IRTG 1792 Discussion Paper 2019-013

Inference of Break-Points in High-Dimensional Time Series

Likai Chen [*]
Weining Wang [*2]
Wei Biao Wu [*3]

[*] Washington University in St. Louis
[*2] Humboldt-Universität zu Berlin
[*3] University of Chicago

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

INFERENCE OF BREAK-POINTS IN HIGH-DIMENSIONAL TIME SERIES

By Likai Chen*, Weining Wang† and Wei Biao Wu‡

Washington University in St. Louis, City, U of London, C.A.S.E.-Center of Applied Statistics and Economics, Humboldt-Universität zu Berlin† and University of Chicago‡*

We consider a new procedure for detecting structural breaks in mean for high-dimensional time series. We target breaks happening at unknown time points and locations. In particular, at a fixed time point our method is concerned with either the biggest break in one location or aggregating simultaneous breaks over multiple locations. We allow for both big or small sized breaks, so that we can 1), stamp the dates and the locations of the breaks, 2), estimate the break sizes and 3), make inference on the break sizes as well as the break dates. Our theoretical setup incorporates both temporal and cross-sectional dependence, and is suitable for heavy-tailed innovations. We derive the asymptotic distribution for the sizes of the breaks by extending the existing powerful theory on local linear kernel estimation and high dimensional Gaussian approximation to allow for trend stationary time series with jumps. A robust long-run covariance matrix estimation is proposed, which can be of independent interest. An application on detecting structural changes of the US unemployment rate is considered to illustrate the usefulness of our method.

1. Introduction. Statistical inference of structural breaks in mean is an important subject to study, which involves estimating the trend functions, detecting and locating abnormal changes, and making inferences. 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 ([24]), recurrent DNA copy number variants in multiple samples ([37]), abrupt changes in household electrical power consumption ([18]) and minimum wage policy changes analysis ([9]), 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 $Y_1, Y_2, \ldots, Y_n$

MSC 2010 subject classifications: Primary 62E17, 62G10
Keywords and phrases: high-dimensional time series, multiple change-points, Gaussian approximation, nonparametric estimation, heavy tailed, long-run covariance matrix

follow the model,

$$Y_t = \mu(t/n) + \epsilon_t, \quad t = 1, 2, \ldots, n, \tag{1.1}$$

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

$$\mu(u) = f(u) + \sum_{i=1}^{K_0} \gamma_i \mathbf{1}_{u \geq u_i}, \tag{1.2}$$

where $K_0 < \infty$ is some 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 $0 < u_1 < u_2 < \ldots < u_{K_0} < 1$ are the time-points of the change-points and $\gamma_k \in \mathbb{R}^p$s 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 on the cross-sectional dimension for any fixed time point k (cf. 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). Compared to the classical change-point settings, the smooth part of the trend functions is zero, i.e. $f \equiv 0$. This means that the trend functions are piece-wise 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 detect the existence of structural breaks, and in case that breaks exist, to identify their change-point u_k, to calibrate sizes of the breaks $|\gamma_k|_\infty$, $1 \leq k \leq K_0$, and to construct confidence intervals for the estimated break points.

Change-point detection for univariate or finite dimensional time series has been intensively studied in the literature, see for instance [1, 2], [6] and [13]. Differently, we shall consider the case of $p \to \infty$. This setting has recently drawn a lot of attention due to the increasing number of applications involving large dimensional data in real practice, see [28], [37], [35], [14], [33], etc. All the aforementioned literature assume the error processes to be temporally or cross-sectionally independent. Such assumptions are too restrictive since for time series dependence is the rule rather than the exception. Moreover the commonly assumed normality assumption for the error term, associated with the light tailedness of the distribution, are often too restrictive for real data. To see this point, [26] documents evidence of the power-law behavior in asset prices and [30] show heavy tailedness in the high frequency asset return data.

Due to technical challenges, the literature is thinner for change-point detection in high dimensional data with dependence. [3] considers detecting a common break-point in time, with linear temporal dependence and no cross-sectional dependence; [22] extends his method by allowing additional factor structure to the noise. In this paper, we characterize our dependence structure through the widely used vector moving average model with infinite many lags (VMA(∞)), which also incorporates the models in [3] and [22]. Specifically, let

$$\epsilon_t = \sum_{k\geq 0} A_k \eta_{t-k}, \tag{1.3}$$

where $\eta_t \in \mathbb{R}^{\tilde{p}}$ are independent and identically distributed (i.i.d.) random vectors with zero mean and identity covariance matrix, and $A_k(k \geq 0)$s are coefficient matrices in $\mathbb{R}^{p\times\tilde{p}}$ such that ϵ_t is a proper random vector, and $p \leq \tilde{p} \leq c_p p$, for some constant $c_p > 1$. If $A_i = 0$ for all $i \geq 1$, then the noise sequences are temporally independent; if $p = \tilde{p}$ and matrices A_i are diagonal, then the sequences become the model in [3] which is spatially independent. The VMA(∞) process is very general and includes most important time series models such as a vector autoregressive moving averages (VARMA) model, i.e.

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

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 $|z| \leq 1$ and $\mathcal{B}$ is the backshift operator.

For multiple change-points detection, various approaches have been invented. A traditional one is through an exhaustive search, which exams all the possible break points combination. Exhaustive search is very time consuming and some dynamic technique and improved versions are invented, see for instance [4, 5] and [21]. Moreover, when $f \equiv 0$, the problem can be reformulated into a regression problem with sparsity in the parameters, and thus can be solved through the LASSO (least absolute shrinkage and selection operator) method, e.g. [19], [32], and [25]. A further alternative is the binary segmentation method and its variants, see for example [28] and [17]. The basic idea is to apply one single change-point detection recursively within each segments. However, it can be difficult to interpret the results and make statistical inference as a multiple testing procedure is involved. In sum, a computationally cheap method tailored to high dimensional non-stationary time series with general dependency structure is needed.

Comparatively, we adopt a two-step nonparametric approach. In the first step, we conduct a "rough" estimation based on a nonparametric moving sum type method. To be more specific, for any fixed time point t, we consider two kernel estimators, $(\hat{\mu}^{(r)}(t/n), \hat{\mu}^{(l)}(t/n))$ of the trends using the left and the right hand side observations respectively. Then we draw conclusion on the existence of a break and obtain a "rough" estimate of the change points locations according to the difference of the two kernel estimates. In the second step, we refine our jump estimates based on a one-dimensional aggregated time series. Utilizing estimates from the first step, the aggregated time series is obtained by a weighted sum of simultaneous observations corresponding to estimated significant jump locations. The weights are determined by the estimated jump sizes. 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 thus would achieve better precision asymptotically. For moving sum type statistic, [20] and [34] consider the univariate situations and [27] apply it for a screening and ranking algorithm. In this paper, we adopt the local linear method for a better boundary performance, and then establish a uniform Gaussian approximation for the gap estimate $|\hat{\mu}^{(r)}(t/n) - \hat{\mu}^{(l)}(t/n)|_\infty$, where $|x|_\infty = \max_{1\le j\le p} |x_i|$.

Our theoretical results extend the Gaussian approximation theory in [11, 12], 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 have derived the corresponding concentration inequalities based on m-dependence approximation of the underlying processes. Compared to existing results on Gaussian approximation for time series, for example [36], our setting works for non-centered Gaussian approximation accommodating our interest for time series with non-smooth trends.

A further challenge is that due to the underlying dependence, our test statistic needs an estimation of the long-run covariance matrix. A simpler version of this problem has been considered in [29] and [23] who allow for a constant mean of the random vector. More generally, [10] consider the high-dimensional situation with smooth trends. However, this does not fit directly to our interest due to the possible existence of the break points, we then propose a robust covariance matrix estimation motivating from the M-estimation method in [8] (cf. Section 2.4). Our long run variance covariance matrix estimation is complementary to the recent article on high dimensional robust matrix method under independence settings in [16].

To summarize, our method greatly improves upon the existing methods on change-point detection for multivariate time series in five folds: (i) the dimension p can increase as n increases, (ii) both temporal and cross-sectional dependence are incorporated, (iii) we do not restrict the noise to follow a specific distribution and allow for the case of heavy-tailedness, (iv) we do not require the mean function to be a piece-wise constant (i.e. $f \equiv 0$), and instead we allow f to vary smoothly, (v) asymptotic distribution is derived for the change points locations to facilitates inference on the break-points. Our procedure is not computationally expensive as we only need to evaluate the statistic for each point t once. Additionally, we consider the estimation of the long-run covariance matrices. The paper is structured as follows. Section 2 presents the model setup, and descriptions of change-points test procedure. Section 3 contains the assumptions and the main theorems for testing the existence of the change-points with results on test sizes and power. In the same section, we also derive the asymptotic distribution of the sizes of the breaks. Simulation results are in Section 4 and an application on US unemployment rate is given in Section 5. Detailed proofs are presented in Supplementary materials.

2. Model and estimation. We now introduce some notations. For a constant $k > 0$ and a vector $v = (v_1, \dots, v_d)^\top \in \mathbb{R}^d$, 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.

We then provide our model setups and our estimation procedure with explanations on how our method works. Considering our observations generated by the model in (1.1) and (1.2), we would like to test the null hypothesis,

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

which corresponds to case of no breaks, against the alternative of the existence of at least one break i.e. $\mathcal{H}_A: \exists k \in 1, \cdots, K_0, \quad \gamma_k \neq 0$.

Next we discuss how to obtain our test statistics. Recall that our trend function $\mu(u)$ can be disentangled into two parts, namely, a smooth transition part $f(u)$ and a jump part $\gamma_i \mathbf{1}\{u \geq u_i\}$. To estimate the trend function $\mu(u)$, 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 $\mu^{(l)}(u)$. Let the gap function

$$\begin{aligned} &J(u) = \mu^{(r)}(u) - \mu^{(l)}(u), \\ &\quad \text{where} \quad \mu^{(r)}(u) = \lim_{t \downarrow u} \mu(t) \quad \text{and} \quad \mu^{(l)}(u) = \lim_{t \uparrow u} \mu(t). \end{aligned}$$

If we assume certain degree of 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 classical nonparametric local linear estimation technique, see [15].

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

$$\hat{\mu}_i^{(l)} = \sum_{t=i-bn}^{i-1} w_{i-t} Y_t \quad \text{and} \quad \hat{\mu}_i^{(r)} = \sum_{t=i+1}^{i+bn} w_{t-i} Y_t, \tag{2.1}$$

with weights

$$w_i = w_{i,b} = w_b(0, i/n), \quad i \geq 1, \quad w_0 = 0. \tag{2.2}$$

The weight functions are defined as

$$\begin{aligned} w_b(u, v) &= \frac{K((v-u)/b)[S_2(u) - (u-v)S_1(u)]}{S_2(u)S_0(u) - S_1^2(u)}, \\ S_l(u) &= \sum_{i=1}^{n} (u - i/n)^l K((i/n - u)/b), \end{aligned} \tag{2.3}$$

where $K(.)$ is a kernel function and b is a bandwidth with $b \to 0$ and $bn \to \infty$. The kernel function plays a critical role in localization, and the bandwidth parameter is pertained to the local window size. Usually we assume that the kernel function $K(.)$ is smooth and has a compact support $[-1, 1]$.

The local linear estimates can precisely evaluate the gap level with a proper choice of the bandwidth. Therefore if there is no jump around the time point $u = i/n$, the gap estimate $\hat{J}(i/n) = \hat{\mu}_i^{(l)} - \hat{\mu}_i^{(r)}$ would be small. Otherwise if for some entity $1 \leq i \leq p$, the gap estimate $|\hat{J}_j(i/n)|$ is too large, there might exist a jump around i/n.

Furthermore, we need to standardize our test statistics in order to get a regular limit distribution. For this purpose, the long-run covariance matrix of the underlying innovations is involved in constructing our test statistic. Recall the definition of the error process as in (1.3), define the sum of the coefficient matrix to be $S = \sum_{k \geq 0} A_k$. The long-run covariance matrix for the error process is

$$\Sigma = SS^\top. \tag{2.4}$$

We denote $\Sigma = (\sigma_{i,j})$, $1 \leq i, j \leq p$, and let the diagonal elements of the variance covariance matrix form a new matrix

$$\Lambda = \operatorname{diag}(\sigma_{1,1}^{1/2}, \sigma_{2,2}^{1/2}, \ldots, \sigma_{p,p}^{1/2}). \tag{2.5}$$

Finally, we suppose that we know the long run variances Λ for the moment. Following the previous intuition of the effect of jumps on the gap statistics $\hat{J}(.)$, we consider the test statistic

$$T_n = \max_{bn+1 \leq i \leq n-bn} |V_i|_\infty, \quad \text{where} \quad V_i = \Lambda^{-1}(\hat{\mu}_i^{(l)} - \hat{\mu}_i^{(r)}). \tag{2.6}$$

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 $1 \leq j \leq p$ can be very different.

REMARK 1. *In practice, we usually do not know the long-run covariance matrix Σ. Its estimation $\hat{\Sigma}$ is discussed in Section 2.4. It is worth noting that all the theorems remain true with Σ replaced by $\hat{\Sigma}$ if additionally conditions in Theorem 5 in Section 3 holds.*

2.1. *Critical values for testing the breaks.* We will proceed with analyzing the asymptotic properties of our test statistics, which gives us critical values of our test. First we analyze the mean of the normalized jump estimators, i.e. $\mathbb{E}V_i$. Intuitively, we can decompose the level of our jump estimator $\mathbb{E}V_i$ by two parts, one is the commonly encountered bias term for the nonparameteric 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 $\tau_k = nu_k$, the definition of w_i in (2.2) for $i = 1, 2, \ldots, bn$, and $w_i = 0$ for $i = 0$ and $i > bn$. We denote Ω_i as a set of indices indicating the break locations within the bn neighborhood around time i, namely $\Omega_i = \big\{k \big| |i - \tau_k| \leq bn, 1 \leq k \leq K_0\big\}$. For a time point i with at least one break around i.e. $\Omega_i \neq \emptyset$, we define the weighted break size to be,

$$d_i = (1 - \sum_{t=1}^{|i-\tau_k|} w_t)\Lambda^{-1}\gamma_k, \quad k = \mathrm{argmin}_{j \in \Omega_i} |i - \tau_j|, \tag{2.7}$$

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

For large n, the cardinality of Ω_i is at most one, i.e. $|\Omega_i| \leq 1$. Since for $k_1 \neq k_2$, $|\tau_{k_1} - \tau_{k_2}| > 2bn$ in view of $\tau_{k_1} - \tau_{k_2} = n(u_{k_1} - u_{k_2}) \asymp n$. Actually we do not require $\tau_{k_1} - \tau_{k_2} \asymp n$ as all the results can be directly extended to the case of $\min_{1 \leq k_1 \neq k_2 \leq K_0} |\tau_{k_1} - \tau_{k_2}| \gg bn$. 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 [15], 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.

$$\max_{bn+1 \leq i \leq n-bn} |\Lambda^{-1}(\hat{f}_i^{(l)} - \hat{f}_i^{(r)})|_\infty = O(b^2). \tag{2.8}$$

Given the definition of our model $Y_i = \mu(i/T) + \epsilon_i$, d_i can be expressed by the estimated gap $(\hat{\mu}_i^{(r)} - \hat{\mu}_i^{(l)})$ with a smooth trend component eliminated,

$$d_i = \mathbb{E}\big\{\Lambda^{-1}\big((\hat{\mu}_i^{(r)} - \hat{\mu}_i^{(l)}) - (\hat{f}_i^{(r)} - \hat{f}_i^{(l)})\big)\big\}. \tag{2.9}$$

Accordingly, the expectation of our jump statistics V_i would be dominated by the part induced by jumps γ_ks, as

$$|\mathbb{E}V_i - d_i|_\infty = |\Lambda^{-1}(\hat{f}_i^{(r)} - \hat{f}_i^{(l)})|_\infty = O(b^2). \tag{2.10}$$

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,

$$V_i - \mathbb{E}V_i = \sum_{l=i-bn}^{i-1} w_{i-l}\Lambda^{-1}\epsilon_l - \sum_{l=i+1}^{i+bn} w_{l-i}\Lambda^{-1}\epsilon_l. \tag{2.11}$$

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

$$G^\diamond_{i,l} = \begin{cases} w_{i-l}\Lambda^{-1}S, & \text{if} \quad i-bn\leq l\leq i-1, \\ -w_{l-i}\Lambda^{-1}S, & \text{if} \quad i+1\leq l\leq i+bn, \end{cases} \tag{2.12}$$

and elsewhere zero. We let $\underline{z}$ be a Gaussian vector in $\mathbb{R}^{n\tilde{p}}$ with zero mean and identity covariance matrix. We set $G^\diamond_{i,\cdot}$ be $(G^\diamond_{i,1}, G^\diamond_{i,2}, \ldots, G^\diamond_{i,n})$. It can be shown that $G^\diamond_{i,\cdot}\underline{z}$ has similar covariance structure as $V_i - \mathbb{E}V_i$. In the special case when ϵ_i are temporally independent, they share the same covariance structure. We shall use the distribution of $|G^\diamond_{i,\cdot}\underline{z}|_\infty$ to approximate the distribution of $|V_i - \mathbb{E}V_i|_\infty$. This approximation is further combined together with the bias term as in (2.10), and 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 around a level d_i,

$$\mathbb{P}(|V_i|_\infty \leq u) \approx \mathbb{P}(|d_i + G^\diamond_{i,\cdot}\underline{z}|_\infty \leq u).$$

We now let the statistics go over all the time points, and recall the definition of $T_n = \max_{bn+1\leq i\leq n-bn}|V_i|_\infty$. The asymptotic distribution is also expected to be approximated well by a multivariate Gaussian distribution.

$$\mathbb{P}(T_n \leq u) \approx \mathbb{P}(|\underline{d} + G^\diamond\underline{z}|_\infty \leq u), \tag{2.13}$$

where we recall that $\underline{d}$ is a long vector by stacking d_i, and $G^\diamond$ is a $(n-2bn)p\times n\tilde{p}$ matrix. The above argument will be rigorously formulated in Theorem 1 in Section 3.

Subsequently we can figure out the critical values and the power of our test. First, we define the block matrix to be

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

Therefore Q only depends on the long-run covariance matrix Σ and the weight functions with the form

(2.14)

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

Define $(Z_i)_{bn+1\leq i\leq n-bn}$ to be a sequence of centered Gaussian vectors in $\mathbb{R}^p$ with covariance matrices $Q_{i,j}$, for $bn+1\leq i,j\leq n-bn$. We let $Z =$

$(Z_{bn+1}^\top, Z_{bn+2}^\top, \ldots, Z_{n-bn}^\top)^\top$. Then Z is a Gaussian vector with zero mean and covariance matrix Q. Thus we have the two Gaussian vectors Z_i and Z with the same distribution as the above Gaussian limits

$$Z_i \overset{d}{=} G_{i,\cdot}^\diamond \underline{z} \quad \text{and} \quad Z \overset{d}{=} G^\diamond \underline{z}. \tag{2.15}$$

Next 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 i.e. the quantile of the Gaussian limit distribution,

$$q_\alpha = \inf_{r \geq 0} \{ r : \mathbb{P}(|Z|_\infty > r) \leq \alpha \}. \tag{2.16}$$

As from the Gaussian approximation results, we have the approximated sizes of the test statistics,

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

Finally we will reject the null hypothesis at the significant level α, if the test statistics exceed the critical value i.e. $T_n > q_\alpha$. Moreover under the alternative hypothesis, $\underline{d} \neq 0$, we can derive the corresponding power (cf. Corollary 1)

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

We can see that the power our test would be depending on the vector $\underline{d}$, whose size is determined by the true jump sizes i.e. γ_ks.

2.2. *Test procedure and break sizes.* We now summarize our procedure in this subsection on estimating the number of change-points, the time stamps, the spatial coordinates and the sizes of the structural breaks. Define the sizes of the break points at time k as the maximum absolute value over the jump vector,

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

Here we normalize γ_k by the diagonal matrix of the long run variances Λ^{-1} for the same reason as V_i in (2.6). 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

$$\delta^\diamond = \min_{1 \leq k \leq K_0} |\Lambda^{-1} \gamma_k|_\infty. \tag{2.17}$$

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

Step 1. For some significance level α, we test the existence of jumps based on the critical value q_α and the test statistics T_n. If we find no significant breaks, then we cannot reject the null $\mathcal{H}_0$ that there exists no break point at significant level α. q_α is selected according to the explanation in the previous section and based on our Gaussian approximation results in Theorem 1. 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 (2.6). Note that the threshold value should be big enough to ensure that we identify breaks with probability approaching 1. 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 $\hat{\tau}_k = \mathrm{argmax}_{\tau\in\mathcal{A}_k}|V_\tau|_\infty$ and $\mathcal{A}_{k+1} = \mathcal{A}_k \setminus \{\tau : |\tau-\hat{\tau}_k| \le 2bn\}$. We denote the maximum number of break points as $\hat{K}_0$, with $\hat{K}_0 = \max_{k\ge1}\{k : \mathcal{A}_k \ne \emptyset\}$.
Step 3. Given the detected break points in Step 2, we calculate the break sizes over time. We denote the window size to be $M = bn$,

$$\hat{\gamma}_k = Y^{(r)}_{\hat{\tau}_k-M} - Y^{(l)}_{\hat{\tau}_k+M} \quad \text{and} \quad \hat{\delta}^\diamond = \min_{1\le k\le\hat{K}_0} |\Lambda^{-1}\hat{\gamma}_k|_\infty. \tag{2.18}$$

It is worth noting that in this algorithm, we only need to calculate once for each point the gap statistics $|V_\tau|_\infty$ hence it is not time consuming regardless of the true number of break-points. 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 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 presence of jumps, it would be of further interest to make inference on the break sizes. We thus introduce the confidence interval induced from the asymptotic distribution of our test statistics. We recall the definition for $Q_{i,j}$ in (2.14). We let $\tilde{Z}$ be a Gaussian vector in $\mathbb{R}^p$ with zero mean and a covariance matrix

$$\tilde{Q} := Q_{bn+1,bn+1} = 2\sum_{t=1}^{bn} w_t^2 \Lambda^{-1}\Sigma\Lambda^{-1}. \tag{2.19}$$

Based on Theorem 2 (ii) and Theorem 3 in Section 3, we can construct the confidence interval for γ_k. We set

$$\alpha = \mathbb{P}(|\tilde{Z}|_\infty \geq q) \quad \text{and} \quad \theta = (\sigma_{1,1}^{1/2}, \sigma_{2,2}^{1/2}, ..., \sigma_{p,p}^{1/2})^\top. \tag{2.20}$$

Then the confidence interval for vector γ_{k^*} at level α is $(\hat{\gamma}_k - q\theta, \hat{\gamma}_k + q\theta)$.

2.3. *Aggregated change-point detection, a further refinement..* The estimation procedure in Section 2.2 is only driven by $|\gamma_k|_\infty$, i.e. the maximum size of jumps at a time point k. Therefore it is only sensitive to the biggest jump across all the time series at the same time. In case there are multiple jumps occurring at the same time, it would be beneficial to modify our procedure to aggregate all of the series with jump.

To this end, we need to modify our procedure to first look at the estimated sizes of the jumps across different time series. In the case that jump happens at multiple locations simultaneously, aggregating all of them would improve the estimation precision. This enlightens us to propose a new two-stage method: firstly, we follow the steps in the previous subsections to detect the "rough" timing of the jumps and the estimation of series that jump at each location; secondly, for each bn neighborhood of a change-point $\hat{\tau}_k$ in 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 us 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

$$\mathcal{S}_k = \{1 \leq j \leq p \mid \gamma_j \neq 0\}. \tag{2.21}$$

Detailed steps of the aggregation are formulated as follows:
Stage 1. Apply Steps 1-3 in Subsection 2.2 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

$$\hat{\mathcal{S}}_k = \big\{1 \leq j \leq p \big| |(\Lambda^{-1}\hat{\gamma}_k)_j| \geq w^\dagger\big\}. \tag{2.22}$$

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 $|t - \hat{\tau}_k| \leq bn$, we let

$$X_t = \sum_{j \in \hat{\mathcal{S}}_k} (\Lambda^{-1}\hat{\gamma}_k)_j (\Lambda^{-1} Y_t)_j. \tag{2.23}$$

Note that after the modification, for all the jump locations, the new time series X_t would only contain positive sized jumps i.e. $\sum_{j\in\hat{\mathcal{S}}_k}(\Lambda^{-1}\gamma_k)_j^2$.

Based on the aggregated time series X_t, the refined change point locations can be detected through a "CUSUM"(cumulative sum) type of test statistics

$$\tilde{\tau}_k = \mathrm{argmax}_{|t-\hat{\tau}_k|\le bn}\Big(\sum_{s=\hat{\tau}_k-bn}^{\hat{\tau}_k+bn} X_s \frac{t-\hat{\tau}_k+bn}{2bn+1} - \sum_{s=\hat{\tau}_k-bn}^{t-1} X_s\Big), \quad k=1,2,\ldots,\hat{K}_0. \tag{2.24}$$

After we update the break points estimation, we can construct confidence intervals for the updated break points 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 variance covariance matrix corresponding to the significant jump locations at time k and 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 i.e.

$$a_k = |\tilde{\gamma}_k|_2 \quad \text{and} \quad \varsigma_k^2 = \tilde{\gamma}_k^\top \tilde{\Sigma}_k \tilde{\gamma}_k. \tag{2.25}$$

Then ς_k^2 is the long run variance for the sequence $\sum_{j\in\mathcal{S}_k}(\Lambda^{-1}\gamma_k)_j(\Lambda^{-1}\epsilon_t)_j$. From Theorem 4 in Section 3, with estimates of a_k and ς_k, we can construct a $100(1-\alpha)\%$ confidence interval for $\tilde{\tau}_k$:

$$\big(\tilde{\tau}_k - \lfloor \hat{q}'_{1-\alpha/2}\rfloor - 1, \tilde{\tau}_k + \lfloor \hat{q}'_{\alpha/2}\rfloor + 1\big), \tag{2.26}$$

where $q'_{1-\alpha/2}$ ($q'_{\alpha/2}$) is $1-\alpha/2$ ($\alpha/2$)th quantile of the limit distribution of the break point $\tilde{\tau}_k$ i.e. $\mathrm{argmax}_r\{-2^{-1}a_k^2|r| + \varsigma_k\mathbb{W}(r)\}$, 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 [31]. Alternatively we can also simulate the critical values.

2.4. *Estimation of the long-run covariance matrix.* In the previous subsections, we assume that Σ is known. However this is unrealistic in practice as we mostly do not know the long run covariance matrix. Therefore, in this subsection, we provide an estimator for it. It is worth noting that due to the jumps, our method shall be very different from the classical covariance matrix estimation.

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

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

We 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. Thus a robust covariance matrix estimation is needed. We borrow the framework of [7], who considers a new robust M- estimation method. We extend the method for estimating our long run covariance matrix.

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

$$\hat{\sigma}_{i,j,k} = m(\xi_{k,i} - \xi_{k-1,i})(\xi_{k,j} - \xi_{k-1,j})/2, \quad k = 1, 2, \ldots, N_1. \tag{2.27}$$

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

$$h_{i,j}(u) = \sum_{k=1}^{N_1} \phi_{\alpha_{i,j}}(\hat{\sigma}_{i,j,k} - u)/N_1, \tag{2.28}$$

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

$$\phi(x) = \begin{cases} \log(2), & x \ge 1, \\ -\log(1 - x + x^2/2), & 0 \le x \le 1, \\ \log(1 + x + x^2/2), & -1 \le x \le 0, \\ -\log(2), & x \le -1. \end{cases} \tag{2.29}$$

Remark 2. *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 envelops of nice form,*

$$-\log(1 - x + x^2/2) \le \phi(x) \le \log(1 + x + x^2/2). \tag{2.30}$$

We set the estimates of the components of the long run covariance matrix $\hat{\sigma}_{i,j}$ 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 matrx,

$$\hat{\Sigma} = (\hat{\sigma}_{i,j})_{1 \le i,j \le p}, \quad \text{and} \quad \hat{\Lambda} = (\hat{\sigma}_{1,1}^{1/2}, \hat{\sigma}_{2,2}^{1/2}, \dots, \hat{\sigma}_{p,p}^{1/2}). \tag{2.31}$$

We denote $\bar{\sigma}_{i,i} = 2\sum_{N_1/4 \le k \le 3N_1/4} \hat{\sigma}_{i,i,k}/N_1$ and let the $\alpha_{i,j}$ in (2.28) be $\bar{\sigma}_{i,i}^{1/2}\bar{\sigma}_{j,j}^{1/2}(m/n)^{1/2}$.

3. Main theorems.

In this section, we present necessary assumptions to guarantee good empirical performance of our method and provide the theoretical foundations of our test procedure. The following condition is to guarantee the smoothness of the trend functions $\mu_j(u)$ when no break occurs.

ASSUMPTION 3.1. *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 $c_f > 0$.*

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

ASSUMPTION 3.2. *The kernel $K(.) \ge 0$ is symmetric with support $[-1,1]$, $|K|_\infty < \infty$ and $\int_{-1}^1 K(x)\mathrm{d}x = 1$. Also assume $K(x)$ has first order derivative with $|K'|_\infty < \infty$. Let $b \to 0$ and $bn \to \infty$.*

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

ASSUMPTION 3.3. *(Lower bound for the long run variance) $\sigma_{j,j} \ge c_\sigma$, $1 \le j \le p$ for some finite constant $c_\sigma > 0$.*

ASSUMPTION 3.4. *(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 3.4 is a very general spatial and temporal dependence condition and embraces many interesting processes. We provide an example as follows.

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

$$\epsilon_t = F_t + Z_t, \ \textit{with}\ Z_t = \sum_{k \ge 0} \Lambda_k \eta_{t-k} \text{ and } F_t = \sum_{k \ge 0} v f_k^\top \eta_{t-k}', \tag{3.1}$$

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* $\sigma_{Z,j} = (\sum_{k\geq 0} \lambda_{k,j})^2$ *and* $\sigma_{F,j} = |\sum_{k\geq 0} f_k|_2^2 v_j^2$, *respectively. If for some constant* $c > 0$,

$$\sum_{k\geq i} |\lambda_{k,j}|/\sigma_{Z,j}^{1/2} \leq ci^{-\alpha} \quad \text{and} \quad \sum_{k\geq i} |f_k|_2 |v_j|/\sigma_{F,j}^{1/2} \leq ci^{-\alpha}, \tag{3.2}$$

then Assumption 3.4 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*

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

ASSUMPTION 3.5. *(Finite moment) The innovations* $\eta_{i,j}$ *are i.i.d. with* $\mu_q = \|\eta_{1,1}\|_q < \infty$ *for some* $q \geq 4$.

ASSUMPTION 3.6. *(Sub-exponential) The innovations* $\eta_{i,j}$ *are i.i.d. with* $\mu_e = \mathbb{E}e^{a_0|\eta_{1,1}|} < \infty$, *for some* $a_0 > 0$.

Assumptions 3.5 and 3.6 put tail assumptions on the distribution of the noise sequences. Given the the above-mentioned conditions, we provide the main Gaussian approximation theorem, which is essential for the asymptotic distribution of our test statistics T_n.

THEOREM 1. *(Gaussian approximation for the test statistics) Under Assumptions 3.1-3.4 and* $b^5 n\log(np) = o(1)$.

(i) If Assumption 3.5 holds and

$$np(bn)^{-q/2}(\log(np))^{3q/2} = o(1), \tag{3.3}$$

(ii) If Assumption 3.6 holds and

$$(bn)^{-1}(\log(np))^{\max\{7,2(1+\beta)/\beta\}} = o(1), \tag{3.4}$$

then we have

$$\sup_{u\in\mathbb{R}} \left|\mathbb{P}(T_n \leq u) - \mathbb{P}(|\underline{d} + Z|_\infty \leq u)\right| \to 0. \tag{3.5}$$

REMARK 3. *(Allowed dimension) For Theorem 1 case (i), we allow* p *to be some polynomial order of* n *whose order depends on the value of* q. *Specifically, for some* $\nu_1 > 0$ *and* $0 < \nu_2 < 1/2$, *assume* $p \asymp n^{\nu_1}$ *and*

$b \asymp n^{-\nu_2}$. If $\nu_1 + \nu_2 < q/2 - 1$ and $\nu_2 > 1/5$, then conditions in case (i) hold. It is easy to see that the bigger the q is, the larger the allowance of the dimension p. The moment condition 3.5 depend 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, i.e. 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 conditions in case (ii) hold.

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

COROLLARY 1. *Under conditions in Theorem 1 (i) or (ii). The testing power is*

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

We can see that the test power is determined by the maximum size of abrupt changes in a localized window. In the following, we provide a few results on the property of the estimated breaks. To ensure a good recovery of the breaks, we need the following assumption on the minimum break size $\delta^\diamond$.

ASSUMPTION 3.7. *Let* $\delta^\diamond \gg \max\left\{\sqrt{\log(pn)/(bn)}, b\right\}$.

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 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 bound the errors of the estimated break locations and the break sizes.

THEOREM 2. *We assume conditions in Theorem 1 (i) or (ii) hold, and Assumption 3.7. If $\delta^\diamond/2 \geq \omega^\dagger \geq 2c'_w(bn)^{-1/2}\sqrt{\log(pn)}$, where c'_w is the constant defined as $(bn\sum_{i=0}^n w_i^2)^{1/2} \to c'_w$, then*

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

(i) indicates that the number of breaks is consistently estimated, (ii) suggests that the estimated break dates u_k can be consistently set 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 is depending on the bandwidth b, sample size n and the dimension of the time series p.

Given the consistency of the break points, we can obtain a distribution theory which facilitates us in making inference on the break sizes.

THEOREM 3. *(Break size inference) Recall that $\tilde{Z}$ is a centered Gaussian vector with covariance matrix (2.19). Assume conditions in Theorem 2 and $b^3 n \log(np) = o(1)$. We have*

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

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.

We further present a few results on the aggregated break point estimation. We recall the set of location of significant break as $\mathcal{S}_k$ defined in (2.21). For the aggregated jump estimation, we define alternatively 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 jump size of the time series. We shall put the same assumption on $\delta^\dagger$ as for $\delta^\diamond$.

ASSUMPTION 3.8. *Let $\delta^\dagger \gg \max\left\{\sqrt{\log(pn)/(bn)}, b\right\}$.*

In the following corollary, we show that we can consistently recovers 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 3.8. If $\delta^\dagger/2 \ge w^\dagger \gg (bn)^{-1/2}\log(np)^{1/2} + b$, then we have*

$$\mathbb{P}(\hat{\mathcal{S}}_k = \mathcal{S}_k,\ k = 1, 2, \ldots, \hat{K}_0) \to 1.$$

In addition, we provide a theorem that allows us making inference on the estimated break dates $\tilde{\tau}_k$ from the aforementioned procedure in Subsection 2.3.

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

$$c_1 \leq \lambda_{\max}(\Lambda^{-1}\Sigma\Lambda^{-1})/\lambda_{\min}(\Lambda^{-1}\Sigma\Lambda^{-1}) \leq c_2. \tag{3.6}$$

Recall definition of a_k and ς_k in (2.25). Then we have
(i) $|\tilde{\tau}_k - \tau_{k^*}| = O_{\mathbb{P}}(\varsigma_k^2/a_k^2)$.
(ii) In addition, if Assumption 3.4 holds with $\beta > 1$, then we have

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

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 $r \leq 0$, and $\mathbb{W}_1$, $\mathbb{W}_2$ are two independent Brownian motions.

REMARK 4. *Since the variance part of the limit distribution satisfies $\varsigma_k^2 \leq |\tilde{\Sigma}_k|_2 a_k$, and the drift part has $a_k \geq |\mathcal{S}_k|\delta^{\dagger 2}$, if $|\tilde{\Sigma}_k|_2/(|\mathcal{S}_k|\delta^{\dagger 2}) = o(1)$ then by Theorem 4 (i) we have $\tilde{\tau}_k \to \tau_k$ in probability. For example, if the noise sequence ϵ_t is spatially independent, then $\Sigma = I_p$ and thus $|\tilde{\Sigma}_k|_2 = 1$. In this case, $|\tilde{\tau}_k - \tau_k| = O_{\mathbb{P}}(|\mathcal{S}_k|^{-1}\delta^{\dagger -2})$, which becomes $o_{\mathbb{P}}(1)$ when $|\mathcal{S}_k|\delta^{\dagger 2} \to \infty$. If $\tilde{\Sigma}_k$ is a d-banded matrix, $|\tilde{\Sigma}_k|_2 \leq (|\tilde{\Sigma}_k|_1|\tilde{\Sigma}_k|_\infty)^{1/2} \leq d$. We can derive that $|\tilde{\tau}_k - \tau_k| = O_{\mathbb{P}}(d|\mathcal{S}_k|^{-1}\delta^{\dagger -2})$.*

In view of Theorem 4 (i), after the aggregating, we can obtain a finer estimation of the change point $\tilde{\tau}_k$. The asymptotic distribution result (ii) can be used for drawing inference for the estimated break points.

Finally, we derive a theorem on the precision of the estimated long run variance covariance estimation.

THEOREM 5. *(Long run variance precision) We assume Assumption 3.4 holds with $\beta \geq 1.5$, and let*

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

Then we have $\varsigma \log(np)^2 \to 0$ in probability under either one of the following two conditions:

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

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 (2.14) with Σ therein replaced by $\hat{\Sigma}$ in (2.31). 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.

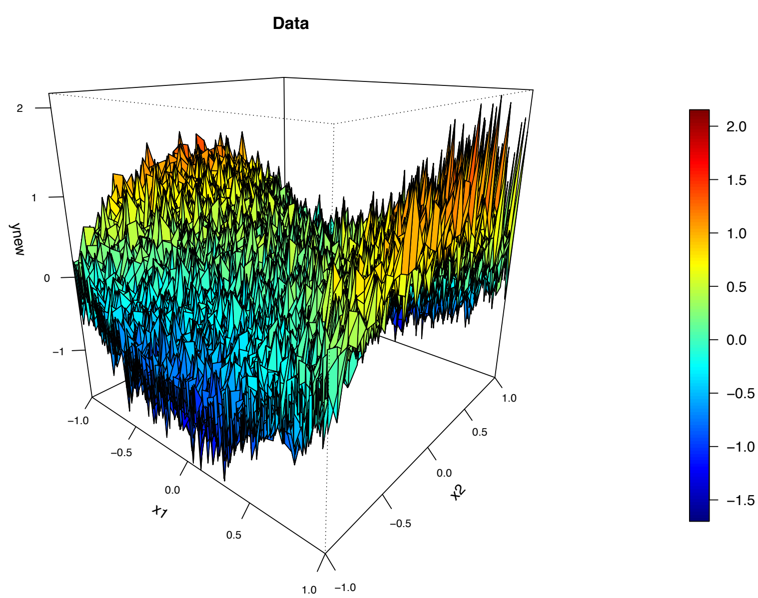
4. Simulation. In this section, we conduct a simulation study to evaluate the accuracy of our method. The discrete version of the model can be written as:

$$y_{it} = u_{it} + \sum_{j=1}^{K_0} \gamma_{jit} \mathbf{1}\{t \geq \tau_j\} + \epsilon_{it}, \tag{4.1}$$

$i = 1, \cdots, p$, $t = 1, \cdots, n$.

We consider different kind of data generating processes. We choose a) $f_i(u) = i^2/p^2 + u^2$, b) $f_i(u) = sin(2\pi u + i/p)$. Let $u_{it} = f_i(t/n)$, $n = 500, 1000$, $p = 20, 30$. ϵ_t is taken to follow either 1) a i.i.d. standard normal distribution 2) a VAR(1) model, with a randomly simulated coefficient matrix (maximum eigenvalue smaller than 1). The break locations are selected to be starting at time point 100 and are distanced by 100, and the break sizes are set to be either i) $\gamma_{jit} = 0.05$ for $i = 5, 10$, or ii) $\gamma_{jit} = (\sqrt{j * t}/(pn))$.

FIG 1. *Visualization of one sample of simulated data with jump in case a),2),ii).*



We use cross validation to select the bandwidth and the block parameter. The detailed testing procedure is summarized as follows in line with the descriptions in Section 2.

Step 1 (**Long run covairance estimation.**) We estimate of the long-run covariance matrix $\hat{\Sigma} = (\hat{\sigma}_{i,j})$ and its diagonal matrix $\hat{\Lambda}$. We first calculate $\hat{\sigma}_{i,j,k}$ in (2.27) and we let $\hat{\sigma}_{i,j}$ be the solution of $h_{i,j}(u) = 0$ as in (2.28).

Step 2 (**Q matrix relates to critical values.**) We construct the block matrix $\hat{Q} = (\hat{Q}_{i,j})$ where $\hat{Q}_{i,j}$ is $Q_{i,j}$ in (2.14) with Σ and Λ therein replaced by $\hat{\Sigma}$ and $\hat{\Lambda}$ respectively.

Step 3 (**Calculating critical values.**) We generate i.i.d. Gaussian vectors $\hat{Z}^{(i)}$, $i = 1, 2, \ldots N$, with the covariance matrix $\hat{Q}$ and we obtain $\hat{q}_\alpha$ which is the empirical $(1-\alpha)$ quantile of the $|\hat{Z}^{(i)}|_\infty$ over several samples and it can be viewed as an estimate of q_α in (2.16).

Step 4 (**Testing the existence of jump.**) We construct $\hat{T}_n$ as T_n in (2.6) with Λ replaced by $\hat{\Lambda}$. We reject the null hypothesis that there is no jump at level α if $\hat{T}_n$ is larger than $\hat{q}_\alpha$.

Step 5 (**Detecting significant of break-points.**) Supposing that in Step 4 we reject the null, we will continue with the following steps. To detect the significant jumps, we construct $|\hat{V}_t|_\infty$ for $t = bn+1, bn+2, \ldots, n-bn$, where $\hat{V}_t$ is same as V_t in (2.6) with Λ therein replaced by $\hat{\Lambda}$. Let $\mathcal{A}_1 = \{\tau \ : |\hat{V}_\tau|_\infty > w^\dagger\}$. $w^\dagger$ can be set as $\hat{q}_\alpha$ with α to be small, for example $w^\dagger = 0.0001$.

Step 6 (**Stamping multiple breaks**) In case of multiple significant breaks in Step 5, we sequentially locate the multiple change-points following steps in Section 2.2. To be more specific, for $k \geq 1$, we let $\hat{\tau}_k = \mathrm{argmax}_{\tau \in \mathcal{A}_k} |\hat{V}_\tau|_\infty$ and $\mathcal{A}_{k+1} = \mathcal{A}_k \setminus \{\tau : |\tau - \hat{\tau}_k| \leq 2bn\}$. Then the estimate of the number of breaks is $\hat{K}_0 = \max_{k\geq 1}\{k : \mathcal{A}_k \neq \emptyset\}$.

Step 7 (**Estimating the sizes of breaks**) We construct $\hat{\gamma}_k$ as in Step 3 in Subsection 2.2. We set the estimates of the sizes of the jumps as $\hat{\delta}_k = |\hat{\Lambda}^{-1}\hat{\gamma}_k|_\infty$ and their minimum as $\hat{\delta}^\diamond = \min_{1\leq k \leq \hat{K}_0} \hat{\delta}_k$.

Step 8 (**Constructing confidence interval for the sizes**) We construct $\tilde{q}_\alpha$ as in (2.20). Let $\hat{\theta} = (\hat{\sigma}_{1,1}^{1/2}, \hat{\sigma}_{2,2}^{1/2}, ..., \hat{\sigma}_{p,p}^{1/2})^\top$. Then the confidence interval for vector γ_{k^*} at level 2α is $(\hat{\gamma}_k - \tilde{q}_\alpha\hat{\theta}, \hat{\gamma}_k + \tilde{q}_\alpha\hat{\theta})$.

Step 9 (**Aggregated jump location estimation and confidence interval construction**) Construct aggregated jump location estimates $\tilde{\tau}_k$ as in (2.24). The confidence interval for τ_k is $(\tilde{\tau}_k - x, \tilde{\tau}_k + x)$, where x is the $1-\alpha/2$ quantile of the distribution $\mathrm{argmax}_r(-2^{-1}\hat{a}_k^2|r| + \hat{\varsigma}_k\mathbb{W}(r))$ and $\hat{a}_k$ (resp. $\hat{\varsigma}_k$) is a_k (resp. ς_k) with Λ, Σ and γ_k replaced by their estimations.

Figure 1 shows the simulated data with the model corresponding to the

TABLE 1
AD averaged over 1000 *samples in different simulation scenarios, and their standard deviations in bracket.*

		$p = 20, n = 500$		$p = 30, n = 1000$	
		1)	2)	1)	2)
a)	i)	0.033 (0.008)	0.047 (0.007)	0.027 (0.004)	0.036 (0.006)
	ii)	0.026 (0.012)	0.037 (0.010)	0.021 (0.005)	0.023 (0.006)
b)	i)	0.037 (0.011)	0.039 (0.012)	0.031 (0.003)	0.036 (0.004)
	ii)	0.022 (0.009)	0.031 (0.010	0.015 (0.003)	0.026 (0.002)

TABLE 2
AM/n averaged over 1000 *samples in different simulation scenarios, and their standard deviations in bracket.*

		$p = 20, n = 500$		$p = 30, n = 1000$	
		1)	2)	1)	2)
a)	i)	0.041 (0.011)	0.045 (0.021)	0.036 (0.010)	0.038 (0.008)
	ii)	0.029 (0.014)	0.054 (0.019)	0.013 (0.008)	0.020 (0.008)
b)	i)	0.025 (0.013)	0.047 (0.022)	0.018 (0.004)	0.034 (0.007)
	ii)	0.047 (0.023)	0.066 (0.029)	0.034 (0.007)	0.038 (0.008)

cases a),1),ii). We evaluate our simulation performance over 1000 samples. We report the averaged difference between the estimated number of breaks and the true break points (AD) ($|\hat{K}_0 - K_0|$) as in Table 1. The averaged distances between the breaks $\sum_{k=1}^{\hat{K}_0} |\tilde{\tau}_k - \tau_k^*|_\infty$ (AM) are shown in Table 2. And the averaged coverage probabilities of the confidence interval for the breaks (AC) are in Table 3 at the confidence level of 90%. As the sample sizes increase, the estimation precision is improved. We can see that our method is robust against different data simulation scenarios, and we can achieve good level of accuracy of our method. In particular the spatial and temporal dependency in the error term would not affect our estimation.

Figure 2 shows the plot of the estimated robust long run covariance matrix (right) against the true one (left). On an overall level, we see that the true correlation matrix has been recovered precisely, as the patterns of these two plots look the same. We also report the distance between our robustly estimated variance covariance matrix and the true one in Table 4. The estimation precision of the long run variance covariance matrix is maintained across different data generating processes.

5. Application. As an application, we analyze the monthly unemployment rate data in 20 states in the USA (namely, Alabama, Arizona, California, Colorado, Florida, Georgia, Illinois, Indiana, Kentucky, Michigan,

TABLE 3
AC in different simulation scenarios over all the estimated break points and samples.

		$p=20, n=500$		$p=30, n=1000$	
		1)	2)	1)	2)
a)	i)	0.653	0.660	0.789	0.753
	ii)	0.689	0.662	0.798	0.776
b)	i)	0.677	0.668	0.799	0.783
	ii)	0.654	0.608	0.776	0.765

TABLE 4
Averaged difference between the variance covariance and the true one. (L_1 norm divided by $p(p-1)/2$).

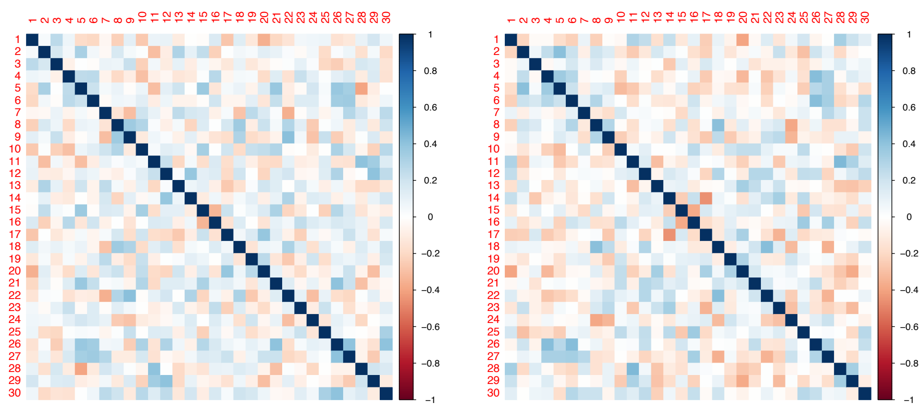
		$p=20, T=500$		$p=30, T=1000$	
		1)	2)	1)	2)
a)	i)	0.006	0.008	0.004	0.007
	ii)	0.008	0.009	0.005	0.007
b)	i)	0.009	0.009	0.003	0.004
	ii)	0.004	0.007	0.004	0.006

Mississippi, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Texas, Virginia, Washington and Wisconsin). The data time span is from Jan, 1976 to Sep 2018, and the data source is Burean of labor statistics from Department of labor in the United States (https://www.bls.gov/). Figure 3 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 think of detect changes induced by some well known exogenous shocks, such as the sub-prime crisis in 2008. It is understood that there will be likely a smooth cyclical trend associated with the unemployment time series, as they mostly raises during recession and falls during period of economics prosperity following the business cycle. To study whether the shock induced by recessions creates significantly structural change in the unemployment rate is of our interest.

Figure 4 shows the estimated robust long run correlation matrix using the method in Section 2.4. 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 5 plots the estimated break points and the confidence intervals around them. We see that the estimated breaks $\tilde{\tau}_k$ using the CUSUM statistics in Section 2.3 pick up the breaks earlier than the estimates obtained

Fig 2. *Visualization of the real (left) and the estimated correlation matrix (using robust estimation method).*



from the non-aggregated method i.e. $\hat{\tau}_k$. We can see that our method can identify important dates such as the financial crisis period starting at Jan, 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 Jan, 1977, Oct, 1981, Jan, 1991 and Oct, 2001. There are a few documented NBER (national bureau of economics research) recession periods, namel Nov 1973- Mar 1975, July 1981-Nov 1982, July 1991- March 1991 and Mar 2001 - Nov 2001. All the breaks dates of 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 of unemployment rate across all the states.

FIG 3. *Plot of Unemployment rate of 20 states in USA*

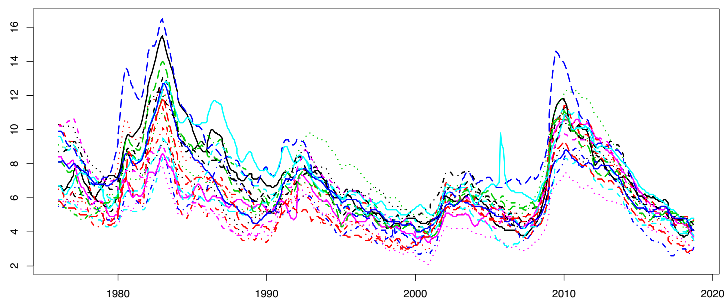


FIG 4. *Plot of estimation of the robust long run correlation matrix, $m = 10$.*

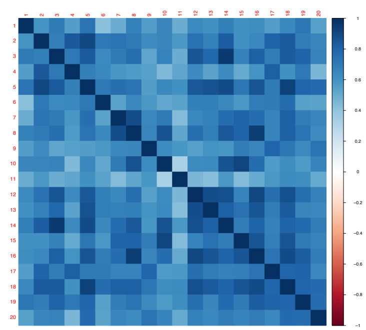
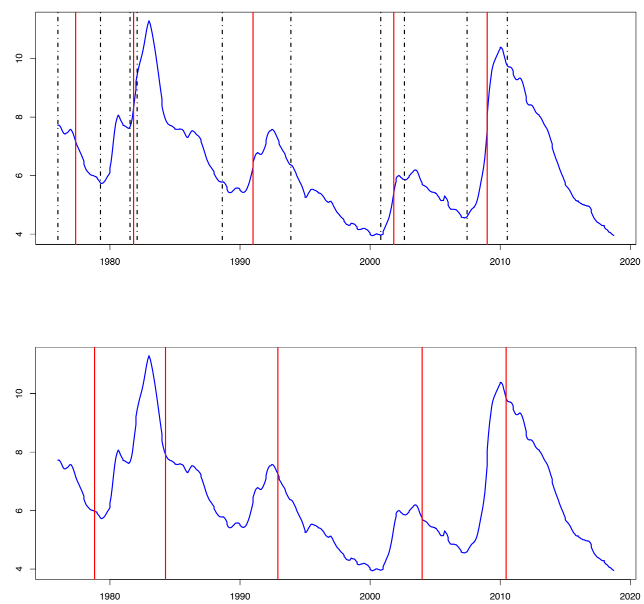


FIG 5. *Plot of estimated break points $\tilde{\tau}_k(\hat{\tau}_k)$ (red lines) and their confidence intervals (dotted black lines). $\tilde{\tau}_k$ (upper panel), $\hat{\tau}_k$ (lower panel). The blue time series line represents the averaged unemployment rate over states.*



References.

[1] Bai, J. (1994). Least squares estimation of a shift in linear processes. *J. Time Ser. Anal. 15*(5), 453–472.

[2] Bai, J. (1997). Estimation of a change point in multiple regression models. *The review of economics and statistics 79*(4), 551–563.

[3] Bai, J. (2010). Common breaks in means and variances for panel data. *J. Econometrics 157*(1), 78–92.

[4] Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

[5] Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics 18*(1), 1–22.

[6] Brodsky, B. E. and B. S. Darkhovsky (1993). *Nonparametric methods in change-point problems*, Volume 243 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht.

[7] Catoni, O. (2012). Challenging the empirical mean and empirical variance: a deviation study. *Ann. Inst. Henri Poincaré Probab. Stat. 48*(4), 1148–1185.

[8] Catoni, O. et al. (2012). Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, Volume 48, pp. 1148–1185. Institut Henri Poincaré.

[9] Chen, L., W. Wang, and W. B. Wu (2017). Dynamic semiparametric factor model with structural breaks. *Priprint*.

[10] Chen, L. and W. B. Wu (2017). Testing for trends in high-dimensional time series. *preprint*.

[11] Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist. 41*(6), 2786–2819.

[12] Chernozhukov, V., D. Chetverikov, and K. Kato (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab. 45*(4), 2309–2352.

[13] Csörgö, M. and L. Horváth (1997). *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. With a foreword by David Kendall.

[14] Enikeeva, F. and Z. Harchaoui (2013). High-dimensional change-point detection with sparse alternatives. *arXiv preprint arXiv:1312.1900*.

[15] Fan, J. and I. Gijbels (1996). *Local polynomial modelling and its applications*, Volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

[16] Fan, J., Q. Li, and Y. Wang (2017). Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *J. R. Stat. Soc. Ser. B. Stat. Methodol. 79*(1), 247–265.

[17] Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist. 42*(6), 2243–2281.

[18] Harlé, F., F. Chatelain, C. Gouy-Pailler, and S. Achard (2016). Bayesian model for multiple change-points detection in multivariate time series. *IEEE Trans. Signal Process. 64*(16), 4351–4362.

[19] Huang, T., B. Wu, P. Lizardi, and H. Zhao (2005). Detection of DNA copy number alterations using penalized least squares regression. *Bioinformatics 21*(20), 3811–3817.

[20] Hušková, M. and A. Slabỳ (2001). Permutation tests for multiple changes. *Kybernetika 37*(5), 605–622.

[21] Jackson, B., J. D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai (2005). An algorithm for optimal parti-

tioning of data on an interval. *IEEE Signal Processing Letters 12*(2), 105–108.
[22] Kim, D. (2011). Estimating a common deterministic time trend break in large panels with cross sectional dependence. *J. Econometrics 164*(2), 310–330.
[23] Lahiri, S. N. (2003). *Resampling methods for dependent data.* Springer Series in Statistics. Springer-Verlag, New York.
[24] Lévy-Leduc, C. and F. Roueff (2009). Detection and localization of change-points in high-dimensional network traffic data. *Ann. Appl. Stat. 3*(2), 637–662.
[25] Li, D., J. Qian, and L. Su (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association 111*(516), 1804–1819.
[26] Mandelbrot, B. (1963). The variation of certain speculative prices. *The journal of business 36*(4), 394–419.
[27] Niu, Y. S. and H. Zhang (2012). The screening and ranking algorithm to detect DNA copy number variations. *Ann. Appl. Stat. 6*(3), 1306–1326.
[28] Olshen, A. B., E. Venkatraman, R. Lucito, and M. Wigler (2004). Circular binary segmentation for the analysis of array-based dna copy number data. *Biostatistics 5*(4), 557–572.
[29] Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling.* Springer Series in Statistics. Springer-Verlag, New York.
[30] Rachev, S. and S. Mittnik (2000). Stable paretian models in finance.
[31] Stryhn, H. (1996). The location of the maximum of asymmetric two-sided brownian motion with triangular drift. *Statistics & Probability Letters 29*(3), 279 – 284.
[32] Tibshirani, R. and P. Wang (2007). Spatial smoothing and hot spot detection for cgh data using the fused lasso. *Biostatistics 9*(1), 18–29.
[33] Wang, T. and R. J. Samworth (2016). High-dimensional changepoint estimation via sparse projection. *arXiv preprint arXiv:1606.06246*.
[34] Wu, W. B. and Z. Zhao (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol. 69*(3), 391–410.
[35] Xie, Y. and D. Siegmund (2013). Sequential multi-sensor change-point detection. *Ann. Statist. 41*(2), 670–692.
[36] Zhang, D., W. B. Wu, et al. (2017). Gaussian approximation for high dimensional time series. *The Annals of Statistics 45*(5), 1895–1919.
[37] Zhang, N. R., D. O. Siegmund, H. Ji, and J. Z. Li (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika 97*(3), 631–645. With supplementary data available online.

Department of Mathematics and Statistics
1 Brookings Dr
St. Louis, MO 63105
E-mail: likai.chen@wustl.edu

Department of economics
City, University of London
Northampton Square
London EC1V 0HB
United Kingdom
E-mail: Weining.wang@city.ac.uk

Department of Statistics
George Herbert Jones Laboratory
5747 S. Ellis Avenue
Chicago, IL 60637
E-mail: wbwu@uchicago.edu

SUPPLEMENTARY MATERIAL TO "INFERENCE OF BREAK-POINTS IN HIGH-DIMENSIONAL TIME SERIES"

APPENDIX A: SOME USEFUL LEMMAS

LEMMA 1 (Basic properties of the weights). *We assume Assumption 3.2. We define $\kappa_i = \int_0^1 x^i K(x)\mathrm{d}x$ with finite κ_1, κ_2, $\kappa_1^2 \neq \kappa_2\kappa_0$. Then by [8], the weights of the local linear estimator take the following form*

$$w_i = \frac{\kappa_2 - \kappa_1 i/(bn)}{\kappa_2\kappa_0 - \kappa_1^2}\frac{K(i/(bn))}{bn} + O((bn)^{-2}).$$

We have the following results which holds uniformly over i. There exist strictly positive constants c_w, c'_w, c''_w only depending on kernel $K(.)$, such that

$$bn \max_{0\le i\le n}|w_i| \le c_w, \quad \max_{|i-j|\le m}|w_i - w_j| \le c_w\frac{m}{(bn)^2}, \quad (bn\sum_{i=0}^{bn} w_i^2)^{1/2} \to c'_w,$$

$$\text{and } \frac{bn}{k}\sum_{i=1}^{k} w_i \ge c''_w, k \le bn. \tag{A.1}$$

PROOF. We only show the last one, since the rest are similar and easier. Note

$$bn\sum_{i=1}^{k} w_i/k = F(k/(bn)) + O((bn)^{-1}), \quad \text{where} \quad F(t) = \frac{\kappa_2\int_0^t K(x)\mathrm{d}x - \kappa_1\int_0^t xK(x)\mathrm{d}x}{(\kappa_2\kappa_0 - \kappa_1^2)t}.$$

Define the numerator function as $g(t) = \kappa_2\int_0^t K(x)\mathrm{d}x - \kappa_1\int_0^t xK(x)\mathrm{d}x$. We can see that $g(0) = 0$, $g(1) > 0$, and the derivative function $g'(t) = (\kappa_2 - \kappa_1 t)K(t)$, which is strictly larger than 0 before κ_2/κ_1 and less than 0 afterwards. Therefore we have $F(x) > 0$ on $(0, 1]$. In addition, we note $F(0+) = \kappa_2 K(0)/(\kappa_2\kappa_0 - \kappa_1^2) > 0$ and $F(1) = 1$. Thus $\inf_{t\in(0,1]} F(t) > 0$ in view of $F(t)$ is a continuous function. □

LEMMA 2 (Burkholder [2], Rio [10]). *Let $q > 1$, $q' = \min\{q, 2\}$. Let $M_T = \sum_{t=1}^T \xi_t$, where $\xi_t \in \mathcal{L}^q$ are martingale differences. Then*

$$\|M_T\|_q^{q'} \le K_q^{q'}\sum_{t=1}^{T}\|\xi_t\|_q^{q'}, \text{ where } K_q = \max((q-1)^{-1}, \sqrt{q-1}).$$

APPENDIX B: ASYMPTOTIC RESULTS FOR GAUSSIAN VECTOR

LEMMA 3 (Comparison). *Let $X = (X_1, X_2, \dots, X_v)^\top$ and $Y = (Y_1, Y_2, \dots, Y_v)^\top$ be two centered Gaussian vectors in $\mathbb{R}^v$ and let $d = (d_1, d_2, \dots, d_v)^\top \in \mathbb{R}^v$. We denote $\Delta = \max_{1\le i,j\le v} |\sigma^X_{i,j} - \sigma^Y_{i,j}|$, where we define $\sigma^X_{i,j} = \mathbb{E}(X_iX_j)$ (resp. $\sigma^Y_{i,j} = \mathbb{E}(Y_iY_j)$). Assume that Y_is have the same variance $\sigma^2 > 0$. Then we have*

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}\big(|X + d|_\infty \le x\big) - \mathbb{P}\big(|Y + d|_\infty \le x\big)\Big| \lesssim \Delta^{1/3}\log(v)^{2/3}, \tag{B.1}$$

where the constant involved in $\lesssim$ only depends on σ.

PROOF. It suffices to show for any $d \in \mathbb{R}^v$,

$$\sup_x \Big|\mathbb{P}(\max_{1\le i\le v}(X_i + d_i) \le x) - \mathbb{P}(\max_{1\le i\le v}(Y_i + d_i) \le x)\Big| \lesssim \Delta^{1/3}\log(v)^{2/3}.$$

To this end, we define

$$F^*_\beta(z) = \beta^{-1}\log\Big(\sum_{j=1}^{v} \exp(\beta(z_j + d_j))\Big).$$

Replace the $F_\beta(\cdot)$ in the proof of Theorem 2 in [5] by $F^*_\beta(z)$. Then by the argument in equation (10) in [5], we have

$$\mathbb{P}\Big(\max_{1\le i\le v}(X_i + d_i) \le x\Big) \le \mathbb{P}\Big(\max_{1\le i\le v}(Y_i + d_i) \le x + \delta + \beta^{-1}\log(v)\Big) + c(\delta^{-2} + \beta\delta^{-1})\Delta,$$

where c is some absolute constant. Then by Lemma 4, we have

$$\begin{aligned}&\mathbb{P}\Big(\max_{1\le i\le v}(X_i + d_i) \le x\Big) - \mathbb{P}\Big(\max_{1\le i\le v}(Y_i + d_i) \le x\Big)\\ \lesssim&(\delta + \beta^{-1}\log(v))\sqrt{\log(v)} + (\delta^{-2} + \beta\delta^{-1})\Delta,\end{aligned}$$

where the constant in $\lesssim$ only depending on σ. Take $\beta = \delta^{-1}\log(v)$ and $\delta = \log(v)^{1/6}\Delta^{1/3}$. Same argument can be applied in the other direction, and the desired result follows. □

LEMMA 4 ([9]). *Let $X = (X_1, X_2, \dots, X_v)^\top$ be a centered Gaussian vector in $\mathbb{R}^v$. Assume $\mathbb{E}(X_i^2) \ge b$ for some $b > 0$ and all $1 \le i \le v$. Then for any $e > 0$ and $d \in \mathbb{R}^v$,*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\Big(\big||X + d|_\infty - x\big| \le e\Big) \le ce\sqrt{\log(v)}, \tag{B.2}$$

where c is some constant depending only on b.

APPENDIX C: PROOF OF GAUSSIAN APPROXIMATION FOR THE TEST STATISTICS

Proof of Theorem 1. The proof of Theorem 1 is quite involved. We shall first provide some intuitive ideas of the proof strategy. We define

$$I_\epsilon := \max_{bn+1\le i\le n-bn} \Big| \sum_{t=i-bn}^{i-1} w_{i-t}\Lambda^{-1}\epsilon_t - \sum_{t=i+1}^{i+bn} w_{t-i}\Lambda^{-1}\epsilon_t + d_i\Big|_\infty. \tag{C.1}$$

By (2.10) and (2.11) we have

$$|T_n - I_\epsilon| \le \max_{bn+1\le i\le n-bn} |\mathbb{E}V_i - d_i|_\infty = O(b^2). \tag{C.2}$$

Thus we only need to work on I_ϵ. For some $m>0$, let a truncated version of the error term be defined as

$$\epsilon_{t,m} = \sum_{k=0}^{m-1} A_k \eta_{t-k}.$$

Consider the m-dependent approximation $I_{\epsilon,m}$ of I_ϵ, where $I_{\epsilon,m}$ is I_ϵ with ϵ_t replaced by $\epsilon_{t,m}$. Then we have $I_\epsilon \approx I_{\epsilon,m}$ for large m. Let $I_{z,m}$ be $I_{\epsilon,m}$ with η_t therein replaced by z_t, where $(z_t, t\in\mathbb{Z})$ are i.i.d. Gaussian vectors with zero mean and identity covariance matrix in $\mathbb{R}^{\tilde p}$. Since $I_{\epsilon,m}$ can be rewritten into the format of the maximum of summation of independent vectors, by the Gaussian approximation theorem in [6], the distributions of $I_{\epsilon,m}$ and $I_{z,m}$ are close. We complete the proof by showing that the distributions of $I_{z,m}$ and $|Z+\underline{d}|_\infty$ are close, and the continuity of the maximum of a non-centered Gaussian distribution.

We now proceed with the formal argument. We shall first focus on case (i). Let $m=(bn)^{1/(\beta+1)}$, for any $\alpha>0$,

$$\mathbb{P}\big((bn)^{1/2}T_n \le u\big) \le \mathbb{P}\big((bn)^{1/2}|T_n - I_{\epsilon,m}| \ge \alpha\big) + \mathbb{P}\big((bn)^{1/2} I_{\epsilon,m} \le u+\alpha\big)$$

and

$$\mathbb{P}\big((bn)^{1/2}|Z+\underline{d}|_\infty \le u\big) = \mathbb{P}\big((bn)^{1/2}|Z+\underline{d}|_\infty \le u+\alpha\big) - \mathbb{P}\big(u < (bn)^{1/2}|Z+\underline{d}|_\infty \le u+\alpha\big).$$

Hence

$$\begin{aligned}
&\sup_{u\in\mathbb{R}}\Big[\mathbb{P}\big((bn)^{1/2}T_n \le u\big) - \mathbb{P}\big((bn)^{1/2}|Z+\underline{d}|_\infty \le u\big)\Big]\\
\le&\mathbb{P}\Big((bn)^{1/2}|T_n - I_{\epsilon,m}| \ge \alpha\Big) + \sup_{u\in\mathbb{R}}\Big|\mathbb{P}(I_{\epsilon,m}\le u) - \mathbb{P}(|Z+\underline{d}|_\infty \le u)\Big|\\
&+\sup_{u\in\mathbb{R}}\mathbb{P}\Big(\big|(bn)^{1/2}|Z+\underline{d}|_\infty - u\big| \le \alpha\Big) =: \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3.
\end{aligned}$$

For the I_1 part, $|T_n - I_{\epsilon,m}| \le |T_n - I_\epsilon| + |I_\epsilon - I_{\epsilon,m}|$. Recall (C.2), then $|T_n - I_\epsilon| \le c_0 b^2$ for some constant $c_0 > 0$. We define $\alpha' = 2c_1 \max\{(bn)^{-1/2}(np)^{1/q}, 1\} m^{-\beta+\beta/q}$, where the constant c_1 is the one to be defined in Lemma 5. Then by Lemma 5, we have

$$\mathbb{P}((bn)^{1/2}|I_\epsilon - I_{\epsilon,m}| \ge \alpha') = o(1).$$

Hence for $\alpha = \alpha' + c_0(bn)^{1/2}b^2$, $\mathrm{I}_1 = o(1)$.

For the I_2 part, we note that

$$\begin{aligned}\mathrm{I}_2 \quad &\le \quad \sup_{u\in\mathbb{R}} |\mathbb{P}(I_{\epsilon,m} \le u) - \mathbb{P}(I_{z,m} \le u)| + \sup_{u\in\mathbb{R}} |\mathbb{P}(I_{z,m} \le u) - \mathbb{P}(|Z + \underline{d}|_\infty \le u)| \\ &=: \quad \mathrm{I}_{21} + \mathrm{I}_{22}.\end{aligned}$$

By Lemma 7 (1), we have $\mathrm{I}_{21} = o(1)$. By Lemma 8, $\mathrm{I}_{22} = o(1)$. Hence $\mathrm{I}_2 = o(1)$.

For the I_3 part, the diagonal entities in bnQ take the same value i.e. $\sigma^{\diamond 2} = 2bn \sum_{i=1}^{bn} w_i^2$, which by (A.1), converges to $2c_w'^2 > 0$, where c_w' is a finite constant. By Lemma 4

$$\mathrm{I}_3 \lesssim \alpha \log(np)^{1/2} = o(1).$$

The desired result follows by combining the I_1-I_3 parts and a similar argument for the other side of the inequality.

For case (ii), we have the same decomposition I_1-I_3. For the I_1 part, we define $\alpha = c_1 \log(np)^{1/2} m^{-\beta} + c_0(bn)^{1/2}b^2$, for some constant $c_1 > 0$. Then by Lemma 6, $\mathrm{I}_1 = o(1)$. For I_2 part, by Lemma 7 (2) and Lemma 8, we have $\mathrm{I}_2 = o(1)$. For I_3, same argument can be applied. Combining the rates of I_1-I_3, we obtain the desired result. □

Lemma 5 and 6 give us concentration inequalities for the m-dependent approximation of I_ϵ.

LEMMA 5 (m-dependent approximation for polynomial case). *Assume conditions in Theorem 1 (i). For some $m > 0$ and $u > 0$, we have*

$$\mathbb{P}\big((bn)^{1/2}|I_\epsilon - I_{\epsilon,m}| \ge u + c_1(bn)^{-1/2}(np)^{1/q}m^{-\beta}\big) \le c_2(e^{-c_3 u^2 m^{2\beta}} + npm^{-q\beta}(bn)^{-q/2}u^{-q}),$$

where c_1, c_2, c_3 are some positive constants only depending on q, c_p, c_w, c_s, μ_q.

PROOF. We note that $I_\epsilon - I_{\epsilon,m}$ can be bounded by

$$\begin{aligned}&|I_\epsilon - I_{\epsilon,m}| \\ \le& \max_{bn+1\le i\le n-bn} \Big(\big| \sum_{t=i-bn}^{i-1} w_{i-t}\Lambda^{-1}(\epsilon_t - \epsilon_{t,m})\big|_\infty + \big| \sum_{t=i+1}^{i+bn} w_{t-i}\Lambda^{-1}(\epsilon_t - \epsilon_{t,m})\big|_\infty \Big) \\ =:& \mathrm{I}_1 + \mathrm{I}_2.\end{aligned}$$

We let $E_{i,l} = \sum_{t=(i-bn)\vee(l+m)}^{i-1} w_{i-t}\Lambda^{-1}A_{t-l}$, then I_1 can be rewritten into

$$\mathrm{I}_1 = \max_{\substack{bn+1\le i\le n-bn \\ 1\le j_1\le p}} \Big| \sum_{\substack{l\le i-m-1 \\ 1\le j_2\le \tilde{p}}} E_{i,l,j_1,j_2}\eta_{l,j_2}\Big|, \tag{C.3}$$

where E_{i,l,j_1,j_2} is the (j_1, j_2)th entity of matrix $E_{i,l}$ and η_{l,j_2} is the j_2th entity of η_l. Since η_{l,j_2}s are independent for different (l, j_2), by Lemma A.2 in [4], for $u > 0$,

$$\mathbb{P}(\sqrt{bn}\mathrm{I}_1 \ge 2\sqrt{bn}\mathbb{E}\mathrm{I}_1 + u) \le e^{-u^2/(3\sigma^2)} + K_q u^{-q} H_q, \tag{C.4}$$

where K_q is some constant only depending on q,

$$\sigma^2 = bn \max_{\substack{bn+1\le i\le n-bn \\ 1\le j_1\le p}} \sum_{\substack{l\le i-m-1 \\ 1\le j_2\le \tilde{p}}} \mathbb{E}(E_{i,l,j_1,j_2}\eta_{l,j_2})^2,$$

and

$$H_q = (bn)^{q/2} \sum_{\substack{l\le n-bn-m-1 \\ 1\le j_2\le \tilde{p}}} \mathbb{E}\left(\max_{\substack{(bn+1)\vee(l+m+1)\le i\le n-bn \\ 1\le j_1\le p}} |E_{i,l,j_1,j_2}\eta_{l,j_2}|^q \right).$$

Then we start to analyze the rates of the objects involved in (C.4). We define $E_{i,l,j_1,\cdot}$ to be the j_1th row of $E_{i,l}$. For the σ^2 part, by Assumption 3.4 and (A.1),

$$|E_{i,l,j_1,\cdot}|_2 \le \sum_{t\ge l+m} w_{i-t}\sigma_{j_1,j_1}^{-1/2}|A_{t-l,j_1,\cdot}|_2 \le c_w c_s m^{-\beta}/(bn), \tag{C.5}$$

and therefore

$$\sum_{l\le n-bn-m-1} |E_{i,l,j_1,\cdot}|_2 \le \sum_{t=i-bn}^{i-1} \sum_{l\le t-m} w_{i-t}\sigma_{j_1,j_1}^{-1/2}|A_{t-l,j_1,\cdot}|_2 \le c_w c_s m^{-\beta}. \tag{C.6}$$

Combining the above arguments and recall that $\mathbb{E}\eta_{i,j}^2 = 1$, we have

$$\sigma^2 \le bn \max_{\substack{bn+1\le i\le n-bn \\ 1\le j_1\le p}} \Big(\sum_{l\le i-m-1} |E_{i,l,j_1,\cdot}|_2 \max_{l\le n-bn-m-1} |E_{i,l,j_1,\cdot}|_2 \Big) \le (c_w c_s)^2 m^{-2\beta}. \tag{C.7}$$

For the H_q part, by Assumption 3.4 and (A.1), $\max_{i,j_1,j_2} |E_{i,l,j_1,j_2}| \leq c_w c_s((1-l) \vee m)^{-\beta}/(bn)$. Recall that $\tilde{p} \leq c_p p$. Then we have

$$\begin{aligned}
H_q &\leq (bn)^{q/2} \sum_{\substack{l \leq n-bn-m-1 \\ 1 \leq j_2 \leq \tilde{p}}} [c_w c_s((1-l) \vee m)^{-\beta}/(bn)]^q \mu_q^q \\
&\leq (c_w c_s)^q \mu_q^q (bn)^{-q/2} \tilde{p} \Big(\sum_{-m \leq l \leq n-bn-m} m^{-\beta q} + \sum_{l < -m} (1-l)^{-\beta q} \Big) \\
&\leq c_0 (bn)^{-q/2} npm^{-\beta q}, \qquad \text{(C.8)}
\end{aligned}$$

where $c_0 = 3c_p (c_w c_s)^q \mu_q^q$.

For $\mathbb{E}\mathrm{I}_1$ part, note that $\mathbb{E}\mathrm{I}_1 \leq \|\mathrm{I}_1\|_q$. By Lemma 2, we have

$$\mathbb{E}\mathrm{I}_1 \leq \Big(\sum_{i,j_1} \mathbb{E}(|\sum_{l,j_2} E_{i,l,j_1,j_2} \eta_{l,j_2}|^q) \Big)^{1/q} \leq \Big(\sum_{i,j_1} \big((q-1) \sum_{l} |E_{i,l,j_1,\cdot}|_2^2 \mu_q^2 \big)^{q/2} \Big)^{1/q}.$$

Thus by (C.5) and (C.6) we have

$$\mathbb{E}\mathrm{I}_1 \lesssim (bn)^{-1/2} m^{-\beta} (np)^{1/q}, \qquad \text{(C.9)}$$

where the constant in $\lesssim$ only depends on c_w, c_s, μ_q, q. Our conclusions follows by applying (C.7), (C.8) and (C.9) into (C.4) and a similar argument for I_2. $\square$

Lemma 6 (m-dependent approximation for exponential case). *We assume conditions in Theorem 1 (ii). We have*

$$\mathbb{P}\big((bn)^{1/2}|I_\epsilon - I_{\epsilon,m}| \geq u\big) \leq \begin{cases} 2npe^{-a_1 m^{2\beta} u^2}, & \text{if} \quad u < a_2 (bn)^{1/2} m^{-\beta}, \\ 2npe^{-a_3 m^{\beta} (bn)^{1/2} u}, & \text{if} \quad u \geq a_2 (bn)^{1/2} m^{-\beta}, \end{cases}$$

where a_1, a_2, a_3 are some positive constants only depending on a_0, c_w, c_s, μ_e.

Proof. Recall the definition of I_1 and I_2 in the proof of Lemma 5. Let $e^* = c_w c_s m^{-\beta}/(bn)$ and $c^* = a_0/e^*$. Then by (C.5), $\mathbb{E}e^{cE_{i,l,j_1,j_2}\eta_{l,j_2}} < \infty$, for any $0 < c \leq c^*$, and we have

$$\begin{aligned}
\mathbb{E}(e^{cI_1}) &\leq \sum_{\substack{bn+1 \leq i \leq n-bn \\ 1 \leq j_1 \leq p}} \mathbb{E} \left(\exp\Big\{ c \sum_{\substack{l \leq n-bn-m-1 \\ 1 \leq j_2 \leq \tilde{p}}} E_{i,l,j_1,j_2} \eta_{l,j_2} \Big\} + \exp\Big\{ -c \sum_{\substack{l \leq n-bn-m-1 \\ 1 \leq j_2 \leq \tilde{p}}} E_{i,l,j_1,j_2} \eta_{l,j_2} \Big\} \right) \\
&=: \mathrm{I}_{11} + \mathrm{I}_{12}.
\end{aligned}$$

Since $\mathbb{E}\eta_{i,j}=0$, for $E_{i,l,j_1,j_2}\neq 0$, we have

$$\begin{aligned}\mathbb{E}(e^{cE_{i,l,j_1,j_2}\eta_{l,j_2}}) &= 1+\frac{\mathbb{E}(e^{cE_{i,l,j_1,j_2}\eta_{l,j_2}}-1-cE_{i,l,j_1,j_2}\eta_{l,j_2})}{c^2E^2_{i,l,j_1,j_2}}c^2E^2_{i,l,j_1,j_2}\\ &\le 1+\frac{\mathbb{E}(e^{c^*e^*|\eta_{l,j_2}|}-1-c^*e^*|\eta_{l,j_2}|)}{(c^*e^*)^2}c^2E^2_{i,l,j_1,j_2}\\ &\le 1+\frac{\mu_e}{a_0^2}c^2E^2_{i,l,j_1,j_2},\end{aligned}$$

where the first inequality is because that for any $x>0$, the function $g_x(t)=(e^{tx}-1-tx)/t^2$ increases on $t\in(0,\infty)$, and $e^t-t\le e^{|t|}-|t|$. We define $c'=\mu_e/a_0^2$, and the rate of I_{11} is derived as follows,

$$\begin{aligned}\mathrm{I}_{11} &\le \sum_{i,j_1}\prod_{l,j_2}(1+c'c^2E^2_{i,l,j_1,j_2}) \le \sum_{\substack{bn+1\le i\le n-bn\\ 1\le j_1\le p}} \exp\Big\{c'c^2\sum_{\substack{l\le n-bn-m-1\\ 1\le j_2\le\tilde{p}}}E^2_{i,l,j_1,j_2}\Big\},\\ &\le np\ \exp\big\{c^2c_1m^{-2\beta}/(bn)\big\},\end{aligned}$$

where $c_1=c'c_w^2c_s^2$, the second inequality is due to $1+x\le e^x$ for any $x\ge 0$, and the last inequality is by (C.7). Same bound can be derived for I_{12}. We note that

$$\mathbb{P}(\mathrm{I}_1\ge u)\le e^{-cu}\mathbb{E}(e^{c\mathrm{I}_1})\le e^{-cu}(\mathrm{I}_{11}+\mathrm{I}_{12}).$$

We define $c^\diamond=bnm^{2\beta}u/(2c_1)$. Hence if $c^\diamond<c^*$, then $\mathbb{P}(\mathrm{I}_1\ge u)\le 2npe^{-u^2m^{2\beta}bn/(4c_1)}$; if $c^\diamond\ge c^*$, then $\mathbb{P}(\mathrm{I}_1\ge u)\le 2npe^{\mu_e bn-a_0/(c_wc_s)bnm^\beta u}$. The proof for I_2 is similar and therefore omitted. $\square$

LEMMA 7. *Let $m\to\infty$ and $m/(bn)\to 0$.*

(1) Assume conditions in Theorem 1 (i), we have

$$\begin{aligned}&\sup_{u\in\mathbb{R}}|\mathbb{P}(I_{\epsilon,m}\le u)-\mathbb{P}(I_{z,m}\le u)|\\ \lesssim&(bn)^{-1/6}\log^{7/6}(pn)+((np)^{2/q}/(bn))^{1/3}\log(pn),\end{aligned}$$

where the constant in $\lesssim$ only depends on c_w, c_w', c_s and μ_q.

(2) Assume conditions in Theorem 1 (ii), we have

$$\sup_{u\in\mathbb{R}}|\mathbb{P}(I_{\epsilon,m}\le u)-\mathbb{P}(I_{z,m}\le u)|\lesssim(bn)^{-1/6}\log(pn)^{7/6},$$

where the constant in $\lesssim$ only depends on c_w, c_w', c_s, a_0 and μ_e.

Proof. First we consider the case of (1). We denote

(C.10)
$$D_{i,l} = \sum_{t=(i-bn)\vee l}^{(i-1)\wedge(m+l-1)} w_{i-t}\Lambda^{-1}A_{t-l}, \quad \text{and} \quad D^*_{i,l} = \sum_{t=(i+1)\vee l}^{(i+bn)\wedge(m+l-1)} w_{t-i}\Lambda^{-1}A_{t-l}.$$

Then $I_{\epsilon,m}$ can be rewritten into

$$I_{\epsilon,m} = \max_{bn+1\le i\le n-bn} \Big| \sum_{i-m+1-bn\le l\le i-1} D_{i,l}\eta_l - \sum_{i-m+2\le l\le i+bn} D^*_{i,l}\eta_l + d_i \Big|_\infty.$$

Let $N_0 = (n-2bn)p$ and $N_1 = (n+m-1)\tilde{p}$. Let $G = (G_{i,l})_{i,l}$, $bn+1 \le i \le n-bn$, $2-m \le l \le n$, be a block matrix in $\mathbb{R}^{N_0\times N_1}$ with

$$\text{(C.11)}\quad G_{i,l} = \begin{cases} D_{i,l} & \text{if} \quad i-m+1-bn \le l \le i-m+1, \\ D_{i,l} - D^*_{i,l} & \text{if} \quad i-m+2 \le l \le i-1, \\ -D^*_{i,l} & \text{if} \quad i \le l \le i+bn, \end{cases}$$

and elsewhere zero. We define d_{i,j_1} to be the j_1th entity of d_i, $N_2 = bnN_1$ and G_{i,l,j_1,j_2} be the (j_1, j_2)th entity of $G_{i,l}$. Then

$$\text{(C.12)}\qquad N_2^{1/2} I_{\epsilon,m} = \max_{\substack{bn+1\le i\le n-bn\\ 1\le j_1\le p}} \Big| \sum_{\substack{2-m\le l\le n\\ 1\le j_2\le \tilde{p}}} g_{i,l,j_1,j_2} + N_2^{1/2} d_{i,j_1} \Big|,$$

$$\text{where } g_{i,l,j_1,j_2} = N_2^{1/2} G_{i,l,j_1,j_2}\eta_{l,j_2}.$$

For any $r \le q$, we denote

$$M_r := \max_{\substack{bn+1\le i\le n-bn\\ 1\le j_1\le p}} \theta_{ij_1,r}, \quad \text{where} \quad \theta^r_{ij_1,r} := \sum_{\substack{2-m\le l\le n\\ 1\le j_2\le \tilde{p}}} \mathbb{E}|g_{i,l,j_1,j_2}|^r/N_1 = \sum_{l=2-m}^{n} |G_{i,l,j_1,\cdot}|_r^r \mu_r^r N_2^{r/2}/N_1.$$

By Assumption 3.4 and (A.1), for any $r \ge 2$,

(C.13)
$$|D_{i,l,j_1,\cdot}|_r \le |D_{i,l,j_1,\cdot}|_2 \le c_w c_s/(bn), \quad \text{and similarly } |D^*_{i,l,j_1,\cdot}|_r \le c_w c_s/(bn).$$

Then by (C.11), $\max_{i,l,j_1} |G_{i,l,j_1,\cdot}|_r \le 2c_w c_s/(bn)$. Since $G_{i,l}$ is zero for $l < i-m+1-bn$ or $l > i+bn$,

$$\text{(C.14)}\qquad M_r \le (4c_w c_s \mu_r)(N_1/bn)^{1/2-1/r}.$$

Especially, for $r=2$, $M_r \le c_1$ where $c_1 = 4c_w c_s \mu_2$. By (C.10), for $i - bn \le l \le i - m$, and $S_m = \sum_{k=0}^{m-1} A_k$,

$$G_{i,l} = D_{i,l} = \sum_{t=l}^{m+l-1} w_{i-t}\Lambda^{-1}A_{t-l} = w_{i-l}\Lambda^{-1}S + w_{i-l}\Lambda^{-1}(S_m - S) + \sum_{t=l}^{l+m-1}(w_{i-t} - w_{i-l})\Lambda^{-1}A_{t-l}.$$

Therefore by Assumption 3.4 and (A.1), we can derive

$$\begin{aligned}\big||G_{i,l,j_1,\cdot}|_2 - w_{i-l}\big| &= \sigma_{j_1,j_1}^{-1/2} w_{i-l}\sum_{k\ge m}|A_{k,j_1,\cdot}|_2 + \sigma_{j_1,j_1}^{-1/2}\sum_{t=l}^{l+m-1}|w_{i-t} - w_{i-l}||A_{t-l,j_1,\cdot}|_2 \\ &\le c_w c_s m^{-\beta}/(bn) + c_w c_s m/(bn)^2.\end{aligned}$$

Note $\sum_{i=1}^m w_i^2 bn = O(m/(bn)) = o(1)$. Thus by (A.1),

(C.15)

$$\min_{\substack{bn\le i\le n-bn\\ 1\le j_1\le p}} \theta_{ij_1,2} \ge \min_{\substack{bn\le i\le n-bn\\ 1\le j_1\le p}} \Big(\sum_{l=i-bn}^{i-m} |G_{i,l,j_1,\cdot}|_2^2 \mu_2^2 N_2/N_1\Big)^{1/2} \ge c_w' \mu_2 - o(1) \ge c_1,$$

some constant $c_1 > 0$. Since $\max_{i,l,j_1,j_2}|G_{i,l,j_1,j_2}| \le 2c_w c_s/(bn)$,

$$\max_{l,j_2}\mathbb{E}(\max_{i,j_1}|g_{i,l,j_1,j_2}|^q) = \max_{i,l,j_1,j_2}|G_{i,l,j_1,j_2}N_2^{1/2}|^q \le (2c_w c_s)(N_1/(bn))^{q/2}.$$

Note

$$B_n := \max\Big\{M_3^3, M_4^2, \big(\max_{l,j_2}\mathbb{E}(\max_{i,j_1}|g_{i,l,j_1,j_2}|^q)\big)^{1/q}\Big\} \lesssim (N_1/(bn))^{1/2},$$

where the constant in $\lesssim$ only depends on c_w, c_s, μ_q. By Proposition 2.1 in [6] we have

$$\begin{aligned}&\sup_{u\in\mathbb{R}}|\mathbb{P}(N_2^{1/2}I_{\epsilon,m} \le u) - \mathbb{P}(N_2^{1/2}I_{z,m} \le u)| \\ \lesssim&\big(B_n^2\log^7(pn)/N_1\big)^{1/6} + \big(B_n^2\log^3(pn)/N_1^{1-2/q}\big)^{1/3} \\ \lesssim&(bn)^{-1/6}\log^{7/6}(pn) + ((np)^{2/q}/(bn))^{1/3}\log(pn).\end{aligned}$$

For part (2), let $M = \log_2(\mu_e) \vee 1$, and $B_n' = (2c_w c_s M/a_0)(N_1/(bn))^{1/2}$. Since $\max_{i,l,j_1,j_2}|G_{i,l,j_1,j_2}| \le 2c_w c_s/(bn)$, we have

$$\max_{i,l,j_1,j_2}\mathbb{E}(e^{g_{i,l,j_1,j_2}/B_n'}) \le 2.$$

Note

$$B_n := \max\{M_3^3, M_4^2, B_n'\} \lesssim (N_1/(bn))^{1/2}.$$

Apply the same argument as for part (1) with this new B_n, then Proposition 2.1 in [6] leads to

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}(N_2^{1/2}I_{\epsilon,m}\le u)-\mathbb{P}(N_2^{1/2}I_{z,m}\le u)\right|\lesssim\left(B_n^2\log^7(pn)/N_1\right)^{1/6}\lesssim(bn)^{-1/6}\log^{7/6}(pn).$$

□

LEMMA 8. *Assume conditions in Theorem 1 (i) or (ii), for $m\to\infty$, $m/(bn)\to 0$,*

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

where the constant in $\lesssim$ only depends on c_w, c_w' and c_s.

PROOF. We recall that $D_{i,l}$, $D^*_{i,l}$ in (C.10), $G=(G_{i,l})$ in (C.11) and $G^\diamond$ in (2.12). It is not hard to see that the covariance matrix for $I_{z,m}$ is $GG^\top$ and the covariance matrix for Z is $Q=G^\diamond G^{\diamond\top}$. We let

$$H^0=(G_{i,l})_{\substack{2-m\le l\le 0\\ bn+1\le i\le n-bn}},\quad\text{and}\quad H^1=(G_{i,l})_{\substack{1\le l\le n\\ bn+1\le i\le n-bn}}.$$

Then $G=(H^0,H^1)$ and

$$\begin{aligned}|GG^\top-G^\diamond G^{\diamond\top}|_{\max}&\le|H^0H^{0\top}|_{\max}+2|(H^1-G^\diamond)G^{\diamond\top}|_{\max}+|(H^1-G^\diamond)(H^1-G^\diamond)^\top|_{\max}\\&=:\mathrm{I}_1+\mathrm{I}_2+\mathrm{I}_3.\end{aligned}$$

By (C.13), $\max_{i,l,j}|G_{i,l,j,\cdot}|_2\le 2c_wc_s/(bn)$. Therefore

$$(bn)\mathrm{I}_1\le(bn)\max_{i_1,i_2,j_1,j_2}\sum_{l=2-m}^{0}|G_{i_1,l,j_1,\cdot}|_2|G_{i_2,l,j_2,\cdot}|_2\le(2c_wc_s)^2m/(bn).$$

Denote $\Delta_{i,l}=G_{i,l}-G^\diamond_{i,l}$. For $i-m+1-bn\le l<i-bn$, $\Delta_{i,l}=D_{i,l}$, and thus $|\Delta_{i,l,j,\cdot}|_2\le c_wc_s/(bn)$. For $i-bn\le l\le i-m+1$, we have

(C.16)
$$\Delta_{i,l}=D_{i,l}-w_{i-l}\Lambda^{-1}S=\sum_{t=l}^{m+l-1}(w_{i-t}-w_{i-l})\Lambda^{-1}A_{t-l}-w_{i-l}\Lambda^{-1}\sum_{t\ge m}A_t.$$

Hence $|\Delta_{i,l,j,\cdot}|_2\le c_wc_sm/(bn)^2+c_wc_sm^{-\beta}/(bn)$. For $i-m+1\le l\le i-1$, $\Delta_{i,l}=D_{i,l}-D^*_{i,l}-w_{i-l}\Lambda^{-1}S$. Then $|\Delta_{i,l,j,\cdot}|_2\le 3c_wc_s/(bn)$. Similarly we can bound $|\Delta_{i,l,j,\cdot}|_2$ for $i \le l \le i+bn$. For the rest l, $\Delta_{i,l} = 0$. We note that $|G^\diamond_{i,l,j,\cdot}| \le c_w c_s/(bn)$. Consequently,

$$(bn)\mathrm{I}_2 \le (bn) \max_{i_1,i_2,j_1,j_2} \sum_{l=1}^{n} |\Delta_{i_1,l,j_1,\cdot}|_2 |G^\diamond_{i_2,l,j_2,\cdot}|_2 \lesssim m/(bn) + m^{-\beta},$$

where the constant in $\lesssim$ only depends on c_w, c_s. Similarly we have $(bn)\mathrm{I}_3 \lesssim m/(bn) + m^{-\beta}$. Combining I_1-I_3,

$$(bn)|GG^\top - G^\diamond G^{\diamond\top}|_{\max} \lesssim m/(bn) + m^{-\beta}.$$

By (A.1), for any j we have $bnQ_{j,j} = 2bn\sum_{i=1}^n w_i^2 \to 2c_w'^2$. Then the desired result follows from Lemma 3. □

APPENDIX D: PROOFS OF CONSISTENCY FOR ESTIMATED BREAKS

Proof of Theorem 2 (i). Note $1-\Phi(x) \le (2\pi)^{-1/2}x^{-1}e^{-x^2/2}$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. Recall $G^\diamond_{i,l}$ in (2.12). Let $G^\diamond_{i,\cdot} = (G^\diamond_{i,1}, G^\diamond_{i,2}, \ldots, G^\diamond_{i,n})$ and $\underline{z}$ be a Gaussian vector in $\mathbb{R}^{np}$ with zero mean and identity covariance matrix. Let $G^\diamond_{i,\cdot,j,\cdot}$ be the jth row of $G^\diamond_{i,\cdot}$, then

(D.1)
$$\mathbb{P}(|G^\diamond \underline{z}|_\infty \ge u) \le \sum_{i=bn}^{n-bn}\sum_{j=1}^{p} \mathbb{P}(|G^{\diamond\top}_{i,\cdot,j,\cdot}\underline{z}| \ge u) \le np(2\pi)^{-1/2}(\sigma/u)e^{-u^2/(2\sigma^2)},$$

where $\sigma = |G^\diamond_{i,\cdot,j,\cdot}|_2 = (2\sum_{t=1}^{bn} w_t^2)^{1/2}$, which by (A.1) converges to $(bn/2)^{-1/2}c_w' > 0$. Thus

$$\mathbb{P}\big(|G^\diamond \underline{z}|_\infty \ge 2c_w' \log(np)^{1/2}(bn)^{-1/2}\big) \to 0. \tag{D.2}$$

Let $\mathcal{S} := \{1 \le i \le n : |i-\tau_k| > bn, \text{ for all } 1 \le k \le K_0\}$. For any $i \in \mathcal{S}$, $d_i = 0$. Hence by Theorem 1,

$$\sup_{u\in\mathbb{R}} \Big| \mathbb{P}\big(\max_{i\in\mathcal{S}} |V_i|_\infty \ge u\big) - \mathbb{P}\big(\max_{i\in\mathcal{S}} |G^\diamond_{i,\cdot}\underline{z}|_\infty \ge u\big)\Big| \to 0. \tag{D.3}$$

Since $\max_{i\in\mathcal{S}} |G^\diamond_{i,\cdot}\underline{z}|_\infty \le |G^\diamond \underline{z}|_\infty$, by (D.2) and (D.3) we have $\mathbb{P}(\max_{i\in\mathcal{S}}|V_i|_\infty \ge \omega^\dagger) \to 0$. Thus we obtain

$$\lim_{n\to\infty} \mathbb{P}\big(\forall t \in \mathcal{A}_1, \exists 1 \le k \le K_0, |t-\tau_k| \le bn\big) = 1. \tag{D.4}$$

Recall that $d_{\tau_k} = \Lambda^{-1}\gamma_k$. Since $|d_{\tau_k} + G^{\diamond}_{\tau_k,\cdot}\underline{z}|_\infty \geq |d_{\tau_k}|_\infty - |G^{\diamond}_{\tau_k,\cdot}\underline{z}|_\infty$, we have

$$\begin{aligned}\mathbb{P}\Big(\min_{1\leq k\leq K_0} |d_{\tau_k} + G^{\diamond}_{\tau_k,\cdot}\underline{z}|_\infty \leq \omega^\dagger\Big) &\leq \mathbb{P}\Big(\max_{1\leq k\leq K_0} |G^{\diamond}_{\tau_k,\cdot}\underline{z}|_\infty \geq \min_{1\leq k\leq K_0} |d_{\tau_k}|_\infty - \omega^\dagger\Big) \\ &\leq \mathbb{P}\big(|G^{\diamond}\underline{z}|_\infty \geq \delta^\diamond - \omega^\dagger\big).\end{aligned}$$

Since $\delta^\diamond \geq 2\omega^\dagger$, $\mathbb{P}(\min_{1\leq k\leq K_0} |d_{\tau_k} + G^{\diamond}_{\tau_k,\cdot}\underline{z}|_\infty \leq \omega^\dagger) \to 0$. Subsequently the break statistics will be bigger than the threshold at the points of break with probability approach 1, $\mathbb{P}(\min_{1\leq k\leq K_0} |V_{\tau_k}|_\infty \leq \omega^\dagger) \to 0$ in view of

$$\sup_{u\in\mathbb{R}} \Big|\mathbb{P}(|V_{\tau_k}|_\infty \leq u) - \mathbb{P}(|d_{\tau_k} + G^{\diamond}_{\tau_k,\cdot}\underline{z}|_\infty \leq u)\Big| \to 0.$$

Therefore we have

$$\mathbb{P}(\tau_k \in \mathcal{A}_1, 1 \leq k \leq K_0) \to 1. \tag{D.5}$$

Let $\mathcal{B}(\tau, r) = \{t : |t-\tau| \leq r\}$. By (D.4) and (D.5), we have

$$\lim_n \mathbb{P}\Big(\{\tau_1, \tau_2, \ldots, \tau_{K_0}\} \subseteq \mathcal{A}_1 \subseteq \cup_{1\leq k\leq K_0} \mathcal{B}(\tau_k, bn)\Big) = 1.$$

Since for $k_1 \neq k_2$, $|\tau_{k_1} - \tau_{k_2}| \gg bn$, for any $k_1 \neq k_2$ and $t \in \mathcal{B}(\tau_{k_1}, bn)$, for all large n, $\mathcal{B}(t, 2bn) \cap \mathcal{B}(\tau_{k_2}, 2bn) = \emptyset$. Thus we complete the proof. $\square$

Proof of (ii). Let $\mu_i^{(l)}$ (resp. $U_i^{(l)}$) be $\hat{\mu}_i^{(l)}$ with Y_i therein replaced by $\mu(i/n)$ (resp. ϵ_i). Similarly we can define $\mu_i^{(r)}$ and $U_i^{(r)}$. Let $\Delta\mu_i = \mu_i^{(l)} - \mu_i^{(r)}$ and $\Delta U_i = U_i^{(l)} - U_i^{(r)}$. Let Δf_i be $\Delta\mu_i$ with μ replaced by f.

Let $\tau \in \{\tau_1, ..., \tau_{K_0}\}$ be some break point associated with jump γ. For any t such that $|t-\tau| \leq bn$, we have $\Delta\mu_t = (1 - \sum_{i=1}^{|t-\tau|} w_i)\gamma + \Delta f_t$. Hence

$$\begin{aligned} V_t &= \Lambda^{-1}\Delta\mu_t + \Lambda^{-1}\Delta U_t \\ &= (1 - \sum_{i=1}^{|t-\tau|} w_i)\Lambda^{-1}\gamma + \Lambda^{-1}\Delta f_t + \Lambda^{-1}\Delta U_\tau + (\Lambda^{-1}\Delta U_t - \Lambda^{-1}\Delta U_\tau). \end{aligned} \tag{D.6}$$

Let $\hat{\tau} = \mathrm{argmax}_{\{t:|t-\tau|\leq bn\}} |V_t|_\infty$. The proceeding proof contains three steps. Step 1. Let $j_\tau = \mathrm{argmax}_j |V_{\hat{\tau},j}|$, where $V_{\hat{\tau},j}$ is the jth entity of $V_{\hat{\tau}}$. This step shows

$$\liminf_n |(\Lambda^{-1}\gamma)_{j_\tau}|/\delta^\diamond \geq 1.$$

We shall show by contradiction. By (C.2), $|\Delta f_t|_\infty = O(b^2)$. If $|(\Lambda^{-1}\gamma)_{j_\tau}| \leq c\delta^\diamond$, for some $c < 1$, then by (D.6), $|V_{\hat{\tau}}|_\infty \leq c\delta^\diamond + O(b^2) + |\Lambda^{-1}\Delta U_{\hat{\tau}}|_\infty$.

Let $\tilde{U}_t$ be U_t with $\eta_{i,j}$ replaced by $z_{i,j}$ where $z_{i,j}$ are i.i.d standard normal random variables. Then $\max_t |\Lambda^{-1}\Delta\tilde{U}_t|_\infty = O_{\mathbb{P}}((bn)^{-1/2}\log(np)^{1/2})$. Then by Gaussian approximation Theorem 1,

$$\max_t |\Lambda^{-1}\Delta U_t|_\infty = O_{\mathbb{P}}((bn)^{-1/2}\log(np)^{1/2}).$$

Since $\delta^\diamond \gg (bn)^{-1/2}\log(np)^{1/2}$, we have $|V_{\hat{\tau}}|_\infty \le c\delta^\diamond(1+o_{\mathbb{P}}(1))$. On the other hand, by (D.6), $|V_\tau|_\infty \ge \delta^\diamond + O(b^2) - |\Lambda^{-1}\Delta U_\tau|_\infty = \delta^\diamond(1+o_{\mathbb{P}}(1))$. These imply $\mathbb{P}(V_{\hat{\tau}} < V_\tau) \to 1$, which is a contradiction.
Step 2. This step shows

$$\text{(D.7)} \qquad \max_t |\Lambda^{-1}\Delta U_\tau - \Lambda^{-1}\Delta U_t|_\infty / |t-\tau|^{1/2} = O_{\mathbb{P}}\big\{\log(np)^{1/2}/(bn)\big\}.$$

Let $t \le \tau$, the other direction can be similarly dealt with. Let $(z_{i,j})$ be i.i.d. standard Gaussian random variables. Define $\Delta\tilde{U}_t$ (resp. $\tilde{\epsilon}_t$) to be ΔU_t (resp. ϵ_t) with $\eta_{i,j}$ therein replaced by $z_{i,j}$. Then by Gaussian approximation Theorem 1, it suffices to show (D.7) with U_t replaced by $\tilde{U}_t$. We note that

$$\Delta\tilde{U}_t - \Delta\tilde{U}_\tau = \sum_{i=t-bn}^{\tau-bn-1} w_{t-i}\tilde{\epsilon}_i + \sum_{i=\tau-bn}^{t-1}(w_{t-i} - w_{\tau-i})\tilde{\epsilon}_i - \sum_{i=t+1}^{\tau-1}(w_{i-t} + w_{\tau-i})\tilde{\epsilon}_i$$
$$- \sum_{i=\tau+1}^{t+bn}(w_{i-t} - w_{i-\tau})\tilde{\epsilon}_i - \sum_{i=t+bn+1}^{\tau+bn} w_{i-\tau}\tilde{\epsilon}_i + w_{\tau-t}\tilde{\epsilon}_t - w_{\tau-t}\tilde{\epsilon}_\tau =: \sum_{k=1}^{7} r_k.$$

For r_1, we have

$$|\Lambda^{-1}r_1|_\infty = \max_{1\le j_1\le p} \Big|\sum_{l\le\tau-bn-1, 1\le j_2\le p} E_{l,j_1,j_2} z_{l,j_2}\Big|, \quad \text{where } E_l = \sum_{i=(t-bn)\vee l}^{\tau-bn-1} w_{t-i}\Lambda^{-1}A_{i-l},$$

with E_{l,j_1,j_2} as the (j_1, j_2)th entity of matrix E_l. Then

$$\max_t \big(|\Lambda^{-1}r_1|_\infty/|t-\tau|^{1/2}\big) = O_{\mathbb{P}}\big\{\log(np)^{1/2}/(bn)\big\}.$$

A similar argument leads to the same bound for r_3 and r_5. For r_2, we can rewrite

$$|\Lambda^{-1}r_2|_\infty = \max_{1\le j_2\le p} \Big|\sum_{l\le t-1} E'_{l,j_1,j_2} z_{l,j_2}\Big|, \quad \text{where} \quad E'_l = \sum_{i=(\tau-bn)\vee l}^{t-1}(w_{t-i} - w_{\tau-i})\Lambda^{-1}A_{i-l}.$$

Then similarly we have $\max_t |\Lambda^{-1}r_2|_\infty/|\tau-t|^{3/2} = O_{\mathbb{P}}\big\{\log(np)^{1/2}/(bn)^2\big\}$. The same bound can be derived for r_4 as well. We obtain (D.7) by summing the above bounds up.

Step 3. Without loss of generality, assume $\gamma_{j_\tau} > 0$. Then by the argument in step 1, with probability tending to 1, $V_{\hat{\tau},j_\tau} > 0$. By Assumption 3.1, we have $|\Delta f_t - \Delta f_\tau|_\infty = O(|t-\tau|/n)$. With probability tending to 1, by (D.6),

$$|V_\tau|_\infty - |V_{\hat{\tau}}|_\infty \geq V_{\tau,j_\tau} - V_{\hat{\tau},j_\tau} \geq \sum_{i=1}^{|\hat{\tau}-\tau|} w_i(\Lambda^{-1}\gamma)_{j_\tau} - O(|\hat{\tau}-\tau|/n) - |\Lambda^{-1}\Delta U_\tau - \Lambda^{-1}\Delta U_{\hat{\tau}}|_\infty.$$

By (A.1), we have $\sum_{i=1}^{|t-\tau|} w_i \geq c''_w|t-\tau|/(bn)$. Hence by Step 1 and Step 2 we further derive

$$|V_\tau|_\infty - |V_{\hat{\tau}}|_\infty \geq c''_w|\hat{\tau}-\tau|\delta^\diamond/(bn) - O(|\hat{\tau}-\tau|/n) - O_{\mathbb{P}}\big(|\tau-\hat{\tau}|^{1/2}\log(np)^{1/2}/(bn)\big).$$

Since $|V_\tau|_\infty < |V_{\hat{\tau}}|_\infty$, we have

$$|\tau - \hat{\tau}| = O_{\mathbb{P}}\{\log(np)/\delta^{\diamond 2}\}.$$

$\square$

Proof of (iii).. Recall the definition of $\mu_t^{(l)}$, $\mu_t^{(r)}$, $U_t^{(l)}$ and $U_t^{(r)}$ in the proof of (ii) and $M = bn$. Since $M \gg \log(np)/\delta^{\diamond 2}$,

$$|\mu^{(l)}_{\hat{\tau}_k - M} - \mu((\hat{\tau}_k - M)/n)|_\infty = |f^{(l)}_{\hat{\tau}_k-M} - f((\hat{\tau}_k - M)/n)|_\infty = O(b).$$

Similarly $|\mu^{(r)}_{\hat{\tau}_k+M} - \mu((\hat{\tau}_k+M)/n)|_\infty = O(b)$. Since $\max_{1\leq j\leq p}|f'_j|$ is bounded,

$$|\mu((\hat{\tau}_k+M)/n) - \mu((\hat{\tau}_k-M)/n) - \gamma_{k^*}|_\infty = |f((\hat{\tau}_k+M)/n) - f((\hat{\tau}_k-M)/n)|_\infty = O(M/n).$$

Hence

$$\begin{aligned} |\Lambda^{-1}(\hat{\gamma}_k - \gamma_{k^*})|_\infty &= \big|\Lambda^{-1}(\mu^{(r)}_{\hat{\tau}_k+M} - \mu^{(l)}_{\hat{\tau}_k-M} - \gamma_{k^*}) + \Lambda^{-1}U^{(r)}_{\hat{\tau}_k+M} - \Lambda^{-1}U^{(l)}_{\hat{\tau}_k-M}\big|_\infty \\ &\leq O(b + M/n) + |\Lambda^{-1}U^{(l)}_{\hat{\tau}_k-M} - \Lambda^{-1}U^{(r)}_{\hat{\tau}_k+M}|_\infty. \qquad \text{(D.8)} \end{aligned}$$

By Gaussian approximation and (D.2) we have

$$\mathbb{P}\big(|\Lambda^{-1}U^{(l)}_{\hat{\tau}_k-M} - \Lambda^{-1}U^{(r)}_{\hat{\tau}_k+M}|_\infty \geq 2c'_w\log(np)^{1/2}/(bn)^{1/2}\big) \to 0.$$

Inserting the above equation into (D.8) and we obtain the desired result. $\square$

APPENDIX E: PROOFS OF LIMIT DISTRIBUTIONS OF BREAK SIZES

Proof of Theorem 3. We recall $M = bn$. Similar to (D.8), we have $|\Lambda^{-1}(\hat{\gamma}_k - \gamma_{k^*}) - \Lambda^{-1}(U^{(r)}_{\hat{\tau}_k+M} - U^{(l)}_{\hat{\tau}_k-M})|_\infty \leq cM/n$. Therefore

$$\begin{aligned} &\sup_{u\in\mathbb{R}}\Big|\mathbb{P}\big((bn)^{1/2}|\Lambda^{-1}(\hat{\gamma}_k - \gamma_{k^*})|_\infty \leq u\big) - \mathbb{P}\big((bn)^{1/2}|\tilde{Z}|_\infty \leq u\big)\Big| \\ \leq &\sup_{u\in\mathbb{R}}\mathbb{P}\big(|(bn)^{1/2}|\tilde{Z}|_\infty - u| \leq c(bn)^{1/2}M/n\big) \\ &+ \sup_{u\in\mathbb{R}}\big|\mathbb{P}(|\Lambda^{-1}(U^{(l)}_{\hat{\tau}_k-M} - U^{(r)}_{\hat{\tau}_k+M})|_\infty \leq u) - \mathbb{P}(|\tilde{Z}|_\infty \leq u)\big| = \mathrm{I}_1 + \mathrm{I}_2. \end{aligned}$$

We note that $(bn)^{1/2}\tilde{Z}_j$ are i.i.d with variance $2(bn)\sum_{t=1}^{bn} w_t^2$, which by (A.1) converges to $2c_w'^2 > 0$. Therefore by Lemma 4,

$$\mathrm{I}_1 = O\{(bn)^{1/2}(M/n)\log(np)^{1/2}\} = o(1).$$

Let $\tilde{G} = (\tilde{G}_1, \tilde{G}_2, \ldots, \tilde{G}_n)$, where $\tilde{G}_l = w_{\hat{\tau}_k - M - l}\Lambda^{-1}S$, if $\hat{\tau}_k - M - bn \leq l \leq \hat{\tau}_k - M - 1$, and $\tilde{G}_l = w_{l-(\hat{\tau}_k+M)}\Lambda^{-1}S$, if $\hat{\tau}_k + M + 1 \leq l \leq \hat{\tau}_k + M + bn$ and elsewhere zero. Let $\underline{z}$ be Gaussian vector in $\mathbb{R}^{n\tilde{p}}$ with zero mean and identity covariance matrix. Then $\tilde{G}\underline{z} \overset{d}{=} \tilde{Z}$. By the same argument as in Theorem 1 with $d_i = 0$ we have

$$\sup_{u\in\mathbb{R}} |\mathbb{P}(|\Lambda^{-1}(U^{(l)}_{\hat{\tau}_k - M} - U^{(r)}_{\hat{\tau}_k + M})|_\infty \leq u) - \mathbb{P}(|\tilde{G}\underline{z}|_\infty \leq u)| = o(1).$$

Thus $\mathrm{I}_2 = o(1)$ and we complete the proof. $\square$

APPENDIX F: PROOF OF AGGREGATED BREAKS ESTIMATION PROPERTIES

Proof of Proof of Theorem 4 (i). We shall condition on the event where $\hat{\mathcal{S}}_k = \mathcal{S}_k$ and $|\hat{\tau}_k - \tau_k| \ll bn$. By Theorem 2 and Corollary 2, the event would take place with probability tending to 1. Denote
$\varepsilon_t = \sum_{j\in\mathcal{S}_k}(\Lambda^{-1}\hat{\gamma}_k)_j(\Lambda^{-1}\epsilon_t)_j$, and $\hat{a}_k = \sum_{j\in\mathcal{S}_k}(\Lambda^{-1}\gamma_k)_j(\Lambda^{-1}\hat{\gamma}_k)_j$. Then we have

$$X_t = \hat{a}_k \mathbf{1}_{t\geq\tau_k} + \sum_{j\in\mathcal{S}_k} f_j(t/n)(\Lambda^{-1}\hat{\gamma}_k)_j + \varepsilon_t.$$

Let

$$D_t = \sum_{s=\hat{\tau}_k - bn}^{\hat{\tau}_k + bn} X_s \frac{t - \hat{\tau}_k + bn}{2bn+1} - \sum_{s=\hat{\tau}_k - bn}^{t-1} X_s.$$

Denote $r = \tilde{\tau}_k - \tau_k$. Assume $r > 0$. By the continuity of f_j, we have

$$\begin{aligned}
&\max_j \Big| \sum_{s=\hat{\tau}_k - bn}^{\hat{\tau}_k + bn} f_j(s/n)\frac{r}{2bn+1} - \sum_{s=\tau_k}^{\tau_k + r - 1} f_j(s/n)\Big| \\
&= \max_j \Big| \sum_{s=\hat{\tau}_k - bn}^{\hat{\tau}_k + bn} \big(f_j(s/n) - f_j(\tau_k)\big)\frac{r}{2bn+1} - \sum_{s=\tau_k}^{\tau_k + r - 1} \big(f_j(s/n) - f_j(\tau_k)\big)\Big| = O(br).
\end{aligned}$$

By above and Theorem 2 (iii), we have

$$D_{\tau_k+r} - D_{\tau_k} = \sum_{s=\hat{\tau}_k-bn}^{\hat{\tau}_k+bn} X_s \frac{r}{2bn+1} - \sum_{s=\tau_k}^{\tau_k+r-1} X_s$$

$$= -\hat{a}_k \frac{bn+\tau_k-\hat{\tau}_k}{2bn+1} r + \sum_{s=\hat{\tau}_k-bn}^{\hat{\tau}_k+bn} \varepsilon_s \frac{r}{2bn+1} - \sum_{s=\tau_k}^{\tau_k+r-1} \varepsilon_s + O_{\mathbb{P}}(|\tilde{\gamma}_k|_1 br). \tag{F.1}$$

By Theorem 2 (iii), $|(\Lambda^{-1}\hat{\gamma}_k)_{j\in\mathcal{S}_k}|_2 = |\tilde{\gamma}_k|_2(1+o_{\mathbb{P}}(1))$. Then together with (3.6), we obtain that the long run variance for ε_k is $\varsigma_k^2(1+o_{\mathbb{P}}(1))$. Hence by Theorem 1 in [7], $\sum_{s=\hat{\tau}_k-bn}^{\hat{\tau}_k+bn} \varepsilon_s = O_{\mathbb{P}}((bn)^{1/2}\varsigma_k)$ and $\sum_{s=\tau_k}^{\tau_k+r} \varepsilon_s = O_{\mathbb{P}}(r^{1/2}\varsigma_k)$. Note $|\tilde{\gamma}_k|_1 \le |\mathcal{S}_k|^{1/2}|\tilde{\gamma}_k|_2$, thus

$$|\tilde{\gamma}_k|_1 b \le |\mathcal{S}_k|^{1/2}|\tilde{\gamma}_k|_2 b \ll |\mathcal{S}_k|^{1/2}|\tilde{\gamma}_k|_2 \delta^\dagger \le |\tilde{\gamma}_k|_2^2 = a_k.$$

Inserting the above equation into (F.1) leads to

$$D_{\tau_k+r} - D_{\tau_k} = -2^{-1}a_k r(1+o_{\mathbb{P}}(1)) + O_{\mathbb{P}}(\varsigma_k r(bn)^{-1/2} + \varsigma_k r^{1/2}).$$

Since $\tilde{\Sigma}_k$ is a covariance matrix with diagonal entities 1, $|\tilde{\Sigma}_k|_2 \le |\mathcal{S}_k|$. Note $\varsigma_k^2 \le |\tilde{\Sigma}_k|_2 a_k$, thus

$$\varsigma_k \le |\mathcal{S}_k|^{1/2} a_k^{1/2}.$$

Then we have

$$\varsigma_k (bn)^{-1/2} \le |\mathcal{S}_k|^{1/2} a_k^{1/2} (bn)^{-1/2} \ll |\mathcal{S}_k|^{1/2} a_k^{1/2} \delta^\dagger \le a_k,$$

where the last inequality is because $a_k = |\tilde{\gamma}_k|_2^2 \ge \delta^{\dagger 2}|\mathcal{S}_k|$. By above we further obtain

$$D_{\tau_k+r} - D_{\tau_k} = -2^{-1}a_k r(1+o_{\mathbb{P}}(1)) + O_{\mathbb{P}}\big(\varsigma_k r^{1/2}\big). \tag{F.2}$$

Since $D_{\tilde{\tau}_k}$ is the maximum, $D_{\tau_k+r} - D_{\tau_k} > 0$. Therefore $r = O_{\mathbb{P}}(\varsigma_k^2/a_k^2)$. By a similar argument for the $r<0$ part, the desired result follows. □

Proof of *(ii)*. Let $\mathcal{F}_t$ be the σ-field generated by $\{\eta_{s,j}, s \le t, 1 \le j \le p\}$. Denote the projection operator $\mathcal{P}_t \cdot = \mathbb{E}(\cdot|\mathcal{F}_t) - \mathbb{E}(\cdot|\mathcal{F}_{t-1})$. Let $\bar{\gamma}_{k,j} = (\Lambda^{-1}\gamma_k)_j$, if $j \in \mathcal{S}_k$, $\bar{\gamma}_{k,j} = 0$ if $j \notin \mathcal{S}_k$. Let (η_t') be an i.i.d copy of (η_t). Then

$$c_t := \|\mathcal{P}_0 \varepsilon_t\|_4 \le \|\bar{\gamma}_k^\top \Lambda^{-1} A_t(\eta_0 - \eta_0')\|_4 \lesssim |\bar{\gamma}_k^\top \Lambda^{-1} A_t|_2 \mu_4,$$

where the last inequality is by Lemma 2 and that $\eta_{0,j}$, $1 \le j \le p$, are i.i.d. By Assumption 3.4,

$$\sum_{s\le t} c_s \lesssim \sum_{s\le t} |\bar{\gamma}_k^\top \Lambda^{-1} A_s|_2 \le \sum_{j=1}^{p}\sum_{s\le t} \bar{\gamma}_{k,j}\sigma_{j,j}^{-1/2}|A_{s,j,\cdot}|_2 \lesssim |\bar{\gamma}_k|_1 t^{-\beta} = |\tilde{\gamma}_k|_1 t^{-\beta}.$$

Thus by Corollary 2.1 in [1], strong invariance principle holds for $\sum_{s\le t}\varepsilon_s$. Thus from (F.1), we have

$$D_{\tau_k+r} - D_{\tau_k} = -2^{-1}a_k r(1+o_{\mathbb{P}}(1)) - \sum_{s=\tau_k}^{\tau_k+r-1}\varepsilon_s \xrightarrow{\mathcal{D}} -2^{-1}a_k r + \varsigma_k \mathbb{B}(r).$$

The $r<0$ part can be similarly dealt with. □

APPENDIX G: PROOF OF PROPERTIES OF ESTIMATED LONG RUN VARIANCE

Proof of Theorem 5 (i). The main idea follows the proof of Proposition 2.4 in [3], however due to the dependence and the break points, our result is much more involved. Let

$$\mathcal{S} = \{k \mid \mathcal{A}_k \text{ or } \mathcal{A}_{k-1} \text{ contains break points}\}.$$

Then by assumption $|\mathcal{S}| \le 2K_0$. We look at estimators without the break point first.

$$\bar{h}_{i,j}(u) = \sum_{k\notin\mathcal{S}} \phi_{\alpha_{i,j}}(\hat{\sigma}_{i,j,k}-u)/N_2, \quad \text{where } N_2 = N_1 - |S|.$$

Let

$$\tilde{\sigma}_{i,j} = \sum_{k\notin\mathcal{S}} \mathbb{E}\hat{\sigma}_{i,j,k}/N_2 \quad \text{and} \quad v_{i,j}^2 = \sum_{k\notin\mathcal{S}} \mathbb{E}\hat{\sigma}_{i,j,k}^2/N_2 - \tilde{\sigma}_{i,j}^2.$$

Define functions

$$B_{i,j}^{+}(u,x) = \tilde{\sigma}_{i,j} - u + \alpha_{i,j}[(\tilde{\sigma}_{i,j}-u)^2 + v_{i,j}^2]/2 + x,$$
$$B_{i,j}^{-}(u,x) = \tilde{\sigma}_{i,j} - u - \alpha_{i,j}[(\tilde{\sigma}_{i,j}-u)^2 + v_{i,j}^2]/2 - x.$$

The proof contains four steps.
Step 1. This step shows that function $\mathbb{E}\bar{h}_{i,j}(u)$ for any i,j satisfies, the expected loss functions have upper and lower envelope functions,

$$B_{i,j}^{-}(u,0) \le \mathbb{E}\bar{h}_{i,j}(u) \le B_{i,j}^{+}(u,0).$$

By (2.30), $\phi(x) \le x + x^2/2$ and thus

$$\mathbb{E}\bar{h}_{i,j}(u) \le \sum_{k\notin\mathcal{S}} \big(\mathbb{E}(\hat{\sigma}_{i,j,k} - u) + \alpha_{i,j}\mathbb{E}(\hat{\sigma}_{i,j,k} - u)^2/2\big)/N_2 = B^{+}_{i,j}(u,0).$$

Similarly we can bound the other side.
Step 2. This step shows for any $x > 0$, the estimated influence function $\bar{h}_{i,j}(u)$ is highly concentrated around its mean, for $C_0 > 0$ and $x \gtrsim (N_2\log(N_2))^{1/2}$,

$$\sum_{i,j=1}^{p} \mathbb{P}\Big(\sup_{|u-\sigma_{i,j}|\le C_0} |\bar{h}_{i,j}(u) - \mathbb{E}\bar{h}_{i,j}(u)| \ge x(\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2})/N_2\Big) \lesssim p^2\big(N_2\log(n)^{q/4}x^{-q/2} + e^{-x^2/(cN_2)}\big), \tag{G.1}$$

where c and the constant in $\lesssim$ are independent of n, p.

First introduce some notation. For any random variable X, denote $\mathbb{E}_0X = X - \mathbb{E}X$, the centering operator. Let $\mathcal{F}_k = (\eta_t,\ t \in \cup_{s\le k}\mathcal{A}_s)$ and $\mathcal{F}_{k,\{s\}}$, $s \le k$, be $\mathcal{F}_k$ with $\eta_t, t \in \mathcal{A}_s$ therein replaced by η'_t, where η'_t are i.i.d copy of η_t. For any random variable $X = h(\mathcal{F}_k)$, let $X_{\{i\}} = h(\mathcal{F}_{k,\{i\}})$. Denote $\Delta\xi_k = \xi_k - \xi_{k-1}$. We now show that the temporal dependence measure decays with polynomial rate. Let $\zeta_{i,j,k}(u) = \phi_{\alpha_{i,j}}(\hat{\sigma}_{i,j,k} - u)$. Since $|\phi'|_\infty \le 1$, we have for any $s \in \mathbb{N}$ and any u,

$$\begin{aligned}
&\|\sup_u |\zeta_{i,j,k}(u) - \zeta_{i,j,k,\{k-s\}}(u)|\|_{q/2} \le \|\hat{\sigma}_{i,j,k} - \hat{\sigma}_{i,j,k,\{k-s\}}\|_{q/2}\\
\le & 2^{-1}m\Big(\|\mathbb{E}_0\Delta\xi_{k,i}(\Delta\xi_{k,j} - \Delta\xi_{k,j,\{k-s\}})\|_{q/2} + \|\mathbb{E}_0(\Delta\xi_{k,i} - \Delta\xi_{k,i,\{k-s\}})\Delta\xi_{k,j,\{k-s\}}\|_{q/2}\Big)\\
=:& 2^{-1}m(\mathrm{I}_1 + \mathrm{I}_2).
\end{aligned}$$

Let $U_{k,i}$ (resp. $f_{k,i}$) be $\xi_{k,i}$ with Y_t replaced by ϵ_t (resp. $f(t/n)$). Then $\xi_{k,i} = \epsilon_{k,i} + f_{k,i}$ when there is no break. Let $\Delta U_{k,i} = U_{k,i} - U_{k-1,i}$ and $\Delta f_{k,i} = f_{k,i} - f_{k-1,i}$. Then we have

$$\begin{aligned}
\mathrm{I}_1 \quad \le \quad & \Big\|\mathbb{E}_0\Delta f_{k,i}(\Delta U_{k,j} - \Delta U_{k,j,\{k-s\}})\Big\|_{q/2} + \Big\|\mathbb{E}_0\Delta U_{k,i}(\Delta U_{k,j} - \Delta U_{k,j,\{k-s\}})\Big\|_{q/2}\\
=: & \ \mathrm{I}_{11} + \mathrm{I}_{12}.
\end{aligned}$$

Since $\max_j |f_j|_\infty < f^*$,

$$\max_{1\le j\le p} |\Delta f_{k,j}| \le f^* m/n. \tag{G.2}$$

Let $E_{k,l,i,\cdot} = \sum_{t=(km+1)\vee l}^{(k+1)m} A_{t-l,i,\cdot}$, where $A_{t-l,i,\cdot}$ is the ith row of matrix A_{t-l}. Then for $s \ge 1$,

$$\begin{aligned}
U_{k,i} &= \sum_{l\le(k+1)m} E_{k,l,i,\cdot}\eta_l/m \\
\text{and} \quad \Delta U_{k,i} - \Delta U_{k,i,\{k-s\}} &= \sum_{l\in\mathcal{A}_{k-s}} (E_{k,l,i,\cdot} - E_{k-1,l,i,\cdot})(\eta_l - \eta'_l)/m.
\end{aligned} \tag{G.3}$$

Since η_l are i.i.d, by Lemma 2, (G.2) and (G.3),

(G.4)
$$\|\Delta f_{k,i}(\Delta U_{k,j}-\Delta U_{k,j,\{k-s\}})\|_{q/2}\le 2f^*c_q\Big(\sum_{l\in\mathcal{A}_{k-s}}(|E_{k,l,i\cdot}|_2+E_{k-1,l,i\cdot}|_2)^2\Big)^{1/2}\mu_{q/2}/n.$$

By Assumption 3.4, we have for any $s>1$,

(G.5)
$$\sum_{l\in\mathcal{A}_{k-s}}|E_{k,l,i,\cdot}|_2^2\lesssim m(m(s-1))^{-2\beta}\sigma_{i,i},\quad\text{and}\quad\sum_{l\le(k+1)m}|E_{k,l,i,\cdot}|_2^2\lesssim m\sigma_{i,i},$$

where the constant in $\lesssim$ only depending on β, c_s. Hence by (G.4) and (G.5),

$$\mathrm{I}_{11}\lesssim m^{1/2}n^{-1}(m(s-1))^{-\beta}\sigma_{i,i}^{1/2},\tag{G.6}$$

where the constant in $\lesssim$ only depends on $\beta, c_s, q, \mu_q, f^*$. By Lemma 2 and (G.5)

$$\begin{aligned}
&\|\mathbb{E}_0U_{k,i}(U_{k,j}-U_{k,j,\{k-s\}})\|_{q/2}\\
=&\Big\|\mathbb{E}_0\Big(\sum_{l_1\le(k+1)m}E_{k,l_1,i,\cdot}\eta_{l_1}\sum_{l_2\in\mathcal{A}_{k-s}}E_{k,l_2,j,\cdot}(\eta_{l_2}-\eta'_{l_2})\Big)\Big\|_{q/2}m^{-2}\\
\lesssim&m^{-2}\Big(\sum_{l_1\le(k+1)m}\sum_{l_2\in\mathcal{A}_{k-s}}|E_{k,l_1,i,\cdot}|_2^2|E_{k,l_2,j,\cdot}|_2^2\Big)^{1/2}\\
\lesssim&m^{-1}(m(s-1))^{-\beta}\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2},
\end{aligned}\tag{G.7}$$

where the constant in $\lesssim$ only depends on μ_q, q, c_s. Thus $\mathrm{I}_{12}\lesssim m^{-1}(m(s-1))^{-\beta}\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2}$. By combining the bounds for I_{11} and I_{12} and a similar argument for I_2, we have

(G.8)
$$\delta_s:=\max_k\big\|\sup_u|\zeta_{i,j,k}(u)-\zeta_{i,j,k,\{k-s\}}(u)|\big\|_{q/2}\lesssim((ms)^{-\beta}\mathbf{1}_{s>1}+\mathbf{1}_{s\le1})\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2},$$

where the constant in $\lesssim$ only depends on $\mu_q, c_s, c_\sigma, \beta, q, f^*$.

Let $\delta:=\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2}x/(2N_2)$ and A_n be the δ net for $\{u:|u-\sigma_{i,j}|\le C_0\}$. Denote $f(u)=\bar h_{i,j}(u)-\mathbb{E}\bar h_{i,j}(u)$. Then by $|\phi'|_\infty\le1$,

$$\sup_{|v-\sigma_{i,j}|\le C_0}\min_{u\in A_n}|f(u)-f(v)|\le\delta.$$

Therefore $|A_n| = 2C_0/\delta = O(n)$ and

$$\begin{aligned}&\mathbb{P}\Big(\sup_{|u-\sigma_{i,j}|\le C_0} |\bar{h}_{i,j}(u) - \mathbb{E}\bar{h}_{i,j}(u)| \ge x(\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2})/N_2\Big)\\ \le&\mathbb{P}\Big(\max_{u\in A_n} |\bar{h}_{i,j}(u) - \mathbb{E}\bar{h}_{i,j}(u)| \ge x(\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2})/(2N_2)\Big).\end{aligned}$$

Desired result follows from Lemma 5.8 in [12].
Step 3. This step shows for the estimator

(G.9)
$$\max_{1\le i,j\le p} |\tilde{\sigma}_{i,j} - \sigma_{i,j}| = O(m^{-\beta/(\beta+1)}\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2} + m^3/n^2), \quad \text{and} \quad v_{i,j}^2 = O(\sigma_{i,i}\sigma_{j,j}),$$

where the convergence is uniform for $1 \le i, j \le p$.

Let $\hat{\epsilon}_{i,j,k}$ be $\hat{\sigma}_{i,j,k}$ with Y_t replaced by ϵ_t and let $\sigma_{i,j}^{\diamond} = \mathbb{E}\hat{\epsilon}_{i,j,1}$. Then by (G.2),

$$|\tilde{\sigma}_{i,j} - \sigma_{i,j}^{\diamond}| \le m \sum_{k\notin\mathcal{S}} |\Delta f_{k,i}||\Delta f_{k,j}|/(2N_2) = O(m^3/n^2). \tag{G.10}$$

Note the convergence in above $O(\cdot)$ and all the followings are uniform for i, j. Let $\rho_{i,j,k} = \mathbb{E}(\epsilon_{0,i}\epsilon_{k,j})$. Then for any $L < m$,

$$\begin{aligned}&|m\mathbb{E}(U_{1,i}U_{1,j}) - \sigma_{i,j}| = \Big|m^{-1} \sum_{-m<k<m} (m - |k|)\rho_{i,j,k} - \sigma_{i,j}\Big|\\ =&O\big(\sum_{|k|\ge L} |\rho_{i,j,k}| + Lm^{-1}\sum_{k\in\mathbb{Z}} |\rho_{i,j,k}|\big).\end{aligned}$$

By Assumption 3.4, $\sum_{|k|\ge L} |\rho_{i,j,k}| \le \sum_{t\in\mathbb{Z},|k|\ge L} |A_{t,i,\cdot}|_2|A_{t+k,j,\cdot}|_2 = O(L^{-\beta}\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2})$. Take $L = m^{1/(\beta+1)}$, then $|m\mathbb{E}(U_{1,i}U_{1,j}) - \sigma_{i,j}| = O(m^{-\beta/(\beta+1)}\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2})$. And similarly $|m\mathbb{E}(U_{1,i}U_{2,j})| = O(m^{-\beta/(\beta+1)}\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2})$. Hence

$$\begin{aligned}\sigma_{i,j}^{\diamond} &= m\big(\mathbb{E}(U_{1,i}U_{1,j}) + \mathbb{E}(U_{2,i}U_{2,j}) - \mathbb{E}(U_{1,i}U_{2,j}) - \mathbb{E}(U_{2,i}U_{1,j})\big)/2\\ &= \sigma_{i,j} + O(m^{-\beta/(\beta+1)}\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2}).\end{aligned}$$

Together with (G.10) we obtain the first part in (G.9).

Since $\hat{\sigma}_{i,j,k} = m(\Delta f_{k,i} + \Delta U_{k,i})(\Delta f_{k,j} + \Delta U_{k,j})/2$, we have $\mathbb{E}\hat{\sigma}_{i,j,k} = m\Delta f_{k,i}\Delta f_{k,j}/2 + \sigma_{i,j}^{\diamond}$. By (G.2) and (G.10),

$$v_{i,j}^2 = \sum_{k\notin\mathcal{S}} \mathbb{E}\hat{\sigma}_{i,j,k}^2/N_2 - \tilde{\sigma}_{i,j}^2 = \sum_{k\notin\mathcal{S}} \mathrm{Var}(\hat{\sigma}_{i,j,k})/N_2 + O(m^3n^{-2}\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2}).$$

Note by (G.3) and (G.5) we have

$$m^2\mathrm{Var}(\Delta f_{k,i}\Delta U_{k,j}) = O(m^3 n^{-2}\sigma_{j,j}), \quad \text{and} \quad m^2\mathrm{Var}(\Delta U_{k,i}\Delta U_{k,j}) = O(\sigma_{i,i}\sigma_{j,j}).$$

Thus $\mathrm{Var}(\hat{\sigma}_{i,j,k}) = O(\sigma_{i,i}\sigma_{j,j})$ and the second part in (G.9) holds.
Step 4. Since $|\mathcal{S}| \le 2K_0$, for any i, j, and $|\phi|_\infty \le \log(2)$,

$$|N_1 h_{i,j}(u)/N_2 - \bar{h}_{i,j}(u)| \le 2\log(2)K_0/(\alpha_{i,j}N_2). \tag{G.11}$$

Combining (G.11), Step 1 and step 2 with $x = N_2/\log^3(np)$, then with probability tending 1, for all $1 \le i, j \le p$, and $|u - \sigma_{i,j}| \le C_0$,

$$B^-_{i,j}(u,\Delta) \le N_1 N_2^{-1} h_{i,j}(u) \le B^+_{i,j}(u,\Delta), \tag{G.12}$$

where

$$\Delta = h\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2} + 2\log(2)K_0/(\alpha_{i,j}N_2) \quad \text{and } h = \log(np)^{-3}. \tag{G.13}$$

Note if

$$\alpha_{i,j}^2 v_{i,j}^2 + 2\alpha_{i,j}\Delta \le 1, \tag{G.14}$$

then $B^+_{i,j}(u,\Delta)$ exists real roots. Denote the smaller one as u^+, which satisfies $u^+ \le \tilde{\sigma}_{i,j} + \alpha_{i,j}v_{i,j}^2 + 2\Delta$. Take $\alpha^*_{i,j} = \alpha_{i,j}\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2}$. By Step 3 and Assumption 3.3, if (G.14), then

$$\sigma_{i,i}^{-1/2}\sigma_{j,j}^{-1/2}(u^+ - \sigma_{i,j}) = O\Big\{m^{-\beta/(\beta+1)} + m^3/n^2 + \alpha^*_{i,j} + h + m/(\alpha^*_{i,j}n)\Big\}. \tag{G.15}$$

Similar bound can be obtained for u^- as well. When (G.12) holds, $u^- \le \hat{\sigma}_{i,j} \le u^+$. Take $\alpha^*_{i,j} = (m/n)^{1/2}$, then with probability greater than $1 - \log(n)^{-1}$,

$$\sigma_{i,i}^{-1/2}\sigma_{j,j}^{-1/2}|\hat{\sigma}_{i,j} - \sigma_{i,j}| \lesssim m^{-\beta/(\beta+1)} + m^3/n^2 + (m/n)^{1/2} + \log(np)^{-3},$$

where the convergence is uniform for all $1 \le i, j \le p$. Since there exists some constant $c_1, c_2 > 0$, such that $c_1 \le \bar{\sigma}_{j,j}/\sigma_{j,j} \le c_2$, and any i, j with probability tending to 1. Thus the desired result follows. □

Proof of (ii). Same argument as for (i), except that we need to replace Step 2 by Step 2' with $x \ll N_2/\log(np)^{2.5}$. Then we obtain the desired result.

Step 2'. This step shows

$$\sum_{i,j=1}^{p} \mathbb{P}\Big(\sup_{|u-\sigma_{i,j}|\le C_0} \big|\bar{h}_{i,j}(u) - \mathbb{E}\bar{h}_{i,j}(u)\big| \ge x(\sigma_{i,i}^{1/2}\sigma_{j,j}^{1/2})/N_2 \Big) \lesssim p^2 n e^{-cxN_2^{-1/2}}, \tag{G.16}$$

where c and the constant in $\lesssim$ are independent of n, p, i, j.

The proof follows similar argument as in Step 2 and Theorem 3 in [11].

$\square$

REFERENCES

[1] Berkes, I., W. Liu, and W. B. Wu (2014). Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.* *42*(2), 794–817.

[2] Burkholder, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. *Astérisque* (157-158), 75–94. Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987).

[3] Catoni, O. (2012). Challenging the empirical mean and empirical variance: a deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* *48*(4), 1148–1185.

[4] Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Testing many moment inequalities. *arXiv preprint arXiv:1312.7614*.

[5] Chernozhukov, V., D. Chetverikov, and K. Kato (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* *162*(1-2), 47–70.

[6] Chernozhukov, V., D. Chetverikov, and K. Kato (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* *45*(4), 2309–2352.

[7] El Machkouri, M., D. Volný, and W. B. Wu (2013). A central limit theorem for stationary random fields. *Stochastic Process. Appl.* *123*(1), 1–14.

[8] Fan, J. and I. Gijbels (1996). *Local polynomial modelling and its applications*, Volume 66 of *Monographs on Statistics and Applied Probability.* Chapman & Hall, London.

[9] Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric aspects of functional analysis*, Volume 1807 of *Lecture Notes in Math.*, pp. 169–187. Springer, Berlin.

[10] Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.* *22*(1), 146–163.

[11] Wu, W.-B. and Y. N. Wu (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* *10*(1), 352–379.

[12] Zhang, D., W. B. Wu, et al. (2017). Gaussian approximation for high dimensional time series. *The Annals of Statistics* *45*(5), 1895–1919.

IRTG 1792 Discussion Paper Series 2019

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Cooling Measures and Housing Wealth: Evidence from Singapore" by Wolfgang Karl Härdle, Rainer Schulz, Taojun Xie, January 2019.

002 "Information Arrival, News Sentiment, Volatilities and Jumps of Intraday Returns" by Ya Qian, Jun Tu, Wolfgang Karl Härdle, January 2019.

003 "Estimating low sampling frequency risk measure by high-frequency data" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

004 "Constrained Kelly portfolios under alpha-stable laws" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

005 "Usage Continuance in Software-as-a-Service" by Elias Baumann, Jana Kern, Stefan Lessmann, February 2019.

006 "Adaptive Nonparametric Community Detection" by Larisa Adamyan, Kirill Efimov, Vladimir Spokoiny, February 2019.

007 "Localizing Multivariate CAViaR" by Yegor Klochkov, Wolfgang K. Härdle, Xiu Xu, March 2019.

008 "Forex Exchange Rate Forecasting Using Deep Recurrent Neural Networks" by Alexander J. Dautel, Wolfgang K. Härdle, Stefan Lessmann, Hsin-Vonn Seow, March 2019.

009 "Dynamic Network Perspective of Cryptocurrencies" by Li Guo, Yubo Tao, Wolfgang K. Härdle, April 2019.

010 "Understanding the Role of Housing in Inequality and Social Mobility" by Yang Tang, Xinwen Ni, April 2019.

011 "The role of medical expenses in the saving decision of elderly: a life cycle model" by Xinwen Ni, April 2019.

012 "Voting for Health Insurance Policy: the U.S. versus Europe" by Xinwen Ni, April 2019.

013 "Inference of Break-Points in High-Dimensional Time Series" by Likai Chen, Weining Wang, Wei Biao Wu, May 2019.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.