Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Inference of Breakpoints in High-dimensional Time Series

Likai Chen, Weining Wang & Wei Biao Wu

To cite this article: Likai Chen, Weining Wang & Wei Biao Wu (2022) Inference of Breakpoints in High-dimensional Time Series, Journal of the American Statistical Association, 117:540, 1951-1963, DOI: 10.1080/01621459.2021.1893178

To link to this article: https://doi.org/10.1080/01621459.2021.1893178

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2022, VOL. 117, NO. 540, 1951–1963: Theory and Methods
https://doi.org/10.1080/01621459.2021.1893178

Inference of Breakpoints in High-dimensional Time Series

Likai Chen[a], Weining Wang[b], and Wei Biao Wu[c]

[a]Department of Mathematics and Statistics, Washington University in St. Louis, MO; [b]Department of Economics and Related Studies, University of York, New York; [c]Department of Statistics, University of Chicago, Chicago, IL

ABSTRACT

For multiple change-points detection of high-dimensional time series, we provide asymptotic theory concerning the consistency and the asymptotic distribution of the breakpoint statistics and estimated break sizes. The theory backs up a simple two-step procedure for detecting and estimating multiple change-points. The proposed two-step procedure involves the maximum of a MOSUM (moving sum) type statistics in the first step and a CUSUM (cumulative sum) refinement step on an aggregated time series in the second step. Thus, for a fixed time-point, we can capture both the biggest break across different coordinates and aggregating simultaneous breaks over multiple coordinates. Extending the existing high-dimensional Gaussian approximation theorem to dependent data with jumps, the theory allows us to characterize the size and power of our multiple change-point test asymptotically. Moreover, we can make inferences on the breakpoints estimates when the break sizes are small. Our theoretical setup incorporates both weak temporal and strong or weak cross-sectional dependence and is suitable for heavy-tailed innovations. A robust long-run covariance matrix estimation is proposed, which can be of independent interest. An application on detecting structural changes of the U.S. unemployment rate is considered to illustrate the usefulness of our method.

ARTICLE HISTORY
Received December 2019
Accepted February 2021

KEYWORDS
Gaussian approximation; Inference of break locations; Multiple change-point detection; Temporal and cross-sectional dependence

1. Introduction

Statistical inference of structural breaks in mean is an important subject to study, and involves estimating the trend functions, detecting and locating abnormal changes and making inferences on the break estimators. Breaks may arise in various applications in different fields, such as in network analysis, biology, engineering, economics and finance, among others. Specific examples are anomaly of network traffic data caused by attacks (Lévy-Leduc and Roueff 2009), recurrent DNA copy number variants in multiple samples (Zhang et al. 2010), abrupt changes in household electrical power consumption (Harlé et al. 2016) and minimum wage policy changes analysis (Chen, Wang, and Wu 2020), etc. In those data scenarios, temporal and cross-sectional dependence for large-dimensional data might pose challenges to statistical analysis.

To formulate our problem, we assume that observation vectors $Y_1, Y_2, \ldots, Y_n$ follow the model,

$$Y_t = \mu(t/n) + \epsilon_t, \quad t = 1, 2, \ldots, n, \tag{1}$$

where $(\epsilon_t)_t$ is a sequence of zero-mean p-dimensional stationary noise vectors and $\mu(\cdot) = (\mu_1(\cdot), \mu_2(\cdot), \ldots, \mu_p(\cdot))^\top : [0,1] \to \mathbb{R}^p$ is a vector of unknown trend functions. In this way, the data-generating process is trend stationary. We will model breaks occurring on the vector of trend functions $\mu(t/n)$. Notably, we assume that the trend function satisfies

$$\mu(u) = f(u) + \sum_{i=1}^{K_0} \gamma_i \mathbf{1}_{u \geq u_i}, \tag{2}$$

where K_0 is an unknown integer representing the number of breaks; $f(\cdot)$ $(f(\cdot) = (f_1(\cdot), f_2(\cdot), \ldots, f_p(\cdot))^\top : [0,1] \to \mathbb{R}^p)$ is a vector of smooth trend functions; u_ks with $0 < u_1 < u_2 < \cdots < u_{K_0} < 1$ are the time stamps of the change-points with $|u_i - u_j| \gg b$, where b is the bandwidth parameter; and $\gamma_k \in \mathbb{R}^p$ are the jump vectors with size $|\gamma_k|_\infty$ ($|.|_\infty$ is the infinity norm) at point u_k. Note that the jump sizes are characterized in terms of the infinity norm; therefore, we do not require simultaneous jumps for all entities $1 \leq j \leq p$, and some coordinates of γ_k can be zero. Namely, we will focus on the largest jump (i.e., $|\gamma_k|_\infty$) happening in the cross-sectional dimension for any fixed time point k (see Theorem 2), and this is of particular interest when the jumps are sparse. In case many series jump at the same time, we further propose a refined method, which aggregates all the contemporaneous jumps (cf. Theorem 4). In most of the change-point settings, the smooth part of the trend functions is zero (i.e., $f \equiv 0$). This means that the trend functions are piecewise constant for each coordinate. In contrast, our model is more flexible and realistic, since we allow the mean functions to vary smoothly instead of staying at the same level between break-points.

The goal of this paper is to provide theory for structural break inference. We first detect the existence of breaks. We then deliver theorems to test for the existence of breaks, identify their change-point u_k, calibrate sizes of the breaks, that is, $|\gamma_k|_\infty$, $1 \leq k \leq K_0$, and construct confidence intervals for the estimated break points. Our theorem allows us to consider a multiple change-point test based on a threshold method on

CONTACT Likai Chen chenlikai1110@gmail.com Department of Mathematics and Statistics, Washington University in St. Louis, MO 63130.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

the maximum of generalized MOSUM statistics. We derive the asymptotic distribution of the test statistics including estimated breaks sizes, and the estimated breakpoint locations (see Theorems 3 and 4 ii)). The results provide solid foundations for conducting statistical inferences for multiple change-point estimation in high dimensional time series. Moreover, we consider a further aggregation step targeting at simultaneous breaks, and also this step gives us finer consistency rates of the break location estimation.

Multiple change-point detection can be classified into two categories, that is, model selection and testing. The traditional model selection method, for example BIC, has the drawback of computational inefficiency, which can be improved by some modified penalization procedure; see, for example, Killick, Fearnhead, and Eckley (2012), and LASSO (least absolute shrinkage and selection operator)-type penalization such as by Tibshirani and Wang (2007), Li, Qian, and Su (2016) and Lee, Seo, and Shin (2016). Regarding multiple change-point detection via testing, a classical method utilizes an exhaustive search, which examines all the possible breakpoints combination. An exhaustive search is very time consuming and some dynamic technique and improved versions are invented; see, for instance, Bai and Perron (1998, 2003) and Jackson et al. (2005). A very popular approach is the binary segmentation introduced in Scott and Knott (1974). However, its power might suffer for certain alternatives. This drawback can be handled by the wild binary segmentation algorithm developed in Fryzlewicz (2014) and sparsified binary segmentation as in Cho and Fryzlewicz (2015). Moreover, Fryzlewicz (2018) recently introduced a bottom-up algorithm to overcome the disadvantage of the classical binary segmentation. Besides, Wu and Zhou (2019) proposed multiscale abrupt change estimation under complex temporal dynamics.

Detection using the MOSUM (moving sum) statistics is another popular way for multiple change-point analysis; see, for example, Hušková and Slabỳ (2001) for iid data; Wu and Zhao (2007) and Eichinger and Kirch (2018) for general temporal dependent data. Preuss, Puchstein, and Dette (2015) dealt with multivariate time series for structural breaks in covariance. A MOSUM procedure has the advantage of computation simplicity and can avoid issues due to multiple testing in multiple break inference. A possible drawback is that MOSUM introduces a new bandwidth parameter. Such an issue can be dealt with through a multi-scale MOSUM, which uses multiple bandwidths; see, for instance Meier, Kirch, and Cho (2019). Eichinger and Kirch (2018) provide a comprehensive theoretical analysis of multiple change-point detection using MOSUM analysis including the distribution theory of the estimated breakpoint. Our work can be viewed as a generalization of their work on the high-dimensional case as we adopt a MOSUM type of statistics in our first step.

Change-point detection for high-dimensional time series has recently drawn a lot of attention due to the increasing number of applications. In particular, we shall consider the case of $p \to \infty$. Even in the simplest setup of a mean-shift model, large p may pose challenge to change-point detection. It is common to consider aggregation, either over the original time series or certain transformed statistics of individual time series and to convert the problem to a one-dimensional analysis. For instance, targeting at sparse breaks, Cho and Fryzlewicz (2015) proposed a sparse binary segmentation which concerns an $\l_1$-based aggregation with a hard threshold, and Wang and Samworth (2018) considered sparse singular value decomposition based on the CUSUM (cumulative sum) statistics. Moreover, there are a few other work looking at l_2-based aggregation of statistic: Bai (2010) evaluates the performance of a least-square estimation of a single breakpoint with distribution theory on the break location estimates without assuming cross sectional dependency; Zhang et al. (2010) extended the method in Olshen et al. (2004), Enikeeva and Harchaoui (2019), and Liu, Gao, and Samworth (2019) regard the detection of change-points in a high-dimensional mean vector as a minimax testing problem. For a single break point in time and targeting at sparse break coordinates, Jirak (2015) studied a CUSUM type statistic for each coordinate and then takes maximum of them, and asymptotic theory is provided to facilitate the simultaneous inferences of the breakpoint estimation. Cho (2016) proposed a double-CUSUM algorithm, etc. For a single change-point in time, distribution theory is still available in a few works, see, for example, Bai (2010). However, Bai (2010) is only concerning cross-sectional independent data. When it comes to multiple change points detection, the majority of the aforementioned work focus on developing novel algorithms, and a complete distribution theory is not readily available due to the complexity of the problem. An exception is Jirak (2015). Compared to Jirak (2015), we are taking a different path in terms of an algorithm using the MOSUM and an aggregation step with refined rates of estimator achieved. We thus provide a new angle to conduct inferences in multiple change-point analysis for high-dimensional time series.

It shall be noted that as there are already many novel algorithms developed, we do not claim a total superiority of ours. The algorithm proposed here is a generalization or modification of the existing methods, which facilitates us to obtain a complete theory and good theoretical rates. Nevertheless, our aggregation step is different and complement to existing algorithms. For example, one main difference with the aggregation step is that our project is based on the estimates in the first step. Cho and Fryzlewicz (2015) and Wang and Samworth (2018) used other approaches to find the best projection direction.

To summarize, we provide theory for a two-step multiple change-point procedure. We prove consistency results as well as distribution theorems for breakpoint location estimation, which is crucial for inference of breakpoints. The aggregation step can help us to achieve good rates of the breakpoint estimation. We deliver general theoretical results that allow heavy-tailed distribution and general spatial–temporal dependency assumption on the error term, and we do not require the mean function to be piece-wise constant (i.e., $f \equiv 0$). The detection procedure is not computationally expensive, as we only need to evaluate the statistic once for each point t. Additionally, we consider the estimation of the long-run covariance matrices. This article is structured as follows. Section 2 constructs a test and delivers its asymptotic performance for testing the existence of change-points. Section 3 introduces the two-step algorithm for inference on break

estimation. The associated consistency and asymptotic distribution theorems are also covered in this section. Long-run covariance matrix estimation is derived in Section 4. Simulation results are in Section A in supplementary materials and an application on U.S. unemployment rate is given in Section 5. Detailed proofs are presented in Section B in the supplementary materials.

Notations: For a constant $k \in \mathbb{N}$ and a vector $v = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$, we denote $|v|_k = (\sum_{i=1}^d |v_i|^k)^{1/k}$, $|v| = |v|_2$ and $|v|_\infty = \max_{i\le d} |v_i|$. For a matrix $A = (a_{ij})_{1\le i\le m, 1\le j\le n}$, we define the spectral norm $|A|_2 = \max_{|v|=1} |Av|$ and the max norm $|A|_{\max} = \max_{i,j} |a_{i,j}|$. For a function f, we denote $|f|_\infty = \sup_x |f(x)|$. We set (a_n) and (b_n) to be positive number sequences. We write $a_n = O(b_n)$ or $a_n \lesssim b_n$(resp. $a_n \asymp b_n$) if there exists a positive constant C such that $a_n/b_n \le C$(resp. $1/C \le a_n/b_n \le C$) for all large n, and we denote $a_n = o(b_n)$ (resp. $a_n \sim b_n$), if $a_n/b_n \to 0$ (resp. $a_n/b_n \to 1$). For two sequences of random variables (X_n) and (Y_n), we write $X_n = o_{\mathbb{P}}(Y_n)$, if $X_n/Y_n \to 0$ in probability.

2. Testing the Existence of Change-points

In this section, we provide a test for the existence of breaks. Considering our observations generated by the model in Equations (1) and (2), we would like to test the null hypothesis,

$$\mathcal{H}_0 : \quad \gamma_1 = \gamma_2 = \ldots = \gamma_{K_0} = 0,$$

which corresponds to the case of no breaks, against the alternative of the existence of at least one break, that is, $\mathcal{H}_A : \exists \quad k \in \{1, \ldots, K_0\}, \quad s.t. \quad \gamma_k \neq 0$. It shall be noted that we do not need to assume the number of breaks (K_0) to be bounded, but to rather restrict on the separation between breakpoints (see Assumption 2.4).

In Section 2.1, we derive our test statistic. Its asymptotic property is given in Section 2.2. In Section 2.3, we derive the performance of the test based on Gaussian approximation, which provides the theoretical foundation for calculating the size and power of the test.

2.1. Test Statistic

In this subsection, we introduce the test statistics and some intuition. Recall that our trend function $\mu(u)$ can be disentangled into two parts, namely a smooth transition part $f(u)$ and a jump part $\gamma_i \mathbf{1}_{u\ge u_i}$. We can define the jump vector at point u as a gap between the right-side function $\mu^{(r)}(u)$ and the left-side function $\mu^{(l)}(u)$, which is

$$J(u) = \mu^{(r)}(u) - \mu^{(l)}(u), \qquad \text{where we define}$$

$$\mu^{(r)}(u) = \lim_{t\downarrow u} \mu(t) \qquad \text{and}$$

$$\mu^{(l)}(u) = \lim_{t\uparrow u} \mu(t).$$

Due to the smoothness of the constitutes of $f(.)$, the gap function $J(u) = 0$ when there is no jump, and $J(u) = \gamma_k$ when $u = u_k$. A natural way to test the existence of change-points is to check whether the gap is zero (i.e., $J(u) = 0$). To this end, we need $\hat{\mu}^{(r)}(u)$ and $\hat{\mu}^{(l)}(u)$, which are estimates of $\mu^{(r)}(u)$ and $\mu^{(l)}(u)$. We propose to adopt the local linear estimation technique, see Fan and Gijbels (1996).

The local linear estimates of $\hat{\mu}^{(l)}(u)$ and $\hat{\mu}^{(r)}(u)$ at the point $u = i/n$ are of the following weighted form:

$$\hat{\mu}_i^{(l)} := \hat{\mu}^{(l)}(i/n) = \sum_{t=i-bn}^{i-1} w_{i-t} Y_t \qquad \text{and} \tag{3}$$

$$\hat{\mu}_i^{(r)} := \hat{\mu}^{(r)}(i/n) = \sum_{t=i+1}^{i+bn} w_{t-i} Y_t,$$

with weights

$$w_i = w_{i,b} = w_b(0, i/n), \quad i \ge 1, \quad w_0 = 0. \tag{4}$$

The weight functions are defined as

$$w_b(u, v) = \frac{K((v-u)/b)[S_2(u) - (u-v)S_1(u)]}{S_2(u)S_0(u) - S_1^2(u)},$$

$$S_l(u) = \sum_{i=1}^n (u - i/n)^l K((i/n - u)/b), \tag{5}$$

where $K(\cdot)$ is a kernel function and b is a bandwidth with $b \to 0$ and $bn \to \infty$. It is worth noting that the estimator in (3) is equivalent to adopting a one-sided kernel function, that is, $K(u)\mathbf{1}_{u\ge 0}$ to fix the boundary estimation issue for the kernel estimation method.

If there is no jump around the time point $u = i/n$, then the gap estimate $\hat{J}(i/n) = \hat{\mu}_i^{(l)} - \hat{\mu}_i^{(r)}$ would be small for all coordinates. Otherwise if for some entity $1 \le j \le p$, the gap estimate $|\hat{J}_j(i/n)|$ is large, there might exist a jump around time i/n at coordinate j. Note that the test statistics is in fact of a MOSUM type, and we replace the uniform kernel for MOSUM by a local linear one to adapt for slowly varying trends $f(u)$ in Equation (2).

To conduct the breakpoint detection with $p \to \infty$, we consider the maximum of the gap statistics. Furthermore, we need to standardize our test statistics in order to get a regular limit distribution. To obtain the long-run variance matrix involved in the standardization, we need to specify the error process, as in model (1). We would like to make a general temporal and cross-sectional dependence assumption. This is a crucial issue, since for time series data, dependence is the rule rather than the exception. Specifically, we let

$$\epsilon_t = \sum_{k\ge 0} A_k \eta_{t-k}, \tag{6}$$

where $\eta_t \in \mathbb{R}^{\tilde{p}}$ are independent and identically distributed (iid) random vectors with zero mean and an identity covariance matrix. $A_k, k \ge 0$, are coefficient matrices in $\mathbb{R}^{p\times\tilde{p}}$ such that ϵ_t is a proper random vector, and $p \le \tilde{p} \le c_p p$, for some constant $c_p > 1$. If $A_i = 0$ for all $i \ge 1$, then the noise sequences are temporally independent; if $p = \tilde{p}$ and matrices A_i are diagonal, then the sequences become the model in Bai (2010), which is spatially independent. The VMA(∞) process is very general and

includes many important time series models such as a vector autoregressive moving averages (VARMA) model, that is,

$$(1-\sum_{j=1}^{s}\Theta_j\mathcal{B}^j)X_i = X_i - \sum_{j=1}^{s}\Theta_j X_{i-j} = \sum_{k=1}^{t}\Xi_k\eta_{i-k},$$

where Θ_j and Ξ_k are real matrices such that $\det(1-\sum_{j=1}^{s}\Theta_j z)$ is not zero for all $|z|\le 1$ and $\mathcal{B}$ is the backshift operator.

Correspondingly, we define the sum of the coefficient matrix to be $S=\sum_{k\ge 0}A_k$. The long run covariance matrix for the error process is

$$\Sigma = SS^\top. \tag{7}$$

We denote $\Sigma=(\sigma_{i,j})$, $1\le i,j\le p$, and

$$\Lambda = \mathrm{diag}(\sigma_{1,1}^{1/2},\sigma_{2,2}^{1/2},\ldots,\sigma_{p,p}^{1/2}). \tag{8}$$

Following the previous intuition of the effect of jumps on the gap statistics $\hat{J}(.)$, we consider the test statistic

$$T_n = \max_{bn+1\le i\le n-bn}|V_i|_\infty, \quad \text{where} \quad V_i=\Lambda^{-1}(\hat{\mu}_i^{(l)}-\hat{\mu}_i^{(r)}). \tag{9}$$

We adopt a supreme type of statistics as it shares good property under certain alternatives. However, we do not claim the strict superiority of our test statistics. When the majority of locations exhibit simultaneous jumps, an l_2 type statistics tends to have better power.

We exclude Y_i in V_i, because that the weights in front of Y_i would be the same for the right side and the left side estimator, and will be canceled when taking the difference. Note that we consider the normalized statistic as multiplying the jump estimates $\hat{J}(i/n)=\hat{\mu}_i^{(l)}-\hat{\mu}_i^{(r)}$ by Λ^{-1} since the long-run variances $\sigma_{j,j}$ for different coordinates $1\le j\le p$ can be very different. We refer to T_n as an infeasible test statistic since Λ is unknown. The estimation of Λ is deferred to Section 4.

2.2. Properties of the Test Statistics

We shall show the asymptotic properties of our test statistics T_n in Equation (9) in this subsection. First we analyze the mean of the normalized jump estimators, that is, $\mathbb{E}V_i$. Intuitively, we can decompose the level of our jump estimator $\mathbb{E}V_i$ into two parts, one is the commonly encountered bias term for the non-parameteric kernel estimators of the smooth trend functions, and the other is induced by jumps on the deterministic trend, which is denoted as d_i. Recall the definition of w_i in Equation (4) for $i=1,2,\ldots,bn$, and $w_i=0$ for $i=0$ and $i>bn$. We denote the location of breaks as $\tau_k=nu_k$ and Ω_i as a set of indices indicating the break locations within the bn neighborhood around time i, namely $\Omega_i=\{k\big||i-\tau_k|\le bn, 1\le k\le K_0\}$. If $\tau_i-\tau_j=n(u_i-u_j)\asymp n$, for any i,j, then for large n, the cardinality of Ω_i is at most one, that is, $|\Omega_i|\le 1$. Actually, this condition can be relaxed to $\min_{1\le i\ne j\le K_0}|\tau_i-\tau_j|\gg bn$. For a time point i where $\Omega_i\ne\emptyset$, we define the weighted break sizes to be,

$$d_i = (1-\sum_{t=1}^{|i-\tau_k|}w_t)\Lambda^{-1}\gamma_k, \quad k=\mathrm{argmin}_{j\in\Omega_i}|i-\tau_j|, \tag{10}$$

and for the rest of locations i, let $d_i=0$. We further stack d_i over all breakpoints that are of interest, which is denoted by $\underline{d}=(d_{bn+1}^\top,d_{bn+2}^\top,\ldots,d_{n-bn}^\top)^\top$. It should be noted that under the null, $\underline{d}=0$.

We denote the smooth part of the local linear estimate as

$$\hat{f}_i^{(l)}=\sum_{t=i-bn}^{i-1}w_{i-t}f(t/n) \quad \text{and} \quad \hat{f}_i^{(r)}=\sum_{t=i+1}^{i+bn}w_{t-i}f(t/n).$$

By Fan and Gijbels (1996), under some smoothness conditions, the bias part of the estimated smooth functions would be of the order b^2, which goes to zero by assumption, that is,

$$\max_{bn+1\le i\le n-bn}|\Lambda^{-1}(\hat{f}_i^{(l)}-\hat{f}_i^{(r)})|_\infty = O(b^2). \tag{11}$$

Given the definition of our model $Y_i=\mu(i/T)+\epsilon_i$, d_i can be expressed as

$$\begin{aligned} d_i &= \mathbb{E}\{\Lambda^{-1}((\hat{\mu}_i^{(r)}-\hat{\mu}_i^{(l)})-(\hat{f}_i^{(r)}-\hat{f}_i^{(l)}))\} \\ &= \mathbb{E}\{V_i-\Lambda^{-1}(\hat{f}_i^{(r)}-\hat{f}_i^{(l)})\}. \end{aligned} \tag{12}$$

Combining Equations (11) and (12), $\mathbb{E}V_i$ can be approximated by the part induced by jumps γ_ks, as

$$|\mathbb{E}V_i-d_i|_\infty = |\Lambda^{-1}(\hat{f}_i^{(r)}-\hat{f}_i^{(l)})|_\infty = O(b^2). \tag{13}$$

Let us now consider the $V_i-\mathbb{E}V_i$ part. We observe that the centered statistics can be expressed as a weighted sum of the error term, namely

$$V_i-\mathbb{E}V_i=\sum_{l=i-bn}^{i-1}w_{i-l}\Lambda^{-1}\epsilon_l-\sum_{l=i+1}^{i+bn}w_{l-i}\Lambda^{-1}\epsilon_l. \tag{14}$$

To approximate its distribution, we introduce a scaling matrix for variance of the limit distribution. Recall $S=\sum_{k\ge 0}A_k$ and define a block matrix $G^\diamond=(G_{i,l}^\diamond)_{bn+1\le i\le n-bn,1\le l\le n}\in\mathbb{R}^{(n-2bn)p\times n\tilde{p}}$ with components as $p\times\tilde{p}$ dimension matrices,

$$G_{i,l}^\diamond=\begin{cases} w_{i-l}\Lambda^{-1}S, & \text{if} \quad i-bn\le l\le i-1, \\ -w_{l-i}\Lambda^{-1}S, & \text{if} \quad i+1\le l\le i+bn, \end{cases} \tag{15}$$

and elsewhere zero. Let $\underline{z}$ be a Gaussian vector in $\mathbb{R}^{n\tilde{p}}$ with zero mean and identity covariance matrix. We set $G_{i,\cdot}^\diamond$ to be $(G_{i,1}^\diamond,G_{i,2}^\diamond,\ldots,G_{i,n}^\diamond)$. It can be shown that $G_{i,\cdot}^\diamond\underline{z}$ has a similar covariance structure as $V_i-\mathbb{E}V_i$. We shall use the distribution of $|G_{i,\cdot}^\diamond\underline{z}|_\infty$ to approximate the distribution of $|V_i-\mathbb{E}V_i|_\infty$. Combining this approximation with the bias term in Equation (13), we shall expect that for each time point i, our normalized break test statistics can be approximated by the maximum of a Gaussian vector centered at d_i, that is,

$$\mathbb{P}(|V_i|_\infty\le u)\approx\mathbb{P}(|d_i+G_{i,\cdot}^\diamond\underline{z}|_\infty\le u).$$

We now let the statistics go over all the time points, and recall $T_n=\max_{bn+1\le i\le n-bn}|V_i|_\infty$. Then we shall expect

$$\mathbb{P}(T_n\le u)\approx\mathbb{P}(|\underline{d}+G^\diamond\underline{z}|_\infty\le u), \tag{16}$$

and equivalently

$$\mathbb{P}(T_n\le u)\approx\mathbb{P}(|\underline{d}+Z|_\infty\le u), \tag{17}$$

where $Z = (Z_{bn+1}^\top, Z_{bn+2}^\top, \ldots, Z_{n-bn}^\top)^\top$ and $(Z_i)_{bn+1\le i\le n-bn}$ is a sequence of centered Gaussian vectors in $\mathbb{R}^p$ with covariance matrices $\mathrm{cov}(Z_i, Z_j) = Q_{i,j}$ of the following form:

$$Q_{i,j} = \varpi_{i,j}\Lambda^{-1}\Sigma\Lambda^{-1} \quad \text{and} \quad \varpi_{i,j} = \sum_{l=1}^{n} w_{|i-l|}w_{|j-l|}\mathrm{sign}(i-l)\mathrm{sign}(j-l). \tag{18}$$

To see the equivalence between (16) and (17), let

$$Q = (Q_{i,j})_{bn+1\le i,j\le n-bn} = G^\diamond G^{\diamond\top}.$$

Then Z is a Gaussian vector with zero mean and covariance matrix Q. Note that

$$Z_i \overset{d}{=} G^\diamond_{i,\cdot}\underline{z} \quad \text{and} \quad Z \overset{d}{=} G^\diamond\underline{z}. \tag{19}$$

This transformation from $G^\diamond\underline{z}$ to Z is to show that the involved Gaussian process only depends on the long-run covariance matrix Σ and the weight functions. We note that Z are not element-wise independent, but with dependency governed by $G^\diamond$. The above argument will be rigorously formulated in Theorem 1 in the next subsection.

2.3. Gaussian Approximation

In this subsection, we provide the formal theory supporting our test. We first present necessary assumptions. The following is to guarantee the smoothness of the trend functions $\mu_j(u)$ when no break occurs.

Assumption 2.1. Function $f_j \in \mathcal{C}^2[0,1]$ with $\max_{1\le j\le p}|f_j'|_\infty \le c_f$, $\max_{1\le j\le p}|f_j''|_\infty \le c_f$ for some constant $c_f > 0$.

Additionally, to ensure the property of our kernel estimation, we need conditions on the kernel function.

Assumption 2.2. The kernel $K(.) \ge 0$ is symmetric with support $[-1,1]$, assume $|K|_\infty < \infty$ and $\int_{-1}^{1} K(x)\mathrm{d}x = 1$. Also assume $K(x)$ has first-order derivative with $|K'|_\infty < \infty$ on $(0,1)$. Let $b \to 0$ and $bn \to \infty$. Denote $\kappa_i = \int_0^1 x^i K(x)\mathrm{d}x$. Assume $\kappa_1^2 \ne \kappa_2\kappa_0$.

We also set conditions on the regularity of the long-run covariance matrix and the dependency strength of the noise sequence.

Assumption 2.3. (Lower bound for the long run variance) $\sigma_{j,j} \ge c_\sigma$, $1 \le j \le p$ for some finite constant $c_\sigma > 0$.

We need enough separation between adjacent breakpoints.

Assumption 2.4. (Separation) Assume $\min_{1\le i,j\le K_0}|\tau_i - \tau_j| \gg bn$.

It is worth noting that Assumption 2.4 implies that the number of breaks K_0 shall not exceed the order $1/b$.

Assumption 2.5. (Dependence strength) $\max_{1\le j\le p}\sum_{k\ge i}|A_{k,j,\cdot}|_2/\sigma_{j,j}^{1/2} \le c_s(i\vee 1)^{-\beta}$, where $\beta > 0$ is some constant and $A_{k,j,\cdot}$ is the jth row of A_k.

Assumption 2.5 is a very general spatial and temporal dependence condition and embraces many interesting processes. It requires an algebraic decay rate of the temporal dependence. However, the cross-sectional dependence does not need to be weak; and in fact, it can be strong such that it has a factor structure. We provide an example as follows.

Example 1. Assume that $\eta_t, \eta_t' \in \mathbb{R}^p$ are iid random vectors with zero mean and covariance matrix I_p. Let

$$\epsilon_t = F_t + Z_t, \text{ with } Z_t = \sum_{k\ge 0}\Lambda_k\eta_{t-k} \text{ and } F_t = \sum_{k\ge 0} v f_k^\top \eta_{t-k}', \tag{20}$$

where $\Lambda_k = \mathrm{diag}(\lambda_{k,1}, \ldots, \lambda_{k,p})$, $v = (v_1, \ldots, v_p)^\top$ and $f_k = (f_{k,1}, \ldots, f_{k,p})^\top$. Here F_t is the factor term and $Z_{t,j}$ are independent for different j. Then the long-run variances for $Z_{t,j}$ and $F_{t,j}$ are $\sigma_{Z,j} = (\sum_{k\ge 0}\lambda_{k,j})^2$ and $\sigma_{F,j} = |\sum_{k\ge 0} f_k|_2^2 v_j^2$, respectively. If for some constant $c > 0$,

$$\sum_{k\ge i}|\lambda_{k,j}|/\sigma_{Z,j}^{1/2} \le ci^{-\alpha} \quad \text{and} \quad \sum_{k\ge i}|f_k|_2|v_j|/\sigma_{F,j}^{1/2} \le ci^{-\alpha}, \tag{21}$$

then Assumption 2.5 holds with $\beta = \alpha$. To see this, we note $|A_{k,j,\cdot}|_2 = (\lambda_{k,j}^2 + |f_k|_2^2 v_j^2)^{1/2}$, and $\sigma_{j,j} = \sigma_{Z,j}^2 + \sigma_{F,j}^2$. Hence,

$$\sum_{k\ge i}|A_{k,j,\cdot}|_2 \le \sum_{k\ge i}(|\lambda_{k,j}| + |v_j||f_k|_2) \le ci^{-\alpha}(\sigma_{Z,j}^{1/2} + \sigma_{F,j}^{1/2}) \le \sqrt{2}ci^{-\alpha}\sigma_{j,j}^{1/2}.$$

Assumption 2.6. (Finite moment) The innovations $\eta_{i,j}$ are iid with $\mu_q = \|\eta_{1,1}\|_q < \infty$ for some $q \ge 4$.

Assumption 2.7. (Subexponential) The innovations $\eta_{i,j}$ are iid with $\mu_e = \mathbb{E}e^{a_0|\eta_{1,1}|} < \infty$, for some $a_0 > 0$.

Assumptions 2.6 and 2.7 put tail assumptions on the distribution of the noise sequences. Given the above-mentioned conditions, we provide the main Gaussian approximation theorem, which is essential for the asymptotic distribution of our test statistics T_n. Our theorem extends the Gaussian approximation theory in Chernozhukov, Chetverikov, and Kato (2013, 2017), which build on the Stein's method and the anti-concentration bounds. Markedly, our theory is developed for modeling dependent data. To this aim, one important technical non-triviality lies in handling the spatial-temporal dependency of the trend stationary high-dimensional processes. We derive the corresponding concentration inequalities based on m-dependence approximation of the underlying processes. Compared to the existing results on Gaussian approximation for time series, for example Zhang et al. (2017), our setting works for noncentered Gaussian approximation that accommodates our interest for time series with breaks.

Theorem 1 (Gaussian approximation). Under Assumptions 2.1–2.5,

(i) if Assumption 2.6 holds, and $np(bn)^{-q/2}(\log(np))^{3q/2} = o(1)$, then for

$$\Delta = (bn)^{-1/6}\log^{7/6}(pn) + (bn)^{-\beta/(3+3\beta)}\log^{2/3}(np) + b^{5/2}n^{1/2}\log^{1/2}(np),$$

we have

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}(T_n \le u) - \mathbb{P}(|\underline{d}+Z|_\infty \le u)\right| \lesssim \Delta + ((np)^{2/q}/(bn))^{1/3}\log(pn),$$

(ii) if Assumption 2.7 holds, and $(bn)^{-1}(\log(np))^{\max\{7,2(1+\beta)/\beta\}} = o(1)$, we have

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}(T_n \le u) - \mathbb{P}(|\underline{d}+Z|_\infty \le u)\right| \lesssim \Delta,$$

where the constants in $\lesssim$ are independent of n, p, b.

If in addition

$$b^5 n\log(np) = o(1), \tag{22}$$

then under both cases, we have

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}(T_n \le u) - \mathbb{P}(|\underline{d}+Z|_\infty \le u)\right| \to 0. \tag{23}$$

Remark 1. (Allowed dimension) One key theoretical insight is that we explicitly show the trade-off between the tail assumption of the innovations and the allowed dimension of the time series p relative to the sample size n in the above theorem. In particular, when we have exponential tail assumption on the distribution of the innovations, we allow an ultra-high dimensional setup indicating p to be at an exponential rate with respect to n. And when we have only finite moment assumptions, we can allow p to be at a polynomial order with respect to n. Specifically, for Theorem 1 case (i), we allow p to be of some polynomial order of n, and its order depends on the value of q. For some $\nu_1 > 0$ and $0 < \nu_2 < 1/2$, assume $p \asymp n^{\nu_1}$ and $b \asymp n^{-\nu_2}$. If $\nu_1 < (1-\nu_2)q/2 - 1$ and $\nu_2 > 1/5$, then Equation (23) holds. It is easy to see that the bigger the moment q is, the larger the allowance of the dimension p. The moment Condition (2.6) depends on q which characterizes the heavy tailedness of the noise, larger q means thinner tails. For case (ii), we can allow p to be exponential in n, that is, the ultra-high dimensional scenario. For instance, for some $\nu_1 > 0$ and $1/5 < \nu_2 < 1$, we can set $p \asymp e^{n^{\nu_1}}$ and $b \asymp n^{-\nu_2}$. If $\nu_1 < 5\nu_2 - 1$ and $\nu_1 \max\{7, 2(1+\beta)/\beta\} < 1 - \nu_2$, then Equation (23) holds.

It is not hard to understand the size and power implication of Theorem 1 to our test. Under the null hypothesis, we have $\underline{d} = 0$, then for any prefixed significant level $\alpha \in (0,1)$, we have the critical value of our test as q_α, that is, the quantile of the Gaussian limit distribution,

$$q_\alpha = \inf_{r\ge 0}\{r : \mathbb{P}(|Z|_\infty > r) \le \alpha\}. \tag{24}$$

As from the Gaussian approximation result in Equation (23), we have the approximated sizes of the test statistics,

$$\left|\mathbb{P}(T_n > q_\alpha) - \mathbb{P}(|Z|_\infty > q_\alpha)\right| \to 0.$$

We shall reject the null hypothesis at the significant level α, if the test statistics exceed the critical value, that is, $T_n > q_\alpha$.

To evaluate our testing power, consider the alternative that if not all $\gamma_k = 0$, then $\underline{d}$ is non-zero. We have the following corollary for the power, which is a straightforward consequence of Theorem 1.

Corollary 1. (Power) Under conditions in Theorem 1 (i) or (ii). The testing power β_α satisfies

$$\beta_\alpha = \mathbb{P}(|\underline{d}+Z|_\infty \ge q_\alpha) + o(1).$$

Thus, we can see that the power of our test would depend on the vector $\underline{d}$. The size of it is determined by the true jump sizes that is, γ_ks. Since the covariance matrix for Z is $Q = (Q_{i,j})$, where $Q_{i,j} = \varpi_{i,j}\Lambda^{-1}\Sigma\Lambda^{-1}$ with $\varpi_{i,j}$ defined in Equation (18). It can be calculated that $\varpi_{i,i} \asymp (bn)^{-1}$, therefore $|Z|_\infty = O_{\mathbb{P}}((bn)^{-1/2}\log(np))$, which tends to zero by Assumptions 2.6 and 2.7. Thus, if $|\underline{d}|_\infty \gg (bn)^{-1/2}\log(np)$, $\beta_\alpha \to 1$ by Corollary 1.

3. Estimation and Inference of Breaks

In this section, we show how to estimate the number of change-points, the time stamps, the spatial coordinates and the sizes of the structural breaks. We summarize the key steps of the adopted two-step procedure for the multiple change-point detection. The main reason for a two-step estimation is to achieve an optimal rate of consistency for our break estimation. The first step can be regarded as an extension of the MOSUM l_∞ aggregation. Namely, in our first step, we conduct a "rough" estimation though a MOSUM type statistic as in Equation (9), and we can draw a conclusion on the existence of a break. In case it exists, we obtain a "rough" estimate of the change-points locations. In the second step, we refine our jump estimates based on a one-dimensional aggregated time series. The aggregation can be viewed as a projection using information on the jump estimators from the first step. To be more specific, within each time region around the kth breakpoint, we can aggregate data by a weighted sum of different coordinates whose weights are determined by the first step jump size estimators ($\hat{\gamma}_k$). Instead of looking at the biggest break at one time point, the aggregated change-point statistics carry more information regarding significant jumps across contemporaneous locations, and would thus provide better precision. In the following, we introduce the first "rough" estimation step and its properties in Section 3.1. We further improve the first step in Section 3.1 through an aggregated statistics , which is proposed and analyzed in Section 3.2.

3.1. The "Rough" Estimation Step

We define the sizes of the breakpoints at time k as

$$|\Lambda^{-1}\gamma_k|_\infty.$$

Here, we normalize γ_k by the long-run standard deviations for the same reason as V_i in Equation (9). Intuitively, the noise fluctuation levels for different locations can be very different, and at one location, a break can be significant due to purely high noise level without normalization. We define the minimum size of breaks over time as

$$\delta^\diamond = \min_{1\le k\le K_0}|\Lambda^{-1}\gamma_k|_\infty. \tag{25}$$

In the following, we outline the steps of our testing, detecting and estimation procedure.

Step 1. For significance level α, we test the existence of jumps based on the critical value q_α in Equation (24). If we find no

significant breaks, then we cannot reject the null $\mathcal{H}_0$. In case our test statistic exceeds the critical value, we reject $\mathcal{H}_0$ and acknowledge the existence of breaks, then we proceed to step 2.

Step 2. To detect the change-points, we collect all the time stamps with the jump statistics $|V_\tau|_\infty$ exceeding a threshold value $w^\dagger$, namely, $\mathcal{A}_1 = \{bn+1 \le \tau \le n-bn : |V_\tau|_\infty > w^\dagger\}$, where V_τ is defined in Equation (9). Let $\hat{\tau}_1$ be the time point τ in $\mathcal{A}_1$ that maximizes the test statistics $|V_\tau|_\infty$. We further eliminate a $2bn$ neighborhood of time points around $\hat{\tau}_1$ from $\mathcal{A}_1$ to create $\mathcal{A}_2$. Then we find the next point in $\mathcal{A}_2$ that maximize $|V_\tau|_\infty$, and repeat the same operation until the set $\mathcal{A}_k$ is empty. Namely, for $k \ge 1$, we let the kth estimated break point be denoted as $\hat{\tau}_k = \text{argmax}_{\tau\in\mathcal{A}_k}|V_\tau|_\infty$ and $\mathcal{A}_{k+1} = \mathcal{A}_k \setminus \{\tau : |\tau - \hat{\tau}_k| \le 2bn\}$. We denote the maximum number of breakpoints as $\hat{K}_0$, with $\hat{K}_0 = \max_{k\ge1}\{k : \mathcal{A}_k \ne \emptyset\}$. It is worth noting that we have chosen $2bn$ to exclude both bn neigborhood of τ and $\hat{\tau}$.

Step 3. Given the detected breakpoints in Step 2, we calculate the break sizes over time. We denote the window size to be $M = bn$,

$$\hat{\gamma}_k = \hat{\mu}^{(l)}_{\hat{\tau}_k - M} - \hat{\mu}^{(r)}_{\hat{\tau}_k + M} \quad \text{and} \quad \hat{\delta}^\diamond = \min_{1\le k\le \hat{K}_0} |\Lambda^{-1}\hat{\gamma}_k|_\infty. \tag{26}$$

It is worth noting that in this algorithm, we only need to calculate the gap statistics $|V_\tau|_\infty$ once for each point. Hence, it is not time consuming regardless of the true number of breakpoints. In Step 1, we test the existence of the breaks. In Step 2, we use the estimated $|V_\tau|_\infty$ for all the points from $bn+1$ to $n-bn$ and select the points that are beyond the threshold w. Intuitively, the points in $\mathcal{A}_1$ would contain the break indices, as well as points in their neighborhood where estimates are contaminated by the breaks. Therefore, in Step 2, we find the local maximums and discard points around them. In Step 3, we estimate the sizes of the change-points and calculate their minimum values.

In the following, we shall provide consistency results of estimates of the break numbers, locations and break sizes in Theorem 2; and derive asymptotic distribution of break sizes in Theorem 3.

We need to first impose the minimum jump size condition on the break size as

Assumption 3.1. Assume the break size satisfies $\delta^\diamond \gg \max\left\{\sqrt{\log(pn)/(bn)},\ b\right\}$.

It can be seen that the break size requirement is related to the dimensionality of the time series, the number of observations available and the bandwidth parameter. The rate $\sqrt{\log(pn)/(bn)}$ is due to the variance of the estimation error with the typical $\log(pn)$ term compensating for the dimensionality of the test statistics and b is due to the incurred bias in the setting of the smooth varying trend. If we assume the trend function to be piecewise constant, then the term b disappears. The larger the sample n, the smaller the requirement for $\delta^\diamond$ due to the better approximation of the trends. In the following theorem, we show that we would asymptotically obtain the right number of breaks. Moreover, we can bind the errors of the estimated break locations and the break sizes. The threshold $\omega^\dagger$ shall be set as a high quantile of its limited Gaussian distribution to ensure the consistent estimation of the breaks.

Theorem 2. We assume conditions in Theorem 1 (*i*) or (*ii*), Assumption 3.1, and Equation (22) hold. If $\min\{\delta^\diamond - \omega^\dagger, \omega^\dagger\} \ge 2c'_w\sqrt{\log(pn)/(bn)}$, where c'_w is the constant defined as the limit $(bn\sum_{i=0}^n w_i^2)^{1/2} \to c'_w$, then

(i) $\mathbb{P}(\hat{K}_0 = K_0) \to 1$.
(ii) under Theorem 1 (i), $|\hat{\tau}_k - \tau_{k*}| = O_\mathbb{P}\{(np)^{2/q}/\delta^{\diamond 2}\}$, and under Theorem 1 (ii), $|\hat{\tau}_k - \tau_{k*}| = O_\mathbb{P}\{\log^2(np)/\delta^{\diamond 2}\}$, uniformly over k, where $k^* = \text{argmin}_i|\hat{\tau}_k - \tau_i|$.
(iii) $|\Lambda^{-1}(\hat{\gamma}_k - \gamma_{k^*})|_\infty = O_\mathbb{P}((bn)^{-1/2}\log(np)^{1/2} + b)$, uniformly over k, which indicates $|\hat{\delta}^\diamond - \delta^\diamond| = O_\mathbb{P}((bn)^{-1/2}\log(np)^{1/2} + b)$.

Result (i) indicates that the number of breaks can be consistently estimated, (ii) suggests that the estimated break dates u_k can be consistently determined in view of $u_k = \tau_k/n$ and (iii) shows that the break sizes can be consistently recovered. The convergence rate of the break sizes dependents on the bandwidth b, sample size n and the dimension of the time series p. It is worth noting that the bias is of order b in (iii), as the difference is taken with a gap of $2M$ as in Equation (26). It shall be noted that the consistency rate of $\hat{\tau}_k$ depends on the break size $\delta^\diamond$, which depends only on the maximum break size for any fixed time. Therefore, having several large breaks simultaneously would not improve the break size estimation. With respect to the condition $\min\{\delta^\diamond - \omega^\dagger, \omega^\dagger\} \ge 2c'_w\sqrt{\log(pn)/(bn)}$, relative to the summary in the Table 1 in Cho (2016), our break size $\delta^\diamond$ is comparable up to the weakest condition $\delta^\diamond(nb)^{1/2} \to \infty$ up to a logarithmic factor. Moreover, when $p = 1$, our requirement of breaksize is similar to the rate as in Theorem 3.2 in Wu and Zhou (2019), namely $\delta^\diamond \ge \sqrt{\log n}/\sqrt{nb}$.

Given the consistency of the breakpoints, we can obtain a distribution theory that facilitates us in making inferences on the break sizes. Let $\tilde{Z}$ be a Gaussian vector in $\mathbb{R}^p$ with zero mean and covariance matrix

$$\tilde{Q} := Q_{bn+1,bn+1} = 2\sum_{t=1}^{bn} w_t^2 \Lambda^{-1}\Sigma\Lambda^{-1}. \tag{27}$$

Theorem 3. (Break size inference) Assume conditions in Theorem 2 and $b^3 n\log(np) = o(1)$. We have

$$\sup_{u\in\mathbb{R}} |\mathbb{P}(|\Lambda^{-1}(\hat{\gamma}_k - \gamma_{k^*})|_\infty \le u) - \mathbb{P}(|\tilde{Z}|_\infty \le u)| \to 0, \quad \text{where } k^* = \text{argmin}_i|\hat{\tau}_k - \tau_i|.$$

This theorem indicates that the maximum of the difference between the estimated jump size $\hat{\gamma}_k$ and the true jump size γ_k can be approximated by the maximum of a Gaussian random vector with the same asymptotic variance-covariance structure. Based on Theorem 2 (ii) and Theorem 3, we can construct joint confidence interval for $\gamma_{k^*,j}$. We set

$$\alpha = \mathbb{P}(|\tilde{Z}|_\infty \ge q) \quad \text{and} \quad \theta = (\sigma_{1,1}^{1/2}, \sigma_{2,2}^{1/2}, ..., \sigma_{p,p}^{1/2})^\top, \tag{28}$$

some $\alpha > 0$. Then as $n \to \infty$ with probability close to $1-\alpha$, we have

$$-q\sigma_{j,j}^{1/2} + \hat{\gamma}_{k,j} \le \gamma_{k^*,j} \le q\sigma_{j,j}^{1/2} + \hat{\gamma}_{k,j} \qquad \forall j. \tag{29}$$

Theorem 3 can be extended to hold uniformly over k by stacking the statistics overall k's. In addition, we see that Theorem 2 and Theorem 3 are closely connected in the sense that we can reach the same threshold by stacking $\hat{\gamma}_k$ overall k's.

3.2. The Refined Aggregation Step

The estimation in the first step is only driven by $|\gamma_k|_\infty$, that is, the maximum size of jumps at a time point τ_k. Therefore, it is only sensitive to the biggest jump across all the time series at the same time. The l_∞ type test potentially have more power for certain alternatives than the l_2 type statistics. However, if the majority or all of the entities exhibit simultaneous jumps, the supremum statistic tends to have lower power than the l_2 statistic.

In case there are multiple simultaneous time series jumps, it would be beneficial to modify our procedure to aggregate all of the series with a jump. This enlightens us to propose a two-stage method: first, we follow the steps in the previous subsections to detect the "rough" timing of the jumps and the estimated jump sizes; second, for each bn neighborhood of a change-point estimate $\hat{\tau}_k$ obtained from step one, we update the change-point estimates according to a newly aggregated time series. The time series is calculated with a weighted sum of simultaneous observations corresponding to significant jump locations and the weights are based on the jump size estimates in the first step. The aggregation returns a one-dimensional time series with richer information on the cross-sectional jumps.

We denote $\mathcal{S}_k$ to be the set of series that jump at location τ_k, that is,

$$\mathcal{S}_k = \{1 \le j \le p \mid \gamma_{k,j} \neq 0\}, \tag{30}$$

where $\gamma_{k,j}$ is the jth coordinate of γ_k. Detailed steps of the aggregation are formulated as follows:

Stage 1. Apply Steps 1-3 in Section 3.1 to obtain $\hat{\tau}_k$ and $\hat{\gamma}_k$, $k = 1, 2, \ldots, \hat{K}_0$. For some $w^\dagger > 0$, let the estimation of $\mathcal{S}_k$ be

$$\hat{\mathcal{S}}_k = \left\{1 \le j \le p \middle| |(\Lambda^{-1}\hat{\gamma}_k)_j| \ge w^\dagger\right\}. \tag{31}$$

In practice, $w^\dagger$ can be chosen to be large enough to ensure that we can detect all the jumps with probability 1 as in Theorem 2.

Stage 2. For $|t - \hat{\tau}_k| \le 2bn$, we let

$$X_t = \sum_{j \in \hat{\mathcal{S}}_k} (\Lambda^{-1}\hat{\gamma}_k)_j (\Lambda^{-1}Y_t)_j. \tag{32}$$

Note that after the modification, for all the jump locations, the new time series X_t would only contain positive sized jumps that is, $\sum_{j\in\hat{\mathcal{S}}_k}(\Lambda^{-1}\hat{\gamma}_k)_j^2$. This step can be understood as a projection of the high-dimensional observations $\Lambda^{-1}Y_t$ according to the direction of $\Lambda^{-1}\hat{\gamma}_k (j \in \hat{\mathcal{S}}_k)$. This is similar to the idea of Wang and Samworth (2018).

Based on the aggregated time series X_t, the refined change-point locations can be detected through a CUSUM type of test statistics, for $k = 1, 2, \ldots, \hat{K}_0$,

$$\tilde{\tau}_k = \underset{|t-\hat{\tau}_k|\le bn}{\operatorname{argmax}} \left(\sum_{s=\hat{\tau}_k-2bn}^{\hat{\tau}_k+2bn} X_s \frac{t - \hat{\tau}_k + 2bn}{4bn+1} - \sum_{s=\hat{\tau}_k-2bn}^{t-1} X_s \right) \tag{33}$$

$$\times \sqrt{\frac{4bn+1}{(t-(\hat{\tau}_k-2bn)+1)(\hat{\tau}_k+2bn-t)}}. \tag{34}$$

After we update the break points estimation, we can construct confidence intervals for the updated breakpoints estimates $\tilde{\tau}_k$. We denote the long-run correlation matrix to be $(\tilde{\sigma}_{i,j})_{i,j} = \Lambda^{-1}\Sigma\Lambda^{-1}$, where Σ is the long-run covariance matrix for ϵ_t. We let $\tilde{\Sigma}_k = (\tilde{\sigma}_{i,j})_{i,j\in\mathcal{S}_k}$ be the sub covariance matrix corresponding to coordinates in $\mathcal{S}_k$ at time τ_k and let the standardized significant break sizes $\tilde{\gamma}_k = (\Lambda^{-1}\gamma_k)_{i\in\mathcal{S}_k}$. We define two objects involved in the limit distributions of the breaks, that is,

$$a_k = |\tilde{\gamma}_k|_2^2 \quad \text{and} \quad \varsigma_k^2 = \tilde{\gamma}_k^\top \tilde{\Sigma}_k \tilde{\gamma}_k. \tag{35}$$

Then ς_k^2 is the long-run variance for the sequence $\sum_{j\in\mathcal{S}_k}(\Lambda^{-1}\gamma_k)_j(\Lambda^{-1}\epsilon_t)_j$. For the aggregated jump estimation, we alternatively define the minimum jump size across different locations and time points as

$$\delta^\dagger = \min_{1\le k\le K_0} \min_{j\in\mathcal{S}_k} |(\Lambda^{-1}\gamma_k)_j|.$$

Then $\delta^\dagger \le \delta^\diamond$ and it functions similarly as $\delta^\diamond$ to capture the identifiable jump size of the time series. We shall put the same assumption on $\delta^\dagger$ as on $\delta^\diamond$. It is worth noting that $\delta^\dagger$ is the minimum jump size to ensure the consistency of our break estimation.

Assumption 3.2. Let $\delta^\dagger \gg \max\left\{\sqrt{\log(pn)/(bn)}, b\right\}$.

In the following corollary, we show that we can consistently recover the locations of the series with a jump for each change-point. It can be directly derived from Theorem 2 (iii).

Corollary 2. We assume conditions in Theorem 1 (i) or (ii) hold, and Assumption 3.2. If $\delta^\dagger/2 \ge w^\dagger \gg (bn)^{-1/2}\log(np)^{1/2} + b$, then we have

$$\mathbb{P}(\hat{\mathcal{S}}_k = \mathcal{S}_k,\ 1 \le k \le K_0) \to 1.$$

In addition, we provide a theorem that allows us to make inference on the estimated break-dates $\tilde{\tau}_k$.

Theorem 4. (Aggregated break estimation) Assume conditions in Corollary 2, and that for some constants $c_1, c_2 > 0$,

$$c_1 \le \lambda_{\max}(\Lambda^{-1}\Sigma\Lambda^{-1})/\lambda_{\min}(\Lambda^{-1}\Sigma\Lambda^{-1}) \le c_2. \tag{36}$$

Recall definition of a_k and ς_k in (35). Then we have for any fixed $1 \le k \le K_0$,

(i) $|\tilde{\tau}_k - \tau_{k^*}| = O_{\mathbb{P}}(\varsigma_k^2/a_k^2)$.

(ii) In addition, if Assumption 2.5 holds with $\beta > 1$, and $1 \ll \varsigma_k^2/a_k^2 \ll bn$, then we have

$$\tilde{\tau}_k - \tau_{k^*} \xrightarrow{\mathcal{D}} (\varsigma_k/a_k)^2 \operatorname{argmax}_r(-2^{-1}|r| + \mathbb{W}(r)),$$

where $\mathbb{W}(r)$ is a two-sided Brownian motion. That is $\mathbb{W}(r) = \mathbb{W}_1(r)$, if $r > 0$, and $\mathbb{W}(r) = \mathbb{W}_2(-r)$, if $r \le 0$. $\mathbb{W}_1, \mathbb{W}_2$ are two independent Brownian motions.

Remark 2. We shall note that the consistency rate of $\tilde{\tau}_k$ is improved compared to the results for $\hat{\tau}$ in Theorem 2 ii). a_k which is an l_2 aggregation of simultaneous significant break sizes, plays a role in the rate of convergence of $\tilde{\tau}_k$. For instance, if we assume that there are s breaks which are of size $\delta > 0$ in the cross-sectional dimensional, then $a_k = s\delta^2$. If moreover there is no cross-sectional correlation, that is, $\tilde{\Sigma}_k = I$, then

we may expect $\tilde{\tau}_k$ to be consistent so long that $1/(s\delta^2) \to 0$, while $\hat{\tau}_k$ can be not consistent. Thus, the rate of $\tilde{\tau}_k$ will be better than $\hat{\tau}$. Moreover, the long-run variance also plays a critical role in the rate of convergence. For example, when the variance part of the limit distribution satisfies $\varsigma_k^2 \leq |\tilde{\Sigma}_k|_2 a_k$, if $|\tilde{\Sigma}_k|_2/a_k = o(1)$ then by Theorem 4 (i), we have $\tilde{\tau}_k \to \tau_k$ in probability. This corresponds to the insight of Bai (2010) and Hansen (2000). We can also see that when the breaks are truly sparse in the cross sectional dimension or the break size for each time series is very small, the l^2 aggregation cannot improve the performance compared to the previous step. Also when there are strong cross-sectional dependence l_2 aggregation will not improve the break estimation performance. Moreover, we also need the aggregated breaksize to shrink to zero ($1 \ll$

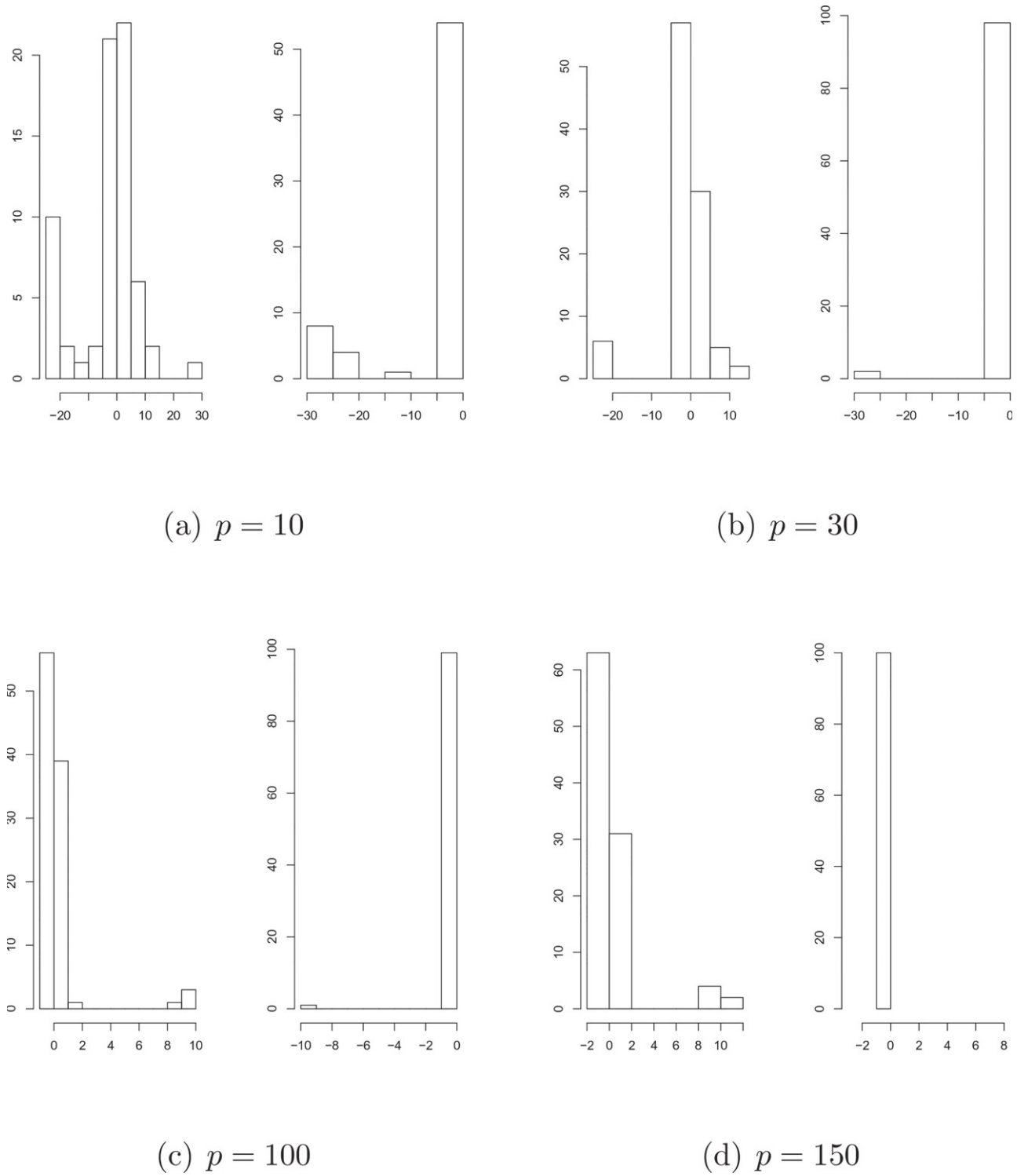

Figure 1. Histogram of $\hat{\tau} - \tau_0$ (left) and $\tilde{\tau} - \tau_0$ (right) for $n = 100$, $p = 10, 30, 100,$ and 150, $K_0 = 1$. The number of breaks in the cross-sectional dimension are $s = 1, 5, 20,$ and 30, respectively, and there are 100 simulation samples. (a) describes the case with $p = 10, s = 1$; (b) describes the case with $p = 30, s = 5$; (c) describes the case with $p = 100, s = 20$; and, (d) describes the case with $p = 150, s = 30$.

ς_k^2/a_k^2) to obtain the limit distribution. If $\tilde{\Sigma}_k$ is a d-banded matrix, $|\tilde{\Sigma}_k|_2 \le (|\tilde{\Sigma}_k|_1|\tilde{\Sigma}_k|_\infty)^{1/2} \le d$. We can derive that $|\tilde{\tau}_k - \tau_k| = O_{\mathbb{P}}(d/a_k)$.

To illustrate the insight of Remark 2, we compare the performance of a simple model with $\mathcal{N}(0, 0.1)$ and one breakpoint placed at $\tau_0 = 50$. Figure 1 shows the histogram of $\hat{\tau} - \tau_0$ and $\tilde{\tau} - \tau_0$, respectively. The jump-size for all breaks are the same, with the value $1.6\sqrt{\log(np)}(\tau_0)^{-1/3}$. As the dimension p grows, we see the significant improvement of the performance of $\tilde{\tau}$ relative to that of $\hat{\tau}$.

From Theorem 4, with estimates of a_k and ς_k, we can construct a $100(1-\alpha)\%$ confidence interval for $\tilde{\tau}_k$:

$$\big(\tilde{\tau}_k - \lfloor \hat{q}'_{1-\alpha/2} \rfloor - 1, \tilde{\tau}_k + \lfloor \hat{q}'_{\alpha/2} \rfloor + 1\big), \tag{37}$$

where $q'_{1-\alpha/2}$ ($q'_{\alpha/2}$) is $1-\alpha/2$ ($\alpha/2$)th quantile of the limit distribution of the break point $\tilde{\tau}_k$, that is, $\text{argmax}_r\{-2^{-1}a_k|r| + \varsigma_k \mathbb{W}(r)\}$ and $\hat{q}'_{\alpha/2}(\hat{q}'_{1-\alpha/2})$ are estimates of the quantiles. $\lfloor \cdot \rfloor$ denotes the floor function. $q'_{1-\alpha/2}$ ($q'_{\alpha/2}$) can be calculated following Stryhn (1996). Alternatively, we can also simulate the critical values.

4. Long-run Covariance Matrix

In the previous sections, we assume that Σ is known. However, this is unrealistic in practice, as we mostly do not know the long-run covariance matrix. Thus, an estimation of the long-run covariance matrix is needed in Gaussian approximation. A simpler version of this problem was considered by Politis, Romano, and Wolf (1999) and Lahiri (2003), who allowed for a constant mean of the random vector. More generally, Chen and Wu (2019) considered the high-dimensional situation with smooth trends. However, this does not fit directly to our interest due to the possible existence of the breakpoints. We then propose a robust covariance matrix estimation motivating from the M-estimation method in Catoni (2012). It is worth noting that due to the jumps, our method shall be different from the classical covariance matrix estimation. Our long-run variance-covariance matrix estimation is complementary to the recent article on high-dimensional robust matrix method under independence settings in Fan, Li, and Wang (2017).

First of all, to account for temporal dependency, we group our observations into blocks of the same size m, for some $m \in \mathbb{N}$. We denote the number of blocks $N_1 = \lfloor (n-m)/m \rfloor$, and the observation indices within a block k is set to be $\mathcal{A}_k = \{t \in \mathbb{N} : km + 1 \le t \le (k+1)m\}$, and we let

$$\xi_k = \sum_{t \in \mathcal{A}_k} Y_t/m,$$

be the average observations within the block $\mathcal{A}_k$. Without jumps, a natural estimate of the long-run covariance matrix is

$$\sum_{k=1}^{N_1} (m/2)(\xi_k - \xi_{k-1})(\xi_k - \xi_{k-1})^\top / N_1.$$

Note that we take the difference $\xi_k - \xi_{k-1}$ to cancel out the trends, as the trend function $\mu(\cdot)$ is smooth, and the aggregated difference between two consecutive blocks can be shown to be of order m/n, which vanishes when $m/n \to 0$. However, this estimator can be greatly contaminated by the jumps, as jumps are not smooth and cannot be canceled out by taking difference. Thus, a robust covariance matrix estimation is needed. We borrow the framework of Catoni (2012), who considered a new robust M- estimation method. We extend the method for estimating our long-run covariance matrix.

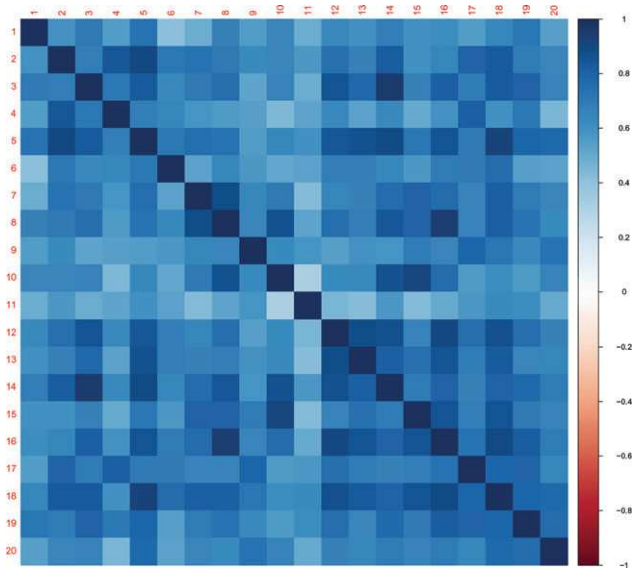

Figure 3. Plot of estimation of the robust long-run correlation matrix; $m = 10$.

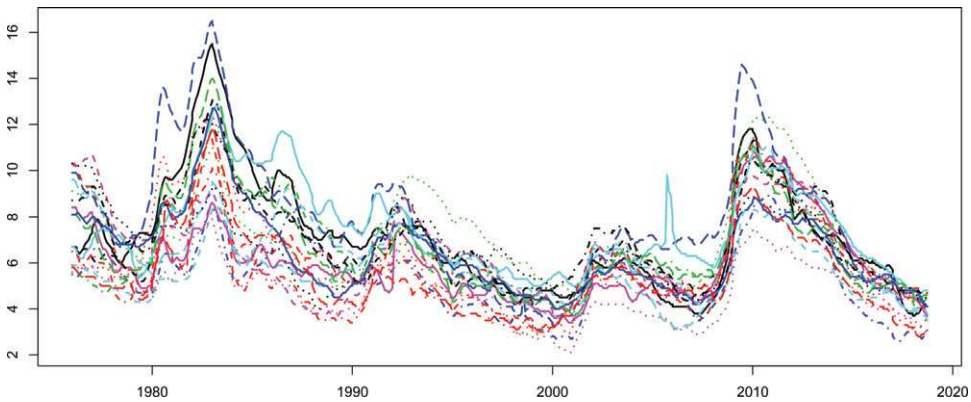

Figure 2. Plot of unemployment rate of 20 U.S. states.

We denote $\xi_k = (\xi_{k,1}, \xi_{k,2}, \ldots, \xi_{k,p})^\top$ and let

$$\hat{\sigma}_{i,j,k} = m(\xi_{k,i} - \xi_{k-1,i})(\xi_{k,j} - \xi_{k-1,j})/2, \quad k = 1, 2, \ldots, N_1. \tag{38}$$

For some $\alpha_{i,j} > 0$, we denote the M- estimation zero function of our variance-covariance matrix to be

$$h_{i,j}(u) = \sum_{k=1}^{N_1} \phi_{\alpha_{i,j}}(\hat{\sigma}_{i,j,k} - u)/N_1, \tag{39}$$

where $\phi_\alpha(x) = \alpha^{-1}\phi(\alpha x)$ and

$$\phi(x) = \begin{cases} \log(2), & x \geq 1, \\ -\log(1 - x + x^2/2), & 0 \leq x \leq 1, \\ \log(1 + x + x^2/2), & -1 \leq x \leq 0, \\ -\log(2), & x \leq -1. \end{cases} \tag{40}$$

Remark 3. Function $|\phi(\cdot)|$ is bounded by $\log(2)$ and is Lipschitz continuous with the Lipschitz constant bounded by 1. Also note that the function has envelopes of nice form,

$$-\log(1 - x + x^2/2) \leq \phi(x) \leq \log(1 + x + x^2/2). \tag{41}$$

We set the estimates of the components of the long-run covariance matrix $\hat{\sigma}_{i,j}$ to be the solution to $h_{i,j}(u) = 0$ (if more than one root, pick one of them). We can collect all the estimates of the variance and covariances and organize them into the variance covariance matrix,

$$\hat{\Sigma} = (\hat{\sigma}_{i,j})_{1 \leq i,j \leq p} \quad \text{and} \quad \hat{\Lambda} = \text{diag}(\hat{\sigma}_{1,1}^{1/2}, \hat{\sigma}_{2,2}^{1/2}, \ldots, \hat{\sigma}_{p,p}^{1/2}). \tag{42}$$

We denote $\bar{\sigma}_{i,i} = 2\sum_{N_1/4 \leq k \leq 3N_1/4} \hat{\sigma}_{i,i,(k)}/N_1$, where $\hat{\sigma}_{i,i,(k)}$ is the ordered statistic of $\hat{\sigma}_{i,i,k}$ as in (38), and let the $\alpha_{i,j}$ in (39) be $\bar{\sigma}_{i,i}^{1/2}\bar{\sigma}_{j,j}^{1/2}(m/n)^{1/2}$.

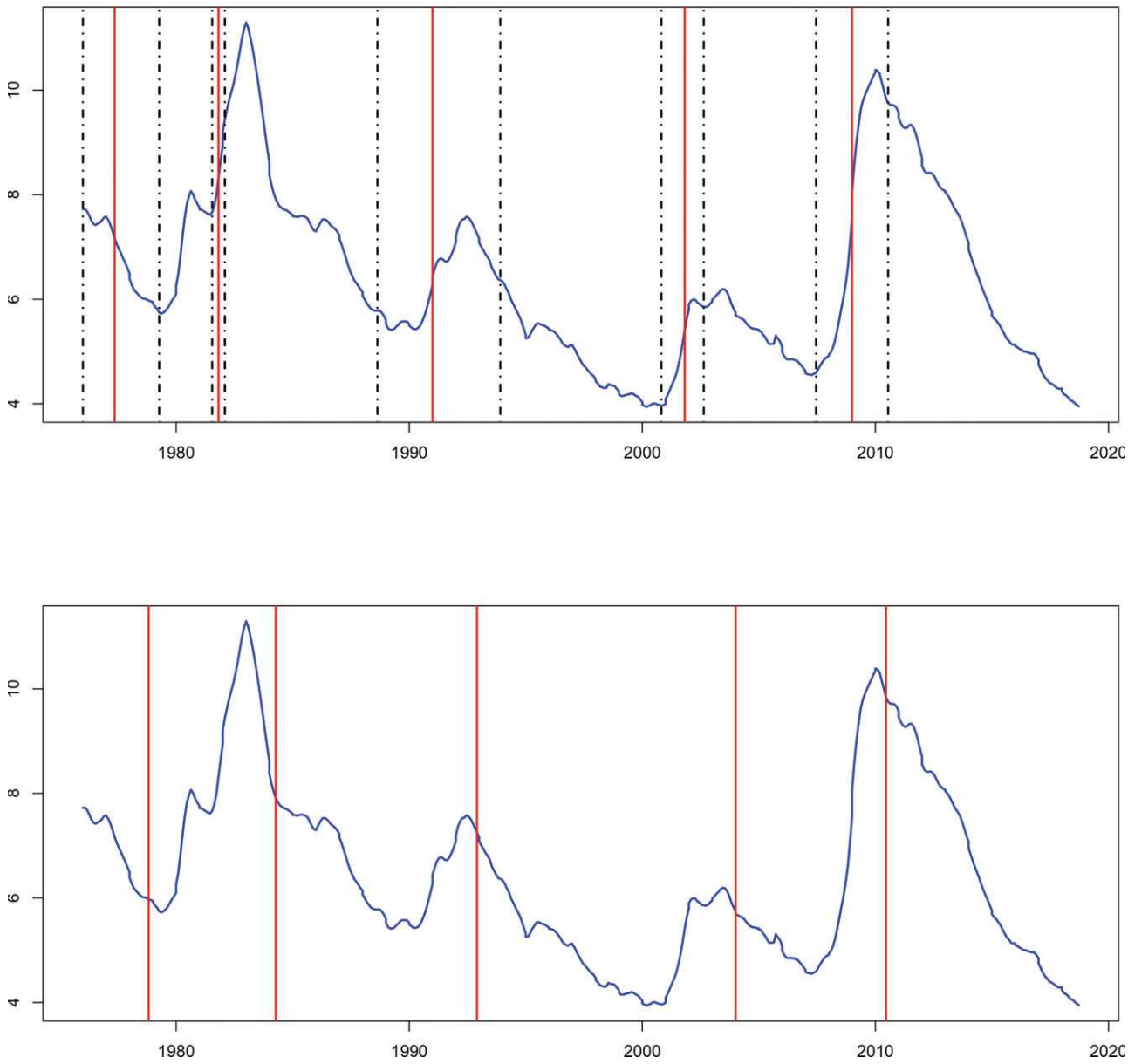

Figure 4. Plot of estimated breakpoints $\tilde{\tau}_k(\hat{\tau}_k)$ (red lines) and their confidence intervals (dotted black lines). $\tilde{\tau}_k$ (upper panel), $\hat{\tau}_k$ (lower panel). The blue time series line represents the average unemployment rate over states under consideration.

Theorem 5. (Long-run variance precision) We assume that Assumption 2.5 holds with $\beta \geq 1.5$ and let

$$\varsigma = |\Lambda^{-1}(\hat{\Sigma} - \Sigma)\Lambda^{-1}|_{\max}.$$

Then for K_0 finite, we have $\varsigma = O_{\mathbb{P}}(n^{-1/4}\log(np))$ under either one of the following two conditions:

(i) Assuming conditions in Theorem 1 (i), $p \leq cn^{\nu}$ with $\nu < q/8 - 1/2$ and some $c > 0$, we take $m = \min\{n^{1-8\nu/(q-4)}, n^{1/2}\}$.
(ii) Assuming conditions in Theorem 1 (ii), we take $m = n^{1/2}$.

By the above theorem, for the diagonal values, we have $\max_{1\leq i\leq p} |\hat{\sigma}_{i,i} - \sigma_{i,i}|/\sigma_{i,i} = o_{\mathbb{P}}(1)$. Let $\hat{Q}$ be the same as Q in (18), with Σ therein replaced by $\hat{\Sigma}$ in Equation (42). We denote $\hat{Z}$ as the Gaussian vector with covariance matrix $\hat{Q}$, then by Theorem 5 and Lemma 3, $|\hat{Z} + d|_{\infty}$ converges to $|Z + d|_{\infty}$ in distribution. Thus, all previous results are still valid with $\hat{\Sigma}$ as well.

5. Application

As an application, we analyze the monthly the unemployment rate data in 20 U.S. states (namely Alabama, Arizona, California, Colorado, Florida, Georgia, Illinois, Indiana, Kentucky, Michigan, Mississippi, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Texas, Virginia, Washington, and Wisconsin). The data time span is from January 1976 to September 2018 ($n = 513$), and the data source is Bureau of Labor Statistics from Department of Labor in the United States (https://www.bls.gov/). Figure 2 displays the 20 time series of unemployment rate. Although from a long time-span and on an overall level, we do not see obvious abrupt structural changes, it would be still of great interest to consider detected changes induced by some well-known exogenous shocks, such as the subprime mortgage crisis in 2007–2008. It is understood that there will be likely a smooth cyclical trend associated with the unemployment time series, as they mostly rise during a recession and fall during periods of economics prosperity, following the business cycle. Further studies on whether the shock induced by recessions creates a significant structural change in the unemployment rate should be performed. We select b according to a cross-validation method, $m = 10$, and $\alpha_{i,j} = \bar{\sigma}_{i,i}^{1/2}\bar{\sigma}_{j,j}^{1/2}(m/bn)^{1/2}$ which varies over different i,j. We have used the estimated 0.999 quantile of the maximum of the Gaussian random variables (as in Equation (19) with correlation matrix replaced by its estimator), which in our case is estimated as 2.10. We refer to the guidance of the selection of tuning parameters as in Remark 4 in the supplementary materials.

Figure 3 shows the estimated robust long-run correlation matrix using the method in Section 4. One sees some significant values in the correlations between residuals in different states. We can see that the correlations across different locations are not negligible, however our method is robust against the underlying spatial–temporal dependency.

Figure 4 plots the estimated breakpoints and the confidence intervals around them. We see that the estimated breaks $\tilde{\tau}_k$ using the CUSUM statistics in Equation (33) pick up the breaks earlier than the estimates obtained from the non-aggregated method, that is, $\hat{\tau}_k$. We can see that our method can identify important dates such as the financial crisis period starting in January, 2009. Moreover, $\tilde{\tau}_k$ tends to detect earlier dates of structure changes than the observed averaged peaks in the time series. Other time-points with significant jumps detected are January 1977, October 1981, January 1991, and October 2001. There are a few recession periods documented by the national bureau of Economics Research, namely November 1973 to March 1975, July 1981 to November 1982, July 1990 to March 1975, July 1981 to November 1982, July 1990 to March 1991, and March 2001 to November 2001. All the break-dates of the unemployment structure happen during or slightly before the recession periods, featuring a close relationship between the structure change of unemployment rate and the economic cycles. This implies that economic recessions indeed bring significant structural changes in unemployment rates across all the states.

Funding

This research is partially supported by NSF-DMS-1916351 and NSF-DMS-2027723.

Supplementary Materials

Simulation results and detailed proofs are provided in the supplementary materials.

References

Bai, J. (2010), "Common Breaks in Means and Variances for Panel Data," *Journal of Econometrics*, 157, 78–92. [1952,1953,1959]
Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models With Multiple Structural Changes," *Econometrica*, 66, 47–78. [1952]
——— (2003), "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics*, 18, 1–22. [1952]
Catoni, O. (2012), "Challenging the Empirical Mean and Empirical Variance: A Deviation Study," *Annales de l'Institut Henri Poincare (B) Probability and Statistics* , 48, 1148–1185. [1960]
Chen, L., Wang, W., and Wu, W. B. (2020), "Dynamic Semiparametric Factor Model With Structural Breaks," *Journal of Business & Economic Statistics*, 1–15. [1951]
Chen, L., and Wu, W. B. (2019), "Testing for Trends in High-dimensional Time Series," *Journal of the American Statistical Association*, 114, 869–881. [1960]
Chernozhukov, V., Chetverikov, D., and Kato, K. (2013), "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-dimensional Random Vectors," *Annals of Statistics*, 41, 2786–2819. [1955]
——— (2017), "Central Limit Theorems and Bootstrap in High Dimensions," *Annals of Probability*, 45, 2309–2352. [1955]
Cho, H. (2016), "Change-point Detection in Panel Data Via Double CUSUM Statistic," *The Electronic Journal of Statistics*, 10, 2000–2038. [1952,1957]
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. [1952]
Eichinger, B., and Kirch, C. (2018), "A MOSUM Procedure for the Estimation of Multiple Random Change Points," *Bernoulli*, 24, 526–564. [1952]
Enikeeva, F., and Harchaoui, Z. (2019), "High-dimensional Change-point Detection Under Sparse Alternatives," *Annals of Statistics*, 47, 2051–2079. [1952]
Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Volume 66 of Monographs on Statistics and Applied Probability, London: Chapman & Hall. [1953,1954]

Fan, J., Li, Q., and Wang, Y. (2017), "Estimation of High Dimensional Mean Regression in the Absence of Symmetry and Light Tail Assumptions," *Journal of the Royal Statistical Society*, Series B, 79, 247–265. [1960]

Fryzlewicz, P. (2014), "Wild Binary Segmentation for Multiple Change-point Detection," *Annals of Statistics*, 42, 2243–2281. [1952]

——— (2018), "Tail-greedy Bottom-up Data Decompositions and Fast Multiple Change-point Detection," *Annals of Statistics*, 46, 3390–3421. [1952]

Hansen, B. E. (2000), Sample splitting and threshold estimation. *Econometrica*, 68, 575–603. [1959]

Harlé, F., Chatelain, F., Gouy-Pailler, C., and Achard, S. (2016), "Bayesian Model for Multiple Change-points Detection in Multivariate Time Series," *IEEE Transaction on Signal Processing*, 64, 4351–4362. [1951]

Hušková, M., and Slabỳ, A. (2001), "Permutation Tests for Multiple Changes," *Kybernetika*, 37, 605–622. [1952]

Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., and Tsai, T. T. (2005), "An Algorithm for Optimal Partitioning of Data on An Interval," *IEEE Signal Processing Letters*, 12, 105–108. [1952]

Jirak, M. (2015), "Uniform Change Point Tests in High Dimension," *Annals of Statistics*, 43, 2451–2483. [1952]

Killick, R., Fearnhead, P., and Eckley, I. A. (2012), "Optimal Detection of Changepoints With a Linear Computational Cost," *Journal of American Statistical Association*, 107, 1590–1598. [1952]

Lahiri, S. N. (2003), *Resampling Methods for Dependent Data*, Springer Series in Statistics, New York: Springer-Verlag. [1960]

Lee, S., Seo, M. H., and Shin, Y. (2016), "The Lasso for High Dimensional Regression With a Possible Change Point," *Journal of the Royal Statistical Society*, Series B, 78, 193–210. [1952]

Lévy-Leduc, C., and Roueff, F. (2009), "Detection and Localization of Change-points in High-dimensional Network Traffic Data," *Annals of Applied Statistics*, 3, 637–662. [1951]

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. [1952]

Liu, H., Gao, C., and Samworth, R. J. (2019), "Minimax Rates in Sparse, High-dimensional Changepoint Detection," arXiv preprint arXiv:1907.10012. [1952]

Meier, A., Kirch, C., and Cho, H. (2019), "mosum: A Package for Moving Sums in Change-point Analysis," Preprint. [1952]

Olshen, A. B., Venkatraman, E., Lucito, R., and Wigler, M. (2004), "Circular Binary Segmentation for the Analysis of Array-based DNA Copy Number Data," *Biostatistics*, 5, 557–572. [1952]

Politis, D. N., Romano, J. P., and Wolf, M. (1999), *Subsampling*. Springer Series in Statistics, New York: Springer-Verlag. [1960]

Preuss, P., Puchstein, R., and Dette, H. (2015), "Detection of Multiple Structural Breaks in Multivariate Time Series," *Journal of American Statistical Association*, 110, 654–668. [1952]

Scott, A. J., and M. Knott (1974), "A Cluster Analysis Method for Grouping Means in the Analysis of Variance," *Biometrics*, 30, 507–512. [1952]

Stryhn, H. (1996), "The Location of the Maximum of Asymmetric Two-sided Brownian Motion With Triangular Drift," *Statistics & Probability Letters*, 29, 279 – 284. [1960]

Tibshirani, R., and Wang, P. (2007), "Spatial Smoothing and Hot Spot Detection for CGH Data Using the Fused Lasso," *Biostatistics*, 9, 18–29. [1952]

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. [1952,1958]

Wu, W., and Zhou, Z. (2019), "Mace: Multiscale Abrupt Change Estimation Under Complex Temporal Dynamics," arXiv preprint arXiv:1909.06307. [1952,1957]

5 *The Electronic Journal of Statistics*, 10, 352–379.

Wu, W. B., and Zhao, Z. (2007), "Inference of Trends in Time Series," *Journal of the Royal Statistical Society*, Series B, 69, 391–410. [1952]

Zhang, D., Wu, W. B. (2017), "Gaussian Approximation for High Dimensional Time Series," *The Annals of Statistics*, 45, 1895–1919. [1955]

Zhang, N. R., Siegmund, D. O., Ji, H., and Li, J. Z. (2010), "Detecting Simultaneous Changepoints in Multiple Sequences," *Biometrika*, 97, 631–645. With supplementary data available online. [1951,1952]