Detection of Multiple Structural Breaks in Large Covariance Matrices

Yu-Ning Li[*]   Degui Li[†]   Piotr Fryzlewicz[‡]

This version: January 17, 2022

Abstract

This paper studies multiple structural breaks in large contemporaneous covariance matrices of high-dimensional time series satisfying an approximate factor model. The breaks in the second-order moment structure of the common components are due to sudden changes in either factor loadings or covariance of latent factors, requiring appropriate transformation of the factor models to facilitate estimation of the (transformed) common factors and factor loadings via the classical principal component analysis. With the estimated factors and idiosyncratic errors, an easy-to-implement CUSUM-based detection technique is introduced to consistently estimate the location and number of breaks and correctly identify whether they originate in the common or idiosyncratic error components. The algorithms of Wild Binary Segmentation for Covariance (WBS-Cov) and Wild Sparsified Binary Segmentation for Covariance (WSBS-Cov) are used to estimate breaks in the common and idiosyncratic error components, respectively. Under some technical conditions, the asymptotic properties of the proposed methodology are derived with near-optimal rates (up to a logarithmic factor) achieved for the estimated breaks. Monte-Carlo simulation studies are conducted to examine the finite-sample performance of the developed method and its comparison with other existing approaches. We finally apply our method to study the contemporaneous covariance structure of daily returns of S&P 500 constituents and identify a few breaks including those occurring during the 2007–2008 financial crisis and the recent coronavirus (COVID-19) outbreak. An R package "BSCOV" is provided to implement the proposed algorithms.

Keywords: Approximate factor models, Binary segmentation, CUSUM, Large covariance matrix, Principal component analysis, Structural breaks.

---

[*]School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, China.

[†]The corresponding author, Department of Mathematics, University of York, YO10 5DD, UK. Email address: degui.li@york.ac.uk.

[‡]Department of Statistics, London School of Economics, WC2A 2AE, UK.

1 Introduction

Estimation of covariance matrices is of fundamental importance in modern multivariate statistics, and has applications in various fields such as biology, economics, finance and social networks. Suppose that $(\mathbf{X}_t : t = 1, \cdots, n)$ is a collection of n observations from a d-dimensional random vector with $\mathsf{E}(\mathbf{X}_t) = \mathbf{0}$ and $\mathsf{E}\left(\mathbf{X}_t \mathbf{X}_t^\intercal\right) = \mathbf{\Sigma}$, where $\mathbf{0}$ is a null vector whose size may change from line to line and $\mathbf{\Sigma}$ is a d $\times$ d positive definite covariance matrix. When the dimension d is fixed or significantly smaller than the sample size n, the population covariance matrix $\mathbf{\Sigma}$ can be well estimated by the conventional sample covariance matrix (c.f., Anderson, 2003):

$$\overline{\mathbf{\Sigma}}_n = \frac{1}{n} \sum_{t=1}^n \left(\mathbf{X}_t - \overline{\mathbf{X}}_n\right)\left(\mathbf{X}_t - \overline{\mathbf{X}}_n\right)^\intercal, \quad \overline{\mathbf{X}}_n = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t. \tag{1.1}$$

However, when d exceeds n, the sample covariance matrix $\overline{\mathbf{\Sigma}}_n$ defined in (1.1) becomes singular. Estimation of such a large covariance matrix is generally challenging and has received increasing attention in recent years. Some techniques have been introduced to regularise covariance matrices and produce reliable estimation (c.f., Wu and Pourahmadi, 2003; Ledoit and Wolf, 2004; Bickel and Levina, 2008a,b; El Karoui, 2008a,b; Lam and Fan, 2009; Rothman, Levina and Zhu, 2009; Cai and Liu, 2011; Fan, Liao and Mincheva, 2013; Huang and Fryzlewicz, 2019). A comprehensive review on large covariance matrix estimation can be found in Pourahmadi (2013), Cai, Ren and Zhou (2016) and Fan, Liao and Liu (2016). All of the aforementioned literature assumes that the large covariance matrix is a constant matrix throughout the entire course of data collection, which may be too restrictive in practical situations. Some recent papers including Chen, Xu and Wu (2013) and Chen and Leng (2016) attempt to relax this restriction and consider a large dynamic covariance matrix estimation by allowing the covariance matrix to evolve smoothly over time (or with an index variable).

The main interest of this paper is to detect and estimate multiple structural breaks in large covariance matrices, to which setting the methodology in Chen, Xu and Wu (2013) and Chen and Leng (2016) is no longer applicable. Structural breaks are very common in many areas such as economics and finance, and may occur for various reasons. If such abrupt structural changes are ignored in covariance matrix estimation, subsequent statistical analyses would lead to invalid inference and misleading conclusions. The classical Binary Segmentation (BS) technique is commonly used to detect structural breaks, and has been extensively studied since its introduction by Vostrikova (1981). For instance, in the setting of univariate mean regression models, theoretical properties, computational algorithms and relevant empirical applications of the traditional BS and its generalised version have been systematically studied by Venkatraman (1992), Bai (1997), Cho and Fryzlewicz (2012), Killick, Fearnhead and Eckley (2012), Fryzlewicz (2014, 2020) and the

2

references therein. In recent years, there has been increasing interest in extending this technique to high-dimensional settings such as high-dimensional time series and large panel data (c.f., Cho and Fryzlewicz, 2015; Jirak, 2015; Cho, 2016; Aston and Kirch, 2018; Wang and Samworth, 2018; Wang *et al.*, 2019; Safikhani and Shojaie, 2020). However, most of the aforementioned literature mainly considers break detection and estimation in the first-order moment structure. Aue *et al.* (2009) use the BS method with mean-corrected cumulative sum (CUSUM) to detect structural breaks in fixed dimensional covariance matrices, and establish the limit distributions for the test statistics and the convergence rate for the estimated break date. Korkas and Fryzlewicz (2017) detect breaks in the second-order moment structure of time series by combining the wavelet-based transformation and the univariate *Wild Binary Segmentation (WBS)* algorithm proposed by Fryzlewicz (2014). However, their technique is not directly applicable to our case when the size of the covariance matrix diverges to infinity as the sample size n increases.

In this paper, we consider the second-order moment structure of the high-dimensional random vector X_t generated by an approximate factor model, i.e.,

$$X_t = \Lambda F_t + \epsilon_t, \ \ t = 1, 2, \cdots, n, \tag{1.2}$$

where $\Lambda = (\lambda_1, \cdots, \lambda_d)^\top$ is a matrix of factor loadings with λ_i being an r-dimensional vector, F_t is an r-dimensional vector of common factors, and $\epsilon_t = (\epsilon_{t1}, \cdots, \epsilon_{td})^\top$ is a d-dimensional vector of idiosyncratic components uncorrelated with F_t. The approximate factor model is very popular in economics and finance, and has become an effective tool in analysing high-dimensional time series (c.f., Chamberlain and Rothschild, 1983; Fama and French, 1992; Stock and Watson, 2002; Bai and Ng, 2002, 2006; Chen *et al.*, 2018). Our main focus is the contemporaneous second-order moment structure, which plays a crucial role in various fields such as the classic mean-variance portfolio choice theory and risk management. Note that

$$\Sigma = \text{Cov}(X_t) = \Lambda \text{Cov}(F_t)\Lambda^\top + \text{Cov}(\epsilon_t) = \Sigma(\Lambda, F) + \Sigma(\epsilon), \tag{1.3}$$

where $\Sigma(\epsilon)$ is the covariance matrix of the idiosyncratic components ϵ_t and $\Sigma(\Lambda, F)$ is the covariance matrix of the common components ΛF_t. In this paper, we aim to study multiple structural breaks in the covariance structure of X_t generated by the approximate factor model, estimating the break points and number of breaks. Structural breaks may occur in either the covariance structure of the common components or that of the idiosyncratic components, indicating that the constant covariance structure in (1.3) is replaced by the following time-varying version:

$$\Sigma_t = \Sigma_t(\Lambda, F) + \Sigma_t(\epsilon), \ \ t = 1, \cdots, n. \tag{1.4}$$

The main contributions of this paper, the fundamental novelty of the proposed methodology and its connection to some recent literature are summarised as follows.

- A transformation mechanism is introduced to convert the approximate factor model with multiple structural breaks to that with constant factor loadings, justifying applicability of the classical Principal Component Analysis (PCA) technique (c.f., Bai and Ng, 2002; Stock and Watson, 2002) to a more general model setting. Note that breaks in the second-order moment structure of the common components are due to changes in factor loadings, covariance of latent factors or factor number. This makes it challenging to directly estimate the latent common factors and factor loadings via PCA. When there are multiple breaks in factor loadings, we transform the original factor model to one with time-invariant factor loadings and a time-varying covariance structure for transformed factors, and then use the PCA method to estimate the transformed factors and factor loadings, which are shown to be consistent with convergence rates comparable to those in the existing literature on PCA for approximate factor models without breaks (c.f., Bai and Ng, 2002; Fan, Liao and Mincheva, 2013). This transformation technique is also used by Han and Inoue (2015) and Baltagi, Kao and Wang (2017, 2021) to estimate and test factor models with structural instability. However, they assume that the number of factors does not change and that the covariance matrix is stable for the idiosyncratic errors. In contrast, we consider a more general setting, allowing changes of factor numbers and multiple breaks in the covariance matrices of both the common and idiosyncratic components.

- With the estimated factors and idiosyncratic errors, we propose an easy-to-implement CUSUM-based detection technique to estimate the location and number of breaks and identify whether they originate in the common or idiosyncratic components. The developed CUSUM statistics are directly computed via the empirical second moments of the estimated factors and idiosyncratic errors. Based on Fryzlewicz (2014)'s WBS algorithm, we propose a *Wild Binary Segmentation for Covariance (WBS-Cov)* to locate multiple breaks in the common components. For breaks in the idiosyncratic error components, we introduce a *Wild Sparsified Binary Segmentation algorithm for Covariance (WSBS-Cov)*, combining WBS-Cov and the *Sparsified Binary Segmentation (SBS)* proposed by Cho and Fryzlewicz (2015) in the high-dimensional mean estimation context. A recent paper by Barigozzi, Cho and Fryzlewicz (2018) combines the wavelet-based transformation and the double CUSUM technique (Cho, 2016) to estimate the break time as well as break number within the second-order moment structure. However, a disadvantage of the wavelet-based transformation on the piecewise-constant "signals", involving selection of the wavelet filter and scale parameter, is that it could destroy the piecewise constancy and subsequently affect the estimation accuracy of the break number and location. When we only focus on a specific feature of the data, data transformation, even

one-to-one transformation, would distort the interpretation of the breaks of the structure. Another recent method to deal with sparsity in high-dimensional break detection is the sparse projection introduced by Wang and Samworth (2018) but their method only detects breaks in the first-order moment structure. Wang, Yu and Rinaldo (2021) extends the BS and WBS techniques to detect multiple breaks in large covariance matrices, but they limit attention to the independent sub-Gaussian random vector and assume that the dimension is divergent at a slow polynomial rate of n. In contrast, we allow the dimension to grow faster, e.g., at an exponential rate of n, a typical setting in big data analysis.

- Simulation studies are implemented to assess the finite-sample performance of the proposed method and its comparison with other competing methods. The simulation results show that our method has superior numerical performance in most settings, whereas the method by Barigozzi, Cho and Fryzlewicz (2018) tends to perform poorly when the number of covariance matrix entries with breaks is small relative to the matrix size. Our simulation also shows that the proposed methods are robust when the factor number is over-estimated.

- The developed model and method are applied to the daily return series of S&P 500 constituent which contain 375 firms over the period from 01/01/2000 to 28/02/2021, and detect five breaks in the covariance structure of the common components and three breaks in the idiosyncratic error components. In particular, we identify the breaks occurring during the 2007–2008 global financial crisis and the recent coronavirus (COVID-19) outbreak. With an additional comparison with break detection result of low-dimensional observable risk factor series, we emphasise that the breaks in covariance structure of returns cannot be fully explained by the breaks in the covariance structure of observed risk factors nor the breaks in loadings of these factors.

The rest of the paper is organised as follows. Section 2 introduces the setting of multiple structural breaks, the PCA estimation and CUSUM-based detection methods, and the WBS-Cov and WSBS-Cov algorithms. Section 3 presents the asymptotic theory for the developed WBS-Cov and WSBS-Cov methods. Section 4 discusses some practical issues in the detection procedure. Sections 5 and 6 give the simulation studies and the empirical application, respectively. Section 7 concludes the paper. The technical assumptions are given in Appendix A and proofs of the main theoretical results are available in Appendices C–E of a supplemental document. An R package "BSCOV" for implementing the proposed methods is available from https://github.com/markov10000/BSCOV. Throughout the paper, we let $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the Euclidean norm of a column vector and the Frobenius norm of a matrix, respectively; let $\mathsf{vech}(\cdot)$ denote the half vectorisation of a symmetric matrix obtained by vectorising only the lower triangular part of the matrix; let $\lfloor \cdot \rfloor$ denote the floor function; let $a \vee b$ and $a \wedge b$ denote $\max\{a, b\}$ and $\min\{a, b\}$, respectively; and let $a_n \asymp b_n$ denote

that $a_n = O(b_n)$ and $b_n = O(a_n)$ hold jointly.

2 Methodology

In this section, we first introduce the setting of multiple structural breaks in the contemporaneous covariance structure of both the common and idiosyncratic components, and transform the approximate factor models to construct PCA estimation of the latent common factors and factor loadings (with appropriate rotation). Using the estimated factors and idiosyncratic errors, we then introduce the CUSUM-based statistics to estimate the location and number of structural breaks as well as the WBS-Cov and WSBS-Cov algorithms.

2.1 Model structure for multiple structural breaks

We consider the following multiple structural breaks in the covariance matrix of the common components, i.e., $\Sigma_t(\Lambda, F)$ in (1.4) is specified as

$$
\Sigma_t(\Lambda, F) = \begin{cases}
\Sigma_1^0(\Lambda, F), & 1 \leqslant t \leqslant \eta_1^c, \\
\Sigma_2^0(\Lambda, F), & \eta_1^c + 1 \leqslant t \leqslant \eta_2^c, \\
\quad \vdots & \quad \vdots \\
\Sigma_{K_1+1}^0(\Lambda, F), & \eta_{K_1}^c + 1 \leqslant t \leqslant n,
\end{cases}
\tag{2.1}
$$

where $\Sigma_k^0(\Lambda, F) \neq \Sigma_{k+1}^0(\Lambda, F)$ for $k = 1, \cdots, K_1$, and $\eta_1^c, \cdots, \eta_{K_1}^c$ are unobservable breaks with the superscript "c" denoting the "*common* component". For convenience, we let $\eta_0^c = 0$ and $\eta_{K_1+1}^c = n$. The above structural breaks may be caused by one or a combination of the following cases,

- Case (i): sudden changes in the factor loadings;

- Case (ii): sudden changes in the covariance matrix for common factors;

- Case (iii): sudden changes in the number of factors.

Let Λ_k^0 be a $d \times r_k$ factor loading matrix and $F_{t,k}$ be the corresponding factor vector with dimension r_k when $\eta_{k-1}^c + 1 \leqslant t \leqslant \eta_k^c$, and write the approximate factor model as

$$
X_t = \Lambda_k^0 F_{t,k} + \epsilon_t, \quad \eta_{k-1}^c + 1 \leqslant t \leqslant \eta_k^c.
\tag{2.2}
$$

Letting $\Sigma_k^0(\Lambda, F) = \Lambda_k^0 \mathrm{Cov}(F_{t,k})(\Lambda_k^0)^\intercal$, the structural break structure (2.1) can be equivalently

written as

$$
\Sigma_t(\Lambda, F) = \begin{cases} \Lambda_1^0 \text{Cov}(F_{t,1})(\Lambda_1^0)^\top, & 1 \leqslant t \leqslant \eta_1^c, \\ \Lambda_2^0 \text{Cov}(F_{t,2})(\Lambda_2^0)^\top, & \eta_1^c + 1 \leqslant t \leqslant \eta_2^c, \\ \quad \vdots & \quad \vdots \\ \Lambda_{K_1+1}^0 \text{Cov}(F_{t,K_1+1})(\Lambda_{K_1+1}^0)^\top, & \eta_{K_1}^c + 1 \leqslant t \leqslant n. \end{cases}
$$

Note that, although $\Sigma_k^0(\Lambda, F) \neq \Sigma_{k+1}^0(\Lambda, F)$ for $k = 1, \cdots, K_1$, Λ_k^0 may be the same as Λ_{k+1}^0 (when case (i) does not occur at the break time). It is worthwhile to point out that case (i) is similar to the approximate factor model with structural breaks, see, for example, Breitung and Eickmeier (2011), Chen, Dolado and Gonzalo (2014), Han and Inoue (2015), Cheng, Liao and Schorfheide (2016), Baltagi, Kao and Wang (2017), Bai, Han and Shi (2020), and Duan, Bai and Han (2021). However, the aforementioned papers consider the case of a single structural break in the factor loadings and usually assume that the covariance of the idiosyncratic error components is time invariant. Ma and Su (2018) consider detecting and estimating multiple structural breaks in the factor loadings with a three-step procedure using the adaptive fused group Lasso. Case (iii) is an important type of structure break in the factor model, which has received much attention in recent years (c.f., Baltagi, Kao and Wang, 2017; Barigozzi, Cho and Fryzlewicz, 2018; Li *et al.*, 2019). Section 2.2 below will give a unified factor model framework covering cases (i)–(iii) which facilitates the construction of PCA estimation, and then proceed to introduce the relevant CUSUM statistics in Section 2.3 to detect the multiple structural breaks in the common component.

The structural breaks in the covariance matrix of the error components are specified as follows

$$
\Sigma_t(\epsilon) = \begin{cases} \Sigma_1^0(\epsilon), & 1 \leqslant t \leqslant \eta_1^e, \\ \Sigma_2^0(\epsilon), & \eta_1^e + 1 \leqslant t \leqslant \eta_2^e, \\ \quad \vdots & \quad \vdots \\ \Sigma_{K_2+1}^0(\epsilon), & \eta_{K_2}^e + 1 \leqslant t \leqslant n, \end{cases}
$$

and $\Sigma_k^0(\epsilon) \neq \Sigma_{k+1}^0(\epsilon)$ for $k = 1, \cdots, K_2$, where $\eta_1^e, \cdots, \eta_{K_2}^e$ are unobservable breaks with the superscript "e" denoting the "idiosyncratic *error* component". For convenience, we let $\eta_0^e = 0$ and $\eta_{K_2+1}^e = n$.

Let

$$
\mathcal{B}^c = \left\{ \eta_1^c, \cdots, \eta_{K_1}^c \right\} \quad \text{and} \quad \mathcal{B}^e = \left\{ \eta_1^e, \cdots, \eta_{K_2}^e \right\}. \tag{2.3}
$$

In this paper, we assume that the minimum length of the sub-intervals separated by η_k^c is of order n, see Assumption 4(ii) in Appendix A. Such a restriction is mainly to ensure that the PCA method is applicable to estimate the latent factors and factor loadings. In contrast, the minimum length of the sub-intervals separated by η_k^e is allowed to be of order smaller than n, see Assumption 4(iii). We do not impose any restriction on how the break points in $\mathcal{B}^c$ and those in $\mathcal{B}^e$ are located with

respect to each other.

2.2 Transformed factor models and PCA estimation

Although the multiple structural breaks formulated in (2.1) and (2.2) may be caused by one or a combination of cases (i)–(iii) described in Section 2.1, we next show that multiple structural breaks in the factor loadings and factor number can be transformed to breaks in the factor covariance structure.

PROPOSITION 2.1. *The approximate factor model (2.2) can be equivalently written as a transformed factor model like*

$$\mathbf{X}_t = \boldsymbol{\Lambda}^\star \mathbf{F}_t^\star + \boldsymbol{\epsilon}_t, \ \ t = 1, \cdots, n, \tag{2.4}$$

where $\boldsymbol{\Lambda}^\star$ denotes the transformed factor loading matrix which is time invariant, and $\mathbf{F}_t^\star$ denotes the transformed factors. In addition, the number of factors in the original factor model and that in the transformed factor model, satisfy the following inequalities,

$$\max_{1 \leqslant k \leqslant K_1+1} \underline{r}_k \leqslant q_0 \leqslant \sum_{k=1}^{K_1+1} \underline{r}_k \leqslant \sum_{k=1}^{K_1+1} r_k, \tag{2.5}$$

where q_0 is the number of transformed factors, $\underline{r}_k$ is the column rank of $\boldsymbol{\Lambda}_k^0$ and r_k is the number of original factors $\mathbf{F}_{t,k}$ when $\eta_{k-1}^c + 1 \leqslant t \leqslant \eta_k^c$.

REMARK 2.1. A representation similar to Proposition 2.1 is also given in Han and Inoue (2015) and Baltagi, Kao and Wang (2017, 2021). Note that construction of $\boldsymbol{\Lambda}^\star$ in the transformation mechanism is not unique due to the factor model identification issue. When each of $\boldsymbol{\Lambda}_k^0$ is of full column rank, we have $\underline{r}_k = r_k$, and consequently the inequalities in (2.5) would be simplified to

$$\max_{1 \leqslant k \leqslant K_1+1} r_k \leqslant q_0 \leqslant \sum_{k=1}^{K_1+1} r_k.$$

The upper bound for q_0 can be achieved if $\boldsymbol{\Lambda}_k^0$, $k = 1, \cdots, K_1 + 1$, are linearly independent. Appendix B in the supplement gives a simple motivating example for the above transformation. Proposition 3.1 in Section 3.1 below will further explore the limiting behavior of $\frac{1}{n} \sum_{t=1}^n \mathbf{F}_t^\star \mathbf{F}_t^{\star\top}$ as $n \to \infty$, which is crucial to application of the classic PCA method to estimate the transformed factors and factor loadings. Through the above transformation, it is sufficient to construct the CUSUM statistic using the estimated transformed factors for all of cases (i) –(iii) discussed in Section 2.1.

As the common factors are latent, we have to obtain their estimates (subject to appropriate

rotation) in practice. By Proposition 2.1, we consider the transformed factor model like (2.4) with time-invariant factor loadings, and apply the PCA estimation technique. Let $\mathcal{X}_n = (\mathbf{X}_1, \cdots, \mathbf{X}_n)^\intercal$ be an $n \times d$ matrix of observations. For the time being, we assume that the number of transformed factors q_0 is known, and will discuss how to determine it later in this subsection. The estimated factors $\widehat{\mathbf{F}}_t$ as well as the estimated idiosyncratic errors $\widehat{\boldsymbol{\epsilon}}_t$ will be used to construct the CUSUM statistics in Sections 2.3 and 2.4 below. The algorithm for PCA estimation is given as follows.

Algorithm 1 Standard PCA

Input: $\mathcal{X}_n, q_0$

1. Let $\widehat{\mathcal{F}}_n = \left(\widehat{\mathbf{F}}_1, \cdots, \widehat{\mathbf{F}}_n \right)^\intercal$ be the $n \times q_0$ matrix consisting of the q_0 eigenvectors (multiplied by $\sqrt{n}$) corresponding to the q_0 largest eigenvalues of $\mathcal{X}_n \mathcal{X}_n^\intercal / (nd)$.

2. The transformed factor loading matrix is estimated as

$$\widehat{\boldsymbol{\Lambda}}_n = \mathcal{X}_n^\intercal \widehat{\mathcal{F}}_n / n =: \left(\widehat{\lambda}_1, \cdots, \widehat{\lambda}_d \right)^\intercal. \tag{2.6}$$

3. The idiosyncratic component $\boldsymbol{\epsilon}_t$ is approximated by

$$\widehat{\boldsymbol{\epsilon}}_t = \mathbf{X}_t - \widehat{\boldsymbol{\Lambda}}_n \widehat{\mathbf{F}}_t =: (\widehat{\epsilon}_{t1}, \cdots, \widehat{\epsilon}_{td})^\intercal \text{ with } \widehat{\epsilon}_{tj} = X_{tj} - \widehat{\lambda}_j^\intercal \widehat{\mathbf{F}}_t, \ 1 \leqslant j \leqslant d, \tag{2.7}$$

where X_{tj} is the j-th element of $\mathbf{X}_t$.

Output: $\widehat{\mathbf{F}}_t, \widehat{\boldsymbol{\epsilon}}_t$ for $t = 1, \cdots, n$

A key issue in the PCA estimation is to appropriately choose the number of latent factors. As the above PCA method focuses on the transformed factor model, our aim is to estimate the number of transformed factors rather than that of the original ones. In general, the number of transformed factors q_0 could be much larger than maximum of $\underline{r}_k$ (or r_k) over $k = 1 \cdots, K_1 + 1$. However, when both $(r_k, k = 1 \cdots, K_1 + 1)$ and K_1 are assumed to be fixed, from (2.5) in Proposition 2.1, we have the true number of transformed factors to be upper bounded by a finite positive integer denoted by $\overline{Q}$. Some existing criteria developed for stable factor models can be applied to the transformed factor model (2.4) to obtain a consistent estimate of q_0. In this paper, we use a commonly-used information criterion proposed by Bai and Ng (2002).

For any $1 \leqslant q \leqslant \overline{Q}$, we let $\widehat{\mathcal{F}}_n(q) = \left[\widehat{\mathbf{F}}_1(q), \cdots \widehat{\mathbf{F}}_n(q) \right]^\intercal$ be the estimated factors obtained in Algorithm 1 replacing q_0 by q. Define

$$V_n(q) \ = \ \min_{\boldsymbol{\Lambda}(q)} \frac{1}{nd} \sum_{j=1}^{d} \sum_{t=1}^{n} \left[X_{tj} - \lambda_j(q) \widehat{\mathbf{F}}_t(q) \right]^2$$

$$= \min_{\mathbf{\Lambda}(q)} \frac{1}{nd} \sum_{t=1}^{n} \left[\mathbf{X}_t - \mathbf{\Lambda}(q)\widehat{\mathbf{F}}_t(q) \right]^\top \left[\mathbf{X}_t - \mathbf{\Lambda}(q)\widehat{\mathbf{F}}_t(q) \right]$$

$$= \sum_{j=q+1}^{d} \mu_j \left(\mathcal{X}_n \mathcal{X}_n^\top/(nd) \right),$$

where $\mathbf{\Lambda}(q) = [\lambda_1(q), \cdots, \lambda_d(q)]^\top$ is a $d \times q$ factor loading matrix and $\mu_j(\cdot)$ denotes the j-th largest eigenvalue. Consequently, we can choose the following objective function:

$$\mathsf{IC}(q) = \log \left[V_n(q) \right] + q \cdot \left(\frac{n+d}{nd} \right) \log(n \wedge d), \tag{2.8}$$

and obtain the estimate $\widehat{q}$ via

$$\widehat{q} = \underset{0 \leqslant q \leqslant \overline{Q}}{\arg \min}\, \mathsf{IC}(q), \tag{2.9}$$

where the common components disappear when $q = 0$. Alternative IC functions with different penalty terms in (2.8) can be found in Bai and Ng (2002) and in Alessi, Barigozzi and Capasso (2010) for the case of large idiosyncratic disturbances. With the asymptotic results given in Theorem 2 of Bai and Ng (2002), we may show that $\widehat{q}$ is a consistent estimator of the true number q_0, indicating that $\widehat{q} = q_0$ whose probability converges to one as the sample size increases.

Although accurate estimation of the factor number may be difficult in practice, slight over-estimation of the (transformed) factor number may be not risky in break detection (c.f., Barigozzi, Cho and Fryzlewicz, 2018). In Appendix F of the supplement, we provide a simulation to show that over-estimating q_0 would not negatively affect accuracy of multiple break detection. It is possible to avoid precisely estimating the factor number by following Barigozzi, Cho and Fryzlewicz (2018)'s method, multiplying the estimated factors and factor loadings to obtain the estimated common components. However, this would lead to a high-dimensional break detection for the common components.

2.3   Break detection in the common components

We start with the Binary Segmentation for Covariance (BS-Cov) technique to detect breaks in the common components and then introduce the WBS-Cov algorithm. As breaks in the factor loadings and factor number can be transformed to those in the covariance of transformed factors after re-formulating the approximate factor model to the one with time-invariant factor loadings, we

next define a CUSUM statistic using the estimated factors $\widehat{F}_t$. For $1 \leqslant l \leqslant s < u \leqslant n$, define

$$\mathbf{C}_{l,u}^{\widehat{F}}(s) = \sqrt{\frac{(s-l+1)(u-s)}{u-l+1}}\left[\frac{1}{s-l+1}\sum_{t=l}^{s}\mathsf{vech}\left(\widehat{F}_t\widehat{F}_t^\intercal\right) - \frac{1}{u-s}\sum_{t=s+1}^{u}\mathsf{vech}\left(\widehat{F}_t\widehat{F}_t^\intercal\right)\right], \quad (2.10)$$

which is a column vector with dimension $q_0(q_0+1)/2$. Maximising $\left\|\mathbf{C}_{l,u}^{\widehat{F}}(s)\right\|_2$ with respect to s, we obtain

$$\widetilde{\eta}_1^c = \arg\max_{l \leqslant s < u}\left\|\mathbf{C}_{l,u}^{\widehat{F}}(s)\right\|_2, \quad (2.11)$$

the first candidate break point (which is not necessarily the estimate of the first break time point η_1^c). If the quantity $\left\|\mathbf{C}_{l,u}^{\widehat{F}}(\widetilde{\eta}_1^c)\right\|_2$ exceeds certain threshold, say ξ_n^c, we split the closed interval $[l,u]$ into two sub-intervals: $[l,\widetilde{\eta}_1^c]$ and $[\widetilde{\eta}_1^c+1,u]$. Then, letting $(l,u) = (l,\widetilde{\eta}_1^c)$ or $(l,u) = (\widetilde{\eta}_1^c+1,u)$, we compute $\mathbf{C}_{l,u}^{\widehat{F}}(s)$ with s in either of the two sub-intervals, maximise $\left\|\mathbf{C}_{l,u}^{\widehat{F}}(s)\right\|_2$ with respect to s, estimate the next candidate break points, and examine whether a further split is needed. We continue this until no more split is needed.

REMARK 2.2. (i) It may be possible to extend (2.11) to a weighted version (c.f., Hariz, Wylie and Zhang, 2007; Cho, 2016; Yu and Chen, 2021)

$$\widetilde{\eta}_1^c = \arg\max_{l \leqslant s < u}\left\{\left[\frac{(s-l+1)(u-s)}{(u-l+1)^2}\right]^\upsilon \left\|\mathbf{C}_{l,u}^{\widehat{F}}(s)\right\|_2\right\},$$

with $-1/2 < \upsilon \leqslant 1/2$. A smaller υ may help detect breaks that are close to boundary. For simplicity, we only consider the unweighted version, i.e., $\upsilon = 0$, otherwise υ would affect the asymptotic analysis.

(ii) In the above BS-Cov, we use the l_2-type aggregation of the CUSUM quantities. Alternatively, we may use different norms in the aggregation such as the l_1-norm and l_∞-norm. It may be also possible to consider a weighted aggregation and seek an optimal projection direction as recommended by Wang and Samworth (2018). However, it is not clear how to implement the optimal direction method in a computationally efficient way for break detection in our model setting, so we do not pursue it in the present paper. In Appendix F of the supplemental document, we compare numerical performance among different types of aggregation on the CUSUM quantities and find that the l_2-type aggregation produces the most accurate detection result.

(iii) The l_2-aggregation in (2.11) ignores possible correlation structure between elements of the vector CUSUM statistic. Such a simplification would not affect estimation consistency of the break number and locations. For break detection in fixed-dimensional covariance matrices, Aue *et al.* (2009) propose a weighted quadratic form of the CUSUM vector using an estimated long-run covariance matrix. This refinement in the l_2-aggregation may improve break estimation efficiency.

However, consistent estimation of the long-run covariance structure for $\mathsf{vech}\left(\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t^\mathsf{T}\right)$ is difficult to obtain when there are unknown multiple breaks.

In order to accurately estimate the break time and delete possible spurious spikes in the CUSUM statistics, instead of using BS-Cov, we next extend the WBS algorithm introduced by Fryzlewicz (2014) to detect the multiple structural breaks in the common components. Let $\widehat{\eta}_1^c, \cdots, \widehat{\eta}_{\widehat{K}_1}^c$ be the estimated break points (arranged in an increasing order) and $\widehat{K}_1$ be the estimate of the unknown break number K_1 obtained in the following WBS-Cov algorithm. A discussion on selection of the threshold ξ_n^c can be found in Section 4. The method to draw random intervals in the WBS-Cov algorithm is the same as that in Fryzlewicz (2014) (see also Algorithm 4 in Section 4). Fryzlewicz (2020) provides an alternative recursive method of drawing random intervals in the WBS algorithm. In general, the intervals drawn in Algorithm 2 do not have to be random but taken over a fixed grid.

Algorithm 2 WBS-Cov algorithm

Input: ξ_n^c, M_n^c, $\widehat{\mathbf{F}}_t$, $t = 1, \cdots, n$

   generate random intervals $[l_m, u_m]$ within $[1, n]$ for $m = 1, \cdots, M_n^c$

   set $\mathcal{W}^c = \{[1, n]\}$ and $\widehat{\mathcal{B}}^c = \emptyset$

   while $\mathcal{W}^c \neq \emptyset$ **do**

     take an element out of $\mathcal{W}^c$ and denote it as $[l, u]$

     $\mathcal{M}_{l,u}^c = \{m : [l_m, u_m] \subset [l, u]\}$

     $(m_0^c, s_0^c) = \underset{m \in \mathcal{M}_{l,u}^c, l_m \leqslant s < u_m}{\arg\max} \left\| \mathbf{C}_{l,u}^{\widehat{\mathbf{F}}}(s) \right\|_2$

     if $\left\| \mathbf{C}_{l_{m_0^c}, u_{m_0^c}}^{\widehat{\mathbf{F}}}(s_0^c) \right\|_2 > \xi_n^c$ **then**

       add s_0^c to $\widehat{\mathcal{B}}^c$

       add $[l, s_0^c]$ and $[s_0^c + 1, u]$ to $\mathcal{W}^c$

     end if

   end while

Output: $\widehat{\mathcal{B}}^c$

2.4 Break detection in the idiosyncratic components

We next turn to detection of multiple structural breaks in the covariance structure of the idiosyncratic vector, which is much more challenging as the dimension involved can be ultra large and $\boldsymbol{\epsilon}_t$ is unobservable. In the existing literature, to obtain reliable estimation of ultra-high dimensional covariance matrices (without breaks), it is often common to impose certain sparsity condition, limiting the number of non-zero entries. Consequently, the thresholding or other generalised

shrinkage methods (Bickel and Levina, 2008a; Rothman, Levina and Zhu, 2009; Cai and Liu, 2011) have been developed to estimate high-dimensional covariance matrices. A similar thresholding technique is also introduced by Cho and Fryzlewicz (2015) to detect multiple structural breaks in high-dimensional time series setting. We next generalise the latter to detect breaks in the large covariance structure of the idiosyncratic error vector, using the approximation of ϵ_t obtained in the PCA algorithm. Define the normalised CUSUM-type statistic using $\widehat{\epsilon}_{tj}$ defined in (2.7):

$$c_{l,u}^{\widehat{\epsilon},\widehat{\sigma}}(s;i,j) = \frac{1}{\widehat{\sigma}_{l,u}(i,j)} \sqrt{\frac{(s-l+1)(u-s)}{(u-l+1)}} \left(\frac{1}{s-l+1} \sum_{t=l}^{s} \widehat{\epsilon}_{ti}\widehat{\epsilon}_{tj} - \frac{1}{u-s} \sum_{t=s+1}^{u} \widehat{\epsilon}_{ti}\widehat{\epsilon}_{tj} \right) \quad (2.12)$$

for $1 \leqslant l \leqslant s < u \leqslant n$ and $1 \leqslant i, j \leqslant d$, where $\widehat{\sigma}_{l,u}(i,j)$ is a properly chosen scaling factor. The value of the scaling factor depends on both the time interval $[l, u]$ and the index pair (i, j). The scaling factors ensure that we can choose a common threshold for each index pair. Note that the scaling factor is not needed for the CUSUM statistic (2.10) to detect breaks in the common components since the estimated factors have been normalised. We choose $\widehat{\sigma}_{l,u}(i,j)$ as the differential error median absolute deviation using $\widehat{\epsilon}_{tj}$, whose detailed construction is given in Section 4. To derive the consistency result for break detection, we assume that the random scaling factors are uniformly bounded away from zero and infinity over $1 \leqslant i, j \leqslant d$ and $1 \leqslant l < u \leqslant n$ with probability tending to one (see Assumption 5 in Appendix A). As the dimension $d \to \infty$ and might be much larger than the sample size n (see Assumption 4(i)), a direct aggregation of the CUSUM statistics $c_{l,u}^{\widehat{\epsilon},\widehat{\sigma}}(s;i,j)$ over i and j might not perform well in detecting the breaks in particular when the sparsity condition is imposed. Hence, we compare $c_{l,u}^{\widehat{\epsilon},\widehat{\sigma}}(s;i,j)$ with a thresholding parameter, say ξ_n^e, and delete the index pair (i,j) when $\max_{l \leqslant t < u} \left| c_{l,u}^{\widehat{\epsilon},\widehat{\sigma}}(t;i,j) \right|$ is smaller than ξ_n^e. For $m \in \mathcal{M}_{l,u}^e$ (to be defined in Algorithm 3) such that $[l_m, u_m]$ is a random sub-interval of $[l, u]$, we define

$$C_{l_m,u_m}^{\widehat{\epsilon}}(s) = \sum_{i=1}^{d} \sum_{j=i}^{d} \left| c_{l_m,u_m}^{\widehat{\epsilon},\widehat{\sigma}}(s;i,j) \right|^2 \mathcal{I} \left(\max_{l \leqslant t < u} \left| c_{l,u}^{\widehat{\epsilon},\widehat{\sigma}}(t;i,j) \right| > \xi_n^e \right), \quad (2.13)$$

where $\mathcal{I}(\cdot)$ is an indicator function. It is easy to see that the truncation component in (2.13) is independent of the random sub-intervals $[l_m, u_m]$. Therefore, if $\max_{l \leqslant t < u} \left| c_{l,u}^{\widehat{\epsilon},\widehat{\sigma}}(t;i,j) \right| \leqslant \xi_n^e$, we have $C_{l_m,u_m}^{\widehat{\epsilon}}(s) = 0$, and the algorithm will no longer search for breaks within the interval $[l, u)$ for the index pair (i, j). The main difference between our method and that in Cho and Fryzlewicz (2015) is that they sparsify the classical Binary Segmentation to locate the break points and estimate the break number, whereas we combine the sparsified CUSUM quantity with the WBS-Cov algorithm introduced as in Section 2.3. Thus we call the proposed algorithm as the *Wild Sparsified Binary Segmentation for Covariance* or *WSBS-Cov*. The detailed algorithm is given as follows.

Note that the sparsified CUSUM statistic $C_{l_m,u_m}^{\widehat{\epsilon}}(s)$ cannot take any value between 0 and $(\xi_n^e)^2$

13

Algorithm 3 WSBS-Cov algorithm

Input: ξ_n^e, M_n^e, $\widehat{\epsilon}_t$, $t = 1, \cdots, n$

  generate random intervals $[l_m, u_m]$ within $[1, n]$ for $m = 1, \cdots, M_n^e$

  set $\mathcal{W}^e = \{[1, n]\}$ and $\widehat{\mathcal{B}}^e = \emptyset$

  while $\mathcal{W}^e \neq \emptyset$ **do**

    take an element out of $\mathcal{W}^e$ and denote it as $[l, u]$

    $\mathcal{M}_{l,u}^e = \{m : [l_m, u_m] \subset [l, u]\}$

    $(m_0^e, s_0^e) = \underset{m \in \mathcal{M}_{l,u}^e, l_m \leqslant s < u_m}{\arg\max} C_{l_m, u_m}^{\widehat{\epsilon}}(s)$

    if $C_{l_{m_0^e}, u_{m_0^e}}^{\widehat{\epsilon}}(s_0^e) > 0$ **then**

      add s_0^e to $\widehat{\mathcal{B}}^e$

      add $[l, s_0^e]$ and $[s_0^e + 1, u]$ to $\mathcal{W}^e$

    end if

  end while

Output: $\widehat{\mathcal{B}}^e$

---

and

$$C_{l_{m_0^e}, u_{m_0^e}}^{\widehat{\epsilon}}(s_0^e) > 0 \Leftrightarrow C_{l_{m_0^e}, u_{m_0^e}}^{\widehat{\epsilon}}(s_0^e) > (\xi_n^e)^2.$$

Hence the thresholding parameter ξ_n^e in WSBS-Cov plays a similar role to ξ_n^c in WBS-Cov. With the WSBS-Cov algorithm, we obtain the estimated break points denoted by $\widehat{\eta}_1^e, \cdots, \widehat{\eta}_{\widehat{K}_2}^e$ (arranged in an increasing order), where $\widehat{K}_2$ is the estimate of the break number K_2. Finally, the sets of break points $\mathcal{B}^c$ and $\mathcal{B}^e$ defined in (2.3) are estimated by

$$\widehat{\mathcal{B}}^c = \left\{ \widehat{\eta}_1^c, \cdots, \widehat{\eta}_{\widehat{K}_1}^c \right\} \quad \text{and} \quad \widehat{\mathcal{B}}^e = \left\{ \widehat{\eta}_1^e, \cdots, \widehat{\eta}_{\widehat{K}_2}^e \right\}.$$

3 Large-sample theory

In this section we establish the large-sample theory for the methods proposed in Section 2 under some technical assumptions listed in Appendix A. The asymptotic properties for the WBS-Cov and WSBS-Cov methods are given in Sections 3.1 and 3.2, respectively.

3.1 Asymptotic theory of WBS-Cov for the common components

We start with the following proposition on some fundamental properties for the transformed factors $\mathbf{F}_t^\star$ and factor loadings $\mathbf{\Lambda}^\star$ in (2.4).

PROPOSITION 3.1. *Suppose that Assumption 2 in Appendix A is satisfied and let $\kappa_n^c = \min_{1 \leqslant k \leqslant K_1+1}(\eta_k^c -$*

η_{k-1}^c). *The transformed factor loading matrix $\mathbf{\Lambda}^\star$ is of full column rank and the transformed factors $\mathbf{F}_t^\star$ satisfy that $\left\|\sum_{t=1}^n \mathbf{F}_t^\star \mathbf{F}_t^{\star\mathsf{T}}\right\|_F = O_P(n)$ when $\kappa_n^c \to \infty$. If, in addition, $\kappa_n^c \asymp n$, there exists a $q_0 \times q_0$ positive definite matrix $\mathbf{\Sigma}_\mathbf{F}$ such that*

$$\frac{1}{n} \sum_{t=1}^n \mathbf{F}_t^\star \mathbf{F}_t^{\star\mathsf{T}} \xrightarrow{P} \mathbf{\Sigma}_\mathbf{F} \quad \text{as } n \to \infty. \tag{3.1}$$

The above proposition shows that the sample second-order moment of the transformed factors converges in probability under some regularity conditions in Assumption 2. The convergence result (3.1) is sometimes given directly as a high-level condition in the literature (e.g., Assumption A1 in Ma and Su, 2018). The restriction $\kappa_n^c \asymp n$ is crucial to ensure positive definiteness of $\mathbf{\Sigma}_\mathbf{F}$. If the order of the minimum distance between breaks is smaller than n, the limit matrix $\mathbf{\Sigma}_\mathbf{F}$ may be singular.

We next study the asymptotic theory for the WBS-Cov algorithm in Section 2.3 to detect breaks in the common components. Define an infeasible CUSUM statistic using the latent transformed factors:

$$\mathbf{C}_{l,u}^{\mathbf{HF}^\star}(s) = \sqrt{\frac{(s-l+1)(u-s)}{u-l+1}} \left[\frac{1}{s-l+1} \sum_{t=l}^s \mathsf{vech}\left(\mathbf{HF}_t^\star \mathbf{F}_t^{\star\mathsf{T}} \mathbf{H}^\mathsf{T}\right) - \frac{1}{u-s} \sum_{t=s+1}^u \mathsf{vech}\left(\mathbf{HF}_t^\star \mathbf{F}_t^{\star\mathsf{T}} \mathbf{H}^\mathsf{T}\right) \right], \tag{3.2}$$

and $\mathbf{H}$ is a $q_0 \times q_0$ rotation matrix defined by

$$\mathbf{H} = \mathbf{\Omega}_{q_0}^{-1} \left(\frac{1}{n} \sum_{t=1}^n \widehat{\mathbf{F}}_t \mathbf{F}_t^{\star\mathsf{T}} \right) \left(\frac{1}{d} \sum_{j=1}^d \lambda_j^\star \lambda_j^{\star\mathsf{T}} \right), \tag{3.3}$$

in which $\mathbf{\Omega}_{q_0}$ is a $q_0 \times q_0$ diagonal matrix with the diagonal elements being the first q_0 largest eigenvalues of $\mathcal{X}_n \mathcal{X}_n^\mathsf{T}/(nd)$ arranged in a descending order, and $\mathbf{\Lambda}^\star = (\lambda_1^\star, \cdots, \lambda_d^\star)^\mathsf{T}$ with $\lambda_j^\star$ being a q_0-dimensional vector of transformed factor loadings.

PROPOSITION 3.2. *Suppose that Assumptions 1–3 and 4(i) in Appendix A are satisfied. If $\kappa_n^c \asymp n$,*

$$\max_{(l,u):\, 1\leqslant l<u\leqslant n} \max_{s:\, l\leqslant s<u} \left\| \mathbf{C}_{l,u}^{\widehat{\mathbf{F}}}(s) - \mathbf{C}_{l,u}^{\mathbf{HF}^\star}(s) \right\|_2 = O_P(1). \tag{3.4}$$

Proposition 3.2 above plays a crucial role in our proofs, indicating that the CUSUM statistic $\mathbf{C}_{l,u}^{\widehat{\mathbf{F}}}(s)$ can be replaced by the infeasible CUSUM statistic $\mathbf{C}_{l,u}^{\mathbf{HF}^\star}(s)$ in the asymptotic analysis. The following theorem establishes the convergence result for the estimates of break points and break number in the covariance structure of the common components.

THEOREM 3.1. *Suppose that Assumptions 1–3 and 4(i)(ii) in Appendix A are satisfied, and there exist two positive constants $\underline{c}_1$ and $\bar{c}_1$ such that the threshold ξ_n^c satisfies*

$$\underline{c}_1 \log^2 n \leqslant \xi_n^c \leqslant \bar{c}_1 \left(\kappa_n^c \underline{\omega}_n^c \right)^{1/2}, \tag{3.5}$$

where κ_n^c is defined in Proposition 3.1 and $\underline{\omega}_n^c$ denotes the minimum break size (in the common components) defined in Assumption 4(ii). Then there exists a positive constant ι^c such that

$$\mathsf{P}\left(\widehat{K}_1 = K_1; \max_{1 \leqslant k \leqslant K_1} |\widehat{\eta}_k^c - \eta_k^c| < \iota^c \varphi_n^c \right) \to 1, \tag{3.6}$$

where $\varphi_n^c = \log^4 n / \underline{\omega}_n^c$.

REMARK 3.1. The convergence result (3.6) shows that the proposed estimator of the break points in the common components via WBS-Cov has the approximation rate of $\varphi_n^c = \log^4 n / \underline{\omega}_n^c$, which can be further simplified to the rate of $\log^4 n$ if the minimum break size $\underline{\omega}_n^c$ is larger than a small positive constant. Note that, even if the factors are observable, the optimal estimation rate of the break points in their covariance structure is $O_P(1)$ (Korostelev, 1988; Aue *et al.*, 2009) when the binary segmentation is used in break detection. Hence, our approximation rate in Theorem 3.1 is nearly optimal up to a fourth-order logarithmic rate. Since we remove the Gaussian assumption on the observations and further allow temporal dependence in the data over time, our approximation rate is slightly slower than that in Theorem 3.2 of Fryzlewicz (2014).

3.2 Asymptotic theory of the WSBS-Cov for the idiosyncratic components

We next present the asymptotic theory of structural break estimation in the idiosyncratic error components. Similarly to that in Section 3.1, we define the following infeasible CUSUM-type statistics using the unobservable ϵ_{tj},

$$c_{l,u}^{\epsilon,\widehat{\sigma}}(s; i, j) = \frac{1}{\widehat{\sigma}_{l,u}(i,j)} \sqrt{\frac{(s-l+1)(u-s)}{(u-l+1)}} \left(\frac{1}{s-l+1} \sum_{t=l}^{s} \epsilon_{ti}\epsilon_{tj} - \frac{1}{u-s} \sum_{t=s+1}^{u} \epsilon_{ti}\epsilon_{tj} \right) \tag{3.7}$$

for $1 \leqslant l \leqslant s < u \leqslant n$ and $1 \leqslant i, j \leqslant d$, where $\widehat{\sigma}_{l,u}(i,j)$ is defined as in (2.12), and

$$C_{l_m,u_m}^{\epsilon}(s) = \sum_{i=1}^{d} \sum_{j=i}^{d} \left| c_{l_m,u_m}^{\epsilon,\widehat{\sigma}}(s; i, j) \right|^2 \mathcal{I} \left(\max_{l \leqslant t < u} \left| c_{l,u}^{\epsilon,\widehat{\sigma}}(t; i, j) \right| > \xi_n^e \right) \tag{3.8}$$

for $m \in \mathcal{M}_{l,u}^e$.

PROPOSITION 3.3. *Suppose that Assumptions 1–3, 4(i) and 5 in Appendix A are satisfied. If $\kappa_n^c \asymp n$,*

$$\max_{(i,j):\, 1\leqslant i,j\leqslant d} \max_{(l,u):\, 1\leqslant l<u\leqslant n} \max_{s:\, l\leqslant s<u} \left| c_{l,u}^{\widehat{\epsilon},\widehat{\sigma}}(s;i,j) - c_{l,u}^{\epsilon,\widehat{\sigma}}(s;i,j) \right| = O_P\left(\sqrt{(\log d)(\log n)} \right). \qquad (3.9)$$

Proposition 3.3 above plays a crucial role in our asymptotic analysis, indicating that the CUSUM statistics $c_{l,u}^{\widehat{\epsilon},\widehat{\sigma}}(s;i,j)$ can be replaced by the infeasible ones $c_{l,u}^{\epsilon,\widehat{\sigma}}(s;i,j)$ which are much easier to handle in the proofs. The following theorem establishes the convergence result for the estimates of the break points and break number in the covariance structure of the idiosyncratic components.

THEOREM 3.2. *Suppose that Assumptions 1–3, 4(i)(iii) and 5 in Appendix A are satisfied, $\kappa_n^c \asymp n$, and there exist two positive constants $\underline{c}_2$ and $\bar{c}_2$ such that*

$$\underline{c}_2 \log^2(dn) \leqslant \xi_n^e \leqslant \bar{c}_2 (\kappa_n^e/n) \cdot (\kappa_n^e \underline{\omega}_n^e)^{1/2}, \qquad (3.10)$$

where $\kappa_n^e = \min_{1\leqslant k\leqslant K_2+1}(\eta_k^e - \eta_{k-1}^e)$ and $\underline{\omega}_n^e$ denotes the minimum break size (in the idiosyncratic components) defined in Assumption 4(iii). Then there exists a positive constant ι^e such that

$$P\left(\widehat{K}_2 = K_2;\, \max_{1\leqslant k\leqslant K_2} |\widehat{\eta}_k^e - \eta_k^e| < \iota^e \varphi_{n,d}^e \right) \to 1, \qquad (3.11)$$

where $\varphi_{n,d}^e = (n/\kappa_n^e)^2 \cdot \left[\log^4(nd)/\underline{\omega}_n^e \right]$.

REMARK 3.2. The rate in (3.11) relies on both n and d, making it substantially different from the convergence result in Theorem 3.1. If the dimension diverges to infinity at a polynomial rate of n (c.f., Barigozzi, Cho and Fryzlewicz, 2018) and assume $\kappa_n^e \asymp n$, the rate $\varphi_{n,d}^e$ becomes $\log^4 n/\underline{\omega}_n^e$, slightly faster than that obtained in Theorem 3 of Barigozzi, Cho and Fryzlewicz (2018) when the breaks are sparse in covariance of the idiosyncratic errors.

4   Practical issues in the detection procedure

In this section, we first extend the so-called *Strengthened Schwarz Information Criterion (SSIC)* to our model setting to determine the number of breaks, and then discuss the choice of the random scaling quantity $\widehat{\sigma}_{l,u}(i,j)$ in (2.12) and the thresholding parameter ξ_n^e in the WSBS-Cov.

The SSIC is proposed by Fryzlewicz (2014) in the context of univariate WBS, and some substantial modifications are needed to make it applicable to our setting. For the break detection in the covariance structure of the common components, when the algorithm proceeds, we estimate the k candidate break points (arranged in an increasing order) denoted by $\widehat{\eta}_{1|k}^c, \cdots, \widehat{\eta}_{k|k}^c$ with the

convention of $\widehat{\eta}_{0|k}^c = 0$ and $\widehat{\eta}_{k+1|k}^c = n$. Letting

$$Z_t^{\widehat{F}} = \text{vech}\left(\widehat{F}_t \widehat{F}_t^{\top}\right) = \left(Z_{t,1}^{\widehat{F}}, \cdots, Z_{t,q_0(q_0+1)/2}^{\widehat{F}}\right)^{\top} \text{ and } Z_{\bullet,j}^{\widehat{F}} = \left(Z_{1,j}^{\widehat{F}}, \cdots, Z_{n,j}^{\widehat{F}}\right)^{\top}$$

for $j = 1, \cdots, q_0(q_0 + 1)/2$, we define the SSIC objective function for the univariate process $Z_{\bullet,j}^{\widehat{F}}$ by

$$\text{SSIC}_j^c(k) = \frac{n}{2} \log \widehat{\sigma}_j^2(k) + kp(n), \quad k = 0, \cdots, \bar{K}, \tag{4.1}$$

where

$$\widehat{\sigma}_j^2(k) = \frac{1}{n} \sum_{t=1}^n \left[Z_{t,j}^{\widehat{F}} - \bar{Z}_{t,j}^{\widehat{F}}(k)\right]^2 \text{ with } \bar{Z}_{t,j}^{\widehat{F}}(k) = \frac{1}{\widehat{\eta}_{i+1|k}^c - \widehat{\eta}_{i|k}^c} \sum_{s=\widehat{\eta}_{i|k}^c+1}^{\widehat{\eta}_{i+1|k}^c} Z_{s,j}^{\widehat{F}}$$

for $\widehat{\eta}_{i|k}^c + 1 \leqslant t \leqslant \widehat{\eta}_{i+1|k}^c$, $i = 0, \cdots, k$, $p(n) = n^{1/2}$, and $\bar{K}$ is a positive constant which is a pre-specified upper bound for the break number. If there exists a positive integer K^c such that

$$\text{SSIC}_j^c(K^c + 1) > \text{SSIC}_j^c(K^c) \ \ \forall j = 1, \cdots, q_0(q_0 + 1)/2, \tag{4.2}$$

no further break exists in any dimension of the transformed factors. Let $\widehat{K}_1$ be the smallest number such that (4.2) is satisfied, and stop the WBS-Cov algorithm in Section 2.3 when the number of breaks reaches $\widehat{K}_1$. By using SSIC, we avoid choosing the tuning parameter ξ_n^c in the WBS-Cov.

Similarly, we can also use the SSIC to obtain the estimated break number for the idiosyncratic error components. Let $\text{SSIC}_{i,j}^e(k)$ be defined similarly to $\text{SSIC}_j^c(k)$ but with $Z_{t,j}^{\widehat{F}}$ and $\bar{Z}_{t,j}^{\widehat{F}}(k)$ replaced by

$$Z_{t,ij}^{\widehat{\epsilon}} := \widehat{\epsilon}_{it} \widehat{\epsilon}_{jt} \text{ and } \bar{Z}_{t,ij}^{\widehat{\epsilon}}(k) = \frac{1}{\widehat{\eta}_{i+1|k}^e - \widehat{\eta}_{i|k}^e} \sum_{s=\widehat{\eta}_{i|k}^e+1}^{\widehat{\eta}_{i+1|k}^e} Z_{s,ij}^{\widehat{\epsilon}} \tag{4.3}$$

for $\widehat{\eta}_{i|k}^e + 1 \leqslant t \leqslant \widehat{\eta}_{i+1|k}^e$, $i = 0, \cdots, k$. Let $\widehat{K}_2$ be the smallest number K^e such that

$$\text{SSIC}_{i,j}^e(K^e + 1) > \text{SSIC}_{i,j}^e(K^e) \tag{4.4}$$

for all $1 \leqslant i,j \leqslant d$, and thus obtain the estimated break number for the idiosyncratic error component. In the construction of the CUSUM statistic in the WSBS-Cov, we need to define the random scaling quantity $\widehat{\sigma}_{l,u}(i,j)$. In the numerical studies, we let $\widehat{\sigma}_{l,u}(i,j)$ be the differential error median absolute deviation which is defined by

$$\widehat{\sigma}_{l,u}(i,j) = \text{med}_{l \leqslant t < u} \left(|(\widehat{\epsilon}_{t+1,i} \widehat{\epsilon}_{t+1,j} - \widehat{\epsilon}_{ti} \widehat{\epsilon}_{tj}) - \text{med}_{l \leqslant t < u} (\widehat{\epsilon}_{t+1,i} \widehat{\epsilon}_{t+1,j} - \widehat{\epsilon}_{ti} \widehat{\epsilon}_{tj})|\right)$$

in which $\text{med}_{l \leqslant t < u}(\cdot)$ denotes the sample median over the time interval $[l, u)$ (e.g., Fryzlewicz,

2014). An advantage of using the median absolute deviation is that it provides robust scale estimation.

When implementing the WBS-Cov and WSBS-Cov algorithms, we have to guarantee sufficient length of the random intervals to facilitate break detection. When l and u are too close, $\widehat{\sigma}_{l,u}(i,j)$ may be unbounded, leading to violation of Assumption 5 in Appendix A. Consequently, the CUSUM statistics on those small intervals would be smaller than the thresholding parameter so that no breaks could be detected. Therefore, we introduce a quantity Δ_n as in Barigozzi, Cho and Fryzlewicz (2018), to control the minimum length of the random intervals. Specifically, Algorithm 4 is used to produce a set of random intervals in WBS-Cov or WSBS-Cov.

Algorithm 4 Drawing random intervals

Input: n, Δ_n and M_n ($M_n = M_n^c$ or M_n^e)
    Draw M_n pairs of random points uniformly from the set $\{1, 2, \cdots, n - 4\Delta_n\}$;
    For each pair, let l_m be the smaller random point, and u_m be the larger one plus $4\Delta_n$.
Output: random intervals $[l_m, u_m]$ for $m = 1, \cdots, M_n$

We set $M_n = M_n^c = M_n^e = 400$ in the numerical studies. By Algorithm 4, the random intervals are produced with length of at least $4\Delta_n$. In addition, we also trim each random interval to keep a distance of Δ_n to both boundaries, indicating that we only maximise the CUSUM statistics over the interval $[l_m + \Delta_n, u_m - \Delta_n]$. Note that Δ_n should be much smaller than κ_n^c and κ_n^e. Thus we choose $\Delta_n = \lfloor (\log^2 n) \wedge (0.25 n^{6/7}) \rfloor$ following the discussion in Barigozzi, Cho and Fryzlewicz (2018).

A crucial issue in constructing the sparsified CUSUM statistic in the WSBS-Cov algorithm is the choice of the thresholding parameter ξ_n^e. Cho and Fryzlewicz (2015) suggest choosing this parameter using the CUSUM statistic based on a stationary process (e.g., a simulated AR(1) process) under the null hypothesis of no breaks. However, in practical applications, for the high-dimensional time series data with latent breaks, it is often difficult to recover the underlying stationary structure. To address this problem, we propose an alternative approach to choose ξ_n^e: (i) pre-detect the breaks (e.g., combining the classical BS-Cov and SSIC without knowing ξ_n^e a priori) and remove the breaks from the high-dimensional covariance matrices through demeaning using the estimated breaks, i.e., generate $Z_{t,ij}^{\widehat{\varepsilon}} - \bar{Z}_{t,ij}^{\widehat{\varepsilon}}(\widetilde{K}_n^e)$ as in (4.3), where $\widetilde{K}_n^e$ is the preliminary estimate of the break number obtained in the pre-detection of the idiosyncratic components; (ii) calculate the CUSUM statistics of the "stationary" process $\left\{ Z_{t,ij}^{\widehat{\varepsilon}} - \bar{Z}_{t,ij}^{\widehat{\varepsilon}}(\widetilde{K}_n^e) \right\}$ for each pair (i,j) and then take the maximum as the chosen thresholding parameter ξ_n^e. This technique is more computationally intensive than any formula-based threshold selection procedures; however, we have found that it performs particularly well in the numerical studies reported in Section 5.

5 Monte-Carlo simulation

In this section, we provide a simulation study to compare the finite-sample performance between the proposed methods and the method proposed by Barigozzi, Cho and Fryzlewicz (2018), and examine the performance of our methods in the weak factor structure. Additional simulation studies are given in Appendix F of the supplement, where the proposed methods are compared with various other competing methods that we designed. Each of these methods differs from the method presented in our main text in only one step. Like ablation studies that are commonly conducted in deep learning research, they help to gain a better understanding of our methods. Since these alternative methods are not designed for practical use, nor are they theoretically justified, we keep them in the supplement.

We consider the following factor model to generate data:

$$X_{ti} = \sum_{j=1}^{r} \lambda_{ij,t} F_{tj} + \sqrt{\theta} \epsilon_{ti}, \quad i = 1, \cdots, d, \quad t = 1, \cdots, n, \tag{5.1}$$

where $\theta > 0$ is used to control the signal-to-noise ratio. The replication number in each simulation cases is set to be $R = 100$. For the 100 simulated samples, we report the estimated number of break(s) as well as the accuracy measure for each break location η_k defined by

$$\mathsf{ACU}_k = 100\% \times \left[\frac{1}{100} \cdot \sum_{i=1}^{100} \mathcal{I} \left(\min_{l=1,\cdots,\widehat{K}(i)} |\widehat{\eta}_l(i) - \eta_k| \leqslant \log n \right) \right], \tag{5.2}$$

for $k = 1, \cdots, K$, where K denotes the break number K_1 or K_2, $\widehat{K}(i)$ is the corresponding break number estimate in the i-th simulated sample, η_k denotes either η_k^c or η_k^e, and $\widehat{\eta}_l(i)$, $l = 1, \cdots, \widehat{K}(i)$, are the break point estimates in the i-th simulated sample,

EXAMPLE 5.1. We use model (5.1) to generate the data in simulation, where $\theta = 0.5$, the number of factors is $r = 5$, the sample size is $n = 400$, and the dimension is $d = 200$. The factor process $F_t = (F_{t1}, \cdots, F_{t5})^\top$ is generated from a multivariate normal distribution $N_5 (0, \Sigma_F^*)$ independently over t, where Σ_F^* is a covariance matrix with one break specified as follows: for $1 \leqslant t \leqslant \eta_1^c = 133$, ϕ_j^F, the square root of the j-th diagonal element of Σ_F^*, is independently generated from a uniform distribution $U(0.5, 1.5)$, and the (i, j)-entry of Σ_F^* is defined as $\phi_i^F \phi_j^F (0.5)^{|i-j|}$ for $1 \leqslant i \neq j \leqslant 5$; for $\eta_1^c < t \leqslant n$, the (1,2) and (2,1)-entries of Σ_F^* change from $0.5\phi_1^F \phi_2^F$ to $0.9\phi_1^F \phi_2^F$, and ϕ_5^F is replaced by $1.3\phi_5^F$ resulting in structural breaks in $(5, j)$ and $(j, 5)$-entries of Σ_F^*. For $1 \leqslant t \leqslant \eta_2^c = 267$, the factor loadings λ_{ij} are independently generated from a uniform distribution $U(-1, 1)$; whereas for $\eta_2^c < t \leqslant n$, the factor loadings corresponding to the first two factors are regenerated by a new uniform distribution $U(-1, 1)$. Therefore, there are two structural breaks in the covariance

structure of the common components covering cases (i) and (ii) discussed in Section 2.1.

The idiosyncratic errors $\boldsymbol{\epsilon}_t$ follow a multivariate normal distribution $N_d\,(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$ independently over t, where ϕ_j, the square root of the j-th diagonal element of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$, is generated from an independent uniform distribution $U(0.5, 1.5)$, and the (i,j)-entry of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is $\phi_i\phi_j(-0.5)^{|i-j|}$ for $1 \leqslant i \neq j \leqslant d$. There are three breaks, $\eta_1^e = \lfloor n/4 \rfloor = 100$, $\eta_2^e = \lfloor n/2 \rfloor = 200$ and $\eta_3^e = \lfloor 3n/4 \rfloor = 300$, in the idiosyncratic error components. At each of the break points, we swap the orders of $\lfloor \rho_1^e d/2 \rfloor$ randomly selected pairs of elements of $\boldsymbol{\epsilon}_t$ with ρ_1^e chosen as 1, 0.5 and 0.1. Note that $\rho_1^e = 0.1$ indicates that the structural breaks are sparse in the high-dimensional error components, whereas $\rho_1^e = 1$ indicates that the breaks are dense.

Table 1: Comparison of break detection results

		Common components					Idiosyncratic error components					
	Methods	# break(%)				ACU_1(%)	ACU_2(%)	# break(%)			ACU_1(%)	ACU_2(%)	ACU_3(%)
		0	1	2	> 2	$\eta_1^c = 133$	$\eta_2^c = 267$	< 3	3	> 3	$\eta_1^e = 100$	$\eta_2^e = 200$	$\eta_3^e = 300$
$\rho_1^e = 1$	W(S)BS-Cov	0	2	98	0	65	93	0	100	0	100	100	100
	BCF	12	86	2	0	0	78	0	100	0	100	100	100
$\rho_1^e = 0.5$	W(S)BS-Cov	0	2	98	0	64	95	5	95	0	100	96	99
	BCF	12	87	1	0	0	81	26	74	0	86	95	83
$\rho_1^e = 0.1$	W(S)BS-Cov	0	3	97	0	64	95	67	33	0	73	58	71
	BCF	11	88	1	0	0	81	100	0	0	3	14	1

Table 1 reports the simulation result, where "W(S)BS-Cov" denotes our proposed methods and "BCF" denotes the method proposed by Barigozzi, Cho and Fryzlewicz (2018) which combines the wavelet-based transformation and the double-CUSUM method. We find that the proposed WBS-Cov method has better finite-sample performance in detecting breaks in the common components with more accurate estimated break number and the higher ACU, whereas "BCF" tends to neglect the first break and thus under-estimate the number of breaks. For the break detection in the idiosyncratic error components, when $\rho_1^e = 1$, the proposed WSBS-Cov method and "BCF" perform equally well. However, when $\rho_1^e = 0.5$ and 0.1, the WSBS-Cov method clearly outperforms "BCF".

We next consider an example with breaks in the weak factor structure. When factor loadings for the transformed factor model in (2.4) satisfies $\mu_k\left(\boldsymbol{\Lambda}^{\star\top}\boldsymbol{\Lambda}^{\star}\right) = o_P(d)$, we call the k-th factor as a weak factor. Although our asymptotic framework (see Assumption 2(ii)) rules out this scenario, we provide the following simulation to examine the effect of weak factors to the proposed break detection methods.

EXAMPLE 5.2. We use model (5.1) with a setting similar to Freyaldenhoven (2021) to generate the data, where the number of factors is $r = 9$, $\theta = 1.5$, the sample size is $n = 400$, and the dimension is $d = 200$. The factor process $\mathbf{F}_t$ is generated from a multivariate normal distribution $N_9\,(\mathbf{0}, \mathbf{I}_9)$ independently over t. The j-th factor only affects d_j elements of the d-dimensional observation vector, indicating that the factor loading matrix is sparse with d_j non-zero entries in its j-th column.

The position of non-zero entries are randomly selected, and their values are drawn from $\mathsf{N}(1,1)$ independently. We set $(d_1, \cdots, d_9) = \left(d, \lfloor d^{0.85} \rfloor, \lfloor d^{0.75} \rfloor, \lfloor d^{2/3} \rfloor, \lfloor d^{2/3} \rfloor, \lfloor d^{0.6} \rfloor, \lfloor d^{1/3} \rfloor, \lfloor d^{1/4} \rfloor, \lfloor \log(d) \rfloor \right)$.

We set two breaks for the common components: $\eta_1^c = 100$ and $\eta_2^c = 300$ and consider 9 cases in the simulation. Case j represents a set-up in which breaks only occur on the j-th column of the factor loading matrix. Specifically, independently for the three periods: $1 \leqslant t \leqslant 100$, $101 \leqslant t \leqslant 300$, and $301 \leqslant t \leqslant 400$, the position of non-zero entries in the j-th factor loading vector is randomly selected with its value independently regenerated. For the idiosyncratic error components, as in Freyaldenhoven (2021), we let $\epsilon_t = \boldsymbol{A}_\epsilon e_t$ and generate $e_t = (e_{t1}, \cdots, e_{td})^\intercal$ by

$$e_{ti} = \rho_e e_{t-1,i} + (1 - \rho_e^2)^{1/2} v_{ti}, \quad v_{ti} = \rho_v v_{t-1,i} + (1 - \rho_v^2)^{1/2} u_{e,ti},$$

where $u_{e,ti}$ are independently drawn from $\mathsf{N}(0,1)$, and $(\rho_e, \rho_v) = (0.3, 0.1)$. We set two break points on $\boldsymbol{A}_\epsilon$: $\eta_1^e = 150$ and $\eta_2^e = 250$. For $1 \leqslant t \leqslant 150$, we first generate a tridiagonal matrix $\widetilde{\boldsymbol{A}}_\epsilon$ with all diagonal entries being one and the other non-zero entries drawn from $\mathsf{U}(0.5, 1.5)$, and then standardise $\widetilde{\boldsymbol{A}}_\epsilon$ by letting each row multiply a constant so that the row sum of squares is one to define $\boldsymbol{A}_\epsilon$. The matrix $\boldsymbol{A}_\epsilon$ is randomly regenerated in the same way for $151 \leqslant t \leqslant 250$ and $251 \leqslant t \leqslant 400$.

As shown by Freyaldenhoven (2021), the classic PCA estimators are consistent only for the "relevant" factors whose corresponding d_j is of an order larger than $d^{1/2}$. Hence, the number of relevant factors is 6 in the simulated weak factor model. By the factor model transformation mechanism in Proposition 2.1, we expect that the number of factors should be 8 for cases 1–6 (6 relevant factors plus 2 extra factors due to factor transformation accommodating breaks), and 6 in cases 7–9. Table 2 reports the simulation results for the break detection. We find that the factor number is over-estimated in all cases. The break numbers and locations in both the common and idiosyncratic components are correctly estimated in general when breaks occur in the relevant factors which are relatively strong (cases 1–3). The break detection results are less accurate when there are breaks in loadings of relevant factors which are relatively weak (cases 4–6). In cases 7–9, since weak factors with breaks in the loadings are not selected as the relevant factors via the information criterion, the proposed WBS-Cov method cannot detect either of the two breaks in the common components.

6 An empirical application

In this section, we use the proposed methods to detect breaks in the contemporaneous covariance structure of the daily returns of S&P 500 constituents.

Table 2: Detection results with breaks on weak factors

	$\widehat{q}$	Common components						Idiosyncratic error components					
		# break(%)				$ACU_1(\%)$	$ACU_2(\%)$	# break(%)				$ACU_1(\%)$	$ACU_2(\%)$
		0	1	2	>2	$\eta_1^c = 100$	$\eta_2^c = 300$	0	1	2	>2	$\eta_1^e = 150$	$\eta_2^e = 250$
Case 1	8.94	0	0	100	0	85	85	0	1	98	1	98	100
Case 2	8.92	0	0	100	0	87	93	0	0	100	0	100	100
Case 3	8.93	0	0	100	0	81	87	0	1	99	0	100	99
Case 4	8.72	3	2	95	0	84	84	0	0	95	5	89	98
Case 5	8.65	1	4	95	0	83	83	0	1	93	6	88	96
Case 6	8.02	21	10	69	0	73	73	0	1	73	26	78	74
Case 7	6.60	100	0	0	0	0	0	0	2	86	12	70	73
Case 8	6.84	100	0	0	0	0	0	0	0	92	8	91	89
Case 9	6.49	100	0	0	0	0	0	0	0	82	18	67	75

6.1 Break detection in covariances

The data are retrieved from Refinitiv Datastream database (formerly Thomson Reuters Datastream), covering the time period from 01/01/2000 to 28/02/2021. We consider 375 firms listed on the S&P 500 index over the entire time period. Let X_{it} denote the demeaned percentage price change (without reinvesting of dividends) of equity i at time t, where $i = 1, \cdots, 375$ and $t = 1, \cdots, 5322$. We fit the factor model (2.2) to the data with $\mathbf{X}_t = (X_{1t}, \cdots, X_{375,t})^\intercal$, allowing structural breaks in the covariance matrices of both the common and idiosyncratic components.

The information criterion introduced in Section 2.2 selects nine factors (with appropriate transformation), i.e., $\widehat{q} = 9$. We set $\Delta_n = 20$ and $M_n^c = M_n^e = 1000$ in the WBS-Cov and WSBS-Cov algorithms. Combined with SSIC, the WBS-Cov algorithm detects five breaks in the covariance matrix of the common components, and the WSBS-Cov algorithm detects three breaks in the idiosyncratic components. We plot the estimated break times on the series of the S&P 500 index and its returns in Figure 1, where the red vertical lines denote breaks in the common components and the blue dotted vertical lines denote breaks in the idiosyncratic components. Among the five breaks in the common components, two occur during the period of a bearish market from 2000 to 2003, two occur in 2008 and 2009 during the global financial crisis, and one occurs recently due to the COVID-19 outbreak. Among the three breaks in the idiosyncratic errors, the first break occurs around the end of 2002, the second and third breaks occur in 2008 and 2009.

6.2 Calibration

We next provide a simulation using the factor model with calibrated parameters from the empirical study, and aim to further examine the proposed break detection methods in finite samples.

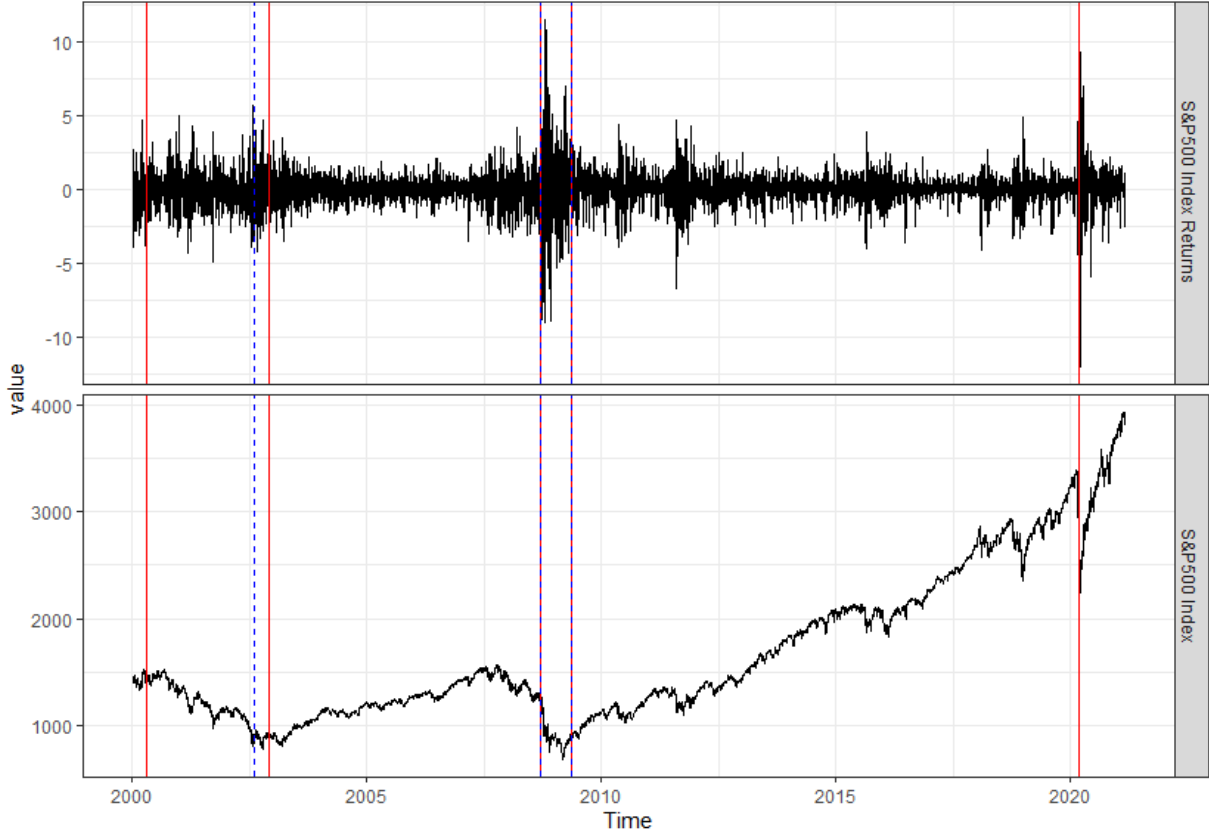- Generate a simulated dataset with $n = 5322$ and $d = 375$ as the real data.

Figure 1: The detected break times in the covariance structure of the 375 daily returns of the S&P 500 constituents from 01/01/2000 to 28/02/2021. The estimated break dates for common components are 24/04/2000, 25/11/2002, 12/09/2008, 08/05/2009 and 06/03/2020 (in red lines); and the estimated break dates for idiosyncratic components are 06/08/2002, 11/09/2008 and 11/05/2009 (in blue dotted lines).

- Consider six sub-periods separated by the five estimated break dates in the common components (see the red lines in Figure 1) and use the PCA estimated factor loadings. The factors are drawn from a standard normal distribution independently over time and across dimension.

- Divide the estimated idiosyncratic errors into four segments split by the three estimated breaks in idiosyncratic components (see the blue dotted lines in Figure 1), and separately estimate their sample covariance matrices. For each segment, generate idiosyncratic errors by a multivariate normal distribution independently over time with mean zero and covariance matrix being the sample covariance matrix.

The above generating mechanism guarantees the simulated covariance structure to be the same as that for the empirical data. As in the simulation study in Section 5, we repeat the simulation procedure 100 times and report the break detection results. For simplicity, we fix $\xi_n^e = 95.99$ in

WSBS-Cov, which is calculated by applying the method in Section 4 to the real data. Table 3 reports the average break number estimates, the number of times that the estimated break is within a distance of $\log(n) \approx 8.6$ to the true break, denoted as ACU, and the root-mean-squared distance of the closest estimated break to the true break, denoted as RMSE. Except for the first two breaks (24/04/2000 and 25/11/2002) in the common components, ACUs are larger than 95% and RMSEs are smaller than 4, indicating that the break dates are accurately detected. The low ACUs and large RMSEs of the first two breaks may be due to the relatively weak break magnitude. The estimated break numbers reported in Table 3 are slightly larger than the true break numbers (5 for the common components and 3 for the idiosyncratic components). In addition, we also fix $q_0 = 12$ in the break detection procedures and obtain similar simulation results, showing that the proposed methods work when the factor number is over-estimated.

Table 3: Detection results on calibrated data

	W(S)BS-Cov		W(S)BS-Cov (fix $q_0 = 12$)	
mean $\widehat{K}_1$	5.01		5.02	
mean $\widehat{K}_2$	3.31		3.08	
	ACU(%)	RMSE	ACU(%)	RMSE
Breaks in common components				
24/04/2000	79	8.721	79	8.731
25/11/2002	88	5.725	85	6.813
12/09/2008	99	1.643	100	1.300
08/05/2009	99	2.502	99	2.240
06/03/2020	97	3.838	97	3.929
Breaks in idiosyncratic error components				
06/08/2002	97	3.425	96	2.466
11/09/2008	100	1.503	100	1.562
11/05/2009	99	2.995	99	1.912

6.3 Comparison with observed risk factors

As a benchmark, we also download eight observable risk factors (over 01/01/2000–28/02/2021) from Kenneth R. French's data library[1] and study breaks in their covariance structure. These risk factors include the excess return (MKT), the size factor (SMB), the value factor (HML), the profitability factor (RMW) and the investment factor (CMA) introduced by Fama and French (1993, 2015); the Momentum factor (MOM) proposed by Carhart (1997) in a four-factor model; and the short term reversal factor (STRev) and the long-term reversal factor (LTRev) which are less famous but equally important (c.f., De Bondt and Thaler, 1987; Lehmann, 1990). They are often used as observable proxies of latent factors in the approximate factor model. We plot patterns of observed factors in Figure 2, from which we can see that they variate drastically during the global

---

[1]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

financial crisis and the COVID-19 outbreak. Let $\mathbf{Z}_t = (Z_{1t}, \cdots, Z_{8t})^\intercal$ denote the vector containing the eight risk factors, $t = 1, \cdots, 5322$. As the dimension of $\mathbf{Z}_t$ is low, there is no need to impose the approximate factor model framework. Using the WBS-Cov algorithm (without any latent factor structure) in Section 2.3, we detect five break dates: 19/04/2001, 01/07/2008, 12/09/2008, 13/05/2009 and 06/03/2020. Note that the first two estimated break dates (19/04/2001 and 01/07/2008) are quite different from the first two break dates (24/04/2000 and 25/11/2002) in the common components obtained by applying the proposed WBS-Cov to the estimated factors, see Figure 1. Table 4 shows the canonical correlation between the estimated factors and observed factors for the entire sample and six sub-samples divided by the estimated break dates in the common components. P-values are reported in the brackets using the F-approximations of Rao's test statistic related to Wilks' Lambda (c.f., Rao, 1973). Some of the minimum canonical correlations are small and insignificant, indicating that the space spanned by the eight observed factors may be not the same as that spanned by the nine PCA estimated factors.

To further analyse whether breaks in the common components can be explained by the observed factors, we run regression of stock returns on the eight observed factors and detect whether breaks exist in covariances of the regression residuals. We apply the factor model (2.2) to the residual vectors and use the information criterion to select six factors. Using WBS-Cov, we detect six breaks in the common components (of the residual vectors): 20/04/2000, 22/11/2002, 07/07/2008, 26/09/2008, 08/05/2009 and 06/03/2020, and one break in the idiosyncratic components which is 12/08/2002. Note that the first two breaks in the common components 20/04/2000 and 22/11/2002 are very close to 24/04/2000 and 25/11/2002, respectively, which are reported in Figure 1. Hence, we may conclude that break detection of a low-dimensional observed risk factor cannot fully capture structural breaks in covariance of price changes in the entire stock market.

7 Conclusion

In this paper, we detect and estimate multiple structural breaks in the contemporaneous covariance structure of the high-dimensional time series vector. We allow high correlation between the time series variables by imposing the approximate factor model, a commonly-used framework in economics and finance. The structural breaks occur in either the common or idiosyncratic error components. When there are breaks in the factor loadings or factor numbers, we transform the factor model in order to make the classic PCA method applicable and subsequently obtain estimates of the transformed factors and approximation of the idiosyncratic errors. The construction of the CUSUM quantities is based on the second-order moments of the estimated factors and errors. The WBS algorithm introduced by Fryzlewicz (2014), combined with SSIC, is extended to the multivariate setting to detect breaks in covariance structure of the common components, estimating

the location and number of breaks. For the idiosyncratic error components, we combine the WBS-Cov algorithm and the sparsified CUSUM statistic to detect structural breaks (which can be either sparse or dense), extending the SBS algorithm proposed by Cho and Fryzlewicz (2015). Under some regularity conditions, we show consistency of the estimated break numbers and derive the near-optimal rates (up to a fourth-order logarithmic factor) for the estimated breaks. Monte-Carlo simulation studies are conducted to examine the numerical performance of the proposed WBS-Cov and WSBS-Cov methods in finite samples and its comparison with the method introduced by Barigozzi, Cho and Fryzlewicz (2018). In addition, we examine performance of the proposed methods when breaks are on weak factors. In the empirical application to the return series of S&P 500 constituent, we detect five break in the common components and three breaks in the idiosyncratic error components by the developed methods.
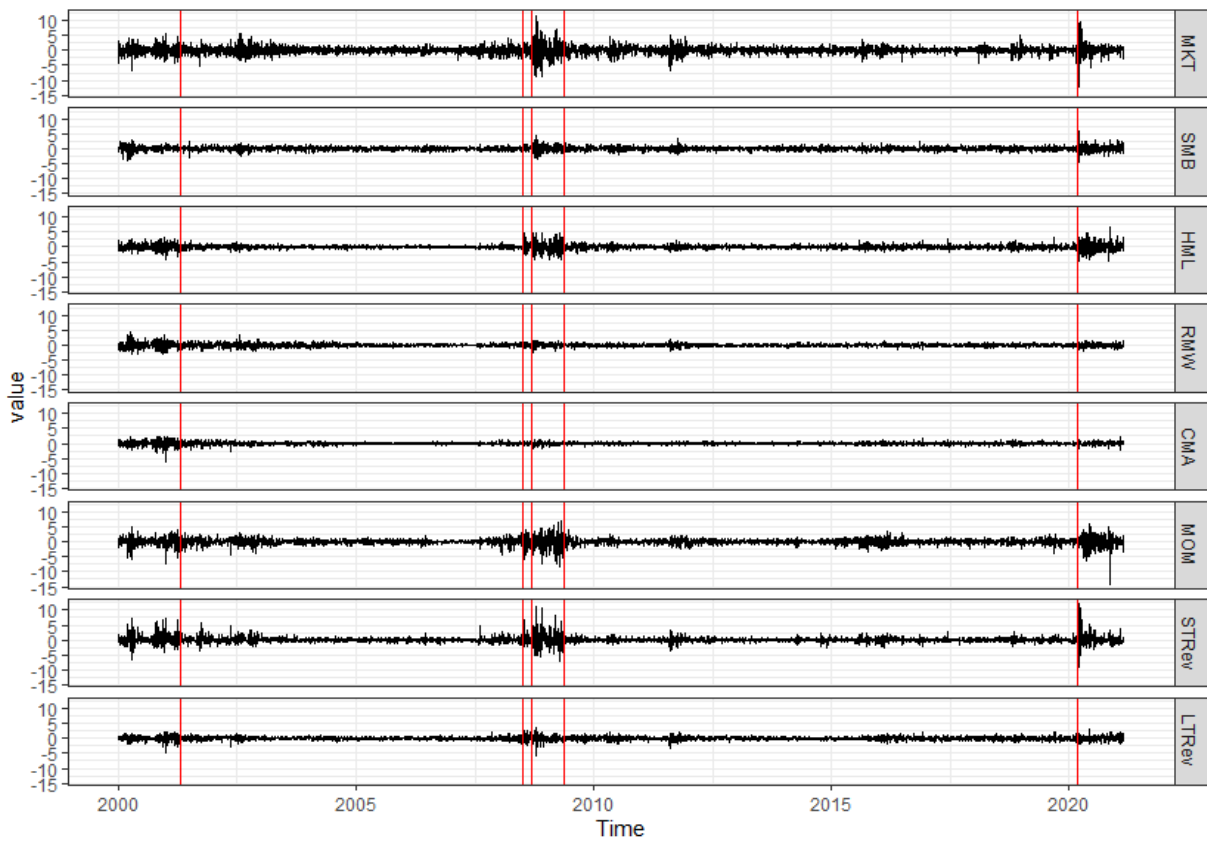

Figure 2: Variation of eight commonly-used risk factors from 01/01/2000 to 28/02/2021. The estimated break dates are 19/04/2001, 01/07/2008, 12/09/2008, 13/05/2009 and 06/03/2020 (in red lines).

Table 4: Canonical correlation between estimated factors and observed factors

	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Total
from	01/01/2000	25/04/2000	26/11/2002	13/09/2008	09/05/2009	07/03/2020	01/01/2000
to	24/04/2000	25/11/2002	12/09/2008	08/05/2009	06/03/2020	28/02/2021	28/02/2021
1st	0.988	0.984	0.993	0.998	0.994	0.996	0.990
p-value 1:8	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2nd	0.964	0.898	0.873	0.947	0.861	0.949	0.850
p-value 2:8	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
3rd	0.742	0.588	0.817	0.877	0.651	0.784	0.726
p-value 3:8	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
4th	0.605	0.508	0.627	0.760	0.553	0.577	0.554
p-value 4:8	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
5th	0.517	0.359	0.444	0.615	0.384	0.383	0.292
p-value 5:8	(0.007)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
6th	0.418	0.280	0.239	0.431	0.256	0.308	0.207
p-value 6:8	(0.129)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
7th	0.259	0.107	0.134	0.398	0.155	0.219	0.119
p-value 7:8	(0.605)	(0.255)	(0.000)	(0.000)	(0.000)	(0.007)	(0.000)
8th	0.051	0.027	0.087	0.256	0.022	0.160	0.039
p-value 8:8	(0.916)	(0.795)	(0.004)	(0.005)	(0.510)	(0.047)	(0.019)

P-values are reported in the brackets using the F-approximations of Rao's test statistics related to Wilks' Lambda.

Acknowledgements

Appendix A: Technical assumptions

The following assumptions are needed to establish the asymptotic theorems in Sections 3.1 and 3.2. Some of the conditions (such as the moment conditions) may be not the weakest possible but can be relaxed at the cost of more lengthy proofs.

ASSUMPTION 1. *(i) For each* $k = 1, \cdots, K_1 + 1$, *the process* $\{(\mathbf{F}_{t,k}, \boldsymbol{\epsilon}_t)\}$ *is* α-*mixing dependent with the mixing coefficient* $\alpha_k(\cdot)$ *satisfying*

$$\max_{1 \leqslant k \leqslant K_1+1} \alpha_k(s) \asymp \rho^s, \ \ 0 < \rho < 1.$$

(ii) For $k = 1, \cdots, K_1 + 1$ *and* $j = 1, \cdots, d$, $\mathsf{E}[\mathbf{F}_{t,k}] = \mathbf{0}$, $\mathsf{E}[\boldsymbol{\epsilon}_t] = \mathbf{0}$ *and* $\mathsf{E}[\epsilon_{tj}\mathbf{F}_{t,k}] = \mathbf{0}$. *In addition,*

$$\max_{1 \leqslant k \leqslant K_1+1} \max_{\eta_{k-1}^c+1 \leqslant t \leqslant \eta_k^c} \mathsf{E}[\|\mathbf{F}_{t,k}\|_2^{\delta_F}] < \infty, \quad \max_{1 \leqslant t \leqslant n} \max_{1 \leqslant j \leqslant d} \mathsf{E}[|\epsilon_{tj}|^{\delta_\epsilon}] < \infty,$$

where $\delta_F > 2$ *and* $\delta_\epsilon > 2$.

ASSUMPTION 2. *(i) For* $k = 1, \cdots, K_1 + 1$, *the covariance matrix* $\mathsf{Cov}(\mathbf{F}_{t,k}) = \boldsymbol{\Sigma}_{F,k}$ *of the latent factors when* $\eta_{k-1}^c + 1 \leqslant t \leqslant \eta_k^c$, *is* $r_k \times r_k$ *positive definite. The positive integers* K_1 *and* r_k, $k = 1, \cdots, K_1 + 1$, *are all bounded.*

(ii) Letting $\lambda_{k,j}^0$ *be the* j-*th factor loading vector of* $\boldsymbol{\Lambda}_k^0$ *defined in (2.2),* $\|\lambda_{k,j}^0\|_2$ *is bounded from above uniformly over* $j = 1, \cdots, d$ *and* $k = 1, \cdots, K_1 + 1$. *In addition, letting* $\bar{\boldsymbol{\Lambda}} = \left[\boldsymbol{\Lambda}_1^0, \cdots, \boldsymbol{\Lambda}_{K_1+1}^0\right]$ *which is a* $d \times \sum_{k=1}^{K_1+1} r_k$ *matrix, the column rank of* $\bar{\boldsymbol{\Lambda}}$ *is* q_0 *and there exist positive constants* $\underline{\mu}$ *and* $\overline{\mu}$ *such that*

$$0 < \underline{\mu} \leqslant \mu_{q_0}\left(\frac{1}{d}\bar{\boldsymbol{\Lambda}}^\intercal \bar{\boldsymbol{\Lambda}}\right) \leqslant \mu_1\left(\frac{1}{d}\bar{\boldsymbol{\Lambda}}^\intercal \bar{\boldsymbol{\Lambda}}\right) \leqslant \overline{\mu} < \infty. \tag{A.1}$$

ASSUMPTION 3. *(i) There exists a positive constant* ι_0 *such that*

$$\max_{1 \leqslant k \leqslant K_1+1} \max_{1 \leqslant t \leqslant n} \mathsf{E}\left[\left\|\sum_{j=1}^d \lambda_{k,j}^0 \epsilon_{tj}\right\|_2^{\delta_\epsilon}\right] \leqslant \iota_0 \cdot d^{\delta_\epsilon/2} \tag{A.2}$$

and

$$\max_{1 \leqslant s,t \leqslant n} \mathsf{E}\left[\left|\sum_{j=1}^d \{\epsilon_{sj}\epsilon_{tj} - \mathsf{E}\left[\epsilon_{sj}\epsilon_{tj}\right]\}\right|^{\delta_\epsilon}\right] \leqslant \iota_0 \cdot d^{\delta_\epsilon/2}. \tag{A.3}$$

(ii) There exists a positive constant $\iota_1 > 0$ *such that*

$$\max_{1 \leqslant j \leqslant d} \max_{1 \leqslant t \leqslant n} \mathsf{E}\left[\exp\left\{\iota_1 \epsilon_{tj}^2\right\}\right] < \infty, \quad \max_{1 \leqslant k \leqslant K_1+1} \max_{\eta_{k-1}^c+1 \leqslant t \leqslant \eta_k^c} \mathsf{E}\left[\exp\left\{\iota_1 \|\mathbf{F}_{t,k}\|_2^2\right\}\right] < \infty. \tag{A.4}$$

ASSUMPTION 4. *(i) There exist positive constants* ι_2, ι_3 *and* ϑ *such that*

$$\iota_2 \log^{5+\vartheta}(d) \leqslant n \leqslant \iota_3 d^{\delta/(\delta+4)},$$

where $\delta = \delta_F \wedge \delta_\epsilon$.

(ii) Letting $\kappa_n^c = \min_{1 \leqslant k \leqslant K_1+1}(\eta_k^c - \eta_{k-1}^c)$,

$$\underline{\omega}_n^c = \frac{1}{d^2} \cdot \min_{1 \leqslant k \leqslant K_1} \left\|\boldsymbol{\Sigma}_{k+1}^0(\boldsymbol{\Lambda}, \mathbf{F}) - \boldsymbol{\Sigma}_k^0(\boldsymbol{\Lambda}, \mathbf{F})\right\|_F^2$$

and

$$\overline{\omega}_n^c = \frac{1}{d^2} \cdot \max_{1 \leqslant k \leqslant K_1} \left\| \mathbf{\Sigma}_{k+1}^0(\mathbf{\Lambda}, \mathbf{F}) - \mathbf{\Sigma}_k^0(\mathbf{\Lambda}, \mathbf{F}) \right\|_F^2,$$

we have

$$\kappa_n^c \asymp n, \quad \frac{\omega_n^c}{\overline{\omega}_n^c} \cdot \frac{\kappa_n^c \omega_n^c}{\log^4 n} \to \infty, \quad M_n^c \to \infty,$$

where M_n^c is the number of random intervals drawn in the WBS-Cov algorithm.

(iii) Letting $\kappa_n^e = \min_{1 \leqslant k \leqslant K_2+1}(\eta_k^e - \eta_{k-1}^e)$,

$$\underline{\omega}_n^e = \min_{1 \leqslant k \leqslant K_2} \min_{(i,j) \in \mathcal{I}_k} \left| \sigma_{k+1|i,j}^e - \sigma_{k|i,j}^e \right|^2, \quad \overline{\omega}_n^e = \max_{1 \leqslant k \leqslant K_2} \max_{(i,j) \in \mathcal{I}_k} \left| \sigma_{k+1|i,j}^e - \sigma_{k|i,j}^e \right|^2,$$

where $\mathcal{I}_k = \left\{ (i,j) : \sigma_{k+1|i,j}^e \neq \sigma_{k|i,j}^e, 1 \leqslant i, j \leqslant d \right\}$ and $\sigma_{k|i,j}^e$ is the (i,j)-entry of $\mathbf{\Sigma}_k^0(\boldsymbol{\epsilon})$, we have

$$\overline{\omega}_n^e \asymp \underline{\omega}_n^e, \quad \frac{\kappa_n^e \underline{\omega}_n^e}{\log^4(nd)} \cdot \left(\frac{\kappa_n^e}{n} \right)^4 \to \infty, \quad \left[1 - (\kappa_n^e/(3n))^2 \right]^{M_n^e} \to 0,$$

where M_n^e is the number of random intervals drawn in the WSBS-Cov algorithm.

ASSUMPTION 5. *There exist $0 < \underline{\sigma} < \overline{\sigma} < \infty$ such that*

$$\underline{\sigma} \leqslant \min_{(l,u):\, 1 \leqslant l < u \leqslant n} \min_{1 \leqslant i,j \leqslant d} \widehat{\sigma}_{l,u}(i,j) \leqslant \max_{(l,u):\, 1 \leqslant l < u \leqslant n} \max_{1 \leqslant i,j \leqslant d} \widehat{\sigma}_{l,u}(i,j) \leqslant \overline{\sigma}$$

with probability approaching one, where $\widehat{\sigma}_{l,u}(i,j)$ is defined in (2.12). In addition, $\widehat{\sigma}_{l,u}(i,j) = \widehat{\sigma}_{l,u}(j,i)$.

The assumption of mixing dependence on the latent common factor and idiosyncratic error process in Assumption 1(i) is very mild and covers some commonly-used vector time series models. Note that the α-mixing dependence condition allows the underlying process to be nonstationary with time-varying covariance structure (e.g., the piece-wise stationary time series process with breaks in the covariance matrix). The moment conditions in Assumptions 1(ii) and 3 are crucial to consistently estimate the factors and factor loadings (with rotation) in the transformed factor model (2.4).

Assumption 2 contains some fundamental conditions for the approximate factor model (2.2), which can be seen as a generalised version of those for the stable factor model without breaks (e.g., Bai and Ng, 2002; Fan, Liao and Mincheva, 2013). These conditions, together with $\kappa_n^c \asymp n$ in Assumption 4(ii), indicate that there exists a nonsingular limiting matrix for $\frac{1}{n}\sum_{t=1}^n \mathbf{F}_t^\star \mathbf{F}_t^{\star\top}$ and $\mathbf{\Lambda}^\star$ is of full column rank (see Proposition 3.1), which are essential to derive the convergence result for

the PCA estimated factors and factor loadings. In addition, Assumption 2 also guarantees that $\overline{\omega}_n^c = O(1)$.

The moment conditions (A.2) and (A.3) in Assumption 3(i) imply that the idiosyncratic errors ε_{tj} are allowed to be weakly dependent over j. Similar high-level moment conditions can be found in Assumption 4 in Bai and Ng (2002) and Assumption 3.4 in Fan, Liao and Mincheva (2013). Assumption 3(ii) is similar to the exponential-tail condition commonly used in high-dimensional data analysis and is needed to allow the dimension d to be divergent at an exponential rate of n.

Assumption 4(i) indicates that both d and n diverge to infinity jointly, which is a typical large panel data setting necessary for consistent estimates of the rotated factors and factor loadings. A very similar assumption is also used in Theorem 1 of Fan, Liao and Mincheva (2013). In particular, Assumption 4(i) shows that d can be divergent to infinity at an exponential rate of n, further relaxing the corresponding condition assumed in Barigozzi, Cho and Fryzlewicz (2018) and Wang, Yu and Rinaldo (2021). Although the restriction $n \leqslant \iota_3 d^{\delta/(\delta+4)}$ indicates that n is of smaller asymptotic order than d, our methodology is still applicable if n is much larger than d by switching the role of n and d and the role of the factor loadings and common factors in the PCA estimation (see Section 3 in Bai and Li, 2012). Assumption 4(ii) is mainly used for the asymptotics of WBS-Cov in the common components, and would be automatically satisfied when $\iota_4 \leqslant \underline{\omega}_n^c \leqslant \overline{\omega}_n^c \leqslant \iota_5$ and $M_n^c = \lfloor \log n \rfloor$, where $\iota_4 < \iota_5$ are two positive constants. Assumption 4(iii) is mainly used to derive the WSBS-Cov asymptotic theory in the idiosyncratic error components. The restriction $\overline{\omega}_n^e \asymp \underline{\omega}_n^e$ is only imposed to facilitate the proofs. The restrictions on $\underline{\omega}_n^c$ and $\underline{\omega}_n^e$ are consistent with the argument in Andrews (1993) and Bai (1997) that the break points can be identified only if the size of breaks, i.e., $(\underline{\omega}_n^c)^{1/2}$ and $(\underline{\omega}_n^e)^{1/2}$ in our paper, is asymptotically larger than $n^{-1/2}$. Assumption 4(iii) also indicates that there might be more frequent breaks in the idiosyncratic error components than in the common components, and the restriction $\frac{\kappa_n^e \underline{\omega}_n^e}{\log^4(nd)} \cdot \left(\frac{\kappa_n^e}{n}\right)^4 \to \infty$ would be automatically satisfied when $\kappa_n^e \asymp n^\Theta$ and $\left(\underline{\omega}_n^e n^{5\Theta-4}\right)/\log^4(nd) \to \infty$ with $\Theta \in (4/5, 1]$. In contrast, the minimum length κ_n^c of the sub-intervals separated by break points in the common components is of order n. Assumption 5 requires the random normalisers in the WSBS-Cov to be uniformly bounded away from zero and infinity with probability approaching one.

In Assumption 2(i), we assume that K_1 and r_k, $k = 1, \cdots, K_1$, are all bounded, which indicates that q_0, the number of transformed factors in (2.4), is upper bounded by Proposition 2.1. This assumption facilitates proofs of the consistent estimation theory for the transformed factors and factor loadings and consistent break estimation theory. It would be an interesting extension by allowing the break numbers K_1 and K_2 to grow slowly as $n \to \infty$. In this case, q_0 would be divergent as well, and the condition of $\kappa_n^c \asymp n$ would be violated. Consequently, we need to amend the estimation and break detection techniques, and modify some technical assumptions

(such as Assumption 4) and the asymptotic proofs. We will further explore this in our future study.

Supplementary document

The supplemental document contains the detailed proofs of the main asymptotic results, a simple motivating example for the factor model transformation as well as additional simulation studies.

References

ALESSI, L., BARIGOZZI, M. AND CAPASSO, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics and Probability Letters* **80**, 1806-1813.

ANDREWS, D.W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**, 821–856.

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis (3rd Edition)*. Wiley Series in Probability and Statistics.

ASTON, J. A. AND KIRCH, C. (2018). High dimensional efficiency with applications to change point tests. *Electronic Journal of Statistics* **12**, 1901–1947.

AUE, A., HÖRMANN, HORVÁTH, L. AND REIMHERR, M. (2009). Break detection in the covariance structure of multivariate time series models. *Annals of Statistics* **37**, 4046–4087.

BAI, J. (1997). Estimating multiple breaks on at a time. *Econometric Theory* **13**, 315–352.

BAI, J., HAN, X. AND SHI, Y., 2020. Estimation and inference of change points in high-dimensional factor models. *Journal of Econometrics* **219**, 66–100

BAI, J. AND LI, K. (2012). Statistical analysis of factor models of high dimension. *Annals of Statistics* **40**, 436–465.

BAI, J. AND NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.

BAI, J. AND NG, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* **74**, 1135–1150.

BALTAGI, B., KAO, C. AND WANG, F. (2017). Identification and estimation of a large factor model with structural instability. *Journal of Econometrics* **197**, 87–100.

BALTAGI, B., KAO, C. AND WANG, F. (2021). Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics* **220**, 349–365.

BARIGOZZI, M., CHO, H. AND FRYZLEWICZ, P. (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics* **206**, 187–225.

BICKEL, P. AND LEVINA, E. (2008a). Covariance regularization by thresholding. *Annals of Statistics* **36**, 2577–2604.

BICKEL, P. AND LEVINA, E. (2008b). Regularized estimation of large covariance matrices. *Annals of Statistics* **36**, 199–227.

DE BONDT, W. F. AND THALER, R. H. (1987). Further evidence on investor overreaction and stock market seasonality. *Journal of Finance* **42**, 557–581.

BREITUNG, J. AND EICKMEIER, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics* **163**, 71–84.

CAI, T. T. AND LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* **106**, 672–684.

CAI, T. T., REN, Z. AND ZHOU, H. (2016). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics* **10**, 1–59.

CARHART, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance* **52**, 57–82.

CHAMBERLAIN, G. AND ROTHSCHILD, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* **51**, 1305–1324.

CHEN, J., LI, D., LINTON, O. AND LU, Z. (2018). Semiparametric ultra-high dimensional model averaging of nonlinear dynamic time series. *Journal of the American Statistical Association* **113**, 919–932.

CHEN, L., DOLADO, J. J. AND GONZALO, J. (2013). Detecting big structural breaks in large factor models. *Journal of Econometrics* **180**, 30–48.

CHEN, X., XU, M. AND WU, W. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Annals of Statistics* **41**, 2994–3021.

CHEN, Z. AND LENG, C. (2016). Dynamic covariance models. *Journal of the American Statistical Association* **111**, 1196–1207.

CHENG, X., LIAO, Z. AND SCHORFHEIDE, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies* **83**, 1511–1543.

CHO, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics* **10**, 2000–2038.

CHO, H. AND FRYZLEWICZ P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica* **22**, 207–229.

CHO, H. AND FRYZLEWICZ P. (2015). Multiple change-point detection for high-dimensional time series via Sparsified Binary Segmentation. *Journal of the Royal Statistical Society Series B* **77**, 475–507.

CHO, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics* **10**, 2000–2038.

DUAN, J., BAI, J. AND HAN, X. (2021). Quasi-maximum likelihood estimation of break point in high-dimensional factor models. Working paper available at https://arxiv.org/pdf/2102.12666.pdf.

EL KAROUI, N. (2008a). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Annals of Statistics* **36**, 2717–2756.

EL KAROUI, N. (2008b). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics* **36**, 2757–2790.

FAMA, E. F. AND FRENCH, K. (1992). The cross-section of expected stock returns. *Journal of Finance* **47**, 427–465.

FAMA, E. F. AND FRENCH, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33**, 3–56.

FAMA, E. F. AND FRENCH, K. (2015). A five-factor asset pricing model. *Journal of Financial Economics* **116**, 1–22.

FAN, J., LIAO, Y. AND LIU, H. (2016). An overview on the estimation of large covariance and precision matrices. *Econometrics Journal* **19**, 1–32.

FAN, J., LIAO, Y. AND MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society, Series B* **75**, 603–680.

FREYALDENHOVEN, S. (2021). Factor models with local factors – Determining the number of relevant factors. Forthcoming in *Journal of Econometrics*.

FRYZLEWICZ, P. (2014). Wild Binary Segmentation for multiple change-point detection. *Annals of Statistics* **42**, 2243–2281.

FRYZLEWICZ, P. (2020). Detecting possibly frequent change-points: Wild Binary Segmentation 2 and steepest-drop model selection. *Journal of the Korean Statistical Society* **49**, 1027–1070.

HAN X. AND INOUE, A. (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory* **31**, 1117–1152.

HARIZ, S. B., WYLIE, J. J. AND ZHANG, Q. (2007). Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences. *Annals of Statistics* **35**, 1802–1826.

HUANG, N. AND FRYZLEWICZ, P. (2019). NOVELIST estimator of large correlation and covariance matrices and their inverses. *Test* **28**, 694–727.

JIRAK, M. (2015). Uniform change point tests in high dimension. *Annals of Statistics* **43**, 2451–2483.

KILLICK, R., FEARNHEAD, P. AND ECKLEY, I. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* **107**, 1590–1598.

KORKAS, K. K. AND FRYZLEWICZ, P. (2017). Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica Sinica* **27**, 287–311.

KOROSTELEV, A.P. (1988). On minimax estimation of a discontinuous signal. *Theory of Probability & Its Applications* **32**, 727–730.

LAM, C. AND FAN, J. (2009). Sparsity and rates of convergence in large covariance matrix estimation. *Annals of Statistics* **37**, 4254–4278.

LEDOIT, O. AND WOLF, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88**, 365–411.

LEHMANN, B. N. (1990). Fads, martingales, and market efficiency. *Quarterly Journal of Economics* **105**, 1–28.

LI, D., QIAN, J. AND SU, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association* **111**, 1804–1819.

LI, J., TODOROV, V., TAUCHEN, G. AND LIN, H. (2019). Rank tests at jump events. *Journal of Business & Economic Statistics* **37**, 312–321.

MA, S. AND SU, L. (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics* **207**, 1–29.

POURAHMADI, M. (2013). *High-Dimensional Covariance Estimation: With High-Dimensional Data*. Wiley.

RAO, C. R (1973). *Linear Statistical Inference and Its Applications (2nd ed.)*. John Wiley and Sons, New York.

ROTHMAN, A. J., LEVINA, E. AND ZHU, J. (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association* **104**, 177–186.

SAFIKHANI, A. AND SHOJAIE, A. (2020). Joint structural break detection and parameter estimation in high-dimensional non-stationary VAR models. Forthcoming in *Journal of the American Statistical Association*.

STOCK, J. H. AND WATSON, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97**, 1167–1179.

VENKATRAMAN, E. S. (1992). Consistency results in multiple change-point problems. *Technical Report No. 24, Department of Statistics, Stanford University*.

VOSTRIKOVA L. YU. (1981). Detecting disorder in multidimensional random processes. *Soviet Mathematics Doklady* **24**, 55–59.

WANG, D., YU, Y. AND RINALDO, A. (2021). Optimal covariance change point localization in high dimensions. *Bernoulli* **27**, 554–575.

WANG, D., YU, Y., RINALDO, A. AND WILLETT, R. (2019). Localizing changes in high-dimensional vector autoregressive processes. Working paper available at https://arxiv.org/abs/1909.06359.

WANG, T. AND SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society Series B* **80**, 57–83.

WU, W. AND POURAHMADI, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90**, 831–844.

YU, M. AND CHEN, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *Journal of the Royal Statistical Society Series B* **83**, 247–270.