

# Detection of Multiple Structural Breaks in Large Covariance Matrices

Yu-Ning Li\*

Degui Li<sup>†</sup>

Piotr Fryzlewicz<sup>‡</sup>

This version: April 19, 2021

## 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. 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. We finally apply our method to detect breaks in the contemporaneous covariance structure of daily returns of S&P 500 constituents. 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.

<sup>†</sup>The corresponding author, Department of Mathematics, University of York, YO10 5DD, UK. Email address: degui.li@york.ac.uk. Li’s research was partly supported by the BA/Leverhulme Small Research Grant funded by the Leverhulme Trust (No. SRG1920/100603).

<sup>‡</sup>Department of Statistics, London School of Economics, WC2A 2AE, UK. Fryzlewicz’s research was supported by the Engineering and Physical Sciences Research Council (No. EP/L01424).

# 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  $(\mathbf{X}_t : t = 1, \dots, n)$  is a collection of  $n$  observations from a  $d$ -dimensional random vector with  $E(\mathbf{X}_t) = \mathbf{0}$  and  $E(\mathbf{X}_t \mathbf{X}_t^\top) = \Sigma$ , where  $\mathbf{0}$  is a null vector whose size may change from line to line and  $\Sigma$  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  $\Sigma$  can be well estimated by the conventional sample covariance matrix (c.f., [Anderson, 2003](#)):

$$\bar{\Sigma}_n = \frac{1}{n} \sum_{t=1}^n (\mathbf{X}_t - \bar{\mathbf{X}}_n) (\mathbf{X}_t - \bar{\mathbf{X}}_n)^\top, \quad \bar{\mathbf{X}}_n = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t. \quad (1.1)$$

However, when  $d$  exceeds  $n$ , the sample covariance matrix  $\bar{\Sigma}_n$  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](#); [Bickel and Levina, 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 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  $\mathbf{X}_t$  generated by an approximate factor model, i.e.,

$$\mathbf{X}_t = \mathbf{\Lambda} \mathbf{F}_t + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \dots, n, \quad (1.2)$$

where  $\mathbf{\Lambda} = (\lambda_1, \dots, \lambda_d)^\top$  is a matrix of factor loadings with  $\lambda_i$  being an  $r$ -dimensional vector,  $\mathbf{F}_t$  is an  $r$ -dimensional vector of common factors, and  $\boldsymbol{\epsilon}_t = (\epsilon_{t1}, \dots, \epsilon_{td})^\top$  is a  $d$ -dimensional vector of idiosyncratic components uncorrelated with  $\mathbf{F}_t$ . 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

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}_t) = \mathbf{\Lambda} \text{Cov}(\mathbf{F}_t) \mathbf{\Lambda}^\top + \text{Cov}(\boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma}(\mathbf{\Lambda}, \mathbf{F}) + \boldsymbol{\Sigma}(\boldsymbol{\epsilon}), \quad (1.3)$$

where  $\Sigma(\epsilon)$  is the covariance matrix of the idiosyncratic components  $\epsilon_t$  and  $\Sigma(\Lambda, F)$  is the covariance matrix of the common components  $\Lambda F_t$ . 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:

$$\Sigma_t = \Sigma_t(\Lambda, F) + \Sigma_t(\epsilon), \quad t = 1, \dots, n. \quad (1.4)$$

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 \(2015\)](#), [Baltagi, Kao and Wang \(2017\)](#) and [Baltagi, Kao and Wang \(2021\)](#) 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 \(2018\)](#) 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 may be divergent at a slow polynomial rate of  $n$ . In contrast, we allow the dimension to grow at an exponential rate of  $n$ , a typical setting in big data analysis.

- Extensive 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 numerical experience, part of which is summarised in [Section 5](#), suggests that our method works well for time series

with dimension  $d = 200$ , provided that the sample size is 200, that the number of change-points is 3, and that the minimum spacing between change-points is 50.

- Furthermore, we apply the developed model and method 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 eight 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 reveal the importance of break detection in idiosyncratic components.

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 and proofs of the main theoretical results are available in 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  $\text{vech}(\cdot)$  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

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

where  $\Sigma_k^0(\Lambda, F) \neq \Sigma_{k+1}^0(\Lambda, F)$  for  $k = 1, \dots, K_1$ , and  $\eta_1^c, \dots, \eta_{K_1}^c$  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  $\Lambda_k^0$  be a  $d \times r_k$  factor loading matrix and  $F_{t,k}$  be the corresponding factor vector with dimension  $r_k$  when  $\eta_{k-1}^c + 1 \leq t \leq \eta_k^c$ , and write the approximate factor model as

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

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

$$\Sigma_t(\Lambda, F) = \begin{cases} \Lambda_1^0 \text{Cov}(F_{t,1})(\Lambda_1^0)^\top, & 1 \leq t \leq \eta_1^c, \\ \Lambda_2^0 \text{Cov}(F_{t,2})(\Lambda_2^0)^\top, & \eta_1^c + 1 \leq t \leq \eta_2^c, \\ \vdots & \vdots \\ \Lambda_{K_1+1}^0 \text{Cov}(F_{t,K_1+1})(\Lambda_{K_1+1}^0)^\top, & \eta_{K_1}^c + 1 \leq t \leq n. \end{cases} \quad (2.3)$$

Note that, although  $\Sigma_k^0(\Lambda, F) \neq \Sigma_{k+1}^0(\Lambda, F)$  for  $k = 1, \dots, K_1$ ,  $\Lambda_k^0$  may be the same as  $\Lambda_{k+1}^0$  (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

$$\Sigma_t(\epsilon) = \begin{cases} \Sigma_1^0(\epsilon), & 1 \leq t \leq \eta_1^e, \\ \Sigma_2^0(\epsilon), & \eta_1^e + 1 \leq t \leq \eta_2^e, \\ \vdots & \vdots \\ \Sigma_{K_2+1}^0(\epsilon), & \eta_{K_2}^e + 1 \leq t \leq n, \end{cases} \quad (2.4)$$

and  $\Sigma_k^0(\epsilon) \neq \Sigma_{k+1}^0(\epsilon)$  for  $k = 1, \dots, K_2$ , where  $\eta_1^e, \dots, \eta_{K_2}^e$  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

$$\mathcal{B}^c = \{\eta_1^c, \dots, \eta_{K_1}^c\} \quad \text{and} \quad \mathcal{B}^e = \{\eta_1^e, \dots, \eta_{K_2}^e\}. \quad (2.5)$$

In this paper, we assume that the minimum length of the sub-intervals separated by  $\eta_k^c$  is of order  $n$ , see Assumption 4(ii) in Appendix A of the supplement. 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  $\eta_k^e$  is allowed to be of order smaller than  $n$ , see Assumption 4(iii) in Appendix A of the supplement.



## 2.2 Transformed factor models and PCA estimation

Although the multiple structural breaks formulated in (2.1) and (2.3) 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*

$$\mathbf{X}_t = \mathbf{\Lambda}^* \mathbf{F}_t^* + \boldsymbol{\epsilon}_t, \quad t = 1, \dots, n, \quad (2.6)$$

where  $\mathbf{\Lambda}^*$  denotes the transformed factor loading matrix which is time invariant, and  $\mathbf{F}_t^*$  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,

$$\max_{1 \leq k \leq K_1+1} \underline{r}_k \leq q_0 \leq \sum_{k=1}^{K_1+1} \underline{r}_k \leq \sum_{k=1}^{K_1+1} r_k, \quad (2.7)$$

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

Note that construction of  $\mathbf{\Lambda}^*$  in the transformation mechanism is not unique due to the factor model identification issue. When each of  $\mathbf{\Lambda}_k^0$  is of full column rank, we have  $\underline{r}_k = r_k$ , and consequently the inequalities in (2.7) would be simplified to

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

The upper bound for  $q_0$  can be achieved if  $\mathbf{\Lambda}_k^0$ ,  $k = 1, \dots, K_1 + 1$ , 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  $\frac{1}{n} \sum_{t=1}^n \mathbf{F}_t^* \mathbf{F}_t^{*\top}$  as  $n \rightarrow \infty$ , which is crucial to application of the classic PCA method to estimate the transformed factors and factor loadings. Through the above transformation, we may 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 (2.6) with time-invariant factor loadings, and apply the PCA estimation technique. Let  $\mathcal{X}_n = (\mathbf{X}_1, \dots, \mathbf{X}_n)^\top$  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  $\widehat{\mathcal{F}}_n = (\widehat{\mathbf{F}}_1, \dots, \widehat{\mathbf{F}}_n)^\top$  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  $\mathcal{X}_n \mathcal{X}_n^\top / (nd)$ .
2. The transformed factor loading matrix is estimated as

$$\widehat{\boldsymbol{\Lambda}}_n = \mathcal{X}_n^\top \widehat{\mathcal{F}}_n / n =: (\widehat{\lambda}_1, \dots, \widehat{\lambda}_d)^\top. \quad (2.8)$$

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

$$\widehat{\boldsymbol{\epsilon}}_t = \mathbf{X}_t - \widehat{\boldsymbol{\Lambda}}_n \widehat{\mathbf{F}}_t =: (\widehat{\boldsymbol{\epsilon}}_{t1}, \dots, \widehat{\boldsymbol{\epsilon}}_{td})^\top \text{ with } \widehat{\boldsymbol{\epsilon}}_{tj} = X_{tj} - \widehat{\lambda}_j^\top \widehat{\mathbf{F}}_t, \quad 1 \leq j \leq d, \quad (2.9)$$

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

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

---

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  $r_k$  (or  $r_k$ ) over  $k = 1, \dots, K_1 + 1$ . However, when both  $(r_k, k = 1, \dots, K_1 + 1)$  and  $K_1$  are assumed to be fixed, from (2.7) in Proposition 2.1, we have the true number of transformed factors to be upper bounded by a finite positive integer denoted by  $\overline{Q}$ . 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). Some existing criteria developed for stable factor models can be applied to the transformed factor model to obtain a consistent estimate of  $q_0$ . We next use the information criterion proposed by Bai and Ng (2002).

For any  $1 \leq q \leq \overline{Q}$ , let  $\widehat{\mathcal{F}}_n(q) = [\widehat{\mathbf{F}}_1(q), \dots, \widehat{\mathbf{F}}_n(q)]^\top$  be the estimated factors obtained

in Algorithm 1 replacing  $q_0$  by  $q$ . Define

$$\begin{aligned}
V_n(q) &= \min_{\Lambda(q)} \frac{1}{nd} \sum_{j=1}^d \sum_{t=1}^n \left[ X_{tj} - \lambda_j(q) \widehat{F}_t(q) \right]^2 \\
&= \min_{\Lambda(q)} \frac{1}{nd} \sum_{t=1}^n \left[ \mathbf{X}_t - \Lambda(q) \widehat{F}_t(q) \right]^\top \left[ \mathbf{X}_t - \Lambda(q) \widehat{F}_t(q) \right] \\
&= \sum_{j=q+1}^d \mu_j (\mathcal{X}_n \mathcal{X}_n^\top / (nd)), \tag{2.10}
\end{aligned}$$

where  $\Lambda(q) = [\lambda_1(q), \dots, \lambda_d(q)]^\top$  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:

$$\text{IC}(q) = \log [V_n(q)] + q \cdot \left( \frac{n+d}{nd} \right) \log(n \wedge d), \tag{2.11}$$

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

$$\widehat{q} = \arg \min_{0 \leq q \leq \overline{Q}} \text{IC}(q), \tag{2.12}$$

where the common components disappear when  $q = 0$ . Alternative IC functions with different penalty terms in (2.11) 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. 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. In Appendix F of the supplement, we report the finite-sample performance of the estimate  $\widehat{q}$  defined in (2.12) and further compare it with other commonly-used methods.

### 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  $\hat{\mathbf{F}}_t$ . For  $1 \leq l \leq s < u \leq n$ , define

$$\mathbf{C}_{l,u}^{\hat{\mathbf{F}}}(s) = \sqrt{\frac{(s-l+1)(u-s)}{u-l+1}} \left[ \frac{1}{s-l+1} \sum_{t=l}^s \text{vech}(\hat{\mathbf{F}}_t \hat{\mathbf{F}}_t^\top) - \frac{1}{u-s} \sum_{t=s+1}^u \text{vech}(\hat{\mathbf{F}}_t \hat{\mathbf{F}}_t^\top) \right], \quad (2.13)$$

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

$$\tilde{\eta}_1^c = \arg \max_{l \leq s < u} \left\| \mathbf{C}_{l,u}^{\hat{\mathbf{F}}}(s) \right\|_2, \quad (2.14)$$

the first candidate break point (which is not necessarily the estimate of the first break time point  $\eta_1^c$ ). If the quantity  $\left\| \mathbf{C}_{l,u}^{\hat{\mathbf{F}}}(\tilde{\eta}_1^c) \right\|_2$  exceeds certain threshold, say  $\xi_n^c$ , we split the closed interval  $[l, u]$  into two sub-intervals:  $[l, \tilde{\eta}_1^c]$  and  $[\tilde{\eta}_1^c + 1, u]$ . Then, letting  $(l, u) = (l, \tilde{\eta}_1^c)$  or  $(l, u) = (\tilde{\eta}_1^c + 1, u)$ , we compute  $\mathbf{C}_{l,u}^{\hat{\mathbf{F}}}(s)$  with  $s$  in either of the two sub-intervals, maximise  $\left\| \mathbf{C}_{l,u}^{\hat{\mathbf{F}}}(s) \right\|_2$  with respect to  $s$ , estimate the next candidate break points, and examine whether a further split is needed. We continue this procedure until no more split is possible. 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  $l_\infty$ -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 Example 5.1, 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.

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  $\hat{\eta}_1^c, \dots, \hat{\eta}_{\hat{K}_1}^c$  be the estimated break points (arranged in an increasing order) and  $\hat{K}_1$  be the estimate of the unknown break number  $K_1$  obtained in Algorithm 2.

A discussion on selection of the threshold  $\xi_n^c$  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

---

**Algorithm 2** WBS-Cov algorithm

---

**Input:**  $\xi_n^c, M_n^c, \hat{\mathbf{F}}_t, t = 1, \dots, n$   
generate random intervals  $[l_m, u_m]$  within  $[1, n]$  for  $m = 1, \dots, M_n^c$   
set  $\mathcal{W}^c = \{[1, n]\}$  and  $\hat{\mathcal{B}}^c = \emptyset$   
**while**  $\mathcal{W}^c \neq \emptyset$  **do**  
  take an element from  $\mathcal{W}^c$  and denote it as  $[l, u]$   
   $\mathcal{M}_{l,u}^c = \{m : [l_m, u_m] \subset [l, u]\}$   
   $(m_0^c, s_0^c) = \arg \max_{m \in \mathcal{M}_{l,u}^c, l_m \leq s < u_m} \left\| \mathbf{C}_{l,u}^{\hat{\mathbf{F}}}(s) \right\|_2$   
  **if**  $\left\| \mathbf{C}_{l_m_0^c, u_{m_0^c}}^{\hat{\mathbf{F}}}(s_0^c) \right\|_2 > \xi_n^c$  **then**  
    add  $s_0^c$  to  $\hat{\mathcal{B}}^c$   
    add  $[l, s_0^c]$  and  $[s_0^c + 1, u]$  to  $\mathcal{W}^c$   
  **end if**  
**end while**  
**Output:**  $\hat{\mathcal{B}}^c$

---

drawn in Algorithm 2 do not have to be random but taken over a fixed grid.

## 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  $\epsilon_t$  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  $\epsilon_t$  obtained in the PCA algorithm. Define the normalised CUSUM-type statistic using  $\hat{\epsilon}_{tj}$  defined in (2.9):

$$c_{l,u}^{\hat{\epsilon}, \hat{\sigma}}(s; i, j) = \frac{1}{\hat{\sigma}_{l,u}(i, j)} \sqrt{\frac{(s-l+1)(u-s)}{(u-l+1)}} \left( \frac{1}{s-l+1} \sum_{t=l}^s \hat{\epsilon}_{ti} \hat{\epsilon}_{tj} - \frac{1}{u-s} \sum_{t=s+1}^u \hat{\epsilon}_{ti} \hat{\epsilon}_{tj} \right) \quad (2.15)$$

for  $1 \leq l \leq s < u \leq n$  and  $1 \leq i, j \leq d$ , where  $\hat{\sigma}_{l,u}(i, j)$  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 (2.13) to detect breaks in the common components since the estimated factors have been normalised. We choose  $\hat{\sigma}_{l,u}(i, j)$  as the differential error median absolute deviation using  $\hat{\epsilon}_{tj}$ , 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  $1 \leq i, j \leq d$  and  $1 \leq l < u \leq n$  with probability tending to one. As the dimension  $d \rightarrow \infty$  and might be much larger than the sample size  $n$ , a direct aggregation of the CUSUM statistics  $c_{l,u}^{\hat{\epsilon}, \hat{\sigma}}(s; i, j)$  over  $i$  and  $j$  might not perform well in detecting the breaks in particular when the sparsity condition is imposed. Hence, we compare  $c_{l,u}^{\hat{\epsilon}, \hat{\sigma}}(s; i, j)$  with a thresholding parameter, say  $\xi_n^e$ , and delete the index pair  $(i, j)$  when  $\max_{l \leq t < u} |c_{l,u}^{\hat{\epsilon}, \hat{\sigma}}(t; i, j)|$  is smaller than  $\xi_n^e$ . 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

$$C_{l_m, u_m}^{\hat{\epsilon}}(s) = \sum_{i=1}^d \sum_{j=i}^d \left| c_{l_m, u_m}^{\hat{\epsilon}, \hat{\sigma}}(s; i, j) \right|^2 \mathcal{J} \left( \max_{l \leq t < u} |c_{l,u}^{\hat{\epsilon}, \hat{\sigma}}(t; i, j)| > \xi_n^e \right), \quad (2.16)$$

where  $\mathcal{J}(\cdot)$  is an indicator function. It is easy to see that the truncation component in (2.16) is independent of the random sub-intervals  $[l_m, u_m]$ . Therefore, if  $\max_{l \leq t < u} |c_{l,u}^{\hat{\epsilon}, \hat{\sigma}}(t; i, j)| \leq \xi_n^e$ , we have  $C_{l_m, u_m}^{\hat{\epsilon}}(s) = 0$ , 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 details are given in Algorithm 3.

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

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

The thresholding parameter  $\xi_n^e$  in WSBS-Cov plays a similar role to  $\xi_n^c$  in WBS-Cov. With the WSBS-Cov algorithm, we obtain the estimated break points denoted by  $\hat{\eta}_1^e, \dots, \hat{\eta}_{\hat{K}_2}^e$

---

**Algorithm 3** WSBS-Cov algorithm

---

**Input:**  $\xi_n^e, M_n^e, \hat{\epsilon}_t, t = 1, \dots, n$   
generate random intervals  $[l_m, u_m]$  within  $[1, n]$  for  $m = 1, \dots, M_n^e$   
set  $\mathcal{W}^e = \{[1, n]\}$  and  $\hat{\mathcal{B}}^e = \emptyset$   
**while**  $\mathcal{W}^e \neq \emptyset$  **do**  
  take an element from  $\mathcal{W}^e$  and denote it as  $[l, u]$   
   $\mathcal{M}_{l,u}^e = \{m : [l_m, u_m] \subset [l, u]\}$   
   $(m_0^e, s_0^e) = \arg \max_{m \in \mathcal{M}_{l,u}^e, l_m \leq s < u_m} C_{l_m, u_m}^{\hat{\epsilon}}(s)$   
  **if**  $C_{l_m_0^e, u_m_0^e}^{\hat{\epsilon}}(s_0^e) > 0$  **then**  
    add  $s_0^e$  to  $\hat{\mathcal{B}}^e$   
    add  $[l, s_0^e]$  and  $[s_0^e + 1, u]$  to  $\mathcal{W}^e$   
  **end if**  
**end while**  
**Output:**  $\hat{\mathcal{B}}^e$

---

(arranged in an increasing order), where  $\hat{K}_2$  is the estimate of the break number  $K_2$ . Finally, the sets of break points  $\mathcal{B}^c$  and  $\mathcal{B}^e$  defined in (2.5) are estimated by

$$\hat{\mathcal{B}}^c = \{\hat{\eta}_1^c, \dots, \hat{\eta}_{\hat{K}_1}^c\} \quad \text{and} \quad \hat{\mathcal{B}}^e = \{\hat{\eta}_1^e, \dots, \hat{\eta}_{\hat{K}_2}^e\}. \quad (2.17)$$

### 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 of the supplemental document. 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  $\mathbf{F}_t^*$  and factor loadings  $\Lambda^*$  in (2.6).

**PROPOSITION 3.1.** *Suppose that Assumption 2 in Appendix A is satisfied and let  $\kappa_n^c = \min_{1 \leq k \leq K_1+1} (\eta_k^c - \eta_{k-1}^c)$ . The transformed factor loading matrix  $\Lambda^*$  is of full column rank and the transformed factors  $\mathbf{F}_t^*$  satisfy that  $\|\sum_{t=1}^n \mathbf{F}_t^* \mathbf{F}_t^{*\top}\|_F = O_P(n)$  when  $\kappa_n^c \rightarrow \infty$ . If, in*

addition,  $\kappa_n^c \asymp n$ , there exists a  $q_0 \times q_0$  positive definite matrix  $\Sigma_F$  such that

$$\frac{1}{n} \sum_{t=1}^n \mathbf{F}_t^* \mathbf{F}_t^{*\top} \xrightarrow{P} \Sigma_F \text{ as } n \rightarrow \infty. \quad (3.1)$$

The above proposition shows that the sample second-order moment of the transformed factors converges in probability, which is often 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  $\Sigma_F$ . If the order of the minimum distance between breaks is smaller than  $n$ , the limit matrix  $\Sigma_F$  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:

$$\mathbf{C}_{l,u}^{\mathbf{H}\mathbf{F}^*}(s) = \sqrt{\frac{(s-l+1)(u-s)}{u-l+1}} \left[ \frac{1}{s-l+1} \sum_{t=l}^s \text{vech}(\mathbf{H}\mathbf{F}_t^* \mathbf{F}_t^{*\top} \mathbf{H}^\top) - \frac{1}{u-s} \sum_{t=s+1}^u \text{vech}(\mathbf{H}\mathbf{F}_t^* \mathbf{F}_t^{*\top} \mathbf{H}^\top) \right], \quad (3.2)$$

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

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

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  $\mathcal{X}_n \mathcal{X}_n^\top / (nd)$  arranged in a descending order, and  $\boldsymbol{\Lambda}^* = (\lambda_1^*, \dots, \lambda_d^*)^\top$  with  $\lambda_j^*$  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  $\kappa_n^c \asymp n$ , then we have*

$$\max_{(l,u): 1 \leq l < u \leq n} \max_{s: l \leq s < u} \left\| \mathbf{C}_{l,u}^{\widehat{\mathbf{F}}} (s) - \mathbf{C}_{l,u}^{\mathbf{H}\mathbf{F}^*} (s) \right\|_2 = O_P(1). \quad (3.4)$$

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  $\mathbf{C}_{l,u}^{\mathbf{H}\mathbf{F}^*}(s)$  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  $\xi_n^c$  satisfies

$$\underline{c}_1 \log^2 n \leq \xi_n^c \leq \bar{c}_1 (\kappa_n^c \underline{\omega}_n^c)^{1/2}, \quad (3.5)$$

where  $\kappa_n^c$  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  $\iota^c$  such that

$$\mathbf{P} \left( \widehat{K}_1 = K_1; \max_{1 \leq k \leq K_1} |\widehat{\eta}_k^c - \eta_k^c| < \iota^c \varphi_n^c \right) \rightarrow 1, \quad (3.6)$$

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 the WBS-Cov algorithm in Section 2.3 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 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  $\epsilon_{tj}$ ,

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

for  $1 \leq l \leq s < u \leq n$  and  $1 \leq i, j \leq d$ , where  $\widehat{\sigma}_{l,u}(i, j)$  is defined as in (2.15), and

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

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  $\kappa_n^e \asymp n$ , then we have*

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

Proposition 3.3 above plays a crucial role in our asymptotic analysis, indicating that the CUSUM statistics  $c_{l,u}^{\hat{e},\hat{\sigma}}(s; i, j)$  can be replaced by the infeasible ones  $c_{l,u}^{e,\hat{\sigma}}(s; i, j)$  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,  $\kappa_n^e \asymp n$ , and there exist two positive constants  $\underline{c}_2$  and  $\bar{c}_2$  such that*

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

where  $\kappa_n^e = \min_{1 \leq k \leq K_2+1} (\eta_k^e - \eta_{k-1}^e)$  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  $\iota^e$  such that

$$\mathbf{P} \left( \hat{K}_2 = K_2; \max_{1 \leq k \leq K_2} |\hat{\eta}_k^e - \eta_k^e| < \iota^e \varphi_{n,d}^e \right) \rightarrow 1, \quad (3.11)$$

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 Theorem 3.1. If the dimension diverges to infinity at a polynomial rate of  $n$  (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  $\hat{\sigma}_{l,u}(i, j)$  in (2.15) and the thresholding parameter

$\xi_n^e$  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 of multivariate and high-dimensional binary segmentation. 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  $\hat{\eta}_{1|k}^c, \dots, \hat{\eta}_{k|k}^c$  with the convention of  $\hat{\eta}_{0|k}^c = 0$  and  $\hat{\eta}_{k+1|k}^c = n$ . Letting

$$\mathbf{Z}_t^{\hat{F}} = \text{vech} \left( \hat{\mathbf{F}}_t \hat{\mathbf{F}}_t^\top \right) = \left( \mathbf{Z}_{t,1}^{\hat{F}}, \dots, \mathbf{Z}_{t,q_0(q_0+1)/2}^{\hat{F}} \right)^\top \quad \text{and} \quad \mathbf{Z}_{\bullet,j}^{\hat{F}} = \left( \mathbf{Z}_{1,j}^{\hat{F}}, \dots, \mathbf{Z}_{n,j}^{\hat{F}} \right)^\top$$

for  $j = 1, \dots, q_0(q_0+1)/2$ , we define the SSIC objective function for the univariate process  $\mathbf{Z}_{\bullet,j}^{\hat{F}}$  by

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

where

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

for  $\hat{\eta}_{i|k}^c + 1 \leq t \leq \hat{\eta}_{i+1|k}^c$ ,  $i = 0, \dots, k$ ,  $p(n) = n^{1/2}$  is a penalty function dependent on  $n$ , 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

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

no further break exists in any dimension of the transformed factors. Let  $\hat{K}_1$  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  $\hat{K}_1$ . An advantage of using the above SSIC is to avoid choosing the tuning parameter  $\xi_n^c$  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  $\mathbf{Z}_{t,j}^{\hat{F}}$  and  $\bar{\mathbf{Z}}_{t,j}^{\hat{F}}(k)$  replaced by

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

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

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

for all  $1 \leq i, j \leq d$ , 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

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

in which  $\text{med}_{l \leq t < u}(\cdot)$  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 of the supplement. 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  $\Delta_n$  as in [Barigozzi, Cho and Fryzlewicz \(2018\)](#), to control the minimum length of the random intervals. Specifically, Algorithm 4 below is used to produce a set of random intervals in WBS-Cov (or WSBS-Cov).

---

**Algorithm 4** Drawing random intervals

---

**Input:**  $n$ ,  $\Delta_n$  and  $M_n$  ( $M_n = M_n^c$  or  $M_n^e$ )

Draw  $M_n$  pairs of random points uniformly from the set  $\{1, 2, \dots, n - 4\Delta_n\}$ ;

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  $m = 1, \dots, M_n$

---

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  $\Delta_n$  to both boundaries, indicating that we only maximise the CUSUM statistics over the interval  $[l_m + \Delta_n, u_m - \Delta_n]$ . Note that  $\Delta_n$  should be much smaller than  $\kappa_n^c$  and  $\kappa_n^e$ . Thus we choose  $\Delta_n = \lfloor (\log^2 n) \wedge (0.25n^{6/7}) \rfloor$  following the discussion in Appendix D of [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  $\xi_n^e$ . [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  $\xi_n^e$ : (i) pre-detect the breaks (e.g., combining the classical BS-Cov and SSIC without knowing  $\xi_n^e$  a priori) and remove the breaks from the high-dimensional covariance matrices through demeaning using the estimated breaks, i.e., generate  $Z_{t,ij}^{\hat{e}} - \bar{Z}_{t,ij}^{\hat{e}}(\tilde{K}_n^e)$  as in (4.3), where  $\tilde{K}_n^e$  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  $\left\{ Z_{t,ij}^{\hat{e}} - \bar{Z}_{t,ij}^{\hat{e}}(\tilde{K}_n^e) \right\}$  for each pair  $(i, j)$  and then take the maximum as the chosen thresholding parameter  $\xi_n^e$ . This technique is more computationally intensive than any formula-based threshold selection procedures; however, we have found that it performs particularly well in practice, as reported in Section 5.

## 5 Simulation studies

In this section we provide simulation studies to compare the finite-sample performance between the proposed methods and various other competing methods. In particular, we compare the numerical performance among the proposed WBS-Cov and WSBS-Cov, BS-Cov and SBS-Cov algorithms, and examine the finite-sample influences of different norms used in aggregation of the CUSUM quantities and various transformation techniques used in construction of the CUSUM statistics.

EXAMPLE 5.1. Consider the following approximate factor model to generate data:

$$X_{ti} = \sum_{j=1}^r \lambda_{ij,t} F_{tj} + \epsilon_{ti}, \quad i = 1, \dots, d, \quad t = 1, \dots, n, \quad (5.1)$$

where  $r = 5$  is the number of common factors, each factor process is generated via an AR(1) model:

$$F_{tj} = \rho_j F_{t-1,j} + u_{tj}, \quad t = 1, \dots, n, \quad (5.2)$$

with  $\rho_j = 0.4 - 0.05(j-1)$  for  $j = 1, \dots, 5$ , and  $u_{tj}$  following a standard normal distribution independently over  $t$  and  $j$ . The sample size is  $n = 200$  and the dimension is  $d = 200$ . In this example, we consider the scenario of a single break in both the common and idiosyncratic components:  $\eta_1^c = \lfloor n/3 \rfloor + 1 = 67$  and  $\eta_1^e = \lfloor 2n/3 \rfloor = 133$ . The factor loadings  $\lambda_{ij,t}$  are first generated from a standard normal distribution independently over  $i$  and  $j$  when  $t$  is from 1 to  $\eta_1^c$ ; whereas after the break point  $\eta_1^c$ , the factor loadings  $\lambda_{ij,t}$  are shifted by a random amount  $N(0, 4)$  as in [Barigozzi, Cho and Fryzlewicz \(2018\)](#). The sudden change on the factor loadings leads to break in the second-order moment structure of the common components. The idiosyncratic errors  $\epsilon_t$  follow a multivariate normal distribution  $N_d(\mathbf{0}, \Sigma_\epsilon)$  independently over  $t$ , where  $\phi_j$ , the square root of the  $j$ -th diagonal element of  $\Sigma_\epsilon$ , is generated from an independent uniform distribution  $U(0.5, 1.5)$ , and the  $(i, j)$ -entry of  $\Sigma_\epsilon$  is  $\phi_i \phi_j (-0.5)^{|i-j|}$  for  $1 \leq i \neq j \leq d$ . After the break point  $\eta_1^e$ , we swap the orders of  $\lfloor \rho_1^e d/2 \rfloor$  randomly selected pairs of elements of  $\epsilon_t$  (c.f., [Cho and Fryzlewicz, 2015](#)) with  $\rho_1^e$  chosen as 0.1, 0.5 or 1. Note that  $\rho_1^e = 0.1$  indicates that the structural breaks are relatively sparse in the high-dimensional error components, whereas  $\rho_1^e = 1$  indicates that the breaks are dense. The replication number in the simulation is  $R = 100$ .

For the 100 simulated samples, we report the estimated number of break(s) as well as the accuracy measure defined by

$$\text{ACU} = 100\% \times \left[ \frac{1}{100} \cdot \sum_{i=1}^{100} \mathcal{J} \left( \min_{k=1, \dots, \hat{K}} |\hat{\eta}_k(i) - \eta_1| \leq \log n \right) \right], \quad (5.3)$$

where  $\eta_1$  denotes either  $\eta_1^c$  or  $\eta_1^e$ ,  $\hat{\eta}_k(i)$  is the corresponding estimate in the  $i$ -th simulated sample, and  $\hat{K}$  denotes the estimated break number  $\hat{K}_1$  or  $\hat{K}_2$ . In addition, to measure the estimation error of the break-point set more precisely, we use the Hausdorff distance:

$$\text{HD}(\mathcal{B}, \hat{\mathcal{B}}) = \max \left\{ \delta(\mathcal{B}, \hat{\mathcal{B}}), \delta(\hat{\mathcal{B}}, \mathcal{B}) \right\} \quad \text{with} \quad \delta(\mathcal{B}, \hat{\mathcal{B}}) = \max_{a \in \mathcal{B}} \left\{ \min_{b \in \hat{\mathcal{B}}} |a - b| \right\}, \quad (5.4)$$

where  $\mathcal{B}$  denotes the set of true breaks in either the common or idiosyncratic components whereas  $\hat{\mathcal{B}}$  denotes its estimate. Note that  $\text{HD}(\mathcal{B}, \hat{\mathcal{B}})$  takes the infinite value in the extreme case when  $\hat{\mathcal{B}} = \emptyset$ . To avoid the latter, we only present the average conditional Hausdorff distance (ACHD) over 100 replications conditional on that the number of breaks is correctly estimated.

Table 1: Comparison of detection results using different BS-based methods

		# break(%)			ACHD	ACU(%)
		< 1	1	> 1		
Breaks in common components						
$\rho_1^e = 1$	BS-Cov	0	99	1	0.202	100
	WBS-Cov	0	99	1	0.202	100
$\rho_1^e = 0.5$	BS-Cov	0	100	0	0.210	100
	WBS-Cov	0	100	0	0.210	100
$\rho_1^e = 0.1$	BS-Cov	0	99	1	0.212	100
	WBS-Cov	0	99	1	0.212	100
Breaks in idiosyncratic error components						
$\rho_1^e = 1$	BS-Cov	0	97	3	0.124	100
	WBS-Cov	0	98	2	0.663	98
	SBS-Cov	0	97	3	0.278	100
	WSBS-Cov	0	99	1	0.313	100
$\rho_1^e = 0.5$	BS-Cov	0	96	4	0.260	99
	WBS-Cov	0	94	6	1.255	95
	SBS-Cov	0	99	1	0.646	98
	WSBS-Cov	0	100	0	0.660	98
$\rho_1^e = 0.1$	BS-Cov	24	72	4	8.306	53
	WBS-Cov	28	70	2	13.229	31
	SBS-Cov	20	80	0	5.175	61
	WSBS-Cov	20	80	0	4.163	61

In Table 1, we compare the proposed WBS-Cov with the classical BS-Cov in detecting breaks in the common components, and compare the proposed WSBS-Cov with the BS-Cov, WBS-Cov and SBS-Cov in detecting breaks in the idiosyncratic components. For the break detection in the common component, the finite-sample performance of WBS-Cov and BS-Cov are the same. For the break detection in the idiosyncratic components, the four methods behave differently in finite samples. When the breaks are sparse in the high-dimensional error covariance matrix ( $\rho_1^e = 0.1$ ), the sparsified detection techniques (WSBS-Cov and SBS-Cov) outperform the non-sparsified ones (BS-Cov and WBS-Cov) in both the break number and location estimation; when the breaks are dense ( $\rho_1^e = 0.5$  and 1), the proposed WSBS-Cov has the best performance in estimating the break number whereas the BS-Cov performs better than the other three methods in estimating the break location.

In Table 2, we examine the finite-sample influence of different norms used in the aggregation of the CUSUM quantities. For the idiosyncratic components, as in (2.16), the CUSUM statistic aggregated with the  $l_1$ -norm is defined by

$$\sum_{i=1}^d \sum_{j=i}^d \left| c_{l_m, u_m}^{\hat{e}, \hat{\sigma}}(s; i, j) \right| \mathcal{J} \left( \max_{l \leq t < u} \left| c_{l, u}^{\hat{e}, \hat{\sigma}}(t; i, j) \right| > \xi_n^e \right)$$

Table 2: Comparison of detection results using different norms in the CUSUM statistics

		# break(%)			ACHD	ACU(%)
		< 1	1	> 1		
Breaks in common components						
$\rho_1^e = 1$	$l_1$	0	99	1	0.697	99
	$l_2$	0	99	1	0.202	100
	$l_\infty$	0	79	21	5.443	64
	op	0	100	0	0.460	99
$\rho_1^e = 0.5$	$l_1$	0	100	0	0.730	100
	$l_2$	0	100	0	0.210	100
	$l_\infty$	0	79	21	4.367	69
	op	0	99	1	0.414	99
$\rho_1^e = 0.1$	$l_1$	0	99	1	0.737	100
	$l_2$	0	99	1	0.212	100
	$l_\infty$	0	78	22	4.962	65
	op	0	100	0	0.490	98
Breaks in idiosyncratic error components						
$\rho_1^e = 1$	$l_1$	0	99	1	0.465	98
	$l_2$	0	99	1	0.313	100
	$l_\infty$	0	90	10	3.600	77
	op	0	89	11	2.652	89
$\rho_1^e = 0.5$	$l_1$	0	97	3	1.062	94
	$l_2$	0	100	0	0.660	98
	$l_\infty$	0	94	6	4.617	69
	op	0	94	6	3.585	74
$\rho_1^e = 0.1$	$l_1$	23	72	5	6.944	50
	$l_2$	20	80	0	4.163	61
	$l_\infty$	23	75	2	8.067	42
	op	32	66	2	6.985	39

and the CUSUM statistic aggregated with the  $l_\infty$ -norm is defined by

$$\max_{1 \leq i \leq j \leq d} \left\{ \left| c_{l_m, u_m}^{\hat{e}, \hat{\sigma}}(s; i, j) \right| \mathcal{J} \left( \max_{l \leq t < u} \left| c_{l, u}^{\hat{e}, \hat{\sigma}}(t; i, j) \right| > \xi_n^e \right) \right\};$$

and the construction is similar for the common components. In addition, we also consider aggregating via the operator norm, as suggested in [Wang, Yu and Rinaldo \(2018\)](#). For the idiosyncratic components, let  $C_{l_m, u_m}^{M, \hat{e}}(s)$  be a  $d \times d$  matrix with the  $(i, j)$ -th entry being

$$c_{l_m, u_m}^{\hat{e}, \hat{\sigma}}(s; i, j) \mathcal{J} \left( \max_{l \leq t < u} \left| c_{l, u}^{\hat{e}, \hat{\sigma}}(t; i, j) \right| > \xi_n^e \right)$$

and then obtain the CUSUM statistic by taking the operator norm of  $C_{l_m, u_m}^{M, \hat{e}}(s)$ . For the common components, the CUSUM statistic is defined by taking the operator norm of the matrix:

$$\sqrt{\frac{(s-l+1)(u-s)}{u-l+1}} \left[ \frac{1}{s-l+1} \sum_{t=l}^s \hat{F}_t \hat{F}_t^\top - \frac{1}{u-s} \sum_{t=s+1}^u \hat{F}_t \hat{F}_t^\top \right].$$



It is obvious from Table 2 that the  $l_2$ -based detection method has the best finite-sample performance with more accurate estimated break number, higher ACU and smaller ACHD. The operator norm based detection method performs well in break detection for the common components, but it performs poorly when breaks are sparse in the idiosyncratic components.

Table 3: Comparison of detection results using different transformations in break detection

		# break(%)			ACHD	ACU(%)
		< 1	1	> 1		
Breaks in common components						
$\rho_1^e = 1$	BCF	0	95	5	0.821	100
	WBS-Cov	0	99	1	0.202	100
	WAVELET	0	92	8	2.065	100
	ADD-MNS	0	99	1	0.313	100
$\rho_1^e = 0.5$	BCF	0	94	6	0.777	100
	WBS-Cov	0	100	0	0.210	100
	WAVELET	0	96	4	2.094	100
	ADD-MNS	0	100	0	0.330	100
$\rho_1^e = 0.1$	BCF	0	96	4	0.865	100
	WBS-Cov	0	99	1	0.212	100
	WAVELET	0	92	8	2.011	100
	ADD-MNS	0	100	0	0.340	100
Breaks in idiosyncratic error components						
$\rho_1^e = 1$	BCF(D)	0	100	0	0.370	100
	BCF	0	100	0	0.330	100
	WSBS-Cov(D)	0	100	0	0.280	99
	WSBS-Cov	0	99	1	0.313	100
	WAVELET	0	93	7	2.484	100
	ADD-MNS	0	83	17	0.458	98
$\rho_1^e = 0.5$	BCF(D)	0	100	0	0.400	100
	BCF	0	100	0	0.430	100
	WSBS-Cov(D)	0	100	0	0.320	100
	WSBS-Cov	0	100	0	0.660	98
	WAVELET	2	89	9	2.034	97
	ADD-MNS	0	90	10	0.956	95
$\rho_1^e = 0.1$	BCF(D)	24	76	0	8.079	55
	BCF	50	50	0	3.640	44
	WSBS-Cov(D)	21	79	0	2.823	65
	WSBS-Cov	20	80	0	4.163	61
	WAVELET	7	77	16	7.364	64
	ADD-MNS	0	87	13	6.080	61

Table 3 reports the simulation result when different transformation techniques are used in construction of the CUSUM statistics. In the table, “BCF” denotes the method proposed by Barigozzi, Cho and Fryzlewicz (2018) which combines the wavelet-based transformation and the double-CUSUM method, “WBS-Cov” denotes the proposed method in Section 2.3, and “WSBS-Cov” denotes the proposed method in Section 2.4. For

structural breaks in the covariance matrix of the error components, we may detect the breaks only for its diagonal elements (variance) rather than all the elements in the high-dimensional covariance matrix in order to save computational time. This is considered in our simulation with “BCF(D)” and “WSBS-Cov(D)” denoting the “BCF” and “WSBS-Cov” methods by only detecting breaks for the diagonal elements. Letting  $a_i$  and  $a_j$  be either the common factors or the idiosyncratic errors, “ADD-MNS” denotes a transformation of  $(a_i + a_j)^2$  and  $(a_i - a_j)^2$  (e.g., [Cho and Fryzlewicz, 2015](#)) in the construction of the CUSUM statistics (instead of  $a_i a_j$  in our proposed method), whereas “WAVELET” denotes the wavelet transformation on  $a_i$  and  $a_j$  (e.g., [Barigozzi, Cho and Fryzlewicz, 2018](#)) in the construction of the CUSUM statistics. The algorithms introduced in Sections 2.3 and 2.4 are used after making the “WAVELET” and “ADD-MINS” transformations. The R package “factorcpt” is used to implement [Barigozzi, Cho and Fryzlewicz \(2018\)](#)’s method.

From the table, the proposed WBS-Cov algorithm has the best finite-sample performance in estimating the break in the common components. In contrast, [Barigozzi, Cho and Fryzlewicz \(2018\)](#)’s method has larger ACHD measure, which might be caused by the use of  $l_1$ -norm aggregation in their double CUSUM algorithm (as we show in Table 2 that the  $l_2$ -norm aggregation is superior to the  $l_1$ -norm aggregation). In terms of the idiosyncratic components, the “WSBS” method has a similar performance to the “BCF” method, and the best performance is from the “WSBS-Cov(D)” method. In terms of “WAVELET” method, we find that the thresholding parameter  $\xi_n^e$  selected in pre-estimation is too small, and thus use  $\sqrt{2}\xi_n^e$  as the threshold. However, this method tends to over-estimate the break number. The “WAVELET” method introduces a systematic error of order  $O(\log \log d)$  (the order of the scale of the wavelet basis used in the algorithm), leading to large ACHD values. The performance of the “ADD-MINS” method in estimating the break location is not as good as the other methods, which might be caused by selection of the thresholding parameter  $\xi_n^e$ .

EXAMPLE 5.2. We still use model (5.1) to generate the data in simulation, where the number of factors is  $r = 5$ , the sample size is  $n = 200$ , the dimension is  $d = 200$  and the replication number is  $R = 100$ . The factor process  $F_t$  is generated from a multivariate normal distribution  $N_5(\mathbf{0}, \Sigma_F^*)$  independently over  $t$ , where  $\Sigma_F^*$  is the covariance matrix with one break specified as follows: for  $1 \leq t \leq \eta_1^c = 67$ ,  $\phi_j^F$ , the square root of the  $j$ -th diagonal element of  $\Sigma_F^*$ , is independently generated from a uniform distribution  $U(0.5, 1.5)$ , and the  $(i, j)$ -entry of  $\Sigma_F^*$  is defined as  $\phi_i^F \phi_j^F (0.5)^{|i-j|}$  for  $1 \leq i \neq j \leq 5$ ; for

$\eta_1^c < t \leq n$ , the (1,2) and (2,1)-entries of  $\Sigma_F^*$  change from  $0.5\phi_1^F\phi_2^F$  to  $0.9\phi_1^F\phi_2^F$ , and  $\phi_5^F$  is replaced by  $1.3\phi_5^F$  resulting in structural breaks in (5, j) and (j, 5)-entries of  $\Sigma_F^*$ . For  $1 \leq t \leq \eta_2^c = 133$ , the factor loadings  $\lambda_{ij}$  are independently generated from a uniform distribution  $U(-1, 1)$ ; whereas for  $\eta_2^c < t \leq n$ , 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.

For the structural breaks in the idiosyncratic error components, we consider two scenarios. In scenario (i), there are two breaks  $\eta_1^e = 67$  and  $\eta_2^e = 133$ , as in Example 5.1 above. At each of the two break points  $\eta_1^e$  and  $\eta_2^e$ , we swap the orders of  $\lfloor \rho_1^e d/2 \rfloor$  randomly selected pairs of elements of  $\epsilon_t$  with  $\rho_1^e$  chosen as 0.2 and 0.8, corresponding to sparse and dense breaks, respectively. In scenario (ii), there are three breaks  $\eta_1^e = \lfloor n/4 \rfloor = 50$ ,  $\eta_2^e = \lfloor n/2 \rfloor = 100$  and  $\eta_3^e = \lfloor 3n/4 \rfloor = 150$ . At each of the break points, as in scenario (i), we swap the orders of  $\lfloor \rho_1^e d/2 \rfloor$  randomly selected pairs of elements of  $\epsilon_t$  with  $\rho_1^e$  chosen as 0.2 and 0.8.

The simulation result for scenario (i) is given in Table 4, where ACU1 and ACU2 are defined as ACU in (5.3) to measure estimation accuracy for the first and second breaks, respectively. We note that a further selection of the penalty in the SSIC objective function may improve our result. The penalty term is now chosen as  $p(n) = \iota_* n^{1/2}$  with  $\iota_* = 1, \sqrt{2}/2$  and  $1/2$ . We compare the finite-sample performance between our method and Barigozzi, Cho and Fryzlewicz (2018)'s method. From Table 4, we find that the proposed WBS-Cov method with  $\iota_* = 1/2$  has the best finite-sample performance with the most accurate estimated break number, highest ACU and smallest ACHD, whereas Barigozzi, Cho and Fryzlewicz (2018)'s method performs poorly in detecting the first break point in the common components, which might be due to the relatively sparse break structure at the first break point. When the breaks in the idiosyncratic components are dense ( $\rho_1^e = 0.8$ ), the proposed WSBS-Cov method with  $\iota_* = 1/2$  has 99% frequency of correctly estimating the true break number, and 98% and 95% for ACU1 and ACU2, respectively, which are comparable to the corresponding results using "BCF". When the breaks in the idiosyncratic components are sparse ( $\rho_1^e = 0.2$ ), the proposed WSBS-Cov algorithm with  $\iota_* = 1/2$  still has 95% frequency of correctly estimating the true number of break points, but "BCF" has only 54%. In addition, from the ACU1 and ACU2 results with  $\rho_1^e = 0.2$ , the WSBS-Cov method (with  $\iota_* = \sqrt{2}/2$  and  $1/2$ ) detects more than 80% of

each of the two break points, whereas “BCF” only detects about 70%.

Table 4: Detection results with 2 breaks in idiosyncratic error components

		# break(%)			ACHD	ACU1(%)	ACU2(%)
		< 2	2	> 2			
Breaks in common components ( $\eta_1^c = 67, \eta_2^c = 133$ )							
$\rho_1^c = 0.8$	BCF	90	10	0	13.400	4	84
	WBS-Cov( $\iota_\star = 1$ )	39	61	0	8.508	36	91
	WBS-Cov( $\iota_\star = \sqrt{2}/2$ )	7	93	0	8.172	57	91
	WBS-Cov( $\iota_\star = 1/2$ )	1	97	2	7.526	61	91
$\rho_1^c = 0.2$	BCF	93	7	0	18.143	2	88
	WBS-Cov( $\iota_\star = 1$ )	35	65	0	8.769	38	94
	WBS-Cov( $\iota_\star = \sqrt{2}/2$ )	10	90	0	8.822	51	94
	WBS-Cov( $\iota_\star = 1/2$ )	1	98	1	9.020	56	94
Breaks in idiosyncratic error components ( $\eta_1^e = 67, \eta_2^e = 133$ )							
$\rho_1^e = 0.8$	BCF	0	100	0	1.210	100	100
	WSBS-Cov( $\iota_\star = 1$ )	5	95	0	1.200	97	93
	WSBS-Cov( $\iota_\star = \sqrt{2}/2$ )	1	99	0	1.343	98	95
	WSBS-Cov( $\iota_\star = 1/2$ )	1	99	0	1.343	98	95
$\rho_1^e = 0.2$	BCF	46	54	0	2.722	72	70
	WSBS-Cov( $\iota_\star = 1$ )	44	56	0	4.982	66	68
	WSBS-Cov( $\iota_\star = \sqrt{2}/2$ )	13	84	3	4.798	84	81
	WSBS-Cov( $\iota_\star = 1/2$ )	1	95	4	3.958	90	86

The simulation result for scenario (ii) is given in Table 5, where ACU1, ACU2 and ACU3 are defined as ACU in (5.3) to measure estimation accuracy for the first, second and third breaks, respectively. In this case, we also choose  $\iota_\star = 1, \sqrt{2}/2$  and  $1/2$  in the penalty term  $p(n) = \iota_\star n^{1/2}$  of the SSIC objective function. From Table 5, we find that the proposed WBS-Cov method with  $\iota_\star = 1/2$  again has the best finite-sample performance in detecting breaks in the common components with the most accurate estimated break number, highest ACU and smallest ACHD. For the break detection in the idiosyncratic error components, when  $\rho_1^e = 0.2$ , the proposed WSBS-Cov method is better than “BCF”; but when  $\rho_1^e = 0.8$ , “BCF” outperforms the WSBS-Cov method.

## 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. 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

Table 5: Detection results with 3 breaks in idiosyncratic error components

		# break(%)			ACHD	ACU1(%)	ACU2(%)	
		< 2	2	> 2				
Breaks in common component ( $\eta_1^c = 67, \eta_2^c = 133$ )								
$\rho_1^e = 0.8$	BCF	89	11	0	13.636	4	86	
	WBS-Cov( $\iota_* = 1$ )	41	59	0	9.254	34	93	
	WBS-Cov( $\iota_* = \sqrt{2}/2$ )	8	92	0	3.661	55	93	
	WBS-Cov( $\iota_* = 1/2$ )	2	96	2	2.855	60	93	
$\rho_1^e = 0.2$	BCF	4	89	7	16.000	3	85	
	WBS-Cov( $\iota_* = 1$ )	40	60	0	8.383	34	91	
	WBS-Cov( $\iota_* = \sqrt{2}/2$ )	10	90	0	8.267	52	91	
	WBS-Cov( $\iota_* = 1/2$ )	1	97	2	7.856	59	91	
		# break(%)			ACHD	ACU1(%)	ACU2(%)	ACU3(%)
		< 3	3	> 3				
Breaks in idiosyncratic error components ( $\eta_1^e = 50, \eta_2^e = 100, \eta_3^e = 150$ )								
$\rho_1^e = 0.8$	BCF	0	100	0	1.330	100	100	100
	WSBS-Cov( $\iota_* = 1$ )	55	45	0	3.956	60	78	88
	WSBS-Cov( $\iota_* = \sqrt{2}/2$ )	44	56	0	3.661	61	90	95
	WSBS-Cov( $\iota_* = 1/2$ )	44	55	1	2.855	61	90	95
$\rho_1^e = 0.2$	BCF	91	9	0	3.000	46	68	41
	WBS-Cov( $\iota_* = 1$ )	95	5	0	11.800	41	26	44
	WBS-Cov( $\iota_* = \sqrt{2}/2$ )	66	34	0	8.853	47	57	70
	WBS-Cov( $\iota_* = 1/2$ )	49	51	0	9.137	51	70	70

change (without reinvesting of dividends) of equity  $i$  at time  $t$ , where  $i = 1, \dots, 375$  and  $t = 1, \dots, 5322$ . We fit the factor model (2.2) to the data with  $\mathbf{X}_t = (X_{1,t}, \dots, X_{375,t})^\top$ , allowing structural breaks in the covariance matrices of both the common and idiosyncratic components.

As suggested in Example 5.2, we choose the penalty term as  $p(n) = \frac{1}{2}n^{1/2}$  in the SSIC objective function defined in Section 4, and select  $M_n^c = M_n^e = 1000$  in the WBS-Cov and WSBS-Cov algorithms. In order to analyse the impact of the recent COVID-19 outbreak on the stock market, we set  $\Delta_n = 20$  which may be relatively small in contrast with the sample size  $n = 5322$ . The information criterion introduced in Section 2.2 selects nine factors (with appropriate transformation), i.e.,  $\hat{q} = 9$ . The WBS-Cov algorithm combined with SSIC detects five breaks in the covariance matrix of the common components, and the WSBS-Cov algorithm detects eight 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 eight breaks in the idiosyncratic

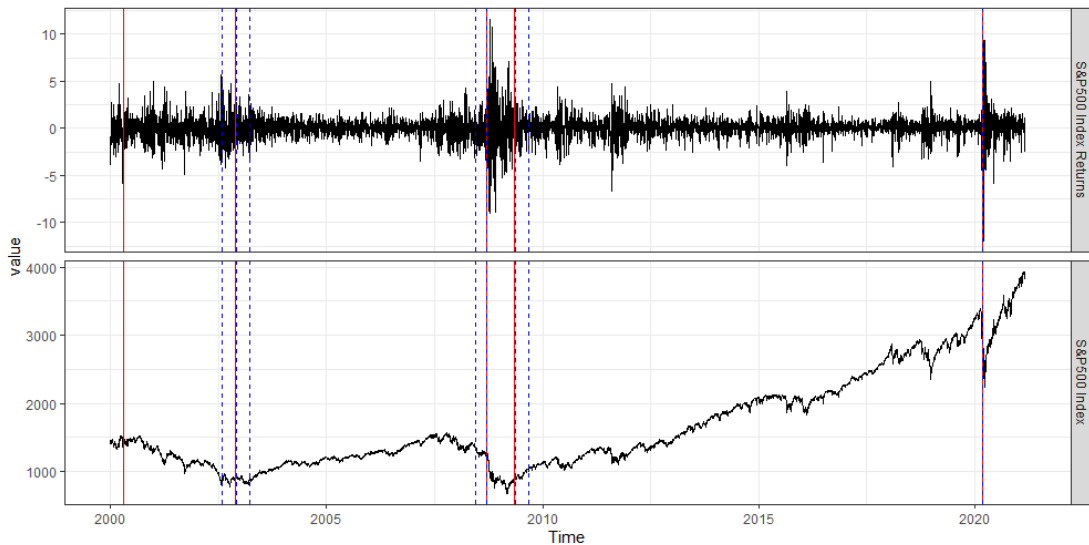


Figure 1: The detected break times in the covariance structure of the 375 daily returns of the S&P 500 constituents from 01/01/2000 to 28/02/2021. The estimated break dates for common components are 24/04/2000, 25/11/2002, 12/09/2008, 08/05/2009 and 06/03/2020 (in red lines); and the estimated break dates for idiosyncratic components are 06/08/2002, 29/11/2002, 25/03/2003, 11/06/2008, 11/09/2008, 11/05/2009, 01/09/2009 and 06/03/2020 (in blue dotted lines).

errors, the first three breaks occur around the end of 2002, the fourth to seventh breaks occur in 2008 and 2009, and the most recent one occurs at the time of COVID-19 outbreak.

As a benchmark, we also download eight observable risk factors (over 01/01/2000–28/02/2021) from Kenneth R. French’s data library<sup>1</sup> 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) and Fama and French (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 (LTRRev) 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 these factors in Figure 2, from which we can see that they variate drastically during the global financial crisis and the COVID-19 outbreak. Let  $\mathbf{Z}_t = (Z_{1t}, \dots, Z_{8t})^\top$  denote the vector containing the eight risk factors,  $t = 1, \dots, 5322$ . As the dimension of  $\mathbf{Z}_t$  is relatively 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

<sup>1</sup>[http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

detect six break dates: 19/04/2001, 17/10/2002, 01/07/2008, 12/09/2008, 13/05/2009 and 06/03/2020. Using the WSBS-Cov algorithm (without any latent factor structure) in Section 2.4, the same six break dates are obtained. Except for the first estimated break, they coincide with the breaks plotted in Figure 2. However, the estimated break number is smaller than that obtained by our method which assumes the factor model structure. This shows that break detection of low-dimensional observable risk factor series may be not sufficient to 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

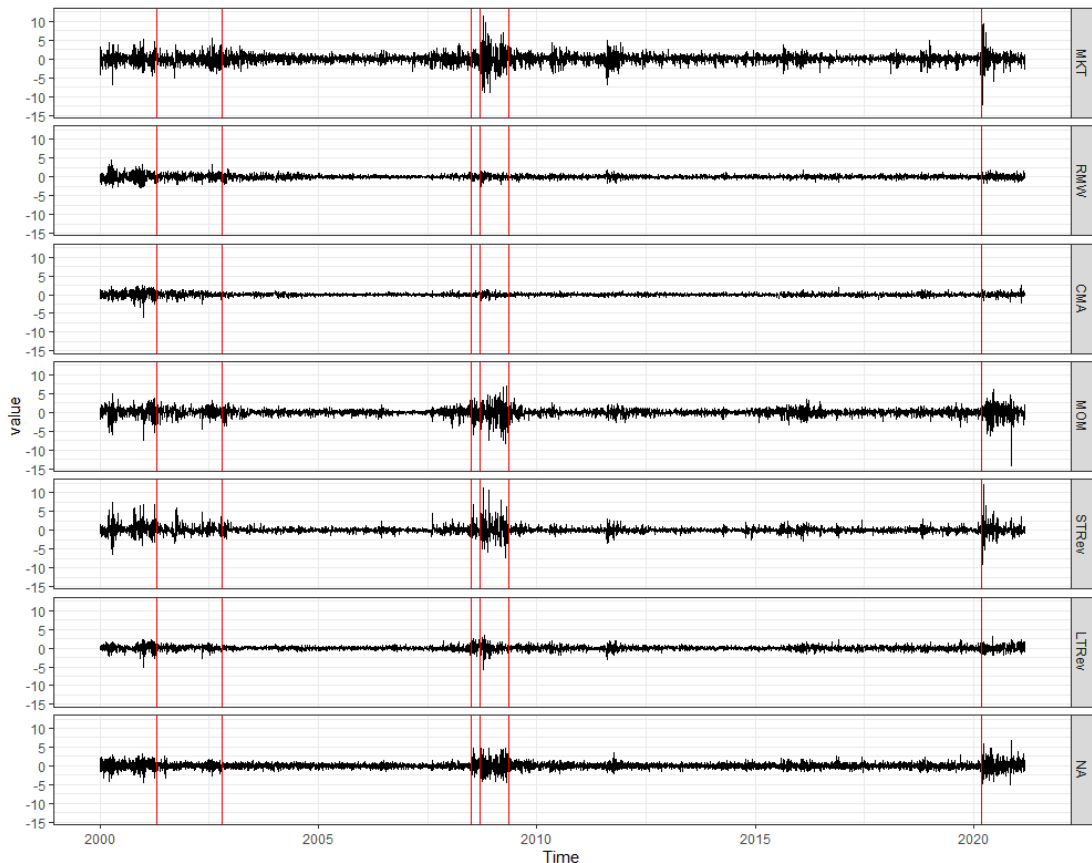


Figure 2: Variation of eight commonly-used risk factors from 01/01/2000 to 28/02/2021. The estimated break dates are 19/04/2001, 17/10/2002, 01/07/2008, 12/09/2008, 13/05/2009 and 06/03/2020 (in red lines).

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, 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 high-dimensional model 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 (available in the supplement), 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 other existing approaches in particular the method introduced by Barigozzi, Cho and Fryzlewicz (2018). In the empirical application to the return series of S&P 500 constituent, we detect five break in the common components and eight breaks in the idiosyncratic error components by the developed methods.

## Supplementary document

The supplemental document contains the technical assumptions, 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.
- 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 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.
- BAI, J., AND ZHOU, G. (2015). Fama-MacBeth two-pass regressions: Improving risk premia estimates. *Finance Research Letters* **15**, 31–40.
- 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 ROTHSCILD, 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.
- 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>.
- 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.
- FAMA, E. F. AND MACBETH, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* **81**, 607–636.
- 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.
- 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. Forthcoming in *Journal of the Korean Statistical Society*.
- HAN X. AND INOUE, A. (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory* **31**, 1117–1152.
- 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.
- LAM, C. AND YAO, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Annals of Statistics* **40**, 694–726.
- 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.
- 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. (2018). Optimal covariance change point localization in high dimensions. Working paper available at <https://arxiv.org/pdf/1712.09912.pdf>.
- 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.