

ℓ^2 INFERENCE FOR CHANGE POINTS IN HIGH-DIMENSIONAL TIME SERIES VIA A TWO-WAY MOSUM

BY JIAQI LI^{1,a}, LIKAI CHEN^{1,b}, WEINING WANG^{2,c} AND WEI BIAO WU^{3,d}

¹Department of Statistics and Data Science, Washington University in St. Louis, ^alijiaqi@wustl.edu, ^blikai.chen@wustl.edu

²Faculty of Economics and Business, University of Groningen, ^cweining.wang@rug.nl

³Department of Statistics, University of Chicago, ^dwbwu@galton.uchicago.edu

We propose an inference method for detecting multiple change points in high-dimensional time series, targeting dense or spatially clustered signals. Our method aggregates moving sum (MOSUM) statistics cross-sectionally by an ℓ^2 -norm and maximizes them over time. We further introduce a novel Two-Way MOSUM, which utilizes spatial-temporal moving regions to search for breaks, with the added advantage of enhancing testing power when breaks occur in only a few groups. The limiting distribution of an ℓ^2 -aggregated statistic is established for testing break existence by extending a high-dimensional Gaussian approximation theorem to spatial-temporal non-stationary processes. Simulation studies exhibit promising performance of our test in detecting nonsparse weak signals. Two applications on equity returns and COVID-19 cases in the United States show the real-world relevance of our algorithms. The R package “L2hdchange” is available on CRAN.

1. Introduction. Change-point analysis is a fundamental problem in various fields of applications: in economics, the break effects of policy are of particular interest [10]; in biology, high-amplitude cofluctuations are utilized in cortical activity to represent dynamics of brain functional connectivity [21, 22]; in network analysis, change-point detection can be employed for the anomaly of network traffic data caused by attacks [32], etc. The above list of scenarios spans a wide range of data structures, including high-dimensional data with temporal and cross-sectional dependence, which pose substantial challenges to change-point analysis. The paper aims to address this issue by providing theory on multiple break inference for high-dimensional time series allowing both temporal and spatial dependence.

There is a sizable literature on high-dimensional change-point detection. Various studies consider data aggregation, and many of them consider ℓ^∞ -based methods; see, for example, [11, 27, 42, 55]. Most aforementioned studies focus on sparse signals, while an ℓ^2 -based approach favors nonsparse weak signals, and this is also the focus of this study. In the meanwhile, the ℓ^2 -type aggregation is quite common in the literature. Zhang et al. [4] evaluates the performance of a least square estimation and establishes a distribution theory for single change-point estimator in panel data without cross-sectional dependence; [58] develops a recursive algorithm based on sums of chi-squared statistics across samples with independent and identically distributed (i.i.d.) Gaussian noises, which could be viewed as an extension of the circular binary segmentation algorithm by [38]. In addition, [5, 8, 24, 25, 35] study ℓ^2 -based cumulative sum (CUSUM) statistics to estimate and make inference for change points in linear regression or panel models. Although their methods primarily concentrate on scenarios with a single break per time series, their approaches bear similarities to ours, with the key difference being our utilization of a moving sum (MOSUM) variant. More recently,

Received October 2023; revised February 2024.

MSC2020 subject classifications. Primary 62E20, 62M10; secondary 62G20.

Key words and phrases. Nonlinear time series, multiple change-point detection, ℓ^2 inference, Two-Way MOSUM, Gaussian approximation, temporal and spatial dependence.

[20] proposes a linear and a scan CUSUM statistic with the minimax bound established for the change-point estimator of i.i.d. Gaussian data; Chen et al. [12] introduce a coordinate-wise likelihood ratio test for online change-point detection for independent Gaussian data and present the response delay rate. In addition to aggregation, there are other well-known techniques for high-dimensional change-point analysis, including the U-statistics as demonstrated by [46, 47, 55], threshold-based approaches proposed by [15, 16] and a projection-based method developed by [49]. In this paper, we consider a maximized ℓ^2 -type test statistic to adapt to different data sets containing signals of distinct temporal-spatial properties and errors with complex dependency structures.

Besides the challenge of change-point test brought by high dimensionality, studies on multiple change-point detection have a long-standing tradition. In general, two broad classes of methods have been developed: model selection and hypothesis testing. Model selection approaches aim to treat change-point signals as parameters and derive estimators for them, such as the PELT algorithm [28] and the fused LASSO penalty [31, 33, 44]. Cho and Kirch [17] propose a localized application of the Schwarz criterion for multiscale change points. Kuchibhotla and others [30, 45, 53] consider change-point analysis for linear regression models featuring varying parameters, encompassing a broad range of nonlinear time series. As for testing, a traditional approach is binary segmentation developed by [41]. Its variants are considered in [6, 38]. Moreover, [23] introduces a wild binary segmentation and [16] proposes a sparsified binary segmentation algorithm. Yu [56] reviews diverse minimax rates in change-point analysis literature.

In the context of testing, MOSUM is a notably popular technique for both univariate and multivariate time series, such as [26] on i.i.d. data, [19, 51] on temporal dependent data and [29] on multivariate time-continuous stochastic processes. MOSUM is attractive due to the simplicity of implementation and an overall control of significance level, which avoids issues in multiple testing. However, for high-dimensional time series, when a MOSUM statistic aggregates all the series by an ℓ^2 -norm, the testing power would suffer if breaks only occur in a portion of them. Hence, in this paper, we propose a novel spatial-temporal moving sum algorithm called Two-Way MOSUM. This method utilizes moving spatial-temporal regions to search for temporal breaks and locate spatial neighborhoods where temporal breaks occur. Such moving regions can be viewed as a generalized concept of the moving windows in previous MOSUM, which can aggregate signals adaptive to cross-sectional group structures to enhance testing power. We emphasize that overlapping groups are allowed in our method and, therefore, the prior knowledge of groups is not a requirement for effective detection of breaks, since one can always search for all possible grouping scenarios. Nevertheless, prior grouping information is available in numerous data applications, which can boost the testing power by decreasing the number of searching windows. See, for example, in neuroscience, regions of interest (ROI) in human brains can be assigned to networks by different functions and the ROIs from the same network will undergo simultaneous functional change points [7]; in finance, stock prices of industries are often grouped by market capitalization and a few number of sectors may experience market shocks at the same time [39]. Note that the MOSUM procedure may not necessarily outperform the CUSUM in testing power. However, it may have better localization property in scenarios with multiple change points.

Although both ℓ^2 aggregation and MOSUM statistics have been well investigated respectively, it is quite challenging to rigorously develop an inference theory for ℓ^2 -based MOSUM statistics to detect the existence of breaks for the high-dimensional data. To be more specific, when we take the maximum of ℓ^2 statistics obtained from all the rolling windows over time, these aggregated statistics are temporally dependent even though the underlying errors may be independent. Most of the previous works concerning ℓ^2 -based statistics only provide inference for the break estimators by assuming the existence of a break, such as an ℓ^2 -type

break location estimator and its inference introduced by [4] for single change-point estimation with cross-sectionally independent errors. To the best of our knowledge, this study is the first to establish the limiting distribution of an ℓ^2 -type test statistic to facilitate the inference for change-point detection, which allows both spatial and temporal dependence.

To summarize, we contribute to the literature in both theory and algorithms. On the theory front, we propose an ℓ^2 -type MOSUM test statistic for multiple break detection in high-dimensional time series, allowing both temporal and spatial dependence. The Gaussian approximation (GA) result under the null is provided as a theoretical foundation to backup our detection of breaks (cf. Theorems 2.1 and 4.1). Correspondingly, we introduce an innovative Two-Way MOSUM statistic to account for spatially-clustered signals (cf. Theorem 3.1). Consistency results of estimators for the number of breaks, temporal and spatial break locations, as well as break sizes are all established (cf. Theorem 2.2 and Proposition 3.1).

Roadmap. The rest of this article is organized as follows. Section 2 is devoted to the test specification and asymptotic properties with cross-sectional independence assumed. Section 3 serves as an extension to the cases where breaks might exist only in a subset of component series (clustered signals). We follow with Section 4 as a generalization to nonlinear time series with spatial space in $\mathbb{Z}^v$, allowing both temporal and cross-sectional dependence. In Section 5, we deliver two empirical applications on testing structural breaks for the stock return and COVID-19 data. The simulation studies and proofs are deferred to the Appendix in the Supplementary Material [34].

Notation. For a vector $v = (v_1, \dots, v_d) \in \mathbb{R}^d$ and $q > 0$, we denote $|v|_q = (\sum_{i=1}^d |v_i|^q)^{1/q}$ and $|v|_\infty = \max_{1 \leq i \leq d} |v_i|$. For $s > 0$ and a random vector X , we say $X \in \mathcal{L}^s$ if $\|X\|_s = [\mathbb{E}(|X|_2^s)]^{1/s} < \infty$, and denote $\mathcal{E}_0(X) = X - \mathbb{E}(X)$. For two positive number sequences (a_n) and (b_n) , we say $a_n = O(b_n)$ or $a_n \lesssim b_n$ (resp., $a_n \asymp b_n$) if there exists $C > 0$ such that $a_n/b_n \leq C$ (resp., $1/C \leq a_n/b_n \leq C$) for all large n and say $a_n = o(b_n)$ or $a_n \gg b_n$ if $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$. Let (X_n) and (Y_n) to be two sequences of random variables. Write $X_n = O_{\mathbb{P}}(Y_n)$ if for $\forall \epsilon > 0$, there exists $C > 0$ such that $\mathbb{P}(|X_n/Y_n| \leq C) > 1 - \epsilon$ for all large n , and say $X_n = o_{\mathbb{P}}(Y_n)$ if $X_n/Y_n \rightarrow 0$ in probability as $n \rightarrow \infty$.

2. Testing and estimating high-dimensional change points. In this section, we propose a test statistic based on an ℓ^2 aggregated MOSUM and investigate its theoretical properties to test the presence of structural breaks. To formulate our model, let $Y_1, \dots, Y_n$ be observed p -dimensional random vectors satisfying

$$(2.1) \qquad Y_t = \mu(t/n) + \epsilon_t, \quad t = 1, \dots, n,$$

where $(\epsilon_t)_t$ is a sequence of p -dimensional stationary errors with zero mean and $\mu(\cdot)$ is a p -dimensional vector of unknown trend functions. Our main interest is to detect the potential change points occurring on the trend function

$$(2.2) \qquad \mu(u) = \mu_0 + \sum_{k=1}^K \gamma_k \mathbf{1}_{u \geq u_k},$$

where $K \in \mathbb{N}$ is the number of structural breaks, which is unknown and could go to infinity as n increases; $u_1, \dots, u_K$ are the time stamps of the breaks with $0 = u_0 < u_1 < \dots < u_K < u_{K+1} = 1$ and the minimum gap $\kappa_n = \min_{0 \leq k \leq K} (u_{k+1} - u_k)$, where $\kappa_n > 0$ is allowed to tend to 0 as $n \rightarrow \infty$; $\mu_0 \in \mathbb{R}^p$ represents the benchmark level when no break occurs and $\gamma_k \in \mathbb{R}^p$ is the jump vector at the time stamp u_k with size $|\gamma_k|_2$.

It should be noted that not all entries of γ_k need to be nonzero, which allows for cases where only a subset of time series experience a jump at the time stamp u_k . In such situations, it is preferable to aggregate only the series with breaks rather than all of them, as it can

improve the testing power. A more detailed discussion of this scenario is given in Section 3 where the Two-Way MOSUM method is introduced. In this section, for readability, we focus solely on the improved MOSUM that aggregates all time series. For brevity, we assume the time series to be linear and cross-sectionally independent throughout Sections 2 and 3, which will be relaxed to nonlinear and cross-sectionally dependent cases in Section 4.

2.1. ℓ^2 -based test statistics. This subsection is devoted to test the null hypothesis:

$$\mathcal{H}_0: \gamma_1 = \gamma_2 = \cdots = \gamma_K = 0,$$

which denotes the case with no breaks, against the alternative $\mathcal{H}_A$: there exists $k \in \{1, \dots, K\}$, such that $\gamma_k \neq 0$. Note that the number of breaks K is allowed to go to infinity under some condition on the separation of breaks. We refer to a detailed discussion below Definition 2.1 in Section 2.3.

The primary reason for testing the presence of structural breaks is to prevent model misspecification. If we apply a change-point algorithm to a data generating process without any actual breaks, we may obtain false break estimates, leading to erroneous conclusions. Therefore, it is necessary to test for the existence of breaks before conducting further analysis. However, previous studies on change points in high-dimensional time series mostly focus on inference for break location estimators, such as Theorem 2.2 in [25], which assumes the existence of breaks. Although there are some available literature on change-point testing for high-dimensional time series (see [11, 27, 48]), there is no existing theory of ℓ^2 -based statistics for testing break existence, which can be necessary and highly beneficial in identifying dense signals.

We define a jump vector $J(u)$ at the time point u as $J(u) = 0$ when no break occurs at the time stamp u , and $J(u) = \gamma_i$ when $u = u_i$ for some $1 \leq i \leq K$. Intuitively, we can test the existence of breaks by evaluating the jump estimate $\hat{J}(i/n) = \hat{\mu}_i^{(l)} - \hat{\mu}_i^{(r)}$, where

$$(2.3) \quad \hat{\mu}_i^{(l)} = \hat{\mu}^{(l)}(i/n) = \frac{1}{bn} \sum_{t=i-bn}^{i-1} Y_t, \quad \hat{\mu}_i^{(r)} = \hat{\mu}^{(r)}(i/n) = \frac{1}{bn} \sum_{t=i}^{i+bn-1} Y_t,$$

are the local averages on the left- and the right-hand sides of the time point i/n , respectively, and b is a bandwidth parameter satisfying $b \rightarrow 0$ and $bn \rightarrow \infty$. Without loss of generality, we assume that bn is an integer. We shall reject the null if $|\hat{J}(i/n)|_2$ is too large. To this end, we shall develop an asymptotic distributional theory, which appears to be highly nontrivial.

Throughout Sections 2 and 3, we assume that there is no dependence between component processes $(\epsilon_{t,j})_{t \in \mathbb{Z}}$, $1 \leq j \leq p$ and we later relax this restriction to allow for weak cross-sectional dependence in Section 4. To make all p components in $\hat{J}(i/n)$ comparable, we need to specify the error process $(\epsilon_t)_{t \in \mathbb{Z}}$ and obtain the long-run variances for standardization. In particular, we model ϵ_t as a vector moving average (VMA) process in Sections 2 and 3, which embraces many important time series models such as a vector autoregressive moving averages (VARMA) model, and we discuss a more general ϵ_t in Section 4 using functional dependence measure, which allows nonlinear forms. Let

$$(2.4) \quad \epsilon_t = \sum_{k \geq 0} A_k \eta_{t-k},$$

where $\eta_t \in \mathbb{R}^{\tilde{p}}$ are i.i.d. random vectors with zero mean and an identity covariance matrix with $p \leq \tilde{p} \leq cp$, for some constant $c > 1$. The coefficient matrices A_k , $k \geq 0$, take values in $\mathbb{R}^{p \times \tilde{p}}$ such that ϵ_t is a proper random vector. Define $\tilde{A}_0 = \sum_{k \geq 0} A_k$. Then the long-run covariance matrix of ϵ_t and the diagonal matrix of the long-run standard deviations are

$$(2.5) \quad \Sigma = \tilde{A}_0 \tilde{A}_0^\top = (\sigma_{i,j})_{i,j=1}^p, \quad \Lambda = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p),$$

respectively, where $\sigma_j^2 = \sigma_{j,j} \geq c_\sigma$, for some constant $c_\sigma > 0$ is the long-run variance of the j th component series. Note that in Sections 2 and 3, we assume $p = \tilde{p}$, and all $A_k, k \geq 0$ are diagonal matrices, which indicates the independence between the component processes $(\epsilon_{t,j})_{t \in \mathbb{Z}}$. We relax this assumption to weak cross-sectional dependence for a general ϵ_t in Section 4, and also provide the results for the same linear ϵ_t in (2.4) as a special case in Appendix B.4 in the Supplementary Material [34].

Following the previous intuition, we test the existence of breaks by evaluating the gap vectors $\hat{J}(\cdot)$. Namely, we standardize $\hat{J}(\cdot)$ by the long-run standard deviations of each time series, that is, for $bn + 1 \leq i \leq n - bn$,

(2.6)
$$V_i = \Lambda^{-1} \hat{J}(i/n) = \Lambda^{-1}(\hat{\mu}_i^{(l)} - \hat{\mu}_i^{(r)}).$$

To conduct the change-point detection with $p \rightarrow \infty$, we take the ℓ^2 aggregation of each V_i in the cross-sectional dimension, that is, $|V_i|_2^2$, to capture dense signals. Note that by model (2.1), the random vector V_i involves both the signal part $\mathbb{E}V_i$ and the error part $V_i - \mathbb{E}V_i$. We define

(2.7)
$$\bar{c} = \sum_{j=1}^p c_j \quad \text{where } c_j = \text{Var}(V_{i,j}),$$

and $V_{i,j} \in \mathbb{R}$ is the j th coordinate of V_i . Since no break exists under the null hypothesis, that is, $\mathbb{E}V_i = 0$, it follows that $|V_i|_2^2 - \bar{c}$ is a centered statistic under the null. The detailed evaluation of $\bar{c}$ is deferred to Remark 2.5. Finally, we move the windows in the temporal direction to find the maximum and formulate our ℓ^2 -based test statistic as follows:

(2.8)
$$\mathcal{Q}_n = \max_{bn+1 \leq i \leq n-bn} (|V_i|_2^2 - \bar{c}).$$

We consider $\mathcal{Q}_n$ as a feasible test statistic by assuming that the long-run standard deviation Λ is known. The estimated long-run variances via a robust M-estimation method proposed by [11] are utilized in Section 5 for applications, and the details are deferred to Appendix A.1 in the Supplementary Material [34].

It is worth noticing that when breaks are sparse in the cross-sectional components, an ℓ^∞ -type statistic, that is, $\mathcal{Q}_{n,\infty} = \max_{bn+1 \leq i \leq n-bn} |V_i|_\infty$, could be more powerful [11] than an ℓ^2 -based one. However, in the presence of weak dense signals, the ℓ^∞ test would have power loss [9] while the ℓ^2 -type statistic can boost the power due to the aggregation of weak signals; see Remark 2.1 for a simple comparison. The current study targets change-point detection with nonsparse or spatially clustered signals. An ℓ^2 -based test statistic $\mathcal{Q}_n$ is therefore being proposed.

REMARK 2.1 (Comparison of ℓ^2 and ℓ^∞ statistics). Here, we present a simulation study to intuitively show that an ℓ^2 -based test statistic is generally more powerful in detecting weak dense signals compared to an ℓ^∞ one, while in the case of sparse signals, an ℓ^∞ type statistic appears better. Specifically, we perform a single change-point detection using test statistics $\mathcal{Q}_n$ and $\mathcal{Q}_{n,\infty}$, respectively, and compare their testing powers with varied proportions of cross-sectional dimensions containing jumps. The errors are generated from MA(∞) models defined in (2.4) with $\eta_t \sim t_9$ and the sample size $n = 100$. We consider $p = 50, 100$ and the window size $bn = 20$. We set the coefficient matrix A_k in (2.4) to be $A_k = \text{diag}(a_1 k^{-3/2}, a_2 k^{-3/2}, \dots, a_p k^{-3/2})$, where $a_1, \dots, a_p$ are uniformly ranging from 0.5 to 0.9. For each time series with a break, the jump size is 1. All the reported powers in Figure 1 are averaged over 1000 replicates. Intuitively, our test statistic $\mathcal{Q}_n$ incorporates the term $|V_i|_2$, which aggregates dense signals in a linear fashion with respect to the number

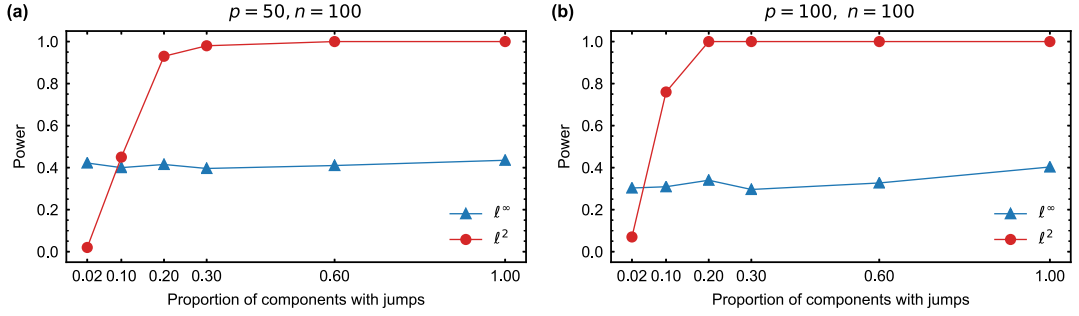


FIG. 1. Power comparison of ℓ^∞ MOSUM and ℓ^2 MOSUM.

of components p , while the standard deviation of weakly dependent random noises is aggregated on the order of $\sqrt{p}$. As a result, an ℓ^2 -type statistic is better suited for identifying dense signals, while its performance for sparse signals may be inferior since aggregating sparse signals will only add noise without any significant signal. After introducing Theorem 2.1, we shall provide a theoretical power comparison (cf. Remark 2.4).

2.2. Asymptotic properties of test statistics. To conduct the test, it is essential to understand the asymptotic behavior of the test statistic $\mathcal{Q}_n$. However, deriving the limiting distribution of $\mathcal{Q}_n$ under the null is highly challenging. This is because, even if the underlying errors are i.i.d., the standardized jump estimator V_i defined in (2.6) is still dependent over i due to the overlapped observations among different moving windows. In this section, we provide an intuition for the theoretical proofs of our first main theorem, which extends the high-dimensional GA for dependent data.

First, recall that the random vector V_i can be decomposed into the expectation $\mathbb{E}V_i$ and the deviation part $V_i - \mathbb{E}V_i$. We have $\mathbb{E}V_i = 0$ for any i under the null hypothesis. Let

$$(2.9) \quad x_{i,j} = (V_{i,j} - \mathbb{E}V_{i,j})^2 - c_j \quad \text{and} \quad X_j = (x_{bn+1,j}, \dots, x_{n-bn,j})^\top,$$

where c_j is defined in (2.7). By (2.6), we can write the test statistic $\mathcal{Q}_n$ under the null into

$$(2.10) \quad \mathcal{Q}_n = \max_{bn+1 \leq i \leq n-bn} \sum_{j=1}^p x_{i,j}.$$

When the errors are cross-sectionally independent, $X_1, \dots, X_p$ are also independent. Therefore, as p goes to infinity, we can apply the high-dimensional GA theorem to (2.10) to derive the asymptotic distribution of $\mathcal{Q}_n$. In this way, the temporal dependence caused by the overlapped moving windows can be properly dealt with. We generalize this result in Section 4 with cross-sectionally dependence allowed between $X_1, \dots, X_p$.

We introduce the centered Gaussian random vector $\mathcal{Z} = (\mathcal{Z}_{bn+1}, \dots, \mathcal{Z}_{n-bn})^\top \in \mathbb{R}^{n-2bn}$ with covariance matrix $\Xi = \mathbb{E}(\mathcal{Z}\mathcal{Z}^\top) \in \mathbb{R}^{(n-2bn) \times (n-2bn)}$, and denote the i th element in $\mathcal{Z}$ by $\mathcal{Z}_i$. Here, $\Xi = (\Xi_{i,i'})_{1 \leq i, i' \leq n-2bn}$ with expression

$$(2.11) \quad \Xi_{i,i'} = p(bn)^{-2} g(|i - i'|/(bn)),$$

where the function $g(\cdot) : [0, \infty) \mapsto \mathbb{R}$ is defined as

$$(2.12) \quad g(\zeta) = \begin{cases} 18\zeta^2 - 24\zeta + 8, & 0 \leq \zeta < 1, \\ 2\zeta^2 - 8\zeta + 8, & 1 \leq \zeta < 2, \\ 0, & \zeta \geq 2. \end{cases}$$

Note that $g(\zeta)$ has bounded second derivatives except for the point $\zeta = 1$. The matrix Ξ is asymptotically equal to the covariance matrix of $\sum_{j=1}^p X_j \in \mathbb{R}^{n-2bn}$. The detailed evaluation of $g(\zeta)$ in (2.12) is deferred to Lemma C.12 in Appendix C in the Supplementary Material [34]. The regime with $0 \leq \zeta < 1$ corresponds to correlations of statistics within length of bn , while $1 \leq \zeta < 2$ is concerning statistics, which are further apart ($bn < |i - i'| < 2bn$) and still have weaker correlations. Finally, $\zeta \geq 2$ suggests that statistics beyond $2bn$ are uncorrelated.

By the GA theorem, we shall expect the asymptotic distribution of $\mathcal{Q}_n$ under the null to be approximated by the maximum coordinate of a centered Gaussian vector $\mathcal{Z}$, that is,

$$(2.13) \quad \mathbb{P}(\mathcal{Q}_n \leq u) \approx \mathbb{P}\left(\max_{bn+1 \leq i \leq n-bn} \mathcal{Z}_i \leq u\right).$$

This result allows us to find the critical value of our proposed test statistic $\mathcal{Q}_n$ by Monte Carlo replicates. We refer to a simulation study in Appendix A.2 in the Supplementary Material [34].

2.3. Gaussian approximation. In this section, we provide a theory, which implies the critical value of the proposed change-point test. We first consider the simplest setting where the errors are assumed to be cross-sectionally independent, and a MOSUM statistic aggregating all the p series is adopted. The novel Two-Way MOSUM follows in Section 3. A generalized case with cross-sectionally dependent errors is investigated in Section 4.

We begin with two necessary assumptions as follows.

ASSUMPTION 2.1 (Finite moment). Assume that the innovations $\eta_{i,j}$ defined in (2.4) are i.i.d. with $\mu_q := \|\eta_{1,1}\|_q < \infty$ for some $q \geq 8$.

ASSUMPTION 2.2 (Temporal dependence). There exist constants $C > 0, \beta > 0$ such that $\max_{1 \leq j \leq p} \sum_{k \geq h} |A_{k,j,\cdot}|_2 / \sigma_j \leq C(1 \vee h)^{-\beta}$, for all $h \geq 0$, where $A_{k,j,\cdot}$ is the j th row of A_k .

ASSUMPTION 2.3 (Cross-sectional independence). Assume that for all $k \geq 0$, the coefficient matrices A_k defined in (2.4) are diagonal matrices.

Assumption 2.1 puts tail assumptions on the moment of the noise sequences in our moving average model (2.4). Assumption 2.2 is a very general and mild temporal dependence condition, which requires the polynomial decay rate of the temporal dependence. It also ensures that the long-run variance is finite. Many interesting processes fulfill this assumption. We refer to a specific example in Appendix B.1 in the Supplementary Material [34]. Note that we introduce Assumption 2.3 for the simplicity to show the GA and it shall be relaxed to allow weak cross-sectional dependence in Section 4.

Provided all the aforementioned conditions, we state our first main theorem as follows.

THEOREM 2.1 (GA with cross-sectional independence). Suppose that Assumptions 2.1–2.3 are satisfied. Then, under the null hypothesis, for $\Delta_0 = (bn)^{-1/3} \log^{2/3}(n)$,

$$\Delta_1 = \left(\frac{n^{4/q} \log^7(pn)}{p} \right)^{1/6}, \quad \Delta_2 = \left(\frac{n^{4/q} \log^3(pn)}{p^{1-2/q}} \right)^{1/3},$$

we have

$$(2.14) \quad \sup_{u \in \mathbb{R}} \left| \mathbb{P}(\mathcal{Q}_n \leq u) - \mathbb{P}\left(\max_{bn+1 \leq i \leq n-bn} \mathcal{Z}_i \leq u\right) \right| \lesssim \Delta_0 + \Delta_1 + \Delta_2,$$

where the constant in $\lesssim$ is independent of n, p, b . If in addition, $\log(n) = o\{(bn)^{1/2}\}$ and

$$(2.15) \quad n^4 p^{2-q} \log^{3q}(pn) \rightarrow 0,$$

then we have

$$(2.16) \quad \sup_{u \in \mathbb{R}} \left| \mathbb{P}(\mathcal{Q}_n \leq u) - \mathbb{P}\left(\max_{bn+1 \leq i \leq n-bn} \mathcal{Z}_i \leq u\right) \right| \rightarrow 0.$$

The symbol $o(1)$ and $\rightarrow 0$ in this theorem and the rest of the paper is in the asymptotic regime $n \wedge p \rightarrow \infty$. It is worth noting that the convergence rate of the GA in Theorem 2.1 depends on the number of cross-sectional components p , and a larger p is no longer a curse when utilizing an ℓ^2 -type test statistic. The intuition behind this is that when applying the GA, we effectively treat our p cross-sectional components as equivalent to the “ n observations” in [13]. As such, a larger p provides more information that can be used to detect change points, which can in turn reduce the approximation error.

Compared to the finite moment condition (E.2) in [13], which assumes that $q \geq 4$, we require $q \geq 8$ since X_j in our test statistic $\mathcal{Q}_n$ is quadratic with regard to the random noise $\epsilon_{t,j}$. As for the GA rate in Proposition 2.1 in [13], our Δ_1 and Δ_2 correspond to their rate with p replaced by n and $B_n = n^{2/q}$. It shall be noted that our additional term Δ_0 is due to an additional step to compare noncentered Gaussian variables (cf. Lemma C.2 in Appendix C).

REMARK 2.2 (Allowed dimension p relative to n). In Theorem 2.1, we can allow p to be of some polynomial order of n , and its order depends on the moment parameter q defined in Assumption 2.1. In particular, let $p \asymp n^{v_1}$, for some $v_1 > 0$. If $v_1 > 4/(q-2)$, then expression (2.16) holds. The larger moment parameter q is, the weaker condition on p we can allow.

REMARK 2.3 (Comment on the convergence rate in Theorem 2.1). We standardize X_j defined in (2.9) and denote it by X_j^* , that is, $X_j^* = bnX_j$. Chernozhukov et al. [14] derives a nearly optimal bound in the case when the smallest eigenvalue σ_*^2 of the covariance matrix of X_j^* is bounded below from zero. However, this sharp approximation rate cannot be achieved in our case. To see this, we consider the simple case where the errors are i.i.d. Then, for any integer h , the $(i, i+h)$ -th element of the covariance matrix $\mathbb{E}(X_j^* X_j^{*\top})$ takes the form of $g(|h|/(bn))$ in (2.11). It shall be noted that $\mathbb{E}(X_j^* X_j^{*\top})$ is a $2bn$ -banded matrix and it is symmetric. Since $g(|h|/(bn))$ has bounded second derivative except for one point, for any four coordinates at the positions $(s, s+2i+1)$, $(s, s+2i+2)$, $(s+2i, s)$ and $(s+2i+1, s)$ in $\mathbb{E}(X_j^* X_j^{*\top})$, $1 \leq s \leq n-2bn$, $0 \leq i \leq \lfloor (n-2bn-s-2)/2 \rfloor$, we can bound them in the following way:

$$(2.17) \quad -g\left(\frac{2i+1}{bn}\right) + g\left(\frac{2i+2}{bn}\right) - g\left(\frac{2i}{bn}\right) + g\left(\frac{2i+1}{bn}\right) = O\left\{\frac{1}{(bn)^2}\right\}.$$

Inspired by (2.17), we define a vector $y = (-1, 1, -1, 1, \dots)^\top \in \mathbb{R}^{n-2bn}$. When $b^2n \rightarrow \infty$, the upper bound of the smallest eigenvalue σ_*^2 of $\mathbb{E}(X_j^* X_j^{*\top})$ tends to 0, that is,

$$(2.18) \quad \sigma_*^2 \leq \frac{y^\top \mathbb{E}(X_j^* X_j^{*\top}) y}{y^\top y} = O\left\{\frac{1}{b^2n}\right\} \rightarrow 0.$$

Therefore, [14] is not applicable to our case and we instead extend the GA in [13] to achieve the rate in Theorem 2.1, which does not require the covariance matrix of X_j^* to be nondegenerate.

Theorem 2.1 implies that, for any level $\alpha \in (0, 1)$, we can choose the threshold value ω to be the quantile of the Gaussian limiting distribution indicated by Theorem 2.1:

$$(2.19) \quad \omega = \inf_{r \geq 0} \left\{ r : \mathbb{P}\left(\max_{bn+1 \leq i \leq n-bn} \mathcal{Z}_i > r\right) \leq \alpha \right\}.$$

We shall reject the null hypothesis if the test statistic $\mathcal{Q}_n$ exceeds the threshold value ω , that is, $\mathcal{Q}_n > \omega$. Under the alternative hypothesis, we show that when the break size is sufficiently large, we can achieve the testing power asymptotically tending to 1 (cf. Corollary 2.1). Recall the trend function defined in (2.2). We denote the break location by $\tau_k = nu_k$, $1 \leq k \leq K$ and introduce the following assumption for the identification of breaks.

DEFINITION 2.1 (Temporal separation). Define the minimum gap between breaks as $\kappa_n = \min_{0 \leq k \leq K} (u_{k+1} - u_k)$, for some $\kappa_n > 0$, and we allow $\kappa_n \rightarrow 0$ as n diverges.

Definition 2.1 concerns the separation of temporal break locations to ensure the consistency of break estimation, and it implies that K cannot be larger than $1/\kappa_n$. When $\kappa_n \rightarrow 0$, K can grow to infinity, which is in line with Assumption (1b) in [53], where they allow the minimum spacing to be a function of n and to vanish as n diverges. In addition, we require the bandwidth parameter b in (2.3) to satisfy $b \ll \kappa_n$ as $n \rightarrow \infty$. Otherwise, if more than one break exists within a window of temporal width bn , the adopted MOSUM statistics might fail to distinguish the break time points in the same window. For any time point i satisfying $|i - \tau_k| \leq bn$, we define the weighted break vectors as

$$(2.20) \quad d_i = \mathbb{E} V_i = \sum_{k=1}^K \left(1 - \frac{|i - \tau_k|}{bn}\right) \Lambda^{-1} \gamma_k \mathbf{1}_{|i - \tau_k| \leq bn},$$

and let $\underline{d} = (d_{bn+1}^\top, d_{bn+2}^\top, \dots, d_{n-bn}^\top)^\top$. Under the alternative hypothesis, there exists at least one break, that is, $\underline{d} \neq 0$. We evaluate our testing power in the following corollary.

COROLLARY 2.1 (Power). *Under Assumptions 2.1–2.3, if (2.15) holds, $b \ll \kappa_n$, and*

$$\max_{1 \leq k \leq K} n(u_{k+1} - u_k) |\Lambda^{-1} \gamma_k|_2^2 \gg \sqrt{p \log(n)},$$

then the testing power $\mathbb{P}(\mathcal{Q}_n > \omega) \rightarrow 1$, as $n \wedge p \rightarrow \infty$.

The symbols $\ll$ and $\gg$ hold here and throughout the rest of the paper in the asymptotic regime that $n \wedge p \rightarrow \infty$. Corollary 2.1 provides a condition for the test power tending to 1. It allows for cases with nontrivial alternatives. Namely, in some component series, the break sizes could be small and tend to 0, as long as the aggregated size of the break is sufficiently large. It is generally challenging to make an exact comparison between different break statistics with different assumptions for complicated panel data. Nevertheless, we can observe that our detection lower bound is quite sharp in the situation where many weak signals with similar magnitude are present. For example, suppose the jump size of each component series is $\vartheta \in \mathbb{R}$ and then we only require $n\kappa_n \vartheta^2 \gg \sqrt{\log(n)/p}$. Compared with Table 1 in [17], which summarizes the procedures under different settings, $\sqrt{\log(n)/p}$ is smaller than all entries in the table since we have $\sqrt{p}$ in the denominator resulting from the ℓ^2 -aggregation.

REMARK 2.4 (Detailed power comparison of ℓ^2 and ℓ^∞ statistics). This remark is a complement to Remark 2.1. Now we clarify these differences by a detailed power comparison in those two cases. First, we consider sparse signals. Suppose that there is only one time series with a single break and the break size is ϑ^* , and we use the MOSUM with bandwidth b for detection. Then to ensure the power tending to 1, $\mathcal{Q}_{n,\infty}$ only requires $\vartheta^* \gg \log^{1/2}(pn)(bn)^{-1/2}$ (see, e.g., [11]), while $\mathcal{Q}_n$ needs a stronger condition by Corollary 2.1 that $\vartheta^* \gg (p \log(n))^{1/4}(bn)^{-1/2}$. Second, we check dense signals. Suppose all series jump with the same size ϑ' . Then, for $\mathcal{Q}_{n,\infty}$, we still need $\vartheta' \gg \log^{1/2}(pn)(bn)^{-1/2}$, while $\mathcal{Q}_n$ only requires $\vartheta' \gg \log^{1/4}(n)p^{-1/4}(bn)^{-1/2}$.

Algorithm 1: ℓ^2 multiple change-point detection via a MOSUM**Data:** Observations $Y_1, Y_2, \dots, Y_n$; bandwidth parameter b ; threshold value ω **Result:** Estimated number of breaks $\hat{K}$; estimated break time stamps $\hat{\tau}_k, k = 1, \dots, \hat{K}$; estimated jump vectors $\hat{\gamma}_k$; estimated minimum break size over time $\hat{\delta}_p$ $\mathcal{Q}_n \leftarrow \max_{bn+1 \leq i \leq n-bn} (|V_i|_2^2 - \bar{c});$ **if** $\mathcal{Q}_n < \omega$ **then** $\hat{K} = 0$; STOP;**else** $k \leftarrow 1; \mathcal{A}_1 \leftarrow \{bn+1 \leq \tau \leq n-bn : (|V_\tau|_2^2 - \bar{c}) > \omega\};$ **while** $\mathcal{A}_k \neq \emptyset$ **do** $\hat{\tau}_k \leftarrow \arg \max_{\tau \in \mathcal{A}_k} (|V_\tau|_2^2 - \bar{c}); \hat{\gamma}_k \leftarrow \hat{\mu}_{\hat{\tau}_k-bn}^{(l)} - \hat{\mu}_{\hat{\tau}_k+bn-1}^{(r)};$ $\mathcal{A}_{k+1} \leftarrow \mathcal{A}_k \setminus \{t : |t - \hat{\tau}_k| \leq 2bn\}; k \leftarrow k + 1;$ **end** $\hat{K} = \max_{k \geq 1} \{k : \mathcal{A}_k \neq \emptyset\}; \hat{\delta}_p \leftarrow \min_{1 \leq k \leq \hat{K}} \|\Lambda^{-1} \hat{\gamma}_k\|_2^2 - \bar{c}^{1/2};$ **end**

2.4. *Estimation of change points.* Based on the theoretical underpinnings of the test in prior subsections, we can now present our detection algorithm (cf. Algorithm 1). To begin with, we explicate our strategy for detecting and estimating change points. Furthermore, we showcase the consistency for the estimators pertaining to the number, time stamps and sizes of breaks. A simulation study can be found in Appendix A.3.

We consider the size of the break at time point τ_k normalized by the long-run standard deviation Λ , that is, $|\Lambda^{-1} \gamma_k|_2$. Define the minimum of normalized break sizes over time as

$$(2.21) \quad \delta_p = \min_{1 \leq k \leq K} |\Lambda^{-1} \gamma_k|_2,$$

which can be viewed as the minimum strength of signals in the setting of Corollary 2.1.

REMARK 2.5 (Comments on the centering term $\bar{c}$). To implement Algorithm 1, we need to provide the centering term $\bar{c}$ defined in (2.7). Due to the weak temporal dependence of $\{\epsilon_t\}$, one can show that

$$(2.22) \quad \bar{c} = 2p/(bn)(1 + O\{1/(bn)\}).$$

In practice, we can simply take the centering term to be $\bar{c} \approx 2p/(bn)$. This still ensures the consistency results in Theorem 2.2 when the window size bn is slightly larger than the dimension p , since the approximation error of using $2p/(bn)$ is of $O\{p/(bn)^2\}$, which is smaller than the order of $(p \log(n))^{1/2}(bn)^{-1}$.

REMARK 2.6 (Selection of bandwidth b). Technically, the larger the bandwidth b is, the more data points could be used. Thus, a larger b would reduce the magnitude of noise, and then the signals shall be easier to detect. However, we also need to restrict b , since the identification condition $b \ll \kappa_n$ below Definition 2.1 requires us to have fewer than one break within each window. Therefore, we suggest starting with a small bandwidth, for example, $b = 1/\sqrt{n}$ to ensure that $bn \rightarrow \infty$. One could continue to increase the bandwidth until the estimated number of breaks decreases. We refer to Appendix A for a simulation study including the influences of different bandwidth parameters, as well as a discussion in Remark A.1 for the potential extension of our method to a multiscale MOSUM.

Now we outline the consistency results of estimators for break numbers, break time stamps and break sizes. We shall impose a minimum requirement of break sizes to guarantee that all of the breaks can be precisely captured and estimated.

ASSUMPTION 2.4 (Signal). Assume $\min_{0 \leq k \leq K} n(u_{k+1} - u_k) |\Lambda^{-1} \gamma_k|_2^2 \gg \sqrt{p \log(n)}$.

We highlight that Assumption 2.4 imposes a moderate condition on the signals, which relates the break separation $(u_{k+1} - u_k)$ to the break signal strength $|\gamma_k|_2^2$, and does not require p series to jump simultaneously. A comprehensive discussion on the advantage of our setting can be found following the consistency results presented below.

THEOREM 2.2 (Temporal consistency). *Let $q \geq 8$. Under Assumptions 2.1–2.4, if (2.15) hold, $b \ll \kappa_n$, $\delta_p^2 \geq 3\omega$, $\omega \gg (p \log(n))^{1/2} (bn)^{-1}$ and $\max_{1 \leq k \leq K} |\Lambda^{-1} \gamma_k|_q / |\Lambda^{-1} \gamma_k|_2 = O(1/K^{1/q})$. Then we have the following results:*

- (i) (Number of breaks). $\mathbb{P}(\hat{K} = K) \rightarrow 1$.
- (ii) (Time stamps of breaks). Let $\Gamma_k = p / (bn |\Lambda^{-1} \gamma_k|_2^2)$. Then we have $\max_{1 \leq k \leq K} |\hat{\tau}_k - \tau_{k^*}| \cdot |\Lambda^{-1} \gamma_k|_2^2 / (1 + \Gamma_k) = O_{\mathbb{P}}\{\log^2(n)\}$, where $k^* = \arg \min_i |\hat{\tau}_k - \tau_i|$. If in addition, $\delta_p^2 / p \gtrsim 1/(bn)$, we have

$$\max_{1 \leq k \leq K} |\hat{\tau}_k - \tau_{k^*}| \cdot |\Lambda^{-1} \gamma_k|_2^2 = O_{\mathbb{P}}\{\log^2(n)\}.$$

- (iii) (Break sizes). $\max_{1 \leq k \leq K} ||\Lambda^{-1}(\hat{\gamma}_k - \gamma_{k^*})|_2^2 - \bar{c}| = O_{\mathbb{P}}\{(p \log(n))^{1/2} (bn)^{-1}\}$, which also implies that $|\hat{\delta}_p - \delta_p| = O_{\mathbb{P}}\{(p \log(n))^{1/4} (bn)^{-1/2}\}$.

The results in Theorem 2.2 are in an asymptotic context as $n \wedge p \rightarrow \infty$. To see the advantage of our method, let us consider the simple case with a jump size $\vartheta \in \mathbb{R}$ for each component series. To achieve the consistency results in (ii), we only require $n\kappa_n \vartheta^2 \gg \sqrt{\log(n)/p}$ by Assumption 2.4 and $K = O(p^{q/2-1})$ by the condition $\max_{1 \leq k \leq K} |\Lambda^{-1} \gamma_k|_q / |\Lambda^{-1} \gamma_k|_2 = O(1/K^{1/q})$. Notably, as discussed following Corollary 2.1, our detection lower bound is weaker compared to those reported in Table 1 by [17]. In (iii), since $\Lambda^{-1} \hat{\gamma}_k$ contains both the signal part $\mathbb{E} V_{\hat{\tau}_k}$ and the error part $V_{\hat{\tau}_k} - \mathbb{E} V_{\hat{\tau}_k}$, we center $|\Lambda^{-1} \hat{\gamma}_k|_2^2$ under the null by subtracting $\bar{c} = \mathbb{E} |V_{\hat{\tau}_k} - \mathbb{E} V_{\hat{\tau}_k}|_2^2$. This guarantees that the break sizes can be estimated consistently. We have $\log(n)$ in the consistency rates in (ii) and (iii) since we take the maximum overall $n - 2bn$ moving windows. It is also worth noticing that, for single change-point detection, [4] achieves $|\hat{\tau}_k - \tau_k| = O_{\mathbb{P}}(1)$ by assuming that δ_p is a constant. Under this condition, by our Theorem 2.2(ii), we can achieve the same consistency rate up to a logarithm factor. Indeed, δ_p does not need to be a constant for consistent estimators (of a single or finitely many breaks) with an order of $O_{\mathbb{P}}(1)$; see, for instance, [25]. Intuitively, a larger δ_p makes the estimation problem easier, and in the same example mentioned earlier, where the break size for each series is denoted as $\vartheta \in \mathbb{R}$, one can see that our δ_p can diverge as p grows, leading to improved consistency rates for $\hat{\tau}_k$.

3. Testing and estimation via a Two-Way MOSUM. The previous section focuses on change-point statistics for data generating processes with dense breaks across all component series. In this section, we propose a novel Two-Way MOSUM to address cases where change points may exist in only a subset of time series. In such case, aggregating all component series using the ℓ^2 -norm dilutes the testing power due to the overwhelming aggregated noises compared to signals. To handle this issue, we construct *temporal-spatial windows* (cf. Definition 3.3) to aggregate series within spatial neighborhoods and maximize over neighborhoods and time. Our method allows for the estimation of both the temporal break stamp and the spatial neighborhood that contains the breaks. We extend the Two-Way MOSUM to nonlinear processes and a general spatial space in Section 4.

3.1. Two-Way MOSUM. As [52] suggested, in certain applications, data streams could represent observations from multiple sensors, where breaks only occur in some but not all of them. However, testing procedures aggregating all sensors include noise from unaffected sensors in detection statistics, leading to poor performance. To address this problem, we introduce cross-sectional neighborhoods comprised of spatially adjacent series, and propose the Two-Way MOSUM to account for spatial group structure and detect existence of breaks. This statistic improves test performance for signals dense within clusters but sparse among them. We emphasize that the Two-Way MOSUM can work even in the absence of a prior clustering information. See the discussion below Definition 3.3 for detailed reasoning after introducing the temporal-spatial moving window.

To define the relative spatial locations of component series, we consider a spatial space $\mathcal{L}_0 = \mathcal{L}_{0,n}$ for all series. In particular, we start with a simple case in this section by assuming a linear ordering in the coordinates, that is, $\mathcal{L}_0 \subset \{1, \dots, p\}$. Then in Section 4, we extend it to a more general space with $\mathcal{L}_0 \subset \mathbb{Z}^v$ for a general v , where the spatial location of each series shall be determined by a v -dimensional vector. We define the linear spatial neighborhood and the corresponding neighborhood-norm as follows.

DEFINITION 3.1 (Linear spatial neighborhood). Let $\mathcal{L}_s \subset \{1, \dots, p\}$ be the set of coordinates in a spatial neighborhood, $1 \leq s \leq S$, where S is the total number of spatial neighborhoods. We denote the size of each $\mathcal{L}_s$ by $|\mathcal{L}_s|$. In particular, we define

$$(3.1) \quad |\mathcal{L}_{\max}| = \max_{1 \leq s \leq S} |\mathcal{L}_s| \quad \text{and} \quad |\mathcal{L}_{\min}| = \min_{1 \leq s \leq S} |\mathcal{L}_s|.$$

DEFINITION 3.2 (Linear nbd-norm). For a p -dimensional vector $v_i = (v_{i1}, \dots, v_{ip})^\top$ with a linear ordering in coordinates, we define the linear neighborhood-norm (nbd-norm) as $|v_i|_{2,s} = (\sum_{j=1}^p v_{i,j}^2 \mathbf{1}_{j \in \mathcal{L}_s})^{1/2}$, $1 \leq s \leq S$.

It shall be noted that the spatial neighborhoods defined in Definition 3.1 can be overlapped, which means that each component series can belong to multiple different spatial neighborhoods. Therefore, these spatial neighborhoods actually can be viewed as an analogue of the temporal moving windows, but moving in the spatial direction and allowing different window sizes (i.e., neighborhood sizes). All $s = 1, 2, \dots, S$ are only the indices of different spatial groups and do not necessarily reflect the spatial order of these groups. The size of each neighborhood could tend to infinity as $p \rightarrow \infty$.

We highlight that the spatial moving windows can be adapted to different data scenarios. Apart from an identification condition, we do not need more knowledge (e.g., clustered breaks) on the spatial structure of the signals. It is worth noting that similar definitions of neighborhoods are considered in the literature; see, for example, [1–3]. These setups exclusively focus on the topological structure of neighborhood with simple Gaussian or i.i.d. assumptions. Comparably, we are more flexible in modeling spatial-temporal dependency of the data. In fact, our spatial windows can be extended to more complicated shapes depending on the demand of applications as long as Assumption 3.1 is satisfied. Many real-life data streams, such as those in geographical or economic contexts, provide prior knowledge about which spatial groups are likely to include breaks, as demonstrated in the geostatistics data examples in Chapter 4 of [18]. Our definition of the temporal-spatial window (cf. Definition 3.3) does not require this prior knowledge, but if it is available, it can be utilized for the more relevant detection and estimation procedures.

Our goal is to model the breaks occurring on the vector of unknown trend functions. When there potentially exists a group structure, we formulate the trend function in (2.1) as

$$(3.2) \quad \mu(u) = \mu_0 + \sum_{r=1}^R \gamma_r \mathbf{1}_{u \geq u_r},$$

where $R \in \mathbb{N}$ is an unknown integer represents the number of localized breaks, which could go to infinity as n or S increases; $u_1, \dots, u_R$ are the time stamps of the breaks with $0 = u_0 < u_1 < \dots < u_R < u_{R+1} = 1$ and $\kappa_n = \min_{0 \leq r \leq R} (u_{r+1} - u_r)$, for some constant $\kappa_n > 0$, where κ_n is allowed to tend to zero as $n \rightarrow \infty$; $\gamma_r = (\gamma_{r,1}, \dots, \gamma_{r,p})^\top \in \mathbb{R}^p$ is the jump vector at the time stamp u_r with $\gamma_{r,j} = 0$ if $j \notin \mathcal{L}_{s_r}$, where s_r is the index of the spatial location of the r th break. We define the break size as $|\gamma_r|_2$. In the rest of this paper, we use (τ_r, s_r) to denote the temporal-spatial location of the r th break.

To test the existence of spatially localized breaks, it suffices to test the null hypothesis

$$\mathcal{H}_0^\diamond: \quad \gamma_r = 0, \quad 1 \leq r \leq R,$$

which denotes the case with no breaks, against the alternative that at least one break exists, that is, $\mathcal{H}_A^\diamond$: there exists $r \in \{1, \dots, R\}$, such that $\gamma_r \neq 0$. This enables us to identify both the time stamps and the spatial neighborhoods with significant breaks. Our Two-Way MOSUM statistic aims to adopt more flexible moving windows to efficiently capture both temporal and spatial information of breaks. To achieve this goal, we shall derive a localized test statistic, which first aggregates the time series within each spatial neighborhood by an ℓ^2 -norm and then take the maximum over all the neighborhoods and time points. Accordingly, we define *temporal-spatial windows* as follows.

DEFINITION 3.3 (Temporal-spatial window). For $bn + 1 \leq i \leq n - bn$, $1 \leq s \leq S$, define an index set $\mathcal{V}_{i,s} = \{(t, l) : t = i, l \in \mathcal{L}_s\}$. Then define the temporal-spatial moving window as $\mathcal{S}_{i,s} = \{\mathcal{V}_{t,l} : i - bn \leq t \leq i + bn - 1, l \in \mathcal{L}_s\} \subset \mathbb{Z}^2$.

Note that the index set $\mathcal{V}_{i,s}$ can be regarded as a vertical line, which is the center of the temporal-spatial moving window $\mathcal{S}_{i,s}$. Specifically, $\mathcal{S}_{i,s}$ spans the neighborhood $\mathcal{L}_s$ in the spatial direction and centered at the time point i with radius bn in the temporal direction. In the rest of this paper, we shall depict the index set $\mathcal{V}_{i,s}$ as a vertical line at time i and neighborhood $\mathcal{L}_s$. We refer to Figure 10 in Appendix B.3 for a more straightforward illustration of this Two-Way moving window. It is worth emphasizing that even without prior knowledge of clusters, the number of potential windows is relatively small, because only the adjacent series can be assigned to the same group, which leads to the number of potential windows at most $O(p^2)$. See a more detailed explanation in Appendix B.3 and consider Figure 9 as an example. In general, for any p time series, there are only at most $O(p^{2v})$ possible windows in a $\mathbb{Z}^v$ space, $v \geq 1$. This fact does not diminish the validity of our results obtained through GA (cf. Theorem 3.1). Consequently, while prior knowledge about the clusters is beneficial for boosting power, it is not a prerequisite.

DEFINITION 3.4 (Influenced set). We define the set of vertical lines influenced by the break located at (τ, s) as

$$(3.3) \quad \mathcal{W}_{\tau,s} = \{\mathcal{V}_{t,l} : 1 \leq t \leq n, 1 \leq l \leq S, \mathcal{S}_{t,l} \cap \mathcal{V}_{\tau,s} \neq \emptyset\}.$$

ASSUMPTION 3.1 (Neighborhood size). Assume that $|\mathcal{L}_{\max}|/|\mathcal{L}_{\min}| \leq c$ holds for some constant $c \geq 1$, where $|\mathcal{L}_{\max}|$ and $|\mathcal{L}_{\min}|$ are defined in Definition 3.1.

Assumption 3.1 requires that the sizes of all spatial neighborhoods do not differ too much, which still allows the flexibility of different neighborhood sizes but in a reasonable range. This assumption embraces many interesting cases in practice. For example, according to the geographical locations, the Centers for Disease Control and Prevention (CDC) divides the states in the United States (US) into four regions with similar spatial sizes, which is also taken into consideration in our application (cf. Section 5); based on the patterns of synchronous

activity and communication between different brain regions, [54] identifies and segregates the brain into seven distinct functional networks with similar scales.

Similar to (2.7), we define the centering term of the statistics as

$$(3.4) \quad \bar{c}_s^\diamond = \sum_{j=1}^p c_{s,j}^\diamond \quad \text{where } c_{s,j}^\diamond = c_j \mathbf{1}_{j \in \mathcal{L}_s}.$$

Following the intuitions that we could adopt temporal-spatial moving windows to account for spatially clustered jumps, we formulate our Two-Way MOSUM test statistic as follows:

$$(3.5) \quad \mathcal{Q}_n^\diamond = \max_{1 \leq s \leq S} \max_{bn+1 \leq i \leq n-bn} \frac{1}{\sqrt{|\mathcal{L}_s|}} (|V_i|_{2,s}^2 - \bar{c}_s^\diamond),$$

where the nbd-norm $|\cdot|_{2,s}$ is introduced in Definition 3.2. Recall (2.9) for $x_{i,j}$. Let

$$(3.6) \quad \begin{aligned} x_{i,s,j}^\diamond &= x_{i,j} \mathbf{1}_{j \in \mathcal{L}_s} \quad \text{and} \\ X_j^\diamond &= (x_{bn+1,1,j}^\diamond, \dots, x_{n-bn,1,j}^\diamond, \dots, x_{bn+1,S,j}^\diamond, \dots, x_{n-bn,S,j}^\diamond)^\top. \end{aligned}$$

Then, under the null hypothesis, we can rewrite $\mathcal{Q}_n^\diamond$ into

$$(3.7) \quad \mathcal{Q}_n^\diamond = \max_{1 \leq s \leq S} \max_{bn+1 \leq i \leq n-bn} \frac{1}{\sqrt{|\mathcal{L}_s|}} \sum_{j=1}^p x_{i,s,j}^\diamond.$$

When the time series are cross-sectionally independent, $X_j^\diamond$ are independent, $1 \leq j \leq p$. Hence, by applying the GA to (3.7), we can properly address the dependence resulting from the overlapped temporal-spatial windows. We also note that cluster-based statistics can be seen as a special case of Two-Way MOSUM as in (3.7) when the prior knowledge of clusters is given, which can also be seen as an extension of the statistic introduced by [27] where one essentially gets back to his idea by taking each component as its own cluster.

We introduce the centered Gaussian vector

$$(3.8) \quad \mathcal{Z}^\diamond = (\mathcal{Z}_{bn+1,1}^\diamond, \dots, \mathcal{Z}_{n-bn,1}^\diamond, \dots, \mathcal{Z}_{bn+1,S}^\diamond, \dots, \mathcal{Z}_{n-bn,S}^\diamond)^\top,$$

with the covariance matrix

$$(3.9) \quad \Xi^\diamond = (\Xi_{i,s,i',s'}^\diamond)_{1 \leq i, i' \leq n-2bn, 1 \leq s, s' \leq S}.$$

Recall the covariance matrix Ξ for the Gaussian vector $\mathcal{Z}$ in (2.11). By (3.6) and (3.7), we similarly define

$$(3.10) \quad \Xi_{i,s,i',s'}^\diamond = (|\mathcal{L}_s| |\mathcal{L}_{s'}|)^{-1/2} \mathbf{1}_{j \in \mathcal{L}_s \cap \mathcal{L}_{s'}} \Xi_{i,i'}.$$

We aim to provide the GA theorem under the null with a Two-Way MOSUM applied (cf. Theorem 3.1). This result shall enable us to find the critical value of our proposed Two-Way MOSUM test statistic. We denote each element in $\mathcal{Z}^\diamond$ by $\mathcal{Z}_\varphi^\diamond$, where

$$(3.11) \quad \varphi = (i, s) \in \mathcal{N} \quad \text{for } \mathcal{N} = \{bn+1, \dots, n-bn\} \times \{1, \dots, S\}.$$

By the GA, the distribution of $\max_{\varphi \in \mathcal{N}} \mathcal{Z}_\varphi^\diamond$ shall approximate the one of our test statistic $\mathcal{Q}_n^\diamond$ under the null with large p , that is,

$$(3.12) \quad \mathbb{P}(\mathcal{Q}_n^\diamond \leq u) \approx \mathbb{P}\left(\max_{\varphi \in \mathcal{N}} \mathcal{Z}_\varphi^\diamond \leq u\right).$$

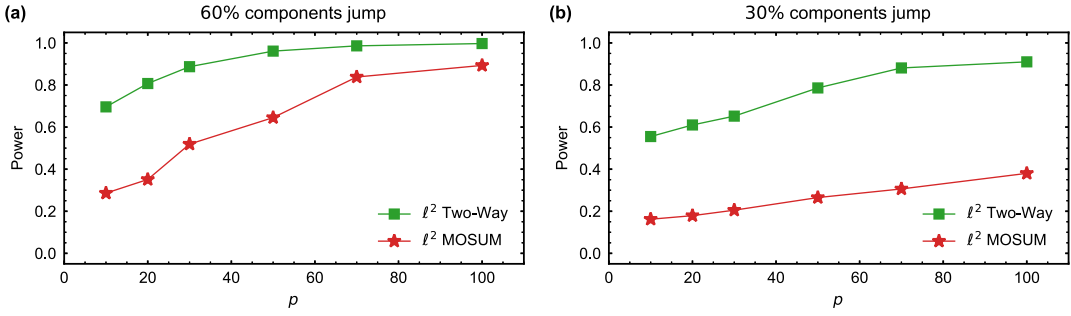


FIG. 2. Power comparison of ℓ^2 MOSUM and Two-Way MOSUM.

REMARK 3.1 (Comparison of ℓ^2 MOSUM and ℓ^2 -clustered Two-Way MOSUM). When the breaks only occur in a portion of time series, aggregation of all p dimensions would cause power loss. This situation is frequently encountered when a spatial group structure exists and only a few groups have breaks. Our proposed Two-Way MOSUM can account for this situation by taking the ℓ^2 -norm within each spatial group. To explicitly show the differences, we present a simulated example with two different proportions of jumps and compare the testing powers in Figure 2. We simulate $n = 100$ observations of $p = 10, 20, 30, 50, 70$ and 100 . The number of spatial groups is $S = 5$ and each group size is $0.3p$. Specifically, we let $\mathcal{L}_s = 2p(s-1)/10 + \{1, 2, \dots, 0.3p\}$, $1 \leq s \leq 4$ and $\mathcal{L}_5 = \{0.7p + 1, 0.7p + 2, \dots, p\}$. In Figure 2(a), two groups $\mathcal{L}_2$ and $\mathcal{L}_5$ contain breaks at the same time $\tau = 50$. In Figure 2(b), breaks only exist in one group $\mathcal{L}_3$ at $\tau = 50$. The errors in both two figures are generated from $\text{MA}(\infty)$ models defined in (2.4) with $\eta_t \sim t_9$ and jump sizes are 0.2 for each dimension. We let the window size $bn = 20$. All the reported powers in Figure 2 are averaged over 1000 samples. We defer a more detailed power comparison to Remark 3.3.

3.2. Gaussian approximation for Two-Way MOSUM. Given the updated statistics for spatially clustered signals, we further formalize our GA theory in this setting. Theoretically, it shall be noted that when signals are dense within the clusters and sparse among clusters, our Two-Way MOSUM improves the one adopted in Section 2.

THEOREM 3.1 (GA for Two-Way MOSUM). Suppose that Assumptions 2.1–2.3 and 3.1 are satisfied. Then, under the null hypothesis, for $\Delta_0^\diamond = (bn)^{-1/3} \log^{2/3}(nS)$,

$$\Delta_1^\diamond = \left(\frac{(nS)^{4/q} \log^7(pn)}{|\mathcal{L}_{\min}|} \right)^{1/6}, \quad \Delta_2^\diamond = \left(\frac{(nS)^{4/q} p^{2/q} \log^3(pn)}{|\mathcal{L}_{\min}|} \right)^{1/3},$$

we have

$$(3.13) \quad \sup_{u \in \mathbb{R}} \left| \mathbb{P}(\mathcal{Q}_n^\diamond \leq u) - \mathbb{P}\left(\max_{\varphi \in \mathcal{N}} \mathcal{Z}_\varphi^\diamond \leq u\right) \right| \lesssim \Delta_0^\diamond + \Delta_1^\diamond + \Delta_2^\diamond,$$

where $\mathcal{N}$ is defined in (3.11), and the constant in $\lesssim$ is independent of n, p, b . If in addition, $\log(nS) = o\{(bn)^{1/2}\}$ and

$$(3.14) \quad (nS)^4 p^2 |\mathcal{L}_{\min}|^{-q} \log^{3q}(pn) \rightarrow 0,$$

then we have

$$(3.15) \quad \sup_{u \in \mathbb{R}} \left| \mathbb{P}(\mathcal{Q}_n^\diamond \leq u) - \mathbb{P}\left(\max_{\varphi \in \mathcal{N}} \mathcal{Z}_\varphi^\diamond \leq u\right) \right| \rightarrow 0.$$

One shall note that the GA results in Theorem 2.1 is a special case of Theorem 3.1. Specifically, when $|\mathcal{L}_{\min}| = p$ and $S = 1$, which indicates that there does not exist any group structure and all p time series belong to the same group, condition (2.15) can be implied by (3.14), and the same convergence rate of GA can be achieved. We extend the above theorem to non-linear processes with general spatial temporal dependency in Theorem 4.1.

REMARK 3.2 (Allowed neighborhood number and size). In Theorem 3.1, we can allow the minimum neighborhood size $|\mathcal{L}_{\min}|$ to be of a polynomial order of the sample size n , and its order depends on the moment parameter q defined in Assumption 2.1. In particular, let $|\mathcal{L}_{\min}| \asymp n^{\nu_2}$, for some $\nu_2 > 0$. Then, if $\nu_2 > 4/(q - 2)$, expression (3.15) holds. The larger the moment parameter q is, the larger minimum group size $|\mathcal{L}_{\min}|$ we can allow. In addition, one shall note that the allowed number of neighborhoods S can be as large as $O(p^2)$, while a bigger S keeps more detailed local structure at the expense of time to inspect more windows.

Next, we consider the alternative hypothesis that there exists at least a break. Since the temporal-spatial moving windows can be overlapped, for the identification of breaks, we pose the following assumption on the separation of break locations.

ASSUMPTION 3.2 (Temporal-spatial separation). For any two breaks located at (τ_r, s_r) and $(\tau_{r'}, s_{r'})$, $1 \leq r \neq r' \leq R$, assume that there does NOT exist any moving window $\mathcal{S}_{\tau, s}$, $bn + 1 \leq \tau \leq n - bn$, $1 \leq s \leq S$, such that $\mathcal{S}_{\tau, s} \cap \mathcal{W}_{\tau_r, s_r} \neq \emptyset$ and $\mathcal{S}_{\tau, s} \cap \mathcal{W}_{\tau_{r'}, s_{r'}} \neq \emptyset$, where $\mathcal{W}(\tau_r, s_r)$ is defined in Definition 3.4.

Assumption 3.2 can be viewed as an analogue of the condition $b \ll \kappa_n$ below Definition 2.1 when we apply a Two-Way MOSUM. To see this, consider the trend function in (2.2). For any two breaks located at τ_k and $\tau_{k'}$, $b \ll \kappa_n$ guarantees that there is no moving window $\mathcal{S}_{\tau, \cdot}$ intersects both $\mathcal{W}_{\tau_k, \cdot}$ and $\mathcal{W}_{\tau_{k'}, \cdot}$, where $\mathcal{S}_{\tau, \cdot}$ (resp., $\mathcal{W}_{\tau, \cdot}$) is the sliding window (resp., influenced set) spanning all p components. This adheres to the separation requirement in Assumption 3.2.

To achieve consistent estimation of the temporal and spatial break locations, for any given significance level $\alpha \in (0, 1)$, we can choose the threshold value $\omega^\diamond$ to be the quantile of the Gaussian limiting distribution indicated by Theorem 3.1, that is,

$$(3.16) \quad \omega^\diamond = \inf_{r \geq 0} \left\{ r : \mathbb{P} \left(\max_{\varphi \in \mathcal{N}} \mathcal{Z}_\varphi^\diamond > r \right) \leq \alpha \right\}.$$

Hence, we shall reject the null hypothesis if $Q_n^\diamond > \omega^\diamond$. To evaluate the power of our localized test, consider the alternative hypothesis that there exists at least a break, that is, $\underline{d} \neq 0$. We provide the power limit of our localized change-point detection in the following corollary.

COROLLARY 3.1 (Power of a Two-Way MOSUM). *Under Assumptions 2.1–2.3, 3.1 and 3.2, if (3.14) holds and*

$$\max_{1 \leq s \leq S} \max_{1 \leq k \leq K} n(u_{k+1} - u_k) |\Lambda^{-1} \gamma_k|_{2, s}^2 \gg \sqrt{|\mathcal{L}_{\min}| \log(nS)},$$

where $|\cdot|_{2, s}$ is defined in Definition 3.2, then the testing power $\mathbb{P}(Q_n^\diamond > \omega^\diamond) \rightarrow 1$.

REMARK 3.3 (Detailed power comparison of ℓ^2 MOSUM and ℓ^2 -clustered Two-Way MOSUM). This comment is complementary to Remark 3.1. Here, we compare the testing power of the MOSUM aggregating all time series and the Two-Way MOSUM. Specifically, we consider a case where p time series belong to S groups and breaks only occur to one group. Suppose that all the series in this group jump with the same size ϑ'' , and we use the (Two-Way) MOSUM with bandwidth b for detection. Then, to ensure the power tending to 1, by Corollary 3.1, $Q_n^\diamond$ only requires $\vartheta'' \gg \log^{1/4}(nS) |\mathcal{L}_{\min}|^{-1/4} (bn)^{-1/2}$, while Q_n needs a stronger condition by Corollary 2.1 that $\vartheta'' \gg (p \log(n))^{1/4} |\mathcal{L}_{\min}|^{-1/2} (bn)^{-1/2}$.

Algorithm 2: ℓ^2 multiple change-point detection via a Two-Way MOSUM

Data: Observations $Y_1, Y_2, \dots, Y_n$; spatial neighborhoods $\mathcal{B}_s, s = 1, \dots, S$; bandwidth parameter b ; threshold value $\omega^\diamond$

Result: Estimated number of breaks $\hat{R}$; estimated break locations $(\hat{\tau}_r, \hat{s}_r), r = 1, \dots, \hat{R}$; estimated jump vectors $\hat{\gamma}_r$; estimated minimum break size $\hat{\delta}_p^\diamond$

$\mathcal{Q}_n^\diamond \leftarrow \max_{1 \leq s \leq S} \max_{bn+1 \leq i \leq n-bn} |\mathcal{B}_s|^{-1/2} (|V_i|_{2,s}^2 - \bar{c}_s^\diamond);$

if $\mathcal{Q}_n^\diamond < \omega^\diamond$ **then**

$\hat{R} = 0$; STOP;

else

$r \leftarrow 1; \mathcal{A}_1^\diamond \leftarrow \{\mathcal{V}_{\tau,s}, bn+1 \leq \tau \leq n-bn, 1 \leq s \leq S : |\mathcal{B}_s|^{-1/2} (|V_\tau|_{2,s}^2 - \bar{c}_s^\diamond) > \omega^\diamond\};$

while $\mathcal{A}_r^\diamond \neq \emptyset$ **do**

$(\hat{\tau}_r, \hat{s}_r) \leftarrow \arg \max_{\mathcal{V}_{\tau,s} \in \mathcal{A}_r^\diamond} |\mathcal{B}_s|^{-1/2} (|V_\tau|_{2,s}^2 - \bar{c}_s^\diamond); \hat{\gamma}_r \leftarrow \hat{\mu}_{\hat{\tau}_r-bn}^{(l)} - \hat{\mu}_{\hat{\tau}_r+bn-1}^{(r)};$

$\mathcal{A}_{r+1}^\diamond \leftarrow \mathcal{A}_r^\diamond \setminus \{\mathcal{V}_{\tau,s} : \text{there exists } bn+1 \leq i \leq n-bn, 1 \leq l \leq S, \text{ such that}$

$\mathcal{S}_{i,l} \cap \mathcal{V}_{\hat{\tau}_r, \hat{s}_r} \neq \emptyset \text{ and } \mathcal{S}_{i,l} \cap \mathcal{V}_{\tau,s} \neq \emptyset\}; r \leftarrow r+1;$

end

$\hat{R} \leftarrow \max_{r \geq 1} \{r : \mathcal{A}_r^\diamond \neq \emptyset\}; \hat{\delta}_p^\diamond \leftarrow \min_{1 \leq r \leq \hat{R}} \|\Lambda^{-1} \hat{\gamma}_r\|_2^2 - \bar{c}^{1/2};$

end

3.3. *Estimation of change points with spatial localization.* Providing the GA for the Two-Way MOSUM statistics, we gather the detailed steps of a change-point estimation procedure in Algorithm 2. Specifically, we extend Algorithm 1 to the cases with cross-sectional localization via Two-Way MOSUM, and we shall expect to obtain spatial locations of change points besides the temporal ones. One follow-up theorem shows the consistency properties of some break statistics in this setup.

We denote the minimum break size over time and spatial neighborhoods by

$$(3.17) \quad \delta_p^\diamond = \min_{1 \leq r \leq R} |\Lambda^{-1} \gamma_r|_2,$$

and assume that this minimum break size is lower bounded as follows.

ASSUMPTION 3.3 (Signal). $\min_{0 \leq r \leq R} n(u_{r+1} - u_r) |\Lambda^{-1} \gamma_r|_2^2 \gg \sqrt{|\mathcal{L}_{\min}| \log(nS)}.$

Let us consider the simple example that within any spatial neighborhood $\mathcal{L}_s, 1 \leq s \leq S$, the jump size of each time series is the same, denoted by $\vartheta \in \mathbb{R}$. Then Assumption 3.3 means $n\kappa_n \vartheta^2 \gg \sqrt{\log(nS)/|\mathcal{L}_{\min}|}$, which is a weaker requirement of the signal strength for each series with breaks similar to Assumption 2.4.

To implement Algorithm 2, by the definition of $\bar{c}_s^\diamond$ in (3.4) and the similar arguments in Remark 2.5, one can take $\bar{c}_s^\diamond = 2|\mathcal{L}_s|/(bn)$, which still ensures the consistency. Also, similar to Algorithm 1, the selection of bandwidth parameter b can follow the suggestions in Remark 2.6, and the long-run variance can be estimated by a robust M-estimation method. The consistency results of the estimated number and temporal-spatial locations of breaks as well as the break sizes are all provided.

PROPOSITION 3.1 (Temporal-spatial consistency). *Let $q \geq 8$. Suppose that Assumptions 2.1–2.3, 3.1–3.3 and condition (3.14) hold. If $|\mathcal{L}_{\min}|^{-1} \delta_p^{\diamond 2} \geq 3\omega^\diamond, \omega^\diamond \gg \log^{1/2}(n)(bn)^{-1}$ and $\max_{1 \leq r \leq R} |\Lambda^{-1} \gamma_r|_q / |\Lambda^{-1} \gamma_r|_2 = O(1/R^{1/q})$, then we have the following results:*

- (i) (Number of breaks). $\mathbb{P}(\hat{R} = R) \rightarrow 1.$

(ii) (*Time stamps of breaks*). $\max_{1 \leq r \leq R} |\hat{\tau}_r - \tau_{r^*}| \cdot |\Lambda^{-1} \gamma_r|_2^2 / (1 + \Phi_r) = O_{\mathbb{P}}\{\log^2(nS)\}$, where $\Phi_r = |\mathcal{L}_{\min}| / (bn |\Lambda^{-1} \gamma_r|_2^2)$ and $r^* = \arg \min_i |(\hat{\tau}_r, \hat{s}_r) - (\tau_i, s_i)|$. If in addition, $\delta_p^{\diamond 2} / |\mathcal{L}_{\min}| \gtrsim 1 / (bn)$, then

$$\max_{1 \leq r \leq R} |\hat{\tau}_r - \tau_{r^*}| \cdot |\Lambda^{-1} \gamma_r|_2^2 / (bn) = O_{\mathbb{P}}\{\log^2(nS)(bn)^{-1}\}.$$

(iii) (*Spatial locations of breaks*). If there exists a constant $c_\gamma > 0$ such that $|\gamma_{r,j}| / |\gamma_{r,j'}| \leq c_\gamma$, for all $1 \leq r \leq R$ and $j, j' \in \mathcal{L}_{s_r}$, then

$$\max_{1 \leq r \leq R} |\mathcal{L}_{\hat{s}_r} \ominus \mathcal{L}_{s_{r^*}}| \cdot |\Lambda^{-1} \gamma_r|_2^2 / |\mathcal{L}_{\min}| = O_{\mathbb{P}}\{\log^2(nS)(bn)^{-1}\},$$

where $\mathcal{L}_{\hat{s}_r} \ominus \mathcal{L}_{s_{r^*}} = (\mathcal{L}_{\hat{s}_r} \setminus \mathcal{L}_{s_{r^*}}) \cup (\mathcal{L}_{s_{r^*}} \setminus \mathcal{L}_{\hat{s}_r})$ and $r^* = \arg \min_i |(\hat{\tau}_r, \hat{s}_r) - (\tau_i, s_i)|$.

(iv) (*Break sizes*). $\max_{1 \leq r \leq R} ||\Lambda^{-1}(\hat{\gamma}_r - \gamma_{r^*})|_2^2 - \bar{c}| = O_{\mathbb{P}}\{(p \log(nS))^{1/2}(bn)^{-1}\}$. This also implies that $|\hat{\delta}_p^\diamond - \delta_p^\diamond| = O_{\mathbb{P}}\{(p \log(nS))^{1/4}(bn)^{-1/2}\}$.

Proposition 3.1(i) indicates the consistency of the estimator for the number of significant breaks; (ii) and (iii) show that we can consistently recover both the spatial break neighborhood $\mathcal{L}_{s_r}$ and the temporal break stamp τ_r ; (iv) suggests that the sizes of break vector γ_r can also be estimated consistently. Note that in Proposition 3.1(ii), $|\hat{\tau}_r - \tau_{r^*}| \cdot |\Lambda^{-1} \gamma_r|_2^2 / (bn)$ indicates the temporal precision, and the spatial precision is represented by $|\mathcal{L}_{\hat{s}_r} \ominus \mathcal{L}_{s_{r^*}}| \cdot |\Lambda^{-1} \gamma_r|_2^2 / |\mathcal{L}_{\min}|$ in (iii). Both results are normalized by their window widths respectively and the two consistency rates are of the same order.

REMARK 3.4 (Comparison of consistency rates with Theorem 2.2). We see that the temporal consistency rate of $|\hat{\tau}_r - \tau_{r^*}| \cdot |\Lambda^{-1} \gamma_r|_2^2$ in Proposition 3.1(ii) is similar to that in Theorem 2.2 except for an additional S term in the log factor. For the break size, the convergence rate of $|\hat{\delta}_p^\diamond - \delta_p^\diamond|$ similarly admits an additional S term in the log factor compared to $|\hat{\delta}_p - \delta_p|$ in Theorem 2.2. Both two S terms result from the maximization over all S spatial neighborhoods in the estimators.

4. Nonlinear time series with cross-sectional dependence. In this section, we present three generalizations. First, we expand the linear series given in (2.4) to accommodate a nonlinear scenario (see equation (4.10)) for a broader range of time-series models. Second, we move beyond the linear ordering in coordinates by introducing a more comprehensive space in the spatial dimension, denoted as $\mathcal{L}_0 \subset \mathbb{Z}^v$ (where $v \geq 1$ is a fixed integer). Lastly, we generalize the GA from earlier sections by allowing weak cross-sectional dependence in the underlying error process. We will begin with the definition of the new spatial space and the nonlinear model, proceed with the conditions for both temporal and cross-sectional dependence structures, and ultimately, present our primary theoretical findings and the rationale behind the proof strategy.

4.1. Multidimensional spatial space. To detect breaks in $\mathcal{L}_0$, we shall first provide a generalized notion of spatial window accordingly. In particular, denote $\mathcal{B}_s$, $1 \leq s \leq S$, as spatial neighborhoods, which is a generalization of $\mathcal{L}_s$ in the previous section. Without loss of generality, we focus on hyperrectangles,

$$(4.1) \quad \mathcal{B}_s = I_{s,1} \times I_{s,2} \times \cdots \times I_{s,v},$$

where $I_{s,r} = I_{s,r,n} = [n_{s,r}^-, n_{s,r}^+]$ is some interval on $\mathbb{Z}$ whose end points $n_{s,r}^-$ and $n_{s,r}^+$ can depend on n . Different $\mathcal{B}_s$ are allowed to be overlapped and S can go to infinity as $p \rightarrow \infty$. We define $I_r = I_{r,n} = [\min_s n_{s,r}^-, \max_s n_{s,r}^+]$ and let $\mathcal{B}_0 = I_1 \times I_2 \times \cdots \times I_v$, which implies

$$(4.2) \quad \bigcup_{1 \leq s \leq S} \mathcal{B}_s \subset \mathcal{B}_0.$$

Suppose that the total number of locations in $\mathcal{B}_0 \cap \mathcal{L}_0$ denoted as $|\mathcal{B}_0 \cap \mathcal{L}_0|$ is $p = p_n$, which can go to infinity as n increases. We consider the time-series model

$$(4.3) \quad Y_t(\ell) = \mu_\ell(t/n) + \epsilon_t(\ell), \quad t = 1, \dots, n, \ell \in \mathcal{B}_0 \cap \mathcal{L}_0.$$

Our main objective is to identify possible change points in the trend functions

$$(4.4) \quad \mu_\ell(u) = \mu_{\ell,0} + \sum_{k=1}^K \gamma_{k,\ell} \mathbf{1}_{\{u \geq u_k, \ell \in \mathcal{B}_{s_k} \cap \mathcal{L}_0\}},$$

where K and $u_1, \dots, u_K$ are defined similar to those in (2.2); $\mu_{\ell,0} \in \mathbb{R}$ represents the benchmark level when no break occurs, and $\gamma_{k,\ell} \in \mathbb{R}$ denotes the jump at time point u_k and location ℓ in the neighborhood $\mathcal{B}_{s_k}$, the k th spatial neighborhood containing breaks.

We then introduce definitions to characterize the mass and volume of the spatial neighborhood $\mathcal{B}_s$. By working with the spatial location ℓ , we can bypass the linear ordering presented in Section 3. This notion of spatial location is similar to the general definition of spatial change-points in a v -dimensional spatial lattice used in studies such as [57] and [36].

DEFINITION 4.1 (Spatial neighborhood). (i) (Mass). Define the mass of the spatial neighborhood $\mathcal{B}_s$ by the number of series in $\mathcal{B}_s$, that is, $|\mathcal{B}_s \cap \mathcal{L}_0|$, where $|\cdot|$ is the number of elements in a Borel set. Denoted by $B_{\min}$ and $B_{\max}$ the sizes of the smallest and biggest spatial neighborhoods, respectively, that is,

$$B_{\min} = \min_{1 \leq s \leq S} |\mathcal{B}_s \cap \mathcal{L}_0|, \quad B_{\max} = \max_{1 \leq s \leq S} |\mathcal{B}_s \cap \mathcal{L}_0|,$$

which satisfy $B_{\max}/B_{\min} \leq c$, for some constant $c \geq 1$. (ii) (Volume). Define the volume of the spatial neighborhood $\mathcal{B}_s$ as $\lambda(\mathcal{B}_s)$, where $\lambda(\cdot)$ is the Lebesgue measure of a Borel set.

Following [37], we introduce the following density assumption on the spatial space $\mathcal{L}_0$ that makes it possible to extend the asymptotic properties in regular space in $\mathbb{Z}^v$ to ones in irregular space.

ASSUMPTION 4.1 (Density of spatial space $\mathcal{L}_0$). Let ℓ_j , $j = 1, \dots, p$, be the spatial locations in $\mathcal{L}_0 \subset \mathbb{Z}^v$ on which $Y_t(\ell_j)$ is observed, $t = 1, \dots, n$. Assume that each ℓ_j can be written as $\ell_j = (A_1 u_{j,1}, \dots, A_v u_{j,v})^\top$. Here, $\mathbf{u}_j = (u_{j,1}, \dots, u_{j,v})^\top$ is a sequence of i.i.d. random vectors with a density function $g(\mathbf{x})$ with a compact support in $[0, 1]^v$. We assume that $A_r \rightarrow \infty$ as $p \rightarrow \infty$, for all $r = 1, \dots, v$. Also, for all $\mathbf{x} \in [0, 1]^v$, $c_1 \leq g(\mathbf{x}) \leq c_2$, for some constants $c_1, c_2 > 0$.

Here, we only require the density function $g(\mathbf{x})$ to be uniformly bounded from both sides, which is a weaker condition compared to Assumption 1 in [37], where they aim to perform Fourier analysis for irregularly spaced data on $\mathbb{R}^v$ and more restrict assumptions such as the existence of higher-order derivatives of $g(\mathbf{x})$ are therefore desired. Differently, our goal is to perform block approximation in the spatial direction to deal with the cross-sectional dependence (cf. Remark C.1), which in fact only requires that, for any hyperrectangle $\mathcal{A} \subset \mathbb{Z}^d$ satisfying $\lambda(\mathcal{A}) \rightarrow \infty$, there exist constants $c_1, c_2 > 0$ such that $c_1 \leq |\mathcal{A} \cap \mathcal{L}_0|/\lambda(\mathcal{A}) \leq c_2$. One can view this condition as a special case of Assumption 4.1. Also, when it breaks down to a simple space with linear ordering, the linear spatial neighborhood $\mathcal{L}_s$ defined in Definition 3.1 can be represented by $\mathcal{L}_s = \mathcal{B}_s \cap \mathcal{L}_0$ and we have $|\mathcal{L}_s| = |\mathcal{B}_s \cap \mathcal{L}_0| = \lambda(\mathcal{B}_s)$, $1 \leq s \leq S$, that is, $g(\mathbf{x}) \equiv 1$ for all $\mathbf{x} \in [0, 1]^v$ with $v = 1$. Concerning the shape of spatial neighborhoods, we pose the following assumption to eliminate the degenerate case, which holds little relevance in the context of spatial statistics.

ASSUMPTION 4.2 (Neighborhood shape). There exists a constant $c \geq 1$, such that for each neighborhood $\mathcal{B}_s$, $\max_{1 \leq r \leq v} (n_{s,r}^+ - n_{s,r}^-) \leq c \min_{1 \leq r \leq v} (n_{s,r}^+ - n_{s,r}^-)$.

It is worth noting that our approach is not limited to hyperrectangles, provided that Assumptions 4.1 and 4.2 are satisfied. In this context, we can have a general concept of aggregation for our statistics within the spatial space.

4.2. *Generalized Two-Way MOSUM.* We now introduce a generalized Two-Way MOSUM, designed to accommodate the multidimensional spatial space constructed in the previous section. Let $\epsilon_t = (\epsilon_t(\ell))_{\ell \in \mathcal{B}_0 \cap \mathcal{L}_0}^\top$, $t \in \mathbb{Z}$. We denote the long-run covariance matrix of $\{\epsilon_t\}_{t \in \mathbb{Z}}$ and the corresponding diagonal matrix by

$$(4.5) \quad \Sigma = (\sigma(\ell_1, \ell_2))_{\ell_1, \ell_2 \in \mathcal{B}_0 \cap \mathcal{L}_0} \quad \text{and} \quad \Lambda = \text{diag}(\sigma(\ell))_{\ell \in \mathcal{B}_0 \cap \mathcal{L}_0},$$

respectively, where $\sigma(\ell) = \sigma(\ell, \ell)$ representing the long-run variance of the component $\epsilon_t(\ell)$. To test for the existence of breaks, we denote $\hat{\mu}_i^{(l)}(\ell) = \sum_{t=i-bn}^{i-1} Y_t(\ell)/(bn)$, $\hat{\mu}_i^{(r)}(\ell) = \sum_{t=i}^{i+bn-1} Y_t(\ell)/(bn)$, and evaluate a jump statistic defined by

$$(4.6) \quad V_i(\ell) = \sigma^{-1}(\ell)(\hat{\mu}_i^{(l)}(\ell) - \hat{\mu}_i^{(r)}(\ell)).$$

Note that $V_i(\ell)$ comprises the signal part $\mathbb{E}[V_i(\ell)]$ and the noise part $V_i(\ell) - \mathbb{E}[V_i(\ell)]$. Under the null hypothesis, where no break exists and $\mathbb{E}[V_i(\ell)] = 0$, we define the centering term of the ℓ^2 -aggregation of $V_i(\ell)$ within the neighborhood $\mathcal{B}_s$ as

$$(4.7) \quad c_{\mathcal{B}_s} = \sum_{\ell \in \mathcal{B}_s \cap \mathcal{L}_0} c(\ell) \quad \text{where } c(\ell) = \text{Var}[V_i(\ell)].$$

Subsequently, we propose the following test statistic:

$$(4.8) \quad \tilde{Q}_n = \max_{1 \leq s \leq S} \max_{bn+1 \leq i \leq n-bn} Q_{i, \mathcal{B}_s} \quad \text{where } Q_{i, \mathcal{B}_s} = \frac{1}{\sqrt{|\mathcal{B}_s \cap \mathcal{L}_0|}} \left(\sum_{\ell \in \mathcal{B}_s \cap \mathcal{L}_0} V_i^2(\ell) - c_{\mathcal{B}_s} \right).$$

Under the null hypothesis $\mathcal{H}_0$, since $\mathbb{E}[V_i(\ell)] = 0$, we can rewrite $Q_{i, \mathcal{B}_s}$ into

$$(4.9) \quad Q_{i, \mathcal{B}_s} = \frac{1}{\sqrt{|\mathcal{B}_s \cap \mathcal{L}_0|}} \sum_{\ell \in \mathcal{B}_s \cap \mathcal{L}_0} x_i(\ell) \quad \text{where } x_i(\ell) = (V_i(\ell) - \mathbb{E}[V_i(\ell)])^2 - c(\ell).$$

It shall be noted that when $v = 1$, the test statistic $\tilde{Q}_n$ reduces to $Q_n^\diamond$ in Section 3.

4.3. *Dependence structure.* Although it is quite convenient to assume that the errors are cross-sectionally i.i.d., it is unrealistic to ignore the spatial dependence. The assumption on cross-sectional independence in previous sections can be relaxed accordingly to allow for a weak spatial dependence case. In this section, we extend the GA in Section 3 to the cases where the underlying errors are allowed to be cross-sectionally weakly dependent. This will allow us to evaluate the critical values of the test statistics $\tilde{Q}_n$ accordingly.

Suppose that the stationary noise process $\{\epsilon_t(\ell)\}_{t \in \mathbb{Z}}$ in (2.1) is of the form:

$$(4.10) \quad \epsilon_t(\ell) = f(\eta_{t-k, \ell-\ell'}; k \geq 0, \ell' \in \mathbb{Z}^v).$$

Here, $\eta_{i,s}$, $i \in \mathbb{Z}$, $s \in \mathbb{Z}^v$ are i.i.d. random variables, and $f(\cdot)$ is an $\mathbb{R}$ -valued measurable function such that $\epsilon_t(\ell)$ is well-defined. We assume throughout the paper that $\mathbb{E}[\epsilon_t(\ell)] = 0$ and $\max_{\ell \in \mathcal{B}_0 \cap \mathcal{L}_0} \|\epsilon_t(\ell)\|_q < \infty$, for some $q \geq 8$. Next, we introduce the functional dependence measures to characterize the temporal and spatial dependence structure of $\epsilon_t(\ell)$. Let

$(\eta'_{i,s})_{i \in \mathbb{Z}, s \in \mathbb{Z}^v}$ be an i.i.d. copy of $(\eta_{i,s})_{i \in \mathbb{Z}, s \in \mathbb{Z}^v}$. Specifically, we consider the *temporal* and *temporal-spatial* coupled versions of $\epsilon_t(\ell)$ defined respectively by

$$\epsilon_t^*(\ell) = f(\eta_{t-k, \ell-\ell'}^*; k \geq 0, \ell' \in \mathbb{Z}^v) \quad \text{and} \quad \epsilon_t^{**}(\ell) = f(\eta_{t-k, \ell-\ell'}^{**}; k \geq 0, \ell' \in \mathbb{Z}^v),$$

where for any $i \geq 0$ and $s \in \mathbb{Z}^v$,

$$\eta_{i,s}^* = \begin{cases} \eta_{i,s} & \text{if } i \neq 0, \\ \eta'_{i,s} & \text{if } i = 0 \end{cases} \quad \text{and} \quad \eta_{i,s}^{**} = \begin{cases} \eta_{i,s} & \text{if } i \neq 0 \text{ and } s \neq 0, \\ \eta'_{i,s} & \text{if } i = 0 \text{ or } s = 0. \end{cases}$$

Following [50], we generalize the functional dependence measures as follows:

$$(4.11) \quad \theta_{t,\ell,q} = \|\epsilon_t(\ell) - \epsilon_t^*(\ell)\|_q, \quad \delta_{t,\ell,q} = \|\epsilon_t(\ell) - \epsilon_t^{**}(\ell)\|_q.$$

Note that $\theta_{t,\ell,q}$ represents the change measure of dependence by perturbing solely in the temporal direction, while $\delta_{t,\ell,q}$ denotes the counterpart, which perturbs in both the temporal and spatial directions.

To account for the temporal and cross-sectional dependence structure of $\{\epsilon_t(\ell)\}_{t \in \mathbb{Z}}$, we shall impose the following assumptions on $\theta_{t,\ell,q}$ and $\delta_{t,\ell,q}$. The assumptions essentially require the algebraic decay of dependence both in the temporal and the spatial directions and are controlling the tail behavior of the noise terms.

ASSUMPTION 4.3 (Finite moment). Assume $\max_{\ell \in \mathcal{B}_0 \cap \mathcal{L}_0} \|\epsilon_t(\ell)\|_q < \infty$, for $q \geq 8$.

ASSUMPTION 4.4 (Temporal dependence). There exist some constants $C > 0$ and $\beta > 0$, such that, for all $h \geq 0$, $\max_{\ell \in \mathcal{B}_0 \cap \mathcal{L}_0} \sum_{k \geq h} \theta_{k,\ell,q} / \sigma(\ell) \leq C(1 \vee h)^{-\beta}$, for $q \geq 8$.

ASSUMPTION 4.5 (Weak cross-sectional dependence). Let $q \geq 8$. Assume that there exist some constants $C' > 0$ and $\xi > 1$, such that, for all $m \geq 0$,

$$(4.12) \quad \sum_{k \geq 0} \left(\sum_{\{\ell \in \mathcal{B}_0 \cap \mathcal{L}_0 : |\ell|_2 \geq m\}} \delta_{k,\ell,q}^2 / \sigma^2(\ell) \right)^{1/2} \leq C'(1 \vee m)^{-\xi}.$$

It shall be noted that we can also switch the temporal and spatial aggregation in the above assumption. Specifically, (4.12) also implies

$$(4.13) \quad \left(\sum_{\{\ell \in \mathcal{B}_0 \cap \mathcal{L}_0 : |\ell|_2 \geq m\}} \Delta_{0,\ell,q}^2 / \sigma^2(\ell) \right)^{1/2} \leq C'(1 \vee m)^{-\xi} \quad \text{where } \Delta_{0,\ell,q} = \sum_{k \geq 0} \delta_{k,\ell,q}.$$

To see this, for any $n \in \mathbb{N}$, we denote the partial sum $\Delta_{n,\ell,q} = \sum_{k=0}^n \delta_{k,\ell,q}$. Let $\underline{\Delta}_{n,q} = (\Delta_{n,\ell,q})_{\{\ell \in \mathbb{Z}^v : |\ell|_2 \geq m\}}^\top$, and $\underline{\delta}_{k,q} = (\delta_{k,\ell,q})_{\{\ell \in \mathbb{Z}^v : |\ell|_2 \geq m\}}^\top$. By the triangle inequality, $|\underline{\Delta}_{n,q}|_2 = |\sum_{k=0}^n \underline{\delta}_{k,q}|_2 \leq \sum_{k=0}^n |\underline{\delta}_{k,q}|_2$. Since $\sum_{k \geq 0} |\underline{\delta}_{k,q}|_2 < \infty$, we let $n \rightarrow \infty$ and achieve the desired result. Furthermore, (4.13) also indicates the decay of cross-sectional long-run correlation as depicted in Lemma 4.1.

LEMMA 4.1 (Decay of long-run correlation). Assume that condition (4.13) holds. Then, for any $\ell_1, \ell_2 \in \mathcal{B}_0 \cap \mathcal{L}_0$, the long-run correlation between $\epsilon_t(\ell_1)$ and $\epsilon_t(\ell_2)$, denoted by $\tilde{\rho}(\ell_1, \ell_2)$, decays at a polynomial rate as $|\ell_1 - \ell_2|_2$ increases, that is,

$$(4.14) \quad \tilde{\rho}(\ell_1, \ell_2) = \sigma(\ell_1, \ell_2) / (\sigma(\ell_1)\sigma(\ell_2)) = O\{| \ell_1 - \ell_2 |_2^{-2\xi}\}.$$

It is worth noticing that the decay assumption regarding spatial correlation is widely prevalent in spatial statistics. See, for example, [43] and [40] follow a similar pattern that the covariance between variables decreases as their corresponding distance in the input space increases.

4.4. *Gaussian approximation with weak cross-sectional dependence.* This subsection is devoted to the GA in the high-dimensional setting under the above mentioned general framework. In the case when p is fixed, we provide an invariance principle in Appendix B.2. When p grows to infinity as n increases, to derive the limiting distribution of the test statistic $\tilde{Q}_n$ under the null, we introduce the centered Gaussian random vector

$$(4.15) \quad \tilde{Z} = (\tilde{Z}_{bn+1,1}, \dots, \tilde{Z}_{n-bn,1}, \dots, \tilde{Z}_{bn+1,S}, \dots, \tilde{Z}_{n-bn,S})^\top,$$

with the covariance matrix

$$(4.16) \quad \tilde{\Sigma} = (\tilde{\Sigma}_{i,s,i',s'})_{1 \leq i, i' \leq n-2bn, 1 \leq s, s' \leq S}.$$

Let $\pi_{s,s',\ell_1,\ell_2} = (|\mathcal{B}_s \cap \mathcal{L}_0| |\mathcal{B}_{s'} \cap \mathcal{L}_0|)^{-1/2} \mathbf{1}_{\ell_1, \ell_2 \in \mathcal{B}_s \cap \mathcal{B}_{s'} \cap \mathcal{L}_0}$, and $\tilde{\Sigma}_{i,s,i+\zeta bn,s'}$ equals to

$$(4.17) \quad (bn)^{-2} \sum_{\ell_1, \ell_2 \in \mathcal{B}_0 \cap \mathcal{L}_0} \pi_{s,s',\ell_1,\ell_2} \begin{cases} (15\zeta^2 - 20\zeta + 8)\tilde{\rho}^2(\ell_1, \ell_2) + 3\zeta^2 - 4\zeta, & 0 < \zeta \leq 1, \\ (3\zeta^2 - 12\zeta + 12)\tilde{\rho}^2(\ell_1, \ell_2) - \zeta^2 + 4\zeta - 4, & 1 < \zeta \leq 2, \\ 0, & \zeta > 2. \end{cases}$$

We defer the detailed evaluation of (4.17) to Lemma C.13. Note that, if for all $\ell, \ell_1, \ell_2 \in \mathcal{B}_0 \cap \mathcal{L}_0$, $\tilde{\rho}(\ell, \ell) = 1$ and $\tilde{\rho}(\ell_1, \ell_2) = 0$, $\ell_1 \neq \ell_2$, which denotes the case with no spatial dependence, then (4.17) is the same as (3.10). Recall $\mathcal{N}$ defined in (3.11). We denote each element in $\tilde{Z}$ by $\tilde{Z}_\varphi$, where $\varphi = (i, s) \in \mathcal{N}$. Similar to the cross-sectionally independent case, we can approximate the limiting distribution of $\tilde{Q}_n$ under the null by the one of $\max_\varphi \tilde{Z}_\varphi$. We refer to Remark C.1 in Appendix C for the proof strategies based on the block approximation.

THEOREM 4.1 (GA with weak cross-sectional dependence). *Suppose that Assumptions 4.1–4.5 hold. Then, under the null hypothesis, for $\tilde{\Delta}_0 = (bn)^{-1/3} \log^{2/3}(nS)$, $\tilde{\Delta}_1 = c_{p,n}^{-(q-4)/(3q)}$, $\tilde{\Delta}_2 = c_{p,n}^{-1/(8v)} \log(pn)$, where*

$$(4.18) \quad c_{p,n} = p^{\frac{-2}{q-4}} B_{\min}^{\frac{q}{q-4}} (nS)^{-\left(\frac{4}{q-4} + \frac{2v}{q\xi}\right)} (\log(pn))^{-\left(\frac{(2+q)v}{2q\xi} + \frac{3q}{q-4}\right)},$$

we have

$$\sup_{u \in \mathbb{R}} \left| \mathbb{P}(\tilde{Q}_n \leq u) - \mathbb{P}\left(\max_{\varphi \in \mathcal{N}} \tilde{Z}_\varphi \leq u\right) \right| \lesssim \tilde{\Delta}_0 + \tilde{\Delta}_1 + \tilde{\Delta}_2,$$

where $\mathcal{N}$ is defined in (3.11), and the constant in $\lesssim$ is independent of n , p , b and S . If in addition, $\log(nS) = o\{(bn)^{1/2}\}$ and $\log^{8v}(pn) = o(c_{p,n})$, then

$$\sup_{u \in \mathbb{R}} \left| \mathbb{P}(\tilde{Q}_n \leq u) - \mathbb{P}\left(\max_{\varphi \in \mathcal{N}} \tilde{Z}_\varphi \leq u\right) \right| \rightarrow 0.$$

REMARK 4.1 (Comparison with Theorem 3.1). Note that when $v = 1$, we have $|\mathcal{L}_{\min}| = B_{\min}$, and if $\xi \rightarrow \infty$, it indicates the cross-sectional independence setting. Hence, Theorem 3.1 can be viewed as a special case of Theorem 4.1. Specifically, $c_{p,n} \rightarrow \infty$ in (4.18) boils down to the condition (3.14), which implies $c_{p,n}^{q-4} = p^{-2} B_{\min}^q (nS)^{-4} \log^{-3q}(pn) \rightarrow \infty$, and we can achieve the same approximation rate up to a logarithm term.

Moving on to the alternative hypothesis, we can set the detection threshold $\tilde{\omega}$ as the critical value of $\max_{\varphi \in \mathcal{N}} \tilde{Z}_\varphi$ determined by the Gaussian limiting distribution presented in Theorem 4.1. Specifically, we set $\tilde{\omega}$ as $\inf_{r \geq 0} \{r : \mathbb{P}(\max_{\varphi \in \mathcal{N}} \tilde{Z}_\varphi > r) \leq \alpha\}$, for significant level $\alpha \in (0, 1)$. We reject the null hypothesis if $\tilde{Q}_n > \tilde{\omega}$. For any time point i that satisfies $|i - \tau_k| \leq bn$, we define the weighted break as $d_i(\ell) = (1 - |i - \tau_k|/(bn))\sigma^{-1}(\ell)\gamma_{k,\ell}$. Under the alternative, where there exists i and ℓ such that $d_i(\ell) \neq 0$, we refer to Corollary B.2 in Appendix B.5 for the power limit of our test. Further algorithm for detecting and identifying breaks can be developed accordingly.

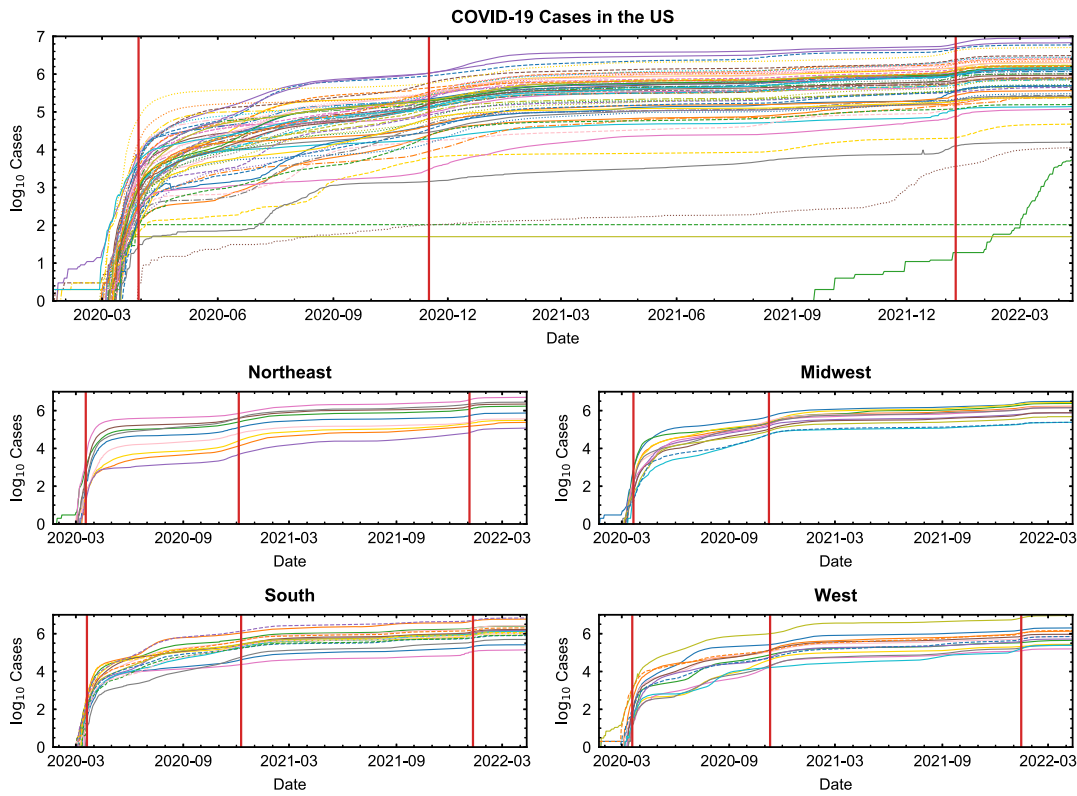


FIG. 3. Top: Algorithm 1 detected three change points which are 2020.03.30, 2020.11.16 and 2022.01.09. Bottom: Algorithm 2 found different change points for four regions: Northeast (2020.03.18, 2020.12.05, 2022.01.04); Midwest (2020.03.21, 2020.11.08); South (2020.03.20, 2020.12.09, 2022.01.10); West (2020.3.19, 2020.11.10, 2022.01.14).

5. Application. This section is devoted to the real-data analysis to illustrate our proposed method for multiple change-point detection. We apply Algorithm 1 to a stock-return data set and use Algorithms 1 and 2 to a COVID-19 data set. Due to the space limit, we defer the results of the stock-return data to Appendix A.4.

Analyzing 812 days of daily COVID-19 case numbers in the US, we identified three significant breaks (Figure 3 top): March 2020 (first outbreak), October 2020 (Delta variant) and December 2021 (Omicron variant). Further, we consider four geographic regions of the US as per the guidelines of the CDC: Northeast, Midwest, South and West. A map of these four regions is available in Figure 8 in Appendix A.5. By our algorithm, each region exhibited different break time stamps, with the Northeast and West experiencing early outbreaks due to major international airports. The Midwest was the first to encounter the Delta variant, while the Northeast initially faced the Omicron variant. Our detection algorithm effectively captured these variations (Figure 3 bottom), demonstrating the efficacy of our proposed testing procedures in identifying breaks over time and across diverse locations. For more detailed information, please refer to Appendix A.5.

Acknowledgments. We express our gratitude to the Co-Editor, the Associate Editor and the reviewers for their insightful comments, which enhanced the quality of this paper.

Funding. Likai Chen’s research is partially supported by the NSF (Grant NSF-2222403). Weining Wang’s research is partially supported by the ESRC (Grant ES/T01573X/1).

Wei Wu's research is partially supported by the NSF (Grants NSF/DMS-1916351, NSF/DMS-2311249, NSF/DMS-2027723).

SUPPLEMENTARY MATERIAL

Supplement to “ ℓ^2 inference for change points in high-dimensional time series via a Two-Way MOSUM” (DOI: [10.1214/24-AOS2360SUPP](https://doi.org/10.1214/24-AOS2360SUPP); .pdf). The Supplementary Material [34] contains all the proofs. A detailed analysis of simulation study and additional information of applications are also presented.

REFERENCES

- [1] ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. [MR2722464 https://doi.org/10.1214/10-AOS817](https://doi.org/10.1214/10-AOS817)
- [2] ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. [MR2797847 https://doi.org/10.1214/10-AOS839](https://doi.org/10.1214/10-AOS839)
- [3] ARIAS-CASTRO, E., CANDÈS, E. J., HELGASON, H. and ZEITOUNI, O. (2008). Searching for a trail of evidence in a maze. *Ann. Statist.* **36** 1726–1757. [MR2435454 https://doi.org/10.1214/07-AOS526](https://doi.org/10.1214/07-AOS526)
- [4] BAI, J. (2010). Common breaks in means and variances for panel data. *J. Econometrics* **157** 78–92. [MR2652280 https://doi.org/10.1016/j.jeconom.2009.10.020](https://doi.org/10.1016/j.jeconom.2009.10.020)
- [5] BAI, J., HAN, X. and SHI, Y. (2020). Estimation and inference of change points in high-dimensional factor models. *J. Econometrics* **219** 66–100. [MR4152786 https://doi.org/10.1016/j.jeconom.2019.08.013](https://doi.org/10.1016/j.jeconom.2019.08.013)
- [6] BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. [MR1616121 https://doi.org/10.2307/2998540](https://doi.org/10.2307/2998540)
- [7] BARNETT, I. and ONNELA, J.-P. (2016). Change point detection in correlation networks. *Sci. Rep.* **6** 18893. <https://doi.org/10.1038/srep18893>
- [8] CHAN, J., HORVÁTH, L. and HUŠKOVÁ, M. (2013). Darling–Erdős limit results for change-point detection in panel data. *J. Statist. Plann. Inference* **143** 955–970. [MR3011306 https://doi.org/10.1016/j.jspi.2012.11.004](https://doi.org/10.1016/j.jspi.2012.11.004)
- [9] CHEN, C. Y.-H., OKHRIN, Y. and WANG, T. (2022). Monitoring network changes in social media. *J. Bus. Econom. Statist.* To appear. <https://doi.org/10.1080/07350015.2021.2016425>
- [10] CHEN, L., WANG, W. and WU, W. B. (2021). Dynamic semiparametric factor model with structural breaks. *J. Bus. Econom. Statist.* **39** 757–771. [MR4272933 https://doi.org/10.1080/07350015.2020.1730857](https://doi.org/10.1080/07350015.2020.1730857)
- [11] CHEN, L., WANG, W. and WU, W. B. (2022). Inference of breakpoints in high-dimensional time series. *J. Amer. Statist. Assoc.* **117** 1951–1963. [MR4528482 https://doi.org/10.1080/01621459.2021.1893178](https://doi.org/10.1080/01621459.2021.1893178)
- [12] CHEN, Y., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional, multiscale online changepoint detection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 234–266. [MR4400396 https://doi.org/10.1111/rssb.12447](https://doi.org/10.1111/rssb.12447)
- [13] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. [MR3693963 https://doi.org/10.1214/16-AOP1113](https://doi.org/10.1214/16-AOP1113)
- [14] CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* **33** 2374–2425. [MR4583674 https://doi.org/10.1214/22-aap1870](https://doi.org/10.1214/22-aap1870)
- [15] CHO, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electron. J. Stat.* **10** 2000–2038. [MR3522667 https://doi.org/10.1214/16-EJS1155](https://doi.org/10.1214/16-EJS1155)
- [16] CHO, H. and FRYZLEWICZ, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 475–507. [MR3310536 https://doi.org/10.1111/rssb.12079](https://doi.org/10.1111/rssb.12079)
- [17] CHO, H. and KIRCH, C. (2022). Two-stage data segmentation permitting multiscale change points, heavy tails and dependence. *Ann. Inst. Statist. Math.* **74** 653–684. [MR4444107 https://doi.org/10.1007/s10463-021-00811-5](https://doi.org/10.1007/s10463-021-00811-5)
- [18] CRESSIE, N. A. C. (2015). *Statistics for Spatial Data*. Wiley Classics Library. Wiley, New York. [MR3559472](https://doi.org/10.1002/9781118134463)
- [19] EICHINGER, B. and KIRCH, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* **24** 526–564. [MR3706768 https://doi.org/10.3150/16-BEJ887](https://doi.org/10.3150/16-BEJ887)
- [20] ENIKEEVA, F. and HARCHAOU, Z. (2019). High-dimensional change-point detection under sparse alternatives. *Ann. Statist.* **47** 2051–2079. [MR3953444 https://doi.org/10.1214/18-AOS1740](https://doi.org/10.1214/18-AOS1740)

- [21] ESFAHLANI, F. Z., JO, Y., FASKOWITZ, J., BYRGE, L., KENNEDY, D. P., SPORNS, O. and BETZEL, R. F. (2020). High-amplitude cofluctuations in cortical activity drive functional connectivity. *Proc. Natl. Acad. Sci. USA* **117** 28393–28401.
- [22] FASKOWITZ, J., ESFAHLANI, F. Z., JO, Y., SPORNS, O. and BETZEL, R. F. (2020). Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. *Nat. Neurosci.* **23** 1644–1654.
- [23] FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. [MR3269979 https://doi.org/10.1214/14-AOS1245](https://doi.org/10.1214/14-AOS1245)
- [24] HORVÁTH, L. and HUŠKOVÁ, M. (2012). Change-point detection in panel data. *J. Time Series Anal.* **33** 631–648. [MR2944843 https://doi.org/10.1111/j.1467-9892.2012.00796.x](https://doi.org/10.1111/j.1467-9892.2012.00796.x)
- [25] HORVÁTH, L., HUŠKOVÁ, M., RICE, G. and WANG, J. (2017). Asymptotic properties of the CUSUM estimator for the time of change in linear panel data models. *Econometric Theory* **33** 366–412. [MR3600047 https://doi.org/10.1017/S0266466615000468](https://doi.org/10.1017/S0266466615000468)
- [26] HUŠKOVÁ, M. and SLABÝ, A. (2001). Permutation tests for multiple changes. *Kybernetika (Prague)* **37** 605–622. [MR1877077](https://doi.org/10.1111/j.1467-9892.2012.00796.x)
- [27] JIRAK, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. [MR3405600 https://doi.org/10.1214/15-AOS1347](https://doi.org/10.1214/15-AOS1347)
- [28] KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. [MR3036418 https://doi.org/10.1080/01621459.2012.737745](https://doi.org/10.1080/01621459.2012.737745)
- [29] KIRCH, C. and KLEIN, P. (2023). Moving sum data segmentation for stochastic processes based on invariance. *Statist. Sinica* **33** 873–892. [MR4575326 https://doi.org/10.5705/ss.202021.0048](https://doi.org/10.5705/ss.202021.0048)
- [30] KUCHIBHOTLA, A. K., BROWN, L. D., BUJA, A., GEORGE, E. I. and ZHAO, L. (2023). Uniform-in-submodel bounds for linear regression in a model-free framework. *Econometric Theory* **39** 1202–1248. [MR4678102 https://doi.org/10.1017/s0266466621000219](https://doi.org/10.1017/s0266466621000219)
- [31] LEE, S., SEO, M. H. and SHIN, Y. (2016). The lasso for high dimensional regression with a possible change point. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 193–210. [MR3453652 https://doi.org/10.1111/rssb.12108](https://doi.org/10.1111/rssb.12108)
- [32] LÉVY-LEDUC, C. and ROUEFF, F. (2009). Detection and localization of change-points in high-dimensional network traffic data. *Ann. Appl. Stat.* **3** 637–662. [MR2750676 https://doi.org/10.1214/08-AOAS232](https://doi.org/10.1214/08-AOAS232)
- [33] LI, D., QIAN, J. and SU, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *J. Amer. Statist. Assoc.* **111** 1804–1819. [MR3601737 https://doi.org/10.1080/01621459.2015.1119696](https://doi.org/10.1080/01621459.2015.1119696)
- [34] LI, J., CHEN, L., WANG, W. and WU, W. B. (2024). Supplement to “ ℓ^2 inference for change points in high-dimensional time series via a Two-Way MOSUM.” <https://doi.org/10.1214/24-AOS2360SUPP>
- [35] LIU, B., QI, Z., ZHANG, X. and LIU, Y. (2022). Change point detection for high-dimensional linear models: A general tail-adaptive approach. Preprint. Available at [arXiv:2207.11532](https://arxiv.org/abs/2207.11532).
- [36] MADRID PADILLA, O. H., YU, Y. and RINALDO, A. (2021). Lattice partition recovery with dyadic CART. *Adv. Neural Inf. Process. Syst.* **34** 26143–26155.
- [37] MATSUDA, Y. and YAJIMA, Y. (2009). Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 191–217. [MR2655530 https://doi.org/10.1111/j.1467-9868.2008.00685.x](https://doi.org/10.1111/j.1467-9868.2008.00685.x)
- [38] OLSHEN, A. B., VENKATRAMAN, E. S., LUCITO, R. and WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572. <https://doi.org/10.1093/biostatistics/kxh008>
- [39] ONNELA, J. P., CHAKRABORTI, A., KASKI, K., KERTÉSZ, J. and KANTO, A. (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Phys. Rev. E* **68** 056110.
- [40] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. [MR2514435](https://doi.org/10.1007/978-1-4612-1494-6)
- [41] SCOTT, A. J. and KNOTT, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics* **30** 507–512.
- [42] SHAO, X. (2010). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 343–366. [MR2758116 https://doi.org/10.1111/j.1467-9868.2009.00737.x](https://doi.org/10.1111/j.1467-9868.2009.00737.x)
- [43] STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in Statistics*. Springer, New York. [MR1697409 https://doi.org/10.1007/978-1-4612-1494-6](https://doi.org/10.1007/978-1-4612-1494-6)
- [44] TIBSHIRANI, R. and WANG, P. (2008). Spatial smoothing and hot spot detection for CGH data using the fused lasso. *Biostatistics* **9** 18–29.
- [45] WANG, D. and ZHAO, Z. (2022). Optimal change-point testing for high-dimensional linear models with temporal dependence. Preprint. Available at [arXiv:2205.03880](https://arxiv.org/abs/2205.03880).

- [46] WANG, R. and SHAO, X. (2020). Hypothesis testing for high-dimensional time series via self-normalization. *Ann. Statist.* **48** 2728–2758. [MR4152119](#) <https://doi.org/10.1214/19-AOS1904>
- [47] WANG, R. and SHAO, X. (2023). Dating the break in high-dimensional data. *Bernoulli* **29** 2879–2901. [MR4632124](#) <https://doi.org/10.3150/22-bej1567>
- [48] WANG, R., ZHU, C., VOLGUSHEV, S. and SHAO, X. (2022). Inference for change points in high-dimensional data via selfnormalization. *Ann. Statist.* **50** 781–806. [MR4405366](#) <https://doi.org/10.1214/21-aos2127>
- [49] WANG, T. and SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 57–83. [MR3744712](#) <https://doi.org/10.1111/rssb.12243>
- [50] WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. [MR2172215](#) <https://doi.org/10.1073/pnas.0506715102>
- [51] WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 391–410. [MR2323759](#) <https://doi.org/10.1111/j.1467-9868.2007.00594.x>
- [52] XIE, Y. and SIEGMUND, D. (2013). Sequential multi-sensor change-point detection. *Ann. Statist.* **41** 670–692. [MR3099117](#) <https://doi.org/10.1214/13-AOS1094>
- [53] XU, H., WANG, D., ZHAO, Z. and YU, Y. (2022). Change point inference in high-dimensional regression models under temporal dependence. Preprint. Available at [arXiv:2207.12453](#).
- [54] YEO, B. T. T., KRIENEN, F. M., SEPULCRE, J., SABUNCU, M. R., LASHKARI, D., HOLLINSHEAD, M., ROFFMAN, J. L., SMOLLER, J. W., ZÖLLEI, L. et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106** 1125–1165.
- [55] YU, M. and CHEN, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 247–270. [MR4250275](#) <https://doi.org/10.1111/rssb.12406>
- [56] YU, Y. (2020). A review on minimax rates in change point detection and localisation. Preprint. Available at [arXiv:2011.01857](#).
- [57] YU, Y., MADRID, O. and RINALDO, A. (2022). Optimal partition recovery in general graphs. In *International Conference on Artificial Intelligence and Statistics* 4339–4358. PMLR.
- [58] ZHANG, N. R., SIEGMUND, D. O., JI, H. and LI, J. Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. [MR2672488](#) <https://doi.org/10.1093/biomet/asq025>