suppose that Assumptions A1–A4 hold. Suppose that S_j contains a break point t_κ^0 for some $\kappa = \kappa(j)$ and $j \in \mathbb{S}_{2b}$. Then*

(i) $N^{-1}\hat{\mathbf{\Delta}}_j^\top\left[(N\tau_j)^{-1}\mathbf{X}_{S_j}\mathbf{X}_{S_j}^\top\right]\hat{\mathbf{\Delta}}_j = V_{N,j} = V_\kappa + O_P(\eta_{N\tau_j}^{-1} + \tau_{j1}/\tau_j)$,

(ii) $N^{-1}\hat{\mathbf{\Delta}}_j^\top\boldsymbol{\alpha}_\kappa^0 = Q_\kappa + O_P(\eta_{N\tau_j}^{-1} + \tau_{j1}/\tau_j)$,

(iii) $H_{j,2} = \Sigma_F^{1/2}\Upsilon_\kappa V_\kappa^{-1/2} + O_P(\eta_{N\tau_j}^{-1} + \tau_{j1}/\tau_j)$,

(iv) $\frac{1}{N}\left\|\hat{\mathbf{\Delta}}_j - \boldsymbol{\alpha}_\kappa^0 H_{j,2}\right\|^2 = \frac{1}{N}\sum_{i=1}^N\left\|\hat{\Delta}_{ij} - H_{j,2}^\top\alpha_{i,\kappa}^0\right\|^2 = O_P(\eta_{N\tau_j}^{-2} + (\tau_{j1}/\tau_j)^2)$,

(v) $\hat{F}_t = \frac{1}{N}\hat{\mathbf{\Delta}}_j^\top\boldsymbol{\alpha}_\kappa^* F_t^* + O_P(N^{-1/2} + \eta_{N\tau_j}^{-1}(\eta_{N\tau_j}^{-1} + \tau_{j1}/\tau_j))$ *for each* $t \in S_j$,

(vi) $\frac{1}{\tau_j}\left\|\hat{\mathbf{F}}_{S_j} - \mathbf{F}_{S_j}^0 H_{j,2}^{\top -1}\right\|^2 = O_P(N^{-1} + \eta_{N\tau_j}^{-2}(\eta_{N\tau_j}^{-1} + \tau_{j1}/\tau_j)^2)$.

Lemma A.3. *Suppose that Assumptions A1–A4 hold. Suppose that S_j contains a break point t_κ^0 for some $\kappa = \kappa(j)$ and $j \in \mathbb{S}_{2c}$. Then*

(i) $N^{-1}\hat{\mathbf{\Delta}}_j^\top\left[(N\tau_j)^{-1}\mathbf{X}_{S_j}\mathbf{X}_{S_j}^\top\right]\hat{\mathbf{\Delta}}_j = V_{N,j} = V_{\kappa*} + O_P(\eta_{N\tau_j}^{-1})$,

(ii) $N^{-1}\hat{\mathbf{\Delta}}_j^\top\boldsymbol{\alpha}_\kappa^* = Q_{\kappa*} + O_P(\eta_{N\tau_j}^{-1})$,

(iii) $H_{j*} = \Sigma_F^{1/2}\Upsilon_{\kappa*}V_{\kappa*}^{-1/2} + O_P(\eta_{N\tau_j}^{-1})$,

(iv) $\frac{1}{N}\left\|\hat{\mathbf{\Delta}}_j - \boldsymbol{\alpha}_\kappa^* H_{j*}\right\|^2 = \frac{1}{N}\sum_{i=1}^N\left\|\hat{\Delta}_{ij} - H_{j*}^\top\alpha_{i\kappa}^*\right\|^2 = O_P(\eta_{N\tau_j}^{-2})$,

(v) $\hat{F}_t = \frac{1}{N}\hat{\mathbf{\Delta}}_j^\top\boldsymbol{\alpha}_\kappa^* F_t^* + O_P(N^{-1/2} + \tau_j^{-1})$ *for each* $t \in S_j$ *and* $j \in \mathbb{S}_{2c}$,

(vi) $\frac{1}{\tau_j}\left\|\hat{\mathbf{F}}_{S_j} - \mathbf{F}_{S_j}^*\frac{1}{N}\boldsymbol{\alpha}_\kappa^{*\top}\hat{\mathbf{\Delta}}_j\right\|^2 = O_P(N^{-1} + \tau_j^{-2})$,

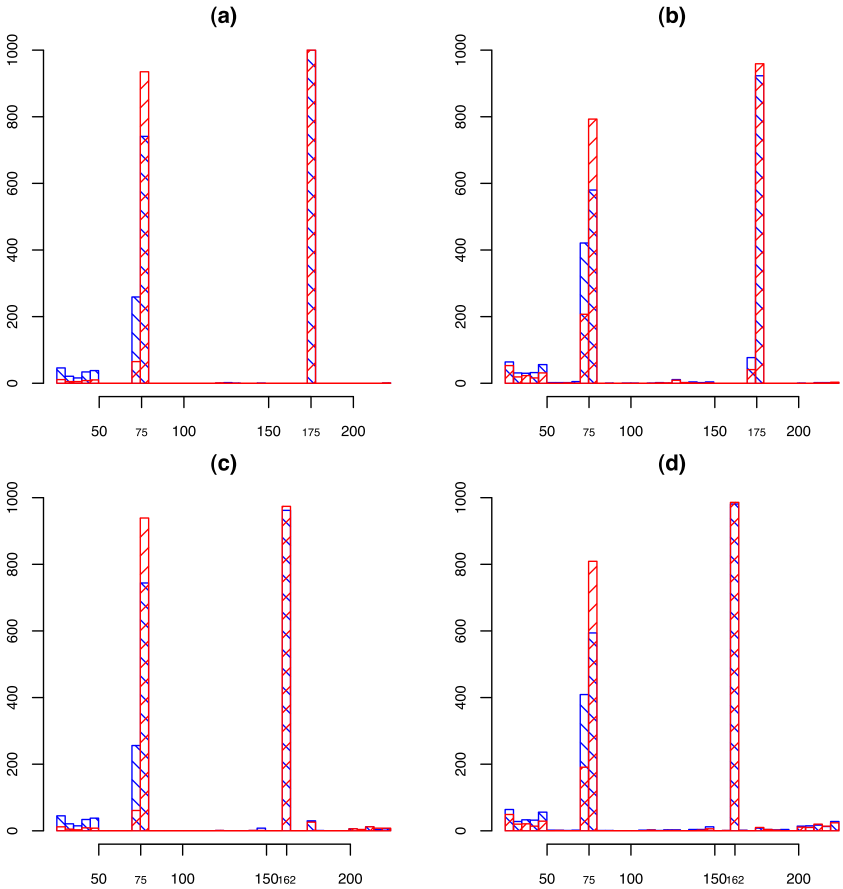

Fig. 5. Plots of the frequency of the estimated breaks among 1000 replications for DGP2 and $T = 250$ and for (a) Case 1 and IID errors, (b) Case 1 and AR errors, (c) Case 2 and IID errors, and (d) Case 2 and AR errors. The blue shaded line with angle = 135 is for $b = 1$ and the red shaded line with angle = 45 is for $b = 2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $V_{\kappa*}$ is the diagonal matrix consisting of the R largest eigenvalues of $\Sigma_{F_\kappa^*}^{1/2}\Sigma_{\Lambda_\kappa^*}\Sigma_{F_\kappa^*}^{1/2}$ in descending order with $\Upsilon_{\kappa*}$ being the corresponding (normalized) $2R \times R$ eigenvector matrix, $Q_{\kappa*} = V_{\kappa*}^{1/2}\Upsilon_{\kappa*}^{\top}\Sigma_{F_\kappa^*}^{-1/2}$, $\Sigma_{F_\kappa^*}$ =diag$\left(c_j\Sigma_F, (1-c_j)\Sigma_F\right)$, $c_j = \tau_{j1}/\tau_j$, and $\Sigma_{\Lambda_\kappa^*} = \lim_{N\to\infty} N^{-1}\boldsymbol{\alpha}_\kappa^{*\top}\boldsymbol{\alpha}_\kappa^*$.

The first part of Proposition 3.2(i) follows from Lemma A.1(iv). For the second part of Proposition 3.2(i), we have by the Cauchy–Schwarz inequality and the submultiplicative property of Frobenius norm,

$$
\begin{aligned}
&\frac{1}{N}\|\hat{\boldsymbol{\Delta}}_j V_{N,j} - \boldsymbol{\alpha}_{\kappa-1}^0 \Sigma_F Q_{\kappa-1}^{\top}\|^2 \\
&\leq \frac{2}{N}\|(\hat{\boldsymbol{\Delta}}_j - \boldsymbol{\alpha}_{\kappa-1}^0 H_{j,1})V_{N,j}\|^2 + \frac{2}{N}\|\boldsymbol{\alpha}_{\kappa-1}^0\left(H_{j,1}V_{N,j} - \Sigma_F Q_{\kappa-1}^{\top}\right)\|^2 \\
&\leq \frac{2}{N}\|\hat{\boldsymbol{\Delta}}_j - \boldsymbol{\alpha}_{\kappa-1}^0 H_{j,1}\|^2 \left\|V_{N,j}\right\|^2 + \frac{2}{N}\left\|\boldsymbol{\alpha}_{\kappa-1}^0\right\|^2 \|H_{j,1}V_{N,j} - \Sigma_F Q_{\kappa-1}^{\top}\|^2 \\
&= O_P(\eta_{N\tau_j}^{-2} + (\tau_{j2}/\tau_j)^2),
\end{aligned}
$$

where the last equality follows from Lemma A.1(i), (iii) and (iv). Analogously, we can apply Lemma A.2 to prove Proposition 3.2(ii).

The first part of Proposition 3.2(iii) follows from Lemma A.3(iv). For the second part of Proposition 3.2(iii), noting that $\boldsymbol{\alpha}_\kappa^* H_j^* = (N\tau_j)^{-1}(\boldsymbol{\alpha}_{\kappa-1}^0 \mathbf{F}_{S_j,1}^{\top}\mathbf{F}_{S_j,1}\boldsymbol{\alpha}_{\kappa-1}^{0\top} + \boldsymbol{\alpha}_\kappa^0 \mathbf{F}_{S_j,2}^{\top}\mathbf{F}_{S_j,2}\boldsymbol{\alpha}_\kappa^{0\top})\hat{\boldsymbol{\Delta}}_j V_{N\tau,j}^{-1}$ by the definitions of $\boldsymbol{\alpha}_\kappa^*$ and H_j^*, we have for any $t \in S_j$ and $j \in \mathbb{S}_{2c}$

$$
\begin{aligned}
&\frac{1}{N}\left\|\hat{\boldsymbol{\Delta}}_j V_{N,j} - (N\tau_j)^{-1}(\boldsymbol{\alpha}_{\kappa-1}^0 \mathbf{F}_{S_j,1}^{\top}\mathbf{F}_{S_j,1}\boldsymbol{\alpha}_{\kappa-1}^{0\top} + \boldsymbol{\alpha}_\kappa^0 \mathbf{F}_{S_j,2}^{\top}\mathbf{F}_{S_j,2}\boldsymbol{\alpha}_\kappa^{0\top})\hat{\boldsymbol{\Delta}}_j\right\|^2 \\
&= \frac{1}{N}\left\|(\hat{\boldsymbol{\Delta}}_j - \boldsymbol{\alpha}_\kappa^* H_j^*)V_{N,j}\right\|^2 \leq \frac{1}{N}\left\|\hat{\boldsymbol{\Delta}}_j - \boldsymbol{\alpha}_\kappa^* H_j^*\right\|^2 \left\|V_{N,j}\right\|^2 = O_P(\eta_{N\tau_j}^{-2})
\end{aligned}
$$

by Lemma A.3(i) and (iv).

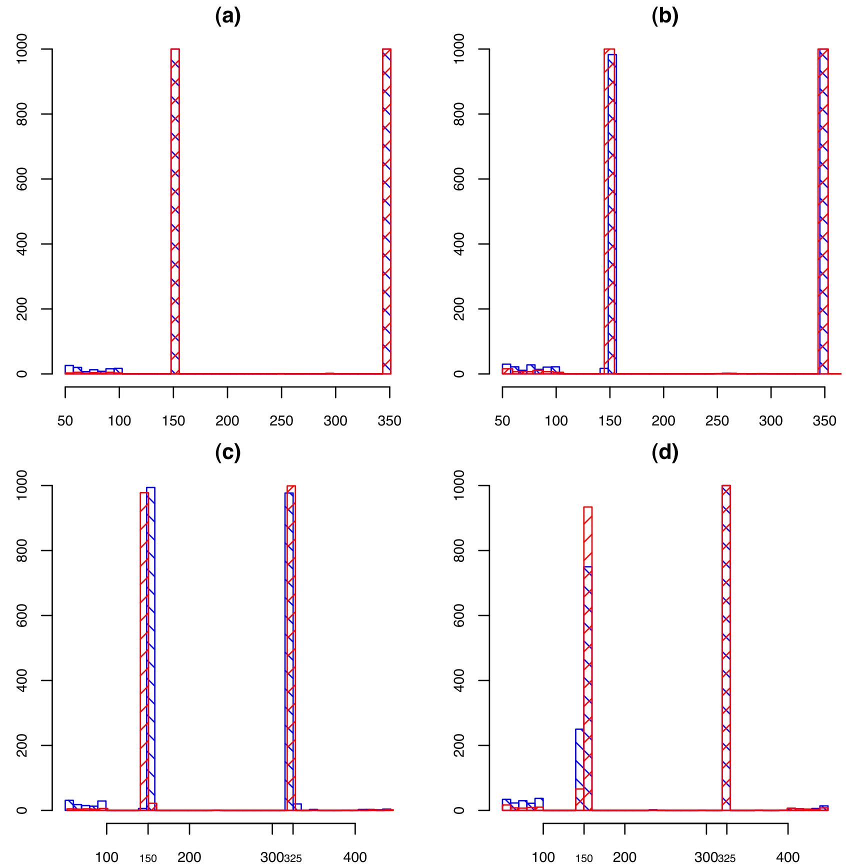

Fig. 6. Plots of the frequency of the estimated breaks among 1000 replications for DGP2 and $T = 500, J + 1 = 10$ and for (a) Case 1 and IID errors, (b) Case 1 and AR errors, (c) Case 2 and IID errors, and (d) Case 2 and AR errors. The blue shaded line with angle $= 135$ is for $b = 1$ and the red shaded line with angle $= 45$ is for $b = 2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A.2. Proof of Proposition 3.3

Recall $\hat{Z}_{jt} = N^{-1/2}\left\|\hat{\boldsymbol{\Delta}}_j V_{N,j}\right\| V_{N,j}^{-1}\hat{F}_t$ and $\hat{\mathbf{Z}}_{S_j} = (\hat{Z}_{jt}, t \in S_j) = N^{-1/2}\hat{\mathbf{F}}_{S_j} V_{N,j}^{-1}\left\|\hat{\boldsymbol{\Delta}}_j V_{N,j}\right\|$ (a $\tau_j \times R$ matrix), where $\hat{\mathbf{F}}_{S_j} = (\hat{F}_t, t \in S_j)$. Let a_j be defined as in Proposition 3.3 . Let

$$\bar{\boldsymbol{\Delta}}_j = \begin{cases} \boldsymbol{\alpha}_\kappa^0 \Sigma_F Q_\kappa^\top & \text{if } j \in \mathbb{S}_1 \\ \boldsymbol{\alpha}_{\kappa-1}^0 \Sigma_F Q_{\kappa-1}^\top & \text{if } j \in \mathbb{S}_{2a} \\ \boldsymbol{\alpha}_\kappa^0 \Sigma_F Q_\kappa^\top & \text{if } j \in \mathbb{S}_{2b} \\ \boldsymbol{\alpha}_\kappa^* H_{\kappa*} V_{\kappa*} & \text{if } j \in \mathbb{S}_{2c} \end{cases}, \quad \bar{H}_j = \begin{cases} \Sigma_F Q_\kappa^\top V_\kappa^{-1} & \text{if } j \in \mathbb{S}_1 \\ \Sigma_F Q_{\kappa-2}^\top V_{\kappa-1}^{-1} & \text{if } j \in \mathbb{S}_{2a} \\ \Sigma_F Q_\kappa^\top V_\kappa^{-1} & \text{if } j \in \mathbb{S}_{2b} \\ \bar{H}_{\kappa*} = \Sigma_{F*}^{1/2} \Upsilon_{\kappa*} V_{\kappa*}^{-1/2} = \Sigma_{F*} Q_{\kappa*}^\top V_{\kappa*}^{-1} & \text{if } j \in \mathbb{S}_{2c} \end{cases},$$

$$\bar{V}_j = \begin{cases} V_\kappa & \text{if } j \in \mathbb{S}_1 \\ V_{\kappa-1} & \text{if } j \in \mathbb{S}_{2a} \\ V_\kappa & \text{if } j \in \mathbb{S}_{2b} \\ V_{\kappa*} & \text{if } j \in \mathbb{S}_{2c} \end{cases} \text{ for some } \kappa = \kappa(j).$$

Note that $\bar{\boldsymbol{\Delta}}_j$ and $\bar{V}_j$ denote the probability limits of $\hat{\boldsymbol{\Delta}}_j V_{N,j}$ and $V_{N,j}$, respectively. Let

$$\mathbf{Z}_{S_j} = \left(Z_{jt}, t \in S_j\right) = \begin{cases} N^{-1/2}\mathbf{F}_{S_j}^0 \bar{H}_j^{\top+} \bar{V}_j^{-1}\left\|\bar{\boldsymbol{\Delta}}_j\right\| & \text{if } j \in \mathbb{S}_1 \cup \mathbb{S}_{2a} \cup \mathbb{S}_{2b} \\ N^{-1/2}\mathbf{F}_{S_j}^* Q_{\kappa*}^\top V_{\kappa*}^{-1}\left\|\bar{\boldsymbol{\Delta}}_j\right\| & \text{if } j \in \mathbb{S}_{2c} \end{cases},$$

where $Z_{jt} = N^{-1/2}\bar{V}_j^{-1}\bar{H}_j^{+}F_t\left\|\bar{\boldsymbol{\Delta}}_j\right\|$ if $j \in \mathbb{S}_1 \cup \mathbb{S}_{2a} \cup \mathbb{S}_{2b}$, and $= N^{-1/2}V_{\kappa*}^{-1}Q_{\kappa*}F_t^*\left\|\bar{\boldsymbol{\Delta}}_j\right\|$ if $j \in \mathbb{S}_{2c}$. Let $\Theta_j^* \equiv N^{1/2}\bar{\boldsymbol{\Delta}}_j / \left\|\bar{\boldsymbol{\Delta}}_j\right\|$.

To prove Proposition 3.3, we need a lemma.

Lemma A.4. *Let* $\mathbf{E}_{S_j}^* = \mathbf{X}_{S_j} - \Theta_j^* \mathbf{Z}_{S_j}^\top$. *Let* $\vartheta_j = \left(\vartheta_{j,1}, \ldots, \vartheta_{j,R}\right)$ *, an* $N \times R$ *matrix, for* $j = 0, 1, \ldots, J$. *Let* $\boldsymbol{\vartheta}_j = N^{-1/2}\mathrm{vec}\left(\vartheta_j\right)$ *and* $J_1 = J + 1$. *Suppose that the conditions in Proposition 3.3 hold. Then for each* $j \in \mathbb{S}$*, we have*

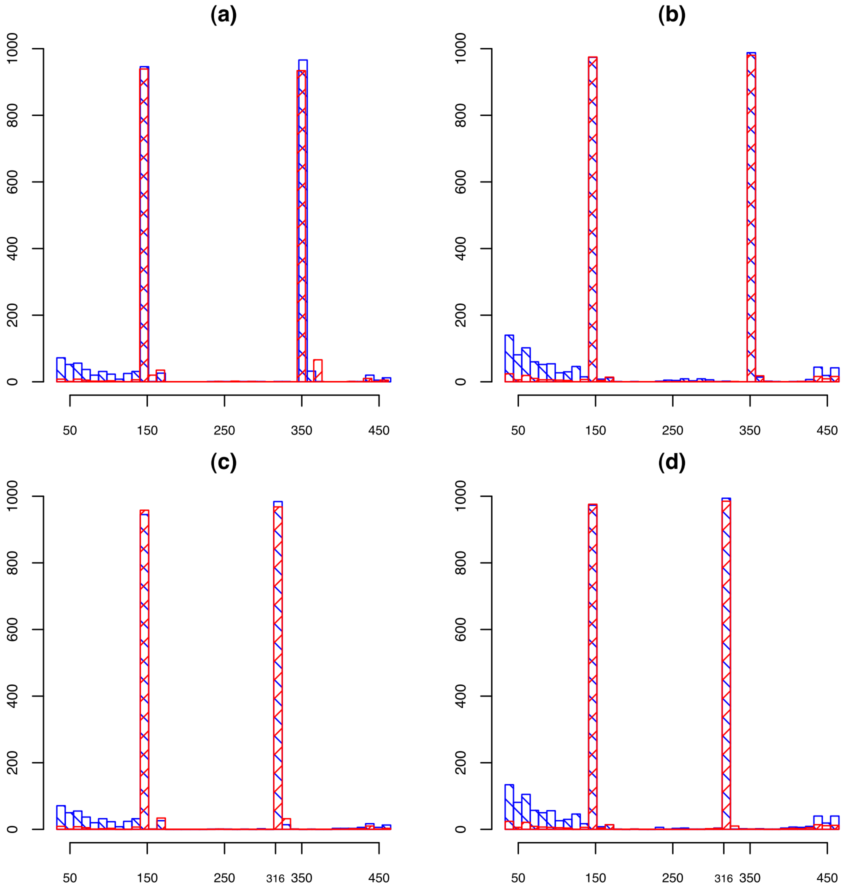

Fig. 7. Plots of the frequency of the estimated breaks among 1000 replications for DGP2 and $T = 500$, $J + 1 = 15$ and for (a) Case 1 and IID errors, (b) Case 1 and AR errors, (c) Case 2 and IID errors, and (d) Case 2 and AR errors. The blue shaded line with angle = 135 is for $b = 1$ and the red shaded line with angle = 45 is for $b = 2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

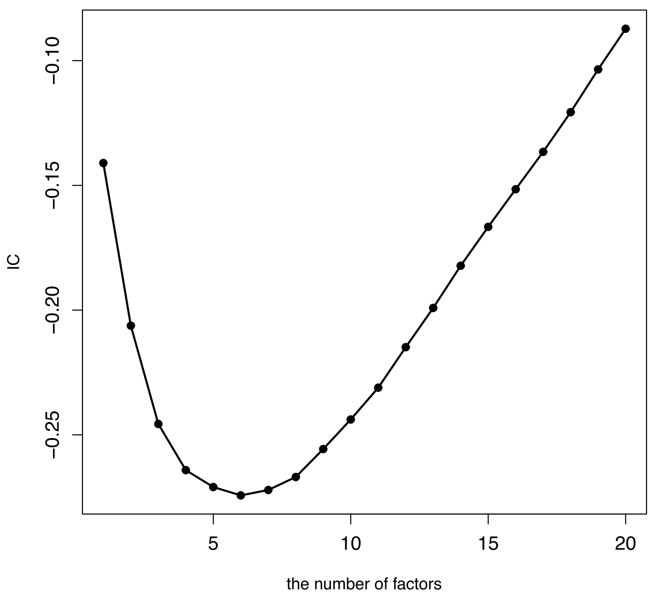

Fig. 8. Plots of the values of IC_{1b} against the number of factors for the real data application.

(i) $\frac{1}{\tau_j a_j^2}\left\|\hat{\mathbf{Z}}_{S_j}-\mathbf{Z}_{S_j}\right\|^2=O_P(1)$ *and* $\frac{1}{J_1}\sum_{j=0}^{J}\frac{1}{\tau_j a_j^2}\left\|\hat{\mathbf{Z}}_{S_j}-\mathbf{Z}_{S_j}\right\|^2=O_P(1)$,

(ii) $\frac{1}{N^{1/2}\tau_j a_j}\left\|\mathbf{E}_{S_j}^{*}(\hat{\mathbf{Z}}_{S_j}-\mathbf{Z}_{S_j})\right\|=O_P(1)$ *and* $\frac{1}{J_1}\sum_{j=0}^{J}\frac{1}{N(\tau_j a_j)^2}\left\|\mathbf{E}_{S_j}^{*}(\hat{\mathbf{Z}}_{S_j}-\mathbf{Z}_{S_j})\right\|^2=O_P(1)$,

(iii) $\frac{1}{N^{1/2}\tau_j a_j}\left\|\mathbf{E}_{S_j}^{*}\mathbf{Z}_{S_j}\right\|=O_P(1)$ *and* $\frac{1}{J_1}\sum_{j=0}^{J}\frac{1}{N(\tau_j a_j)^2}\left\|\mathbf{E}_{S_j}^{*}\mathbf{Z}_{S_j}\right\|^2=O_P(1)$,

(iv) $\frac{1}{N\tau_j}tr\left[\vartheta_j\left(\hat{\mathbf{Z}}_{S_j}\hat{\mathbf{Z}}_{S_j}^{\top}-\mathbf{Z}_{S_j}\mathbf{Z}_{S_j}^{\top}\right)\vartheta_j^{\top}\right]=o_P(1)\left\|\boldsymbol{\vartheta}_j\right\|^2$,

(v) $\frac{1}{N\tau_j a_j}tr\left[(\mathbf{X}_{S_j}-\Theta_j^{*}\hat{\mathbf{Z}}_{S_j}^{\top})^{\intercal}\vartheta_j\hat{\mathbf{Z}}_{S_j}^{\top}-(\mathbf{X}_{S_j}-\Theta_j^{*}\mathbf{Z}_{S_j}^{\top})^{\intercal}\vartheta_j\mathbf{Z}_{S_j}^{\top}\right]=O_P(1)\left\|\boldsymbol{\vartheta}_j\right\|$.

Proof of Proposition 3.3. (i) Let $\Theta_j=\Theta_j^{*}+a_j\vartheta_j$. Let

$$\begin{aligned}\Gamma_{NT,\gamma}\left(\{\Theta_j\}\right)&=\frac{1}{2N}\sum_{j=0}^{J}\frac{1}{\tau_j}\sum_{t\in S_j}(\mathbf{X}_t-\Theta_j\hat{Z}_{jt})^{\intercal}(\mathbf{X}_t-\Theta_j\hat{Z}_{jt})+\gamma\sum_{j=1}^{J}w_j\left\|\Theta_j-\Theta_{j-1}\right\|\\&=\frac{1}{2N}\sum_{j=0}^{J}\frac{1}{\tau_j}\mathrm{tr}\left[(\mathbf{X}_{S_j}-\Theta_j\hat{\mathbf{Z}}_{S_j}^{\top})^{\intercal}(\mathbf{X}_{S_j}-\Theta_j\hat{\mathbf{Z}}_{S_j}^{\top})\right]+\gamma\sum_{j=1}^{J}w_j\left\|\Theta_j-\Theta_{j-1}\right\|.\end{aligned}$$

Let c_j, $j=0,1,\ldots,J$ be arbitrary positive constants that do not depend on (N,T). Our aim is to show that for any given $\epsilon>0$, there exists a large constant L such that for sufficiently large (N,T) we have

$$P\left\{\inf_{N^{-1/2}\|\vartheta_j\|=c_jL,\ j=0,1,\ldots,J}\Gamma_{NT,\gamma}\left(\{\Theta_j^{*}+a_j\vartheta_j\}\right)>\Gamma_{NT,\gamma}\left(\{\Theta_j^{*}\}\right)\right\}\geq 1-\epsilon. \tag{A.1}$$

This implies that w.p.a.1 there is a local minimum $\{\tilde{\Theta}_j\}$ such that the estimator $\{\tilde{\Theta}_j\}$ lies inside the ball $\{\{\Theta_j^{*}+a_j\vartheta_j\}: N^{-1/2}\left\|\vartheta_j\right\|\leq c_jL\}$. Then we have $N^{-1/2}\left\|\tilde{\Theta}_j-\Theta_j^{*}\right\|=O_P(a_j)$ for $j=0,1,\ldots,J$.

Let $D\left(\{\vartheta_j\}\right)=\Gamma_{NT,\gamma}(\{\Theta_j^{*}+a_j\vartheta_j\})-\Gamma_{NT,\gamma}\left(\{\Theta_j^{*}\}\right)$. Noting that $\mathbf{X}_{S_j}-\Theta_j\hat{\mathbf{Z}}_{S_j}^{\top}=(\mathbf{X}_{S_j}-\Theta_j^{*}\hat{\mathbf{Z}}_{S_j}^{\top})-a_j\vartheta_j\hat{\mathbf{Z}}_{S_j}^{\top}$, we have

$$\begin{aligned}D\left(\{\vartheta_j\}\right)&=\frac{1}{2N}\sum_{j=0}^{J}\frac{1}{\tau_j}\mathrm{tr}\left[(\mathbf{X}_{S_j}-\Theta_j\hat{\mathbf{Z}}_{S_j}^{\intercal})^{\intercal}(\mathbf{X}_{S_j}-\Theta_j\hat{\mathbf{Z}}_{S_j}^{\intercal})-(\mathbf{X}_{S_j}-\Theta_j^{*}\hat{\mathbf{Z}}_{S_j}^{\intercal})^{\intercal}(\mathbf{X}_{S_j}-\Theta_j^{*}\hat{\mathbf{Z}}_{S_j}^{\intercal})\right]\\&\quad+\gamma\sum_{j=1}^{J}w_j\left\{\left\|\Theta_j-\Theta_{j-1}\right\|-\left\|\Theta_j^{*}-\Theta_{j-1}^{*}\right\|\right\}\\&=\frac{1}{2N}\sum_{j=0}^{J}\frac{a_j^2}{\tau_j}\mathrm{tr}\left[\hat{\mathbf{Z}}_{S_j}\vartheta_j^{\intercal}\vartheta_j\hat{\mathbf{Z}}_{S_j}^{\intercal}\right]-\frac{1}{N}\sum_{j=0}^{J}\frac{a_j}{\tau_j}\mathrm{tr}\left[(\mathbf{X}_{S_j}-\Theta_j^{*}\hat{\mathbf{Z}}_{S_j}^{\intercal})^{\intercal}\vartheta_j\hat{\mathbf{Z}}_{S_j}^{\intercal}\right]\\&\quad+\gamma\sum_{j=1}^{J}w_j\left\{\left\|\Theta_j-\Theta_{j-1}\right\|-\left\|\Theta_j^{*}-\Theta_{j-1}^{*}\right\|\right\}\\&=\frac{1}{2N}\sum_{j=0}^{J}\frac{a_j^2}{\tau_j}\mathrm{tr}\left[\vartheta_j\mathbf{Z}_{S_j}^{\intercal}\mathbf{Z}_{S_j}\vartheta_j^{\intercal}\right]-\frac{1}{N}\sum_{j=0}^{J}\frac{a_j}{\tau_j}\mathrm{tr}\left[(\mathbf{X}_{S_j}-\Theta_j^{*}\mathbf{Z}_{S_j}^{\top})^{\intercal}\vartheta_j\mathbf{Z}_{S_j}^{\intercal}\right]\\&\quad+\frac{1}{2N}\sum_{j=0}^{J}\frac{a_j^2}{\tau_j}\mathrm{tr}\left[\vartheta_j\left(\hat{\mathbf{Z}}_{S_j}^{\intercal}\hat{\mathbf{Z}}_{S_j}-\mathbf{Z}_{S_j}^{\intercal}\mathbf{Z}_{S_j}\right)\vartheta_j^{\intercal}\right]\\&\quad-\frac{1}{N}\sum_{j=0}^{J}\frac{a_j}{\tau_j}\mathrm{tr}\left[(\mathbf{X}_{S_j}-\Theta_j^{*}\hat{\mathbf{Z}}_{S_j}^{\intercal})^{\intercal}\vartheta_j\hat{\mathbf{Z}}_{S_j}^{\intercal}-(\mathbf{X}_{S_j}-\Theta_j^{*}\mathbf{Z}_{S_j}^{\top})^{\intercal}\vartheta_j\mathbf{Z}_{S_j}^{\intercal}\right]\\&\quad+\gamma\sum_{j=1}^{J}w_j\left\{\left\|\Theta_j-\Theta_{j-1}\right\|-\left\|\Theta_j^{*}-\Theta_{j-1}^{*}\right\|\right\}\\&\equiv D_1\left(\{\vartheta_j\}\right)-D_2\left(\{\vartheta_j\}\right)+D_3\left(\{\vartheta_j\}\right)-D_4\left(\{\vartheta_j\}\right)+D_5\left(\{\vartheta_j\}\right),\ \text{say}.\end{aligned}$$

Recall that $\boldsymbol{\vartheta}_j=N^{-1/2}\mathrm{vec}(\vartheta_j)$ and $\mathbf{E}_{S_j}^{*}=\mathbf{X}_{S_j}-\Theta_j^{*}\mathbf{Z}_{S_j}^{\intercal}$. Let $\mathbf{A}_j=\frac{1}{\tau_j}\mathbf{Z}_{S_j}^{\intercal}\mathbf{Z}_{S_j}$ and $\mathbf{B}_j=\frac{1}{N^{1/2}\tau_j a_j}\mathrm{vec}(\mathbf{E}_{S_j}^{*}\times\mathbf{Z}_{S_j})$. Apparently, $\left\|\mathbf{A}_j\right\|=O_P(1)$. By Lemma A.4(iii), $\left\|\mathbf{B}_j\right\|=O_P(1)$ for $j\in\mathbb{S}$. Noting that $\mathrm{tr}(B_1B_2)=\mathrm{vec}(B_2^{\intercal})^{\intercal}\,\mathrm{vec}(B_1)$ and $\mathrm{tr}(B_1B_2B_3)=\mathrm{vec}(B_1)^{\intercal}\,(B_2\otimes\mathbb{I})\mathrm{vec}(B_3^{\intercal})$ for any conformable matrices B_1,B_2,B_3 and an identity matrix $\mathbb{I}$ (see, e.g., Bernstein

(2005, p. 247 and p. 253)), we have

$$D_1\left(\{\vartheta_j\}\right) = \frac{1}{2N}\sum_{j=0}^{J}\frac{a_j^2}{\tau_j}\mathrm{tr}\left[\vartheta_j \mathbf{Z}_{S_j}^{\top}\mathbf{Z}_{S_j}\vartheta_j^{\top}\right] = \frac{1}{2}\sum_{j=0}^{J} a_j^2 \boldsymbol{\vartheta}_j^{\top}\left(\mathbf{A}_j\otimes\mathbb{I}_R\right)\boldsymbol{\vartheta}_j, \text{ and}$$

$$D_2\left(\{\vartheta_j\}\right) = \frac{1}{N}\sum_{j=0}^{J}\frac{a_j^2}{\tau_j a_j}\mathrm{tr}\left[\vartheta_j \mathbf{Z}_{S_j}^{\top}\mathbf{E}_{S_j}^{*\top}\right] = \sum_{j=0}^{J} a_j^2 \mathbf{B}_j^{\top}\boldsymbol{\vartheta}_j.$$

By Lemma A.4(iv)–(v),

$$D_3\left(\{\vartheta_j\}\right) = \frac{1}{2N}\sum_{j=0}^{J}\frac{a_j^2}{\tau_j}\mathrm{tr}\left[\vartheta_j\left(\hat{\mathbf{Z}}_{S_j}\hat{\mathbf{Z}}_{S_j}^{\top}-\mathbf{Z}_{S_j}\mathbf{Z}_{S_j}^{\top}\right)\vartheta_j^{\top}\right] = \sum_{j=0}^{J} o_P\left(a_j^2\right)\left\|\boldsymbol{\vartheta}_j\right\|^2, \text{ and}$$

$$\begin{aligned} D_4\left(\{\vartheta_j\}\right) &= \frac{1}{N}\sum_{j=0}^{J}\frac{a_j^2}{\tau_j a_j}\mathrm{tr}\left[(\mathbf{X}_{S_j}-\Theta_j^{*}\hat{\mathbf{Z}}_{S_j}^{\top})^{\top}\vartheta_j\hat{\mathbf{Z}}_{S_j}^{\top}-(\mathbf{X}_{S_j}-\Theta_j^{*}\mathbf{Z}_{S_j}^{\top})^{\top}\vartheta_j\mathbf{Z}_{S_j}^{\top}\right] \\ &= \sum_{j=0}^{J} O_P\left(a_j^2\right)\left\|\boldsymbol{\vartheta}_j\right\|. \end{aligned}$$

To study $D_5\left(\{\vartheta_j\}\right)$, we define the event $\mathcal{E}_{NT} = \{j-1, j\in\mathbb{S} : j-1\in\mathbb{S}_2 \text{ and } j\in\mathbb{S}_2\}$. Let $\mathcal{E}_{NT}^c$ denote the complement of $\mathcal{E}_{NT}$. Noting that $T/(J+1) \ll |I_{\min}|$, we have $P\left(\mathcal{E}_{NT}^c\right)\to 1$ as $(N,T)\to\infty$. Conditional on the event $\mathcal{E}_{NT}^c$,

$$\begin{aligned} D_5\left(\{\vartheta_j\}\right) &= \gamma\sum_{j\in\mathbb{S}_1, j-1\in\mathbb{S}_1} w_j\left\|\Theta_j-\Theta_{j-1}\right\| + \gamma\left\{\sum_{j\in\mathbb{S}_1, j-1\in\mathbb{S}_{2a}\cup\mathbb{S}_{2c}} + \sum_{j\in\mathbb{S}_1, j-1\in\mathbb{S}_{2b}}\right. \\ &\quad \left. + \sum_{j\in\mathbb{S}_{2a}, j-1\in\mathbb{S}_1} + \sum_{j\in\mathbb{S}_{2b}\cup\mathbb{S}_{2c}, j-1\in\mathbb{S}_1}\right\} w_j\left\{\left\|\Theta_j-\Theta_{j-1}\right\| - \left\|\Theta_j^{*}-\Theta_{j-1}^{*}\right\|\right\} \\ &\equiv D_{5,1}\left(\{\vartheta_j\}\right) + \sum_{l=2}^{5} D_{5,l}\left(\{\vartheta_j\}\right), \text{ say,} \end{aligned}$$

where, e.g., $\sum_{j\in\mathbb{S}_1, j-1\in\mathbb{S}_1} = \sum_{j=1,\, j\in\mathbb{S}_1, j-1\in\mathbb{S}_1}^{J}$. Apparently, $D_{5,1}\left(\{\vartheta_j\}\right)\geq 0$. Noting that when $j\in\mathbb{S}_1$ and $j-1\in\mathbb{S}_{2b}$, $\Theta_j^{*}=\Theta_{j-1}^{*}$ and $D_{5,3}\left(\{\vartheta_j\}\right) = \gamma\sum_{j\in\mathbb{S}_1, j-1\in\mathbb{S}_{2b}} w_j\left\|\Theta_j-\Theta_{j-1}\right\|\geq 0$. Similarly, when $j-1\in\mathbb{S}_1$ and $j\in\mathbb{S}_{2a}$, $\Theta_j^{*}=\Theta_{j-1}^{*}$ and

$$D_{5,4}\left(\{\vartheta_j\}\right) = \gamma\sum_{j\in\mathbb{S}_{2a}, j-1\in\mathbb{S}_1} w_j\left\|\Theta_j-\Theta_{j-1}\right\|\geq 0.$$

When $j\in\mathbb{S}_1$, $j-1\in\mathbb{S}_{2a}\cup\mathbb{S}_{2c}$, $\Theta_j^{*}-\Theta_{j-1}^{*}\neq 0$ and

$$\begin{aligned} \left|D_{5,2}\left(\{\vartheta_j\}\right)\right| &\leq \gamma\sum_{j\in\mathbb{S}_1, j-1\in\mathbb{S}_{2a}\cup\mathbb{S}_{2c}} a_j w_j\left\|\vartheta_j-\vartheta_{j-1}\right\| \\ &\leq 2\gamma\max_{j\in\mathbb{S}_1, j-1\in\mathbb{S}_{2a}\cup\mathbb{S}_{2c}} w_j\sum_{j=0}^{J} a_j\left\|\mathrm{vec}\left(\vartheta_j\right)\right\| = O_P((N\tau)^{1/2}\gamma)\sum_{j=0}^{J} a_j^2\left\|\boldsymbol{\vartheta}_j\right\|, \end{aligned}$$

where we use the fact that $\max_{j\in\mathbb{S}_1, j-1\in\mathbb{S}_{2a}\cup\mathbb{S}_{2c}} w_j = O_P(1)$ and $a_j^{-1} = O\left(\tau^{1/2}\right)$. By the same token, we can show that $\left|D_{5,5}\left(\{\vartheta_j\}\right)\right| = O_P((N\tau)^{1/2}\gamma)\sum_{j=0}^{J} a_j^2\left\|\boldsymbol{\vartheta}_j\right\|$.

Consequently, we have show that

$$\begin{aligned} D\left(\{\vartheta_j\}\right) &\geq \frac{1}{2}\sum_{j=0}^{J} a_j^2\boldsymbol{\vartheta}_j^{\top}\left(\mathbf{A}_j\otimes I_R\right)\boldsymbol{\vartheta}_j - \sum_{j=0}^{J} a_j^2\mathbf{B}_j^{\top}\boldsymbol{\vartheta}_j \\ &\quad - O_P((N\tau)^{1/2}\gamma+1)\sum_{j=0}^{J} a_j^2\left\|\boldsymbol{\vartheta}_j\right\| + \text{s.m.} \\ &\geq \sum_{j=0}^{J} a_j^2\left\{\frac{1}{2}\mu_{\min}\left(\mathbf{A}_j\right)\left\|\boldsymbol{\vartheta}_j\right\|^2 - [\left\|\mathbf{B}_j\right\| + O_P(1)]\left\|\boldsymbol{\vartheta}_j\right\|\right\} + \text{s.m.}, \end{aligned}$$

where s.m. denotes terms that are of smaller order than the preceding displayed terms. Noting that $\mu_{\min}\left(\mathbf{A}_j\right) \geq c > 0$ and $\left\|\mathbf{B}_j\right\| = O_P(1)$, by allowing $\left\|\boldsymbol{\vartheta}_j\right\| = N^{-1/2}\left\|\vartheta_j\right\|$ sufficiently large, the linear term $[\left\|\mathbf{B}_j\right\| + O_P(1)]\left\|\boldsymbol{\vartheta}_j\right\|$ will be dominated by the quadratic term $\frac{1}{2}\mu_{\min}\left(\mathbf{A}_j\right)\left\|\boldsymbol{\vartheta}_j\right\|^2$. This implies that $N^{-1/2}\left\|\vartheta_j\right\|$ has to be stochastically bounded for each j in order for $D\left(\left\{\vartheta_j\right\}\right)$ to be minimized. That is, (A.1) must hold for some large positive constant L and $N^{-1/2}\left\|\tilde{\Theta}_j - \Theta_j^*\right\| = O_P(a_j)$ for $j = 0, 1, \ldots, J$.

(ii) Let $\tilde{\vartheta}_j = a_j^{-1}(\tilde{\Theta}_j - \Theta_j^*)$. Then we can follow the analysis of (i) and show that

$$
\begin{aligned}
0 &\geq \frac{1}{J_1} D(\{\tilde{\vartheta}_j\}) \\
&\geq \frac{1}{J_1}\sum_{j=0}^{J} a_j^2 \left\{\frac{1}{2}\mu_{\min}\left(\mathbf{A}_j\right)\left\|\tilde{\boldsymbol{\vartheta}}_j\right\|^2 - [\left\|\mathbf{B}_j\right\| + O_P(1)]\left\|\tilde{\boldsymbol{\vartheta}}_j\right\|\right\} + \text{s.m.},
\end{aligned}
$$

where $O_P(1)$ holds uniformly in j. Then by the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
0 \geq\ & \frac{1}{2}\min_{0\leq j\leq J}\mu_{\min}\left(\mathbf{A}_j\right)\frac{1}{J_1}\sum_{j=0}^{J} a_j^2\left\|\tilde{\boldsymbol{\vartheta}}_j\right\|^2 \\
& - \left\{\frac{1}{J_1}\sum_{j=0}^{J} a_j^2[\left\|\mathbf{B}_j\right\| + O_P(1)]^2\right\}^{1/2}\left\{\frac{1}{J_1}\sum_{j=0}^{J} a_j^2\left\|\tilde{\boldsymbol{\vartheta}}_j\right\|^2\right\}^{1/2} + \text{s.m.}
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\frac{1}{NJ_1}\sum_{j=0}^{J}\left\|\tilde{\Theta}_j - \Theta_j^*\right\|^2 &= \frac{1}{J_1}\sum_{j=0}^{J} a_j^2\left\|\tilde{\boldsymbol{\vartheta}}_j\right\|^2 \\
&= O_P(\frac{1}{J_1}\sum_{j=0}^{J} a_j^2\left\|\mathbf{B}_j\right\|^2) + O_P(\frac{1}{J_1}\sum_{j=0}^{J} a_j^2)O_P(1).
\end{aligned}
$$

Note that

$$
\begin{aligned}
\frac{1}{J_1}\sum_{j=0}^{J} a_j^2\left\|\mathbf{B}_j\right\|^2 &= \frac{1}{J_1}\sum_{j\in\mathbb{S}_1} a_j^2\left\|\mathbf{B}_j\right\|^2 + \frac{1}{J_1}\sum_{j\in\mathbb{S}_2} a_j^2\left\|\mathbf{B}_j\right\|^2 \\
&= O_P\left(\eta_{N\tau}^{-2}\right)\frac{1}{J_1}\sum_{j\in\mathbb{S}_1}\left\|\mathbf{B}_j\right\|^2 + O_P\left(m/J\right) \\
&= O_P\left(\eta_{N\tau}^{-2} + m/J\right).
\end{aligned}
$$

Analogously, we can show that $\frac{1}{J_1}\sum_{j=0}^{J} a_j^2 = O_P\left(\eta_{N\tau}^{-2} + m/J\right)$. It follows that $\frac{1}{NJ_1}\sum_{j=0}^{J}\left\|\tilde{\Theta}_j - \Theta_j^*\right\|^2 = O_P\left(\eta_{N\tau}^{-2} + m/J\right)$.

(iii) Define $\mathcal{S} = \left\{j \in \mathbb{S} : \Theta_j^* - \Theta_{j-1}^* \neq 0\right\}$ and $\mathcal{S}^c = \left\{j \in \mathbb{S} : \Theta_j^* - \Theta_{j-1}^* = 0\right\}$. We focus on the case where $|\mathcal{S}| \geq 1$ which implies that $[1, T]$ contains at least one break. We will show that

$$
\Pr\left\{\left\|\tilde{\Theta}_j - \tilde{\Theta}_{j-1}\right\| = 0 \text{ for all } j,\ j-1 \in \mathbb{S}_1\right\} \to 1 \text{ as } (N, T) \to \infty. \tag{A.2}
$$

Suppose that to the contrary, $\tilde{\beta}_j = \tilde{\Theta}_j - \tilde{\Theta}_{j-1} \neq 0$ for some j such that $j, j-1 \in \mathbb{S}_1$ for sufficiently large (N, T). Then exists $r \in \{1, 2, \ldots, R\}$ such that $\|\tilde{\beta}_{j,r}\| = \max\{\|\tilde{\beta}_{j,l}\|, l = 1, \ldots, R\}$, where $\tilde{\beta}_{j,r}$ denotes the rth column of $\tilde{\beta}_j$. Without loss of generality (Wlog), we assume that $r = R$. Then $\|\tilde{\beta}_{j,R}\|/\|\tilde{\beta}_j\| \geq 1/\sqrt{R}$. To consider the first order condition (FOC) with respect to (wrt) Θ_j, $j = 1, \ldots, J$, we distinguish three cases: (a) $2 \leq j \leq J-1$, (b) $j = J$, and (c) $j = 1$. For $j = 1, \ldots, J$, let $\varrho_j = \left(\varrho_{j,1}, \ldots, \varrho_{j,R}\right)$ where

$$
\varrho_{j,r} = \tilde{\beta}_{j,r}/\left\|\tilde{\beta}_{j+1}\right\| \text{ if } \left\|\tilde{\beta}_{j+1}\right\| \neq 0 \text{ for } r = 1, \ldots, R, \text{ and } \left\|\varrho_j\right\| \leq 1 \text{ otherwise.} \tag{A.3}
$$

In case (a), we consider two subcases: (a1) $j+1 \in \mathbb{S}_{2b} \cup \mathbb{S}_{2c}$ and (a2) $j+1 \in \mathbb{S}_1 \cup \mathbb{S}_{2a}$. In either subcase, the FOC wrt $\Theta_{j,R}$ for the minimization problem in (2.2) is given by

$$
\begin{aligned}
\mathbf{0}_{N\times 1} =\ & \frac{a_j}{\sqrt{N}}(\mathbf{X}_{S_j} - \tilde{\Theta}_j\hat{\mathbf{Z}}_{S_j}^{\intercal})\hat{\mathbf{Z}}_{S_j,R} - a_j\tau_j N^{1/2}\gamma w_j\tilde{\beta}_{j,R}/\left\|\tilde{\beta}_j\right\| + a_j\tau_j N^{1/2}\gamma w_{j+1}\varrho_{j+1,R} \\
=\ & \frac{a_j}{\sqrt{N}}\left[(\Theta_j^* - \tilde{\Theta}_j)\mathbf{Z}_{S_j}^{\intercal} + \tilde{\Theta}_j(\mathbf{Z}_{S_j}^{\intercal} - \hat{\mathbf{Z}}_{S_j}) + \mathbf{E}_{S_j}^*\right]\hat{\mathbf{Z}}_{S_j,R} - a_j\tau_j N^{1/2}\gamma w_j\tilde{\beta}_{j,R}/\left\|\tilde{\beta}_j\right\| \\
& + a_j\tau_j N^{1/2}\gamma w_{j+1}\varrho_{j+1,R}
\end{aligned}
$$

$$\begin{aligned}
&= \frac{a_j}{\sqrt{N}}(\Theta_j^* - \tilde{\Theta}_j)\mathbf{Z}_{S_j}^{\intercal}\mathbf{Z}_{S_j,R} + \frac{a_j}{\sqrt{N}}\mathbf{E}_{S_j}^{*}\mathbf{Z}_{S_j,R} + \frac{a_j}{\sqrt{N}}(\Theta_j^* - \tilde{\Theta}_j)\mathbf{Z}_{S_j}^{\intercal}(\hat{\mathbf{Z}}_{S_j,R} - \mathbf{Z}_{S_j,R}) \\
&\quad + \frac{a_j}{\sqrt{N}}\tilde{\Theta}_j(\mathbf{Z}_{S_j}^{\intercal} - \hat{\mathbf{Z}}_{S_j})\hat{\mathbf{Z}}_{S_j,R} + \frac{a_j}{\sqrt{N}}\mathbf{E}_{S_j}^{*}(\hat{\mathbf{Z}}_{S_j,R} - \mathbf{Z}_{S_j,R}) + a_j\tau_j N^{1/2}\gamma w_j \tilde{\beta}_{j,R}/\left\|\tilde{\beta}_j\right\| \\
&\quad + a_j\tau_j N^{1/2}\gamma w_{j+1}\varrho_{j+1,R} \\
&\equiv B_{1,j}(R) + B_{2,j}(R) + B_{3,j}(R) + B_{4,j}(R) + B_{5,j}(R) - B_{6,j}(R) + B_{7,j}(R), \text{ say,} \qquad \text{(A.4)}
\end{aligned}$$

where $\hat{\mathbf{Z}}_{S_j,R}$ and $\mathbf{Z}_{S_j,R}$ denote the R -th columns of $\hat{\mathbf{Z}}_{S_j}$ and $\mathbf{Z}_{S_j}$, respectively. By part (i), $\left\|B_{1,j}(R)\right\| \le \tau_j a_j^2(N^{-1/2}a_j^{-1}\|\Theta_j^* - \tilde{\Theta}_j\|)\|\frac{1}{\tau_j}\mathbf{Z}_{S_j}^{\intercal}\mathbf{Z}_{S_j,R}\| = O_P(1)$ where we use the fact $a_j = \eta_{N\tau_j}^{-1} = O(\tau_j^{-1/2})$ for $j \in \mathbb{S}_1$ under Assumption A6(i). Similarly, by Lemma A.4(iii), $\left\|B_{2,j}(R)\right\| = O_P(1)$. By the submultiplicative property of the Frobenius norm, part (i) and Lemma A.4(i)–(ii), we have that for $j \in \mathbb{S}_1$,

$$\begin{aligned}
\left\|B_{3,j}(R)\right\| &\le (\tau_j a_j^3)(N^{-1/2}a_j^{-1})\left\|\Theta_j^* - \tilde{\Theta}_j\right\|\tau_j^{-1/2}\left\|\mathbf{Z}_{S_j}\right\|\tau_j^{-1/2}a_j^{-1}\left\|\hat{\mathbf{Z}}_{S_j,R} - \mathbf{Z}_{S_j,R}\right\| \\
&= O_P(\tau_j a_j^3) = o_P(1), \\
\left\|B_{4,j}(R)\right\| &\le (\tau_j a_j^2)N^{-1/2}\left\|\tilde{\Theta}_j\right\|\tau_j^{-1/2}\left\|\hat{\mathbf{Z}}_{S_j,R}\right\|\tau_j^{-1/2}a_j^{-1}\left\|\mathbf{Z}_{S_j} - \hat{\mathbf{Z}}_{S_j}\right\| = O_P(\tau_j a_j^2) = O_P(1), \\
\left\|B_{5,j}(R)\right\| &\le (\tau_j a_j^2)N^{-1/2}\tau_j^{-1}a_j^{-1}\left\|\mathbf{E}_{S_j}^{*}(\hat{\mathbf{Z}}_{S_j,R} - \mathbf{Z}_{S_j,R})\right\| = O_P(\tau_j a_j^2) = O_P(1).
\end{aligned}$$

In addition,

$$\left\|B_{6,j}(R)\right\| \ge a_j\tau_j N^{1/2}\gamma w_j\left\|\tilde{\beta}_{j,R}\right\|/\left\|\tilde{\beta}_j\right\| \ge a_j\tau_j N^{1/2}\gamma w_j/\sqrt{R}, \qquad \text{(A.5)}$$

which is explosive in probability at rate $(N\tau)^{1/2}\gamma\eta_{N\tau}^{\varkappa}$ under Assumption A6(ii).

To determine the probability order of $B_{7,j}(R)$, we consider two subcases. In subcase (a1) $j+1 \in \mathbb{S}_{2b} \cup \mathbb{S}_{2c}$, $N^{-1/2}\|\tilde{\beta}_{j+1}\| \xrightarrow{P} \lim_{N\to\infty}N^{-1/2}\|\Theta_{j+1}^* - \Theta_j^*\| \neq 0$, implying that $w_{j+1} = O_P(1)$ and $\left\|B_{7,j}(R)\right\| \le a_j\tau_j N^{1/2}\gamma w_{j+1} = O_P((N\tau)^{1/2}\gamma) = O_P(1)$. Consequently, $\left\|B_{6,j}(R)\right\| \gg \sum_{l=1}^{5}\left\|B_{l,j}(R)\right\| + \left\|B_{7,j}(R)\right\|$ and (A.4) cannot hold for sufficiently large (N, T). Then we conclude that w.p.a.1 $\tilde{\beta}_j = \tilde{\Theta}_j - \tilde{\Theta}_{j-1}$ must lie in a position where $\left\|\Theta_j - \Theta_{j-1}\right\|$ is not differentiable with respect to Θ_j in subcase (a1). In this case we can apply the subdifferential calculus and the fact that 0 belongs to the subdifferential of the objective function wrt Θ_j to obtain

$$\begin{aligned}
\mathbf{0}_{N\times R} &= \frac{a_j}{\sqrt{N}}(\mathbf{X}_{S_j} - \tilde{\Theta}_j\hat{\mathbf{Z}}_{S_j}^{\intercal})\hat{\mathbf{Z}}_{S_j} - a_j\tau_j N^{1/2}\gamma w_j\varrho_j + a_j\tau_j N^{1/2}\gamma w_{j+1}\varrho_{j+1} \\
&= \frac{a_j}{\sqrt{N}}(\Theta_j^* - \tilde{\Theta}_j)\mathbf{Z}_{S_j}^{\intercal}\mathbf{Z}_{S_j} + \frac{a_j}{\sqrt{N}}\mathbf{E}_{S_j}^{*}\mathbf{Z}_{S_j} + \frac{a_j}{\sqrt{N}}(\Theta_j^* - \tilde{\Theta}_j)\mathbf{Z}_{S_j}^{\intercal}(\hat{\mathbf{Z}}_{S_j} - \mathbf{Z}_{S_j}) \\
&\quad + \frac{a_j}{\sqrt{N}}\tilde{\Theta}_j(\mathbf{Z}_{S_j}^{\intercal} - \hat{\mathbf{Z}}_{S_j})\hat{\mathbf{Z}}_{S_j} + \frac{a_j}{\sqrt{N}}\mathbf{E}_{S_j}^{*}(\hat{\mathbf{Z}}_{S_j} - \mathbf{Z}_{S_j}) - a_j\tau_j N^{1/2}\gamma w_j\varrho_j + a_j\tau_j N^{1/2}\gamma w_{j+1}\varrho_{j+1} \\
&\equiv B_{1,j} + B_{2,j} + B_{3,j} + B_{4,j} + B_{5,j} - B_{6,j} + B_{7,j}, \text{ say,} \qquad \text{(A.6)}
\end{aligned}$$

for some ϱ_j and ϱ_{j+1} that are defined as in (A.3).[2] Following the above analyses of $B_{l,j}(R)$ for $l = 1, 2, 3, 4, 5$ and 7, we have $\sum_{l=1}^{5}\left\|B_{l,j}\right\| + \left\|B_{7,j}\right\| = O_P(1)$. Then (A.6) implies that we must have $\left\|B_{6,j}\right\| = a_j\tau_j N^{1/2}\gamma w_j\left\|\varrho_j\right\| = O_P(1)$.

In subcase (a2) $j+1 \in \mathbb{S}_1 \cup \mathbb{S}_{2a}$. First, we observe that in order for the FOC wrt $\Theta_{j,R}$ in (A.4) to hold, $\|B_{7,j}(R)\| = a_j\tau_j N^{1/2}\gamma w_{j+1}\|\varrho_{j+1,R}\|$ must be explosive at the same rate as $\left\|B_{6,j}(R)\right\|$. In addition, we must have $\left\|B_{7,j}(R)\right\| = \left\|B_{6,j}(R)\right\| + O_P(1)$ and hence $\left\|B_{7,j}(R)\right\| / \left\|B_{6,j}(R)\right\| \xrightarrow{P} 1$ as $(N, T) \to 1$. Next, we consider the FOC wrt $\Theta_{j+1,R}$:

$$\begin{aligned}
\mathbf{0}_{N\times 1} &= \frac{a_{j+1}}{\sqrt{N}}(\mathbf{X}_{S_{j+1}} - \tilde{\Theta}_{j+1}\hat{\mathbf{Z}}_{S_{j+1}}^{\intercal})\hat{\mathbf{Z}}_{S_{j+1},R} - a_{j+1}\tau_{j+1}N^{1/2}\gamma w_{j+1}\varrho_{j+1,R} + a_{j+1}\tau_{j+1}N^{1/2}\gamma w_{j+2}\varrho_{j+2,R} \\
&= B_{1,j+1}(R) + B_{2,j+1}(R) + B_{3,j+1}(R) + B_{4,j+1}(R) + B_{5,j+1}(R) - B_{6,j+1}(R) + B_{7,j+1}(R). \qquad \text{(A.7)}
\end{aligned}$$

Noting that $B_{6,j+1}(R) = \frac{a_{j+1}\tau_{j+1}}{a_j\tau_j}B_{7,j}(R) \asymp \frac{a_{j+1}}{a_j}B_{7,j}(R)$ and $\left\|B_{7,j}(R)\right\| / \left\|B_{6,j}(R)\right\| \xrightarrow{P} 1$, this implies that both $\left\|B_{6,j+1}(R)\right\|$ and $\left\|B_{7,j+1}(R)\right\|$ must explode at the same rate, $\left\|B_{7,j+1}(R)\right\| = \left\|B_{6,j+1}(R)\right\| + O_P(1)$, and $\left\|B_{7,j+1}(R)\right\| / \left\|B_{6,j+1}(R)\right\| \xrightarrow{P} 1$ if $j+2 \in \mathbb{S}_1 \cup \mathbb{S}_{2a}$. Deducting this way until $j+\imath \in \mathbb{S}_1 \cup \mathbb{S}_{2a}$ but $j+\imath+1 \in \mathbb{S}_{2b} \cup \mathbb{S}_{2c}$ for some $\imath \ge 1$. By assumption, if the interval $S_{j+\imath+1}$ contains a break so that $j+\imath+1 \in \mathbb{S}_{2b} \cup \mathbb{S}_{2c}$, then the intervals $S_{j+\imath-1}$ and $S_{j+\imath}$ cannot contain a break (so

[2] Here we abuse the notation slightly for $B_{6,j}$: its rth column is given by $a_j\tau_j N^{1/2}\gamma w_j\tilde{\beta}_{j,R}/\|\tilde{\beta}_j\|$ as long as $\|\tilde{\beta}_j\| \neq 0$ and $a_j\tau_j N^{1/2}\gamma w_j\varrho_{j,r}$ for some $\varrho_{j,r}$ with $\|\varrho_j\| \le 1$ otherwise. Similarly for $B_{7,j}$. We can regard $B_{\ell,J}$ as the matrix version of $B_{\ell,J}(R)$: $B_{\ell,J} = \left(B_{\ell,J}(1), \ldots, B_{\ell,J}(R)\right)$ where $\ell = 1, 2, \ldots, 7$, and $B_{\ell,J}(r)$ are defined analogously to $B_{\ell,J}(R)$ in (A.4) by considering the FOC wrt to $\Theta_{j,r}$ instead.

that we must have $j+\iota-1, j+\iota \in \mathbb{S}_1$). In addition, for $j'=1,\ldots,\iota$ we have

$$\begin{aligned}
\left\|B_{7,j+j'}(R)\right\| &\geq \left\|B_{6,j+j'}(R)\right\| - \sum_{l=1}^{5}\left\|B_{l,j+j'}(R)\right\| = \frac{a_{j+j'}\tau_{j+j'}}{a_{j+j'-1}\tau_{j+j'-1}}\left\|B_{7,j+j'-1}(R)\right\| - \sum_{l=1}^{5}\left\|B_{l,j+j'}(R)\right\| \\
&\geq \frac{a_{j+j'}\tau_{j+j'}}{a_{j+j'-1}\tau_{j+j'-1}}\left(\frac{a_{j+j'-1}\tau_{j+j'-1}}{a_{j+j'-2}\tau_{j+j'-2}}\left\|B_{7,j+j'-2}(R)\right\| - \sum_{l=1}^{5}\left\|B_{l,j+j'-1}(R)\right\|\right) \\
&\quad - \sum_{l=1}^{5}\left\|B_{l,j+j'}(R)\right\| \\
&\geq \cdots \\
&\geq \left(\Pi_{i=j+1}^{j+j'}\frac{a_i\tau_i}{a_{i-1}\tau_{i-1}}\right)\left\|B_{7,j}(R)\right\| - \sum_{s=1}^{j'-1}\left(\Pi_{i=j+j'+1-s}^{j+j'}\frac{a_i\tau_i}{a_{i-1}\tau_{i-1}}\right)\sum_{l=1}^{5}\left\|B_{l,j+j'-s}(R)\right\| \\
&\quad - \sum_{l=1}^{5}\left\|B_{l,j+j'}(R)\right\|.
\end{aligned}$$

Noting that $\Pi_{i=j+1}^{j+\iota}\frac{a_i\tau_i}{a_{i-1}\tau_{i-1}} = \frac{a_{j+\iota}\tau_{j+\iota}}{a_j\tau_j} \geq \underline{C}$ and $\max_{1\leq s\leq \iota-1}\Pi_{i=j+\iota+1-s}^{j+\iota}\frac{a_i\tau_i}{a_{i-1}\tau_{i-1}} = \max_{1\leq s\leq \iota-1}\frac{a_{j+\iota}\tau_{j+\iota}}{a_{j+\iota-s}\tau_{j+\iota-s}} \leq \bar{C}$ for some constants $\underline{C}, \bar{C} > 0$, we have

$$\frac{1}{J_1}\left\|B_{7,j+\iota}(R)\right\| \geq \frac{1}{J_1}\left(\underline{C}\left\|B_{7,j}(R)\right\| - \left(\bar{C}+1\right)\sum_{s=1}^{\iota}\sum_{l=1}^{5}\left\|B_{l,j+s}(R)\right\|\right) \to \infty \text{ in probability}$$

because $\left\|B_{7,j}(R)\right\|/J_1$ is divergent in probability at the same speed as $\left\|B_{6,j}(R)\right\|/J_1$ under Assumption A8(ii) and we can readily show that $\frac{1}{J_1}\sum_{s=1}^{\iota}\sum_{l=1}^{5}\left\|B_{l,j+s}(R)\right\| \leq \frac{1}{J_1}\sum_{s=1}^{J}\sum_{l=1}^{5}\left\|B_{l,j+s}\right\| = O_p(1)$. But when $j+\iota-1, j+\iota \in \mathbb{S}_1$, and $j+\iota+1 \in \mathbb{S}_{2b}\cup\mathbb{S}_{2c}$, the analysis in subcase (a1) applies to the FOC wrt $\Theta_{j+1,R}$ –(A.6) holds with j replaced by $j+\iota$ which forces $\left\|B_{6,j+\iota}(R)\right\| = a_{j+\iota}\tau_{j+\iota} N^{1/2}\gamma w_{j+\iota}\|\varrho_{j+1,R}\| = O_P(1)$. In short, a contradiction would arise unless there is no point after $j+1$ that belongs to $\mathbb{S}_{2b}\cup\mathbb{S}_{2c}$. Similarly, if there is a point in $\{j+1,\ldots,J\}$ that belongs to $\mathbb{S}_{2a}$, we denote it as $j+\iota$ for some $\iota \geq 1$. Then by assumption, $j+\iota-2, j+\iota-1, j+\iota+1, j+\iota+2 \in \mathbb{S}_1$, and we can apply arguments as used in subcase (a1) to derive a contradiction based on the FOC wrt $\Theta_{j+\iota}$. Hence $S_{j+1},\ldots,S_J$ cannot contain any break as long as (A.5) holds. Third, we consider the FOC wrt $\Theta_{j-1,R}$, i.e., (A.7) holds with $j+1$ replaced by $j-1$. Noting that $B_{6,j}(R) = \frac{a_j\tau_j}{a_{j-1}\tau_{j-1}}B_{7,j-1}(R) \asymp \frac{a_j}{a_{j-1}}B_{7,j-1}(R)$ and $B_{6,j}(R)$ is explosive by (A.5), we must have: $\left\|B_{6,j-1}(R)\right\|$ and $\left\|B_{7,j-1}(R)\right\|$ explode the same rate, $\left\|B_{7,j-1}(R)\right\| = \left\|B_{6,j-1}(R)\right\| + O_P(1)$, and $\left\|B_{7,j-1}(R)\right\|/\left\|B_{6,j-1}(R)\right\| \xrightarrow{P} 1$ if $j-2 \in \mathbb{S}_1\cup\mathbb{S}_{2b}$. Deducting this way until $j-\iota \in \mathbb{S}_1\cup\mathbb{S}_{2b}$ but $j-\iota-1 \in \mathbb{S}_{2a}\cup\mathbb{S}_{2c}$ for some $\iota \geq 2$. Again, when $j-\iota-1 \in \mathbb{S}_{2a}\cup\mathbb{S}_{2c}$, the interval $S_{j-\iota-1}$ contains a break so that the neighboring intervals $S_{j-\iota}$ and $S_{j-\iota+1}$ cannot contain a break. So the FOC wrt $\Theta_{j-\iota,R}$ suggests that $\left\|B_{6,j-\iota}(R)\right\|$ and $\left\|B_{7,j-\iota}(R)\right\|$ are explosive at the same rate such that $\left\|B_{7,j-\iota}(R)\right\| = \left\|B_{6,\iota-1}(R)\right\| + O_P(1)$ and $\left\|B_{7,\iota-1}(R)\right\|/\left\|B_{6,\iota-1}(R)\right\| \xrightarrow{P} 1$. Similarly, the FOC wrt $\Theta_{j-\iota-1}$ suggests that in the latter case $\left\|B_{6,j-\iota-1}(R)\right\| = O_P((N\tau)^{1/2}\gamma) = O_P(1)$ but $\left\|B_{7,j-\iota-1}(R)\right\| = \frac{a_{j-\iota-1}\tau_{j-\iota-1}}{a_{j-\iota}\tau_{j-\iota}}\left\|B_{6,j-\iota}(R)\right\|$ is explosive. So the FOC in this last case cannot be satisfied and a contradiction would arise unless there is no break point before $j-1$ for $j\geq 2$. But if there is no point before $j-1$ and after j that belongs to $\mathcal{S}$, there will be no break point in the time interval $[1,T]$, contradicting to the requirement that we have at least one break contained in $[1,T]$. Consequently, w.p.a.1 $\tilde{\beta}_j = \tilde{\Theta}_j - \tilde{\Theta}_{j-1}$ must lie in a position where $\left\|\Theta_j - \Theta_{j-1}\right\|$ is not differentiable with respect to Θ_j in subcase (a2).

Now, we consider case (b). Note that only one term in the penalty component ($\gamma\sum_{j=1}^{J}w_j \parallel \Theta_j - \Theta_{j-1} \parallel$) is involved with Θ_J. Suppose that $\tilde{\beta}_J \neq 0$ for sufficiently large (N,T) (note that $J\in\mathbb{S}_1$ under our assumption) and wlog $\|\tilde{\beta}_{J,R}\|/\|\tilde{\beta}_J\| \geq 1/\sqrt{R}$. Then the FOC wrt $\Theta_{J,R}$ is given by

$$\begin{aligned}
\mathbf{0}_{N\times 1} &= \frac{a_J}{\sqrt{N}}(\mathbf{X}_{S_J} - \tilde{\Theta}_J\hat{\mathbf{Z}}_{S_J}^{\top})\hat{\mathbf{Z}}_{S_J,R} - a_J\tau_J N^{1/2}\gamma w_J\tilde{\beta}_{J,R}/\left\|\tilde{\beta}_J\right\| \\
&= B_{1,J}(R) + B_{2,J}(R) + B_{3,J}(R) + B_{4,J}(R) + B_{5,J}(R) - B_{6,J}(R). \qquad (A.8)
\end{aligned}$$

As in case (a), we can readily show that $\sum_{l=1}^{5}\left\|B_{l,J}(R)\right\| = O_P(1)$ and $\left\|B_{6,J}(R)\right\|$ is explosive in probability at the rate $(N\tau)^{1/2}\gamma\eta_{N\tau}^{\varkappa}$. So the above FOC cannot hold and $\tilde{\beta}_J = \tilde{\Theta}_J - \tilde{\Theta}_{J-1}$ must be in a position where $\left\|\Theta_J - \Theta_{J-1}\right\|$ is not differentiable with respect to Θ_J. Analogously, we can show that in case (c), $\tilde{\beta}_1 = \tilde{\Theta}_1 - \tilde{\Theta}_0$ must be in a position where $\|\Theta_1 - \Theta_0\|$ is not differentiable with respect to Θ_0

In the case where $|\mathcal{S}| = 0$ so that $[1,T]$ contains no break. First, suppose that $\tilde{\beta}_J = \tilde{\Theta}_J - \tilde{\Theta}_{J-1} \neq 0$ and wlog $\|\tilde{\beta}_{J,R}\|/\|\tilde{\beta}_J\| \geq 1/\sqrt{R}$. Then the FOC wrt $\Theta_{J,R}$ is given by (A.8); following the above analysis for case (b), we have $\sum_{l=1}^{5}\|B_{l,J}(R)\| = O_P(1)$ and $\|B_{6,J}(R)\|$ is explosive in probability at the rate $(N\tau)^{1/2}\gamma\eta_{N\tau}^{\varkappa}$. This implies that the equality in (A.8) cannot occur in large samples and $\Pr\{\|\tilde{\Theta}_J - \tilde{\Theta}_{J-1}\| = 0\} \to 1$ as $(N,T) \to \infty$. In this case, by the subdifferential calculus, $\mathbf{0}_{N\times 1}$ belongs to the subdifferential of the objective function wrt Θ_J:

$$\begin{aligned}\mathbf{0}_{N\times R} &= \frac{a_J}{\sqrt{N}}(\mathbf{X}_{S_J} - \tilde{\Theta}_J\hat{\mathbf{Z}}_{S_J}^{\intercal})\hat{\mathbf{Z}}_{S_J} - a_J\tau_J N^{1/2}\gamma w_J \varrho_J \\ &= B_{1,J} + B_{2,J} + B_{3,J} + B_{4,J} + B_{5,J} - B_{6,J},\end{aligned} \tag{A.9}$$

for some ϱ_J with $\left\|\varrho_J\right\| \le 1$. In particular, the rth column of $B_{6,J}$ is now given by $a_J\tau_J N^{1/2}\gamma w_J \varrho_{J,r}$. Since $\sum_{l=1}^{5}\left\|B_{l,J}\right\| = O_P(1)$, we must have $\left\|B_{6,J}\right\| = O_P(1)$. Next, we consider the FOC wrt Θ_{J-1}:

$$\begin{aligned}\mathbf{0}_{N\times R} &= \frac{a_{J-1}}{\sqrt{N}}(\mathbf{X}_{S_{J-1}} - \tilde{\Theta}_{J-1}\hat{\mathbf{Z}}_{S_{J-1}}^{\intercal})\hat{\mathbf{Z}}_{S_{J-1}} - a_{J-1}\tau_{J-1}N^{1/2}\gamma w_{J-1}\varrho_{J-1} + a_{J-1}\tau_{J-1}N^{1/2}\gamma w_J\varrho_J \\ &= B_{1,J-1} + B_{2,J-1} + B_{3,J-1} + B_{4,J-1} + B_{5,J-1} - B_{6,J-1} + B_{7,J-1}.\end{aligned} \tag{A.10}$$

Noting that $B_{7,J-1} = \frac{a_{J-1}\tau_{J-1}}{a_J\tau_J}B_{6,J} = O_P(1)$ and $\sum_{l=1}^{5}\left\|B_{l,J-1}\right\| = O_P(1)$, we must have

$$\left\|B_{6,J-1}\right\| \le \left\|B_{7,J-1}\right\| + \sum_{l=1}^{5}\left\|B_{l,J-1}\right\| = \frac{a_{J-1}\tau_{J-1}}{a_J\tau_J}\left\|B_{6,J}\right\| + \sum_{l=1}^{5}\left\|B_{l,J-1}\right\| = O_P(1)$$

in order for the first equality in (A.10) to hold. Then $\tilde{\beta}_{J-1} = \tilde{\Theta}_{J-1} - \tilde{\Theta}_{J-2}$ must be in a position where $\left\|\Theta_{J-1} - \Theta_{J-2}\right\|$ is not differentiable with respect to Θ_{J-1}. Deducting this way until $j=1$, we must have

$$\begin{aligned}\left\|B_{6,j}\right\| &\le \left\|B_{7,j}\right\| + \sum_{l=1}^{5}\left\|B_{l,j}\right\| = \frac{a_j\tau_j}{a_{j+1}\tau_{j+1}}\left\|B_{6,j+1}\right\| + \sum_{l=1}^{5}\left\|B_{l,j}\right\| \\ &\le \frac{a_j\tau_j}{a_{j+1}\tau_{j+1}}\left(\frac{a_{j+1}\tau_{j+1}}{a_{j+2}\tau_{j+2}}\left\|B_{6,j+2}\right\| + \sum_{l=1}^{5}\left\|B_{l,j+1}\right\|\right) + \sum_{l=1}^{5}\left\|B_{l,j}\right\| \\ &\le \cdots \\ &\le \left(\Pi_{i=j}^{J-1}\frac{a_i\tau_i}{a_{i+1}\tau_{i+1}}\right)\left\|B_{6,J}\right\| + \sum_{s=1}^{J-j-1}\left(\Pi_{i=j+1}^{j+s}\frac{a_i\tau_i}{a_{i+1}\tau_{i+1}}\right)\sum_{l=1}^{5}\left\|B_{l,j+s}\right\| + \sum_{l=1}^{5}\left\|B_{l,j}\right\| \\ &\text{for } j = J-2,\ldots,1.\end{aligned}$$

Noting that when $|\mathcal{S}| = 0$, $a_j\tau_j = \eta_{N\tau_j}^{-1}\tau_j = \max(N^{-1/2}, \tau_j^{-1/2})\tau_j$ and $\tau_{j-1} = \lfloor Tj/J_1\rfloor - \lfloor T(j-1)/J_1\rfloor$ for $j = 1,\ldots,J_1$, it is easy to argue that both $\max_{1\le j\le J-2}\Pi_{i=j}^{J-1}\frac{a_i\tau_i}{a_{i+1}\tau_{i+1}}$ and $\max_{1\le j\le J-2}\max_{1\le s\le J-j-1}\Pi_{i=j}^{j+s}\frac{a_i\tau_i}{a_{i+1}\tau_{i+1}}$ are bounded above by a constant C. [E.g., if $\tau_{j-1} = T/J_1$ for each j, $C = 1$.] Then

$$\frac{1}{J}\max_{1\le j\le J-1}\left\|B_{6,j}\right\| \le (C+1)\frac{1}{J}\sum_{j=1}^{J}\sum_{l=1}^{5}\left\|B_{l,J-j}\right\| + \frac{C}{J}\left\|B_{6,J}\right\| = O_P(1)$$

because we can readily show that $\frac{1}{J}\sum_{j=1}^{J}\sum_{l=1}^{5}\left\|B_{l,J-j}\right\| = O_P(1)$. Under Assumption A8(ii), it is still true that $(N\tau)^{1/2}\gamma\eta_{N\tau}^{\varkappa}/J \to \infty$ as $(N,T) \to \infty$. This implies that all $j = J-2, J-3,\ldots,1$, $\tilde{\beta}_j = \tilde{\Theta}_j - \tilde{\Theta}_{j-1}$ must be in a position where $\left\|\Theta_j - \Theta_{j-1}\right\|$ is not differentiable with respect to Θ_j and thus (A.2) also holds in this case. ■

A.3. Proof of Proposition 4.1

Let $\mathbb{S}$, $\mathbb{S}_1$, $\mathbb{S}_{2a}$, $\mathbb{S}_{2b}$, and $\mathbb{S}_{2c}$ be as defined in Section 3.1. Recall that $\boldsymbol{\alpha}_\kappa^* = (\boldsymbol{\alpha}_{\kappa-1}^0, \boldsymbol{\alpha}_\kappa^0)$, $\mathbf{F}_{S_j}^* = (F_{v_j}^*,\ldots,F_{v_{j+1}}^*)^{\intercal}$, and $F_t^* = (F_t^{0\intercal}1_{jt}, F_t^{0\intercal}\bar{1}_{jt})^{\intercal}$, where $1_t = 1\left\{v_j \le t < t_\kappa^0\right\}$, $\bar{1}_t = 1\left\{t_\kappa^0 \le t < v_{j+1}\right\}$. Define $H_j^{(R)} \equiv (\frac{1}{\tau_j}\mathbf{F}_{S_j}^{0\intercal}\mathbf{F}_{S_j}^0)\times(\frac{1}{N}\boldsymbol{\alpha}_\kappa^{0\intercal}\hat{\boldsymbol{\Delta}}_j^{(R)})$, an $R_0\times R$ matrix, and $H_{j*}^{(R)} \equiv (\frac{1}{\tau_j}\mathbf{F}_{S_j}^{*\intercal}\mathbf{F}_{S_j}^*)\times(\frac{1}{N}\boldsymbol{\alpha}_\kappa^{*\intercal}\hat{\boldsymbol{\Delta}}_j^{(R)})$, an $2R_0\times R$ matrix. Similarly, let $H_{j,\ell}^{(R)} \equiv (\frac{1}{\tau_j}\mathbf{F}_{S_j,\ell}^{0\intercal}\mathbf{F}_{S_j,\ell}^0)(\frac{1}{N}\boldsymbol{\alpha}_{\kappa+\ell-2}^{0\intercal}\hat{\boldsymbol{\Delta}}_j^{(R)})$ for $\ell = 1, 2$. Let $J_1 = J+1$ and $\tau = \min_{0\le j\le J}\tau_j$. Define

$$\bar{\boldsymbol{\Delta}}_j^{(R)} = \begin{cases}\boldsymbol{\alpha}_\kappa^0 H_j^{(R)} & \text{if } j \in \mathbb{S}_1 \\ \boldsymbol{\alpha}_{\kappa-1}^0 H_{j,1}^{(R)} & \text{if } j \in \mathbb{S}_{2a} \\ \boldsymbol{\alpha}_\kappa^0 H_{j,2}^{(R)} & \text{if } j \in \mathbb{S}_{2b} \\ \boldsymbol{\alpha}_\kappa^* H_{j*}^{(R)} & \text{if } j \in \mathbb{S}_{2c}\end{cases} \text{ and } \boldsymbol{\Delta}_j^0 = \begin{cases}\boldsymbol{\alpha}_\kappa^0 & \text{if } j \in \mathbb{S}_1 \\ \boldsymbol{\alpha}_{\kappa-1}^0 & \text{if } j \in \mathbb{S}_{2a} \\ \boldsymbol{\alpha}_\kappa^0 & \text{if } j \in \mathbb{S}_{2b} \\ \boldsymbol{\alpha}_\kappa^* & \text{if } j \in \mathbb{S}_{2c}\end{cases} \text{ for some } \kappa = \kappa(j). \tag{A.11}$$

To prove Proposition 3.3, we need the following three lemmas. More precisely, Lemmas A.5 and A.6 are used in the proof of Lemma A.7, which is used to prove Proposition 4.1. The proofs of these three lemmas are provided in the online Supplementary Material.

Lemma A.5. *Suppose that Assumptions A1–A4 and A8 hold. Then for any $R \geq 1$, there exist $R_0 \times R$ matrices $\{H_j^{(R)}, H_{j,1}^{(R)}, H_{j,2}^{(R)}\}$ and $2R_0 \times R$ matrices $\{H_{j*}^{(R)}\}$ with rank$(H_j^{(R)}) = \min\{R, R_0\}$, rank$(H_{j,\ell}^{(R)}) = \min\{R, R_0\}$ with $\ell = 1, 2$, and rank$(H_{j*}^{(R)}) = \min\{R, 2R_0\}$ such that*

(i) $\sum_{j\in\mathbb{S}_1} N^{-1}\left\|\breve{\boldsymbol{\Delta}}_j^{(R)} - \bar{\boldsymbol{\Delta}}_j^{(R)}\right\|^2 = O_P\left(\eta_{N\tau}^{-2}|\mathbb{S}_1|\right)$,

(ii) $\max_{j\in\mathbb{S}_1} N^{-1}\left\|\breve{\boldsymbol{\Delta}}_j^{(R)} - \bar{\boldsymbol{\Delta}}_j^{(R)}\right\|^2 = O_P\left(\eta_{N\tau}^{-2}\ln T\right)$,

(iii) $\max_{j\in\mathbb{S}_1}\left\|N^{-1}\breve{\boldsymbol{\Delta}}_j^{(R)\intercal}\breve{\boldsymbol{\Delta}}_j^{(R)} - N^{-1}\bar{\boldsymbol{\Delta}}_j^{(R)\top}\bar{\boldsymbol{\Delta}}_j^{(R)}\right\| = O_P(\eta_{N\tau}^{-1}(\ln T)^{1/2})$,

where $H_j^{(R)}, H_{j,1}^{(R)}, H_{j,2}^{(R)}$ and $H_{j*}^{(R)}$ are implicitly defined in $\bar{\boldsymbol{\Delta}}_j^{(R)}$ in (A.11).

Lemma A.6. *Suppose that Assumptions A1–A4 and A8 hold and $R > R_0$. Write the Moore–Penrose generalized inverse of $H_j^{(R)}$ as $H_j^{(R)+} = \begin{pmatrix} H_j^{(R)+}(1) \\ H_j^{(R)+}(2) \end{pmatrix}$, where $H_j^{(R)+}(1)$ and $H_j^{(R)+}(2)$ are $R_0 \times R_0$ and $(R - R_0) \times R_0$ matrices, respectively. Let $V_{N,j}^{(R)}$ denote an $R \times R$ diagonal matrix consisting of the R largest eigenvalues of the $N \times N$ matrix $\left(N\tau_j\right)^{-1}\boldsymbol{X}_{S_j}\boldsymbol{X}_{S_j}^\intercal$ where the eigenvalues are ordered in decreasing order along the main diagonal line. Write $\hat{\boldsymbol{\Delta}}_j^{(R)} = [\hat{\boldsymbol{\Delta}}_j^{(R)}(1), \hat{\boldsymbol{\Delta}}_j^{(R)}(2)]$ and $H_j^{(R)} = [H_j^{(R)}(1), H_j^{(R)}(2)]$, where $\hat{\boldsymbol{\Delta}}_j^{(R)}(1), \hat{\boldsymbol{\Delta}}_j^{(R)}(2), H_j^{(R)}(1)$, and $H_j^{(R)}(2)$ are $N \times R_0$, $N \times (R - R_0)$, $R_0 \times R_0$, and $R_0 \times (R - R_0)$ matrices, respectively. Write $V_{N,j}^{(R)} = diag(V_{N,j}^{(R)}(1), V_{N,j}^{(R)}(2))$, where $V_{N,j}^{(R)}(1)$ denotes the upper left $R_0 \times R_0$ submatrix of $V_{N,j}^{(R)}$. Then*

(i) $\max_{j\in\mathbb{S}_1} N^{-1}\left\|\hat{\boldsymbol{\Delta}}_j^{(R)}(1) - \boldsymbol{\alpha}_\kappa^0 H_j^{(R)}(1)V_{N,j}^{(R)}(1)^{-1}\right\|^2 = O_P\left(\eta_{N\tau}^{-2}\ln T\right)$ *and* $\max_{j\in\mathbb{S}_1}\left\|H_j^{(R)}(2)\right\|^2 = O_P\left(\tau^{-1}\ln T + N^{-1}\right)$,

(ii) $\max_{j\in\mathbb{S}_1}\left\|H_j^{(R)+}(1)\right\| = O_P(1)$ *and* $\max_{j\in\mathbb{S}_1}\left\|H_j^{(R)+}(2)\right\| = O_P(\tau^{-1/2}(\ln T)^{1/2} + N^{-1/2})$,

(iii) $|\mathbb{S}_1|^{-1}\sum_{j\in\mathbb{S}_1}\left(N\tau_j\right)^{-1} tr\{\mathbf{F}_{S_j}^0 H_j^{(R)+\intercal}(\breve{\boldsymbol{\Delta}}_j^{(R)} - \boldsymbol{\alpha}_\kappa^0 H_j^{(R)})^\intercal \mathbf{E}_{S_j}\} = O_P\left(\eta_{N\tau}^{-2}\right)$,

(iv) $|\mathbb{S}_1|^{-1}\sum_{j\in\mathbb{S}_1}\left(N\tau_j\right)^{-1}\left\|(\breve{\boldsymbol{\Delta}}_j^{(R)} - \boldsymbol{\alpha}_\kappa^0 H_j^{(R)})H_j^{(R)+}\mathbf{F}_{S_j}^{0\intercal}\right\|^2 = O_P\left(\eta_{N\tau}^{-2}\right)$.

Lemma A.7. *Suppose that Assumptions A1–A4 and A8 hold. Then*

(i) $V\left(R, \{\breve{\boldsymbol{\Delta}}_j^{(R)}\}\right) - V\left(R, \{\bar{\boldsymbol{\Delta}}_j^{(R)}\}\right) = O_P\left(\eta_{N\tau}^{-1}(\ln T)^{1/2} + mJ^{-1}\right)$ *for each R with $1 \leq R \leq R_0$,*

(ii) there exists a constant $c_R > 0$ such that plim inf$_{(N,T)\to\infty}\left[V\left(R, \{\bar{\boldsymbol{\Delta}}_j^{(R)}\}\right) - V\left(R, \left\{\boldsymbol{\Delta}_j^0\right\}\right)\right] \geq c_R$ for each R with $1 \leq R < R_0$,

(iii) $V\left(R, \{\breve{\boldsymbol{\Delta}}_j^{(R)}\}\right) - V\left(R_0, \{\breve{\boldsymbol{\Delta}}_j^{(R_0)}\}\right) = O_P\left(mJ^{-1} + \eta_{N\tau}^{-2}\right)$ *for each R with $R \geq R_0$,*

where $\boldsymbol{\Delta}_j^0, j = 0, 1, \ldots, J$, are defined in (A.11).

Proof of Proposition 4.1. Let $V(R) = V(R, \{\breve{\boldsymbol{\Delta}}_j^{(R)}\})$ for all R. Note that $IC_1(R) - IC_1(R_0) = \ln[V(R)/V(R_0)] + (R - R_0)\rho_{1NT}$. We discuss two cases: (i) $R < R_0$, and (ii) $R > R_0$.

In case (i), $V(R)/V(R_0) > 1 + \epsilon_0$ for some $\epsilon_0 > 0$ w.p.a.1 by Lemma A.7(i) and (ii). It follows that $\ln[V(R)/V(R_0)] \geq \epsilon_0/2$ w.p.a.1. Noting that $(R - R_0)\rho_{1NT} \to 0$ under Assumption A6, this implies that $IC_1(R) - IC_1(R_0) \geq \epsilon_0/4$ w.p.a.1. Consequently, we have $P(IC_1(R) - IC_1(R_0) > 0) \to 1$ for any $R < R_0$ as $(N, T) \to \infty$. In case (ii), we apply Lemma A.7 (iii) and Assumption A6 to obtain

$$\begin{aligned} P(IC_1(R) - IC_1(R_0) > 0) &= P\{\ln[V(R)/V(R_0)] + (R - R_0)\rho_{1NT} > 0\} \\ &= P\left\{O_P(1) + (R - R_0)\rho_{1NT}/(mJ^{-1} + \eta_{N\tau}^{-2}) > 0\right\} \to 1 \end{aligned}$$

for any $R > R_0$ as $(N, T) \to \infty$. Consequently, the minimizer of $IC_1(R)$ can only be achieved at $R = R_0$ w.p.a.1. That is, $P(\hat{R} = R_0) \to 1$ for any $R \in [1, R_{\max}]$ as $(N, T) \to \infty$. ■

A.4. Proof of Proposition 4.2

Let $\mathcal{T}_{m^0}^0 = \{t_1^0, \ldots, t_{m^0}^0\}$, the collection of the true m^0 break dates. Let $\mathbb{T}_m$ consist of $\mathcal{T}_m = \{t_1, \ldots, t_m\}$ such that $2 \leq t_1 < \cdots < t_m \leq T$, $t_0 = 1$ and $t_{m+1} = T + 1$; and let $\bar{\mathbb{T}}_m$ consist of $\mathcal{T}_m = \{t_1, \ldots, t_m\}$ such that $\mathcal{T}_{m^0}^0 \subset \mathcal{T}_m$, $2 \leq t_1 < \cdots < t_m \leq T$ for $m^0 < m \leq m_{\max}$. As in Section 4.2, we define

$$\hat{\sigma}(\mathcal{T}_m) = \frac{1}{NT}\sum_{j=1}^{m+1}\sum_{t=t_{j-1}}^{t_j-1}\left[\mathbf{X}_t - \hat{\boldsymbol{\alpha}}_j(\mathcal{T}_m)\hat{F}_t(\mathcal{T}_m)\right]^\intercal\left[\mathbf{X}_t - \hat{\boldsymbol{\alpha}}_j(\mathcal{T}_m)\hat{F}_t(\mathcal{T}_m)\right],$$

where $(\hat{\boldsymbol{\alpha}}_\kappa(\mathcal{T}_m), \{\hat{F}_t(\mathcal{T}_m)\}) = \arg\min_{\boldsymbol{\alpha}_\kappa, \{F_t\}}\sum_{t\in I_\kappa}(\mathbf{X}_t - \boldsymbol{\alpha}_\kappa F_t)^\intercal(\mathbf{X}_t - \boldsymbol{\alpha}_\kappa F_t)$ subject to the constraints that $N^{-1}\boldsymbol{\alpha}_\kappa^\intercal\boldsymbol{\alpha}_\kappa = \mathbb{I}_R$ and $\mathbf{F}_{I_\kappa}^\intercal\mathbf{F}_{I_\kappa} =$ diagonal. To prove Proposition 4.2, we need the following two lemmas.

Lemma A.8. *Suppose that the conditions in Proposition 4.2 hold. Then there exists a positive constant $\underline{c}_\lambda$ such that*

$$\min_{0\leq m<m^0}\inf_{\mathcal{T}_m\in\mathbb{T}_m} c_{1NT}^{-1}\left[\hat{\sigma}^2(\mathcal{T}_m)-\hat{\sigma}^2(\mathcal{T}_{m^0}^0)\right]\geq \underline{c}_\lambda+o_P(1).$$

Lemma A.9. *Suppose that the conditions in Proposition 4.2 hold. Then we have*

$$\max_{m^0<m\leq m_{\max}}\sup_{\mathcal{T}_m\in\bar{\mathbb{T}}_m} c_{2NT}^{-1}\left|\hat{\sigma}^2(\mathcal{T}_m)-\hat{\sigma}^2(\mathcal{T}_{m^0}^0)\right|=O_P(1)$$

where $c_{2NT}=N^{-1}+I_{\min}^{-1}+m_{\max}T^{-1}$.

Proof of Proposition 4.2. Let $\bar{\sigma}_{NT}^2=\frac{1}{NT}\sum_{t=1}^N \mathbf{e}_t^\intercal\mathbf{e}_t$. Denote $\Gamma=[0,\gamma_{\max}]$, which is divided into three subsets Γ_0, Γ_- and Γ_+ as follows

$$\Gamma_0=\left\{\gamma\in\Gamma:\hat{m}_\gamma=m^0\right\},\ \Gamma_-=\left\{\gamma\in\Gamma:\hat{m}_\gamma<m^0\right\},\ \text{and}\ \Gamma_+=\left\{\gamma\in\Gamma:m^0<\hat{m}_\gamma\leq m_{\max}\right\}.$$

Let $\gamma^0\equiv\gamma_{NT}^0$ denote an element in Γ_0 that also satisfies the conditions on γ in Assumption A6(ii). Let $\hat{t}_j(\gamma)$ be the estimates of the break dates in the third stage when the tuning parameter γ is applied in the second stage Lasso procedure. By Propositions 3.3 and 3.4, for any $\gamma^0\in\Gamma_0$ we have $\hat{m}_{\gamma^0}=m^0$ w.p.a.1. and $\lim_{(N,T)\to\infty}\Pr(\hat{t}_j(\gamma^0)=t_j^0,\ j=1,\ldots,m^0)=1$. It is easy to show that $\hat{\sigma}^2(\mathcal{T}_{m^0}^0)=\bar{\sigma}_{NT}^2+O_P\left(\eta_{N\tau}^{-2}\right)\xrightarrow{P}\sigma_0^2$, and $IC_2\left(\gamma^0\right)=\log\hat{\sigma}^2(\hat{\mathcal{T}}_{\hat{m}_{\gamma^0}}(\gamma^0))+\rho_{2NT}\left(m^0+1\right)\xrightarrow{P}\ln(\sigma_0^2)$, where $\sigma_0^2=\lim_{(N,T)\to\infty}\frac{1}{NT}\sum_{t=1}^N E(\mathbf{e}_t^\intercal\mathbf{e}_t)$ and the second convergence holds because $\rho_{2NT}\left(m^0+1\right)=o(1)$ by Assumption A10(ii) and $\hat{\mathcal{T}}_{\hat{m}_{\gamma^0}}(\gamma^0)=\mathcal{T}_{m^0}^0$ w.p.a.1. We next consider the cases of under- and over-fitted models separately.

Case 1 (Under-Fitted Model with $\hat{m}_\gamma<m^0$): By Lemma A.8 and Assumption A10(ii),

$$\begin{aligned}\Pr\left(\inf_{\gamma\in\Gamma_-}IC_2(\gamma)>IC_2\left(\gamma^0\right)\right)&=\Pr\left(\inf_{\gamma\in\Gamma_-}c_{1NT}^{-1}\left[\ln\left(\hat{\sigma}^2(\hat{\mathcal{T}}_{\hat{m}_\gamma}(\gamma))/\hat{\sigma}^2(\mathcal{T}_{m^0}^0)\right)+\rho_{2NT}\left(\hat{m}_\gamma-m^0\right)\right]>0\right)\\&\geq\Pr\left(\underline{c}_\lambda/2+o_P(1)>0\right)\to 1.\end{aligned}$$

Case 2 (Over-Fitted Model with $\hat{m}_\gamma>m^0$): For given $\mathcal{T}_m=\{T_1,\ldots,T_m\}\in\mathbb{T}_m$, we let $\bar{\mathcal{T}}_{m^*+m^0}=\left\{\bar{T}_1,\bar{T}_2,\ldots,\bar{T}_{m^*+m^0}\right\}$ denote the union of $\mathcal{T}_m$ and $\mathcal{T}_{m^0}^0$ with elements ordered in non-descending order: $2\leq\bar{T}_1<\bar{T}_2<\cdots<\bar{T}_{m^*+m^0}\leq T$ for some $m^*\in\{0,1,\ldots,m\}$. In view of the fact that $\hat{\sigma}^2(\bar{\mathcal{T}}_{m^*+m^0})\leq\hat{\sigma}^2(\mathcal{T}_m)$ for all $\mathcal{T}_m\in\mathbb{T}_m$, $c_{2NT}^{-1}\left[\hat{\sigma}^2(\bar{\mathcal{T}}_{m^*+m^0})-\bar{\sigma}_{NT}^2\right]=O_P(1)$ uniformly in $\mathcal{T}_m\in\mathbb{T}_m$ by Lemma A.9, and $c_{2NT}^{-1}\rho_{2NT}\to\infty$ by Assumption A10(iii), we have

$$\begin{aligned}&\Pr\left(\inf_{\gamma\in\Gamma_+}IC_2(\gamma)>IC_2(\gamma^0)\right)\\\geq\ &\Pr\left(\min_{m^0<m\leq m_{\max}}\inf_{\mathcal{T}_m\in\mathbb{T}_m}\left\{c_{2NT}^{-1}\left[\ln\left(\hat{\sigma}^2(\mathcal{T}_m)/\hat{\sigma}^2(\mathcal{T}_{m^0})\right)\right]+c_{2NT}^{-1}\rho_{2NT}\left(m-m^0\right)\right\}>0\right)\\\geq\ &\Pr\left(\min_{m^0<m\leq m_{\max}}\inf_{\mathcal{T}_m\in\mathbb{T}_m}\left\{c_{2NT}^{-1}\left[\ln\left(\hat{\sigma}^2(\bar{\mathcal{T}}_{m^*+m^0})/\hat{\sigma}^2(\mathcal{T}_{m^0})\right)\right]+c_{2NT}^{-1}\rho_{2NT}\left(m-m^0\right)\right\}>0\right)\\\to\ &1.\end{aligned}$$

We have completed the proof of Proposition 4.2. ■

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2018.06.019.

References

Bai, J., 1997. Estimating multiple breaks one at a time. Econometric Theory 13, 315–352.
Bai, J., 2003. Inferential theory for factors models of large dimensions. Econometrica 71, 135–173.
Bai, J., Li, K., 2012. Statistical analysis of factor models of high dimension. Ann. Statist. 40, 436–465.
Bai, J., Li, K., 2014. Theory and methods of panel data models with interactive effects. Ann. Statist. 42, 142–170.
Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.
Bai, J., Perron, P., 1998. Estimating and testing liner models with multiple structural changes. Econometrica 66, 47–78.
Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. J. Appl. Econometrics 18, 1–22.
Baltagi, B.H., Kao, C., Wang, F., 2016. Estimating and testing high dimensional factor models with multiple structural changes. Working Paper, Dept. of Economics, Syracuse University.
Baltagi, B.H., Kao, C., Wang, F., 2017. Identification and estimation of a large factor models with structural stability. J. Econometrics 197, 87–100.
Bernstein, D.S., 2005. Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory. Princeton University Press, Princeton.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3, 1–122.
Brandom, J.B., Plagborg-Mollerb, M., Stock, J.H., Watson, M.W., 2013. Consistent factor estimation in dynamic factor models with structural instability. J. Econometrics 177, 289–304.
Breitung, J., Eickmeier, S., 2011. Testing for structural breaks in dynamic factor models. J. Econometrics 163, 71–84.
Chen, L., 2015. Estimation of the common break date in large factor models. Econom. Lett. 131, 70–74.
Chen, L., Dolado, J., Gonzalo, J., 2014. Detecting big structural breaks in large factor models. J. Econometrics 180, 30–48.
Cheng, L., Liao, Z., Schorfheide, F., 2016. Shrinkage estimation of high-dimensional factor models with structural instabilities. Rev. Econom. Stud. 83, 1511–1543.
Corradi, V., Swanson, N.R., 2014. Testing for structural stability of factor augmented forecasting models. J. Econometrics 182, 100–118.
Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 75, 603–680.
Fan, J., Liao, Y., Shi, X., 2015. Risks of large portfolios. J. Econometrics 186, 367–387.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. The generalized dynamic factor model. J. Amer. Statist. Assoc. 100, 830–840.
Hallin, M., Liska, R., 2007. Determining the number of factors in the general dynamic factor models. J. Amer. Statist. Assoc. 103, 603–617.
Han, X., Inoue, A., 2015. Tests for parameter instability in dynamic factor models. Econometric Theory 31, 1117–1152.
Hansen, B.E., 2001. The new econometrics of structural change: dating breaks in U.S. labour productivity. J. Econ. Perspect. 15, 117–128.
Kliesen, K.L., 2003. The 2001 recession: how was it different and what developments may have caused it? Review –Federal Reserve Bank of St. Louis; St. Louis 85, pp. 23-38.
Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: inference for the number of factors. Ann. Statist. 40, 694–726.
Li, D., Qian, J., Su, L., 2016. Panel data models with interactive fixed effects and multiple structural breaks. J. Amer. Statist. Assoc. 111, 1804–1819.
Moon, H., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83, 1543–1579.
Onatski, A., 2012. Asymptotics of the principal components estimator of large factor models with weakly influential factors. J. Econometrics 168, 244–258.
Qian, J., Su, L., 2016. Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. J. Econometrics 191, 86–109.
Qian, J., Su, L., 2016. Shrinkage estimation of regression models with multiple structural changes. Econometric Theory 32, 1376–1433.
Shi, Y., 2015. Structural changes in high dimensional factor models. Working Paper, Dept. of Economics, Texas A&M University.
Stock, J.H., Watson, M.W., 1998. Diffusion indexes. NBER Working Paper 6702.
Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97, 1167–1179.
Stock, J.H., Watson, M.W., 2009. Forecasting in dynamic factor models subject to structural instability. In: Castle, J., Shephard, N. (Eds.), The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F Hendry. Oxford University Press, pp. 1–57.
Su, L., Wang, X., 2017. On time-varying factor models: estimation and testing. J. Econometrics 198, 84–101.
Tibshirani, R., Saunders, S., Rosset, M., Zhu, J., Knight, K., 2005. Sparsity and smoothness via the fused Lasso. J. R. Stat. Soc. Ser. B Stat. Methodol. 67, 91–108.
Wang, W., Fan, J., 2017. Asymptotics of empirical eigen-structure for high dimensional spiked covariance. Ann. Statist. 45, 1342–1374.
Yamamoto, Y., Tanaka, S., 2015. Testing for factor loading structural change under common breaks. J. Econometrics 189, 187–206.
Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B Stat. Methodol. 68, 49–67.
Zou, H., 2006. The adaptive Lasso and its oracle properties. J. Amer. Statist. Assoc. 101, 1418–1429.