Journal of Multivariate Analysis 198 (2023) 105224

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

False discovery rate approach to dynamic change detection

Lilun Du [a,*], Mengtao Wen [b]

[a] Department of Management Sciences, City University of Hong Kong, Hong Kong, China
[b] School of Statistics and Data Science, LPMC and KLMDASR, Nankai University, China

ARTICLE INFO

Article history:
Received 2 August 2022
Received in revised form 14 July 2023
Accepted 14 July 2023
Available online 1 August 2023

AMS 2020 subject classifications:
primary 62H15
secondary 62L10

Keywords:
High-dimensional data streams
Multiple epidemic changes
Multiple testing
Penalized methods
Sequential detection

ABSTRACT

In multiple data stream surveillance, the rapid and sequential identification of individuals whose behavior deviates from the norm has become particularly important. In such applications, the state of a stream can alternate, possibly multiple times, between a null state and an alternative state. To balance the ability to detect two types of changes, that is, a change from the null to the alternative and back to the null, we propose a new multiple testing procedure based on a penalized version of the generalized likelihood ratio test statistics for change detection. The false discovery rate (FDR) at each time point is shown to be controlled under some mild conditions on the dependence structure of data streams. A data-driven approach is developed for selection of the penalization parameter. Its advantage is demonstrated via simulation and a data example.

1. Introduction

Sequential analysis is an important field with general applications to biomedicine, economics, and engineering. Consider the classic case where one receives independent and continuous observations sequentially, and a possible abrupt change may occur in the mean of observations. A common criterion for detecting such a change is minimizing the expected detection delay subject to a constraint on the average run length to a false alarm. Some techniques, including the Shewart chart [32], the CUSUM chart [25], and the Shiryaev–Roberts procedure [26], have been shown to control the average run length to a false alarm, and be optimal under different scenarios. When a change is signaled by the monitoring chart, the chart is usually restarted by setting it to some values below the signaling threshold.

However, sometimes, we keep monitoring the data stream even if a change is alarmed, in the context of which the mean of observation experiences a departure from the prespecified level for a period of time, then it comes back to the prespecified level, and this may happen many times in a whole monitoring process. Nowadays, high-dimensional data streams often arrive simultaneously, and we usually want to process data streams concurrently. Thus, we consider the joint data generating model for m data streams, the means of which experience recurrent epidemic changes heterogeneously. Suppose that m streams of observations $\mathbb{Z}_1, \mathbb{Z}_2, \ldots$ are collected over time, where $\mathbb{Z}_t = (Z_{1,t}, \ldots, Z_{m,t})^\top$ are independent over t, and each follows a multivariate distribution with mean vector $(\mu_{1,t}, \ldots, \mu_{m,t})^\top$ and the same covariance matrix $\Sigma = (\sigma_{i_1,i_2})_{m\times m}$. For the ith stream, we consider the recurrent one-sided epidemic change model:

$$\begin{cases} \mu_{i,t} = \mu_{i,0}, & t \notin [b_{i,k}, e_{i,k}], \\ \mu_{i,t} > \mu_{i,0}, & t \in [b_{i,k}, e_{i,k}], \end{cases} \tag{1}$$

* Corresponding author.
E-mail address: lilundu@cityu.edu.hk (L. Du).

https://doi.org/10.1016/j.jmva.2023.105224

where $\mu_{i,0}$ is the prespecified mean level for the ith stream, and $[b_{i,k}, e_{i,k}]$ for $k \in \{1, 2, \ldots\}$ are a series of unknown non-overlapping change periods in the ith data stream, and the strength of the signals, $\mu_{i,t}$, depending on i and t, is allowed to be heterogeneous within and across streams. Here, we consider the one-sided changes and assume that the initial parameters $\mu_{i,0} = 0$ and $\sigma_{i,i} = 1$ for $i \in \{1, \ldots, m\}$; the cases of two-sided changes and unknown initial parameters are discussed in Section 4. Note that the correlations σ_{i_1,i_2} $(i_1 \neq i_2)$ between data streams are not needed. We call $Z_{i,t}$ in the null state if $\mu_{i,t} = 0$ and in the alternative state if $\mu_{i,t} > 0$. We use $\vartheta_{i,t}$ to denote the state of $Z_{i,t}$, where $\vartheta_{i,t} = 1$ for the alternative state and $\vartheta_{i,t} = 0$ for the null state. An ideal goal is to make decisions $\delta_{i,t} = \delta_{i,t}(\mathbb{Z}_1, \ldots, \mathbb{Z}_t)$ to recover the signals $\vartheta_{i,t} = 1$.

An illustrative example of the considered setting is fund investment. If we want to invest a sum of money in a fund, the most ideal case is to make investment when the fund's returns are positive and withdrawal when the fund's returns become zero, and repeat this process for each positive return period. We usually want to operate investments and withdrawals for a large pool of funds rather than a single one to further expand profits. Therefore, the problem boils down to discovering the positive change periods of funds. Note that even if changes are signaled by the monitoring charts, the stock market remains operational. Thus, the scenario of interest is that the stream can switch, perhaps many times, between a null state and alternative states.

A direct idea for recovering $\vartheta_{i,t} = 1$ is to use the Shewart chart [32] giving that $\delta_{i,t} = 1$ if $Z_t \geq c$ for some constant c, but the power of the Shewart is low when the magnitude of change is not large. Other methods, such as CUSUM and the Shiryaev–Roberts procedure, employ accumulative statistics based on $\mathbb{Z}_1, \ldots, \mathbb{Z}_t$ to give $\delta_{i,t} = 1$ if the statistic for the ith data stream at time t exceeds a given threshold. If we directly use the accumulative methods, after signaling a change, these methods would increase their statistics with receiving more samples from alternative states to make more significant rejections. However, such accumulation makes the statistic extremely large after just experiencing a long change period, and this will lead to a slow falling back below the threshold. Besides the aforementioned problems occurred in each data stream, the multiplicity also arises when high-dimensional data streams are considered.

The notation of the false discovery rate (FDR) was introduced in the seminal work of Benjamini and Hochberg [1] for handling the multiplicity, then they proposed the Benjamini–Hochberg procedure to control the multiplicity while providing considerable detection power. Previous studies on change detection, including [9,15,19,22,35,40], consider various applications of the Benjamini–Hochberg procedure for statistical process control or surveillance. However, the issues aforementioned still persist, because these approaches only focus on the change from null to alternative but cannot provide sufficient protection against the changes from alternative to null, and thus yield excessive false positive results in our setting. Gandy and Lau [13] proposed a variant of CUSUM chart to alleviate such long return delays, with the form

$$S_{i,0} = 0, \quad S_{i,t} = \min\{\max(S_{i,t-1} + Z_{i,t} - k_i, 0), H_i\}, \tag{2}$$

where k_i is the reference value and $H_i > 0$ is a constant to specify an upper boundary for the ith data stream. They apply the Benjamini–Hochberg procedure [1] to decide $\delta_{i,t}$ while controlling the false discovery rate at the nominal level at each time t. Although [18] provided some discussions on the effect of the upper bound constant H_i under a univariate monitoring setting, the selection of tuning parameters in (2) remains unsolved in practice, and even becomes more difficult owing to the consideration of high-dimensional data streams.

Another possible solution is to restart the monitoring process for the data stream whose monitoring statistic alarms a change. Recently, for the restarting CUSUM in the sequential multiple change-point setting, [21,44] provided non-asymptotic analysis on the probability to a false alarm and the upper bound of the detection delay with high probability. Their methods are valid for a single data stream, but in our setting of high-dimensional data streams, how to control the multiplicity for their methods remains unsolved and needs new construction and theory. Sequential change detection problems in multiple data streams were considered by [4,5,24,30,31,34,42,45], amongst others. Their settings differ completely from ours because they aim to minimize the overall expected delay while controlling the average run length to the null hypothesis that none of the data streams experience changes.

To balance the ability to detect two types of changes, that is, changes from the null to the alternative and then back to the null, we propose a new procedure based on a penalized version of the generalized likelihood ratio (GLR)-based scheme [33]. The new method does not need to pre-choose reference values like the modified CUSUM of [18]. The penalization term is constructed to detect the change-points $e_{i,k}$, and the penalization parameter is determined by a pre-specified power loss level relative to the GLR without penalization. Under some mild conditions on the dependence structure of data streams, the FDR is shown to be controlled at each time point. A data-driven approach to the selection of the penalization parameter is developed, which gives the new method an edge over existing methods in terms of FDR control and detection delay. Some practical guidelines including a robust variant of the proposed procedure are also provided. Both numerical studies and real data analysis support the proposed procedure.

The remainder of this paper is organized as follows. Our new penalized approach and an asymptotic approach for choosing the penalization parameter are described in Section 2. In Section 3, we investigate the theoretical properties of the proposed method. We further provide some practical guidelines in Section 4, including a robust variant of the proposed procedure. Section 5 provides extensive simulation studies and a real data example on fund selection as an illustration. Section 6 concludes the paper. Technical details are provided in the Appendix. Some additional numerical studies can be found in the supplementary material.

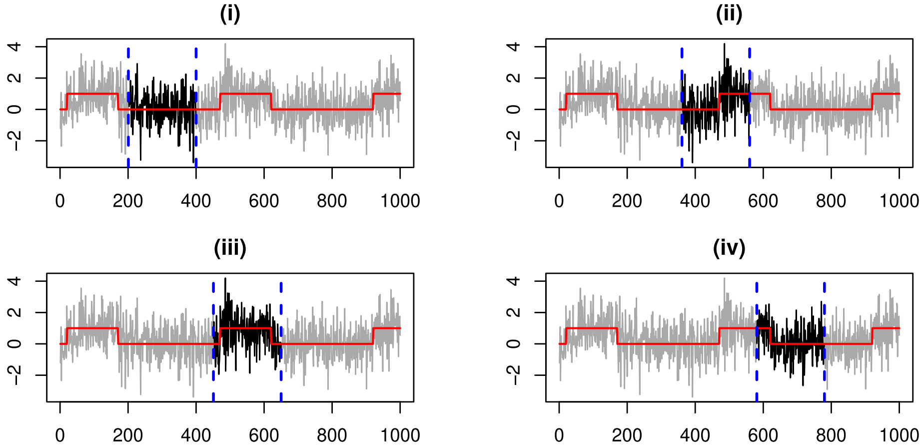

Fig. 1. An example of 1000 observations from $\mathcal{N}(\mu_t, 1)$, where $\mu_t = 1$ for $t \in [21, 170] \cup [471, 620] \cup [921, 1000]$ and $\mu_t = 0$ otherwise. The red line indicates the mean of Z_t. Two blue dashed vertical lines are the margins of the sliding window with window size 200: (top, left) $t = 400$; (top, right) $t = 560$; (bottom, left) $t = 650$; (bottom, right) $t = 780$.

2. Methodology

2.1. A penalized approach

In a sequential setting, where the epidemic changes may happen many times, a sliding window is useful to infer the current states because, in a proper window with size $(w+1)$, a single change structure may be observed. Roughly speaking, for the ith data stream, four cases may happen in a proper window: (i) $\mu_{i,j} = 0$ for all $j \in [t - w, t]$; (ii) a change from null to alternative; (iii) an epidemic change from null to alternative and back to null; and (iv) a change from alternative to null. See Fig. 1 for illustration. To recover the signals $\vartheta_{i,t} = 1$, only case (ii) should be declared as in a change period, and cases (iii) and (iv) should be not declared because a change period has passed and the current state at time t is null. In this context, with the target only at case (ii), it is difficult for the traditional change-point detection methods designed for either the abrupt or epidemic change, because these methods lack the ability to differentiate cases (ii) and (iii).

Next, we introduce the proposed penalization approach. We consider the generalized likelihood ratio (GLR) statistic [33]:

$$T^b_{i,t} = \max_{\tau=\{0,\ldots,w\}} \frac{\sum_{j=t-\tau}^{t} Z_{i,j}}{\sqrt{\tau+1}} \tag{3}$$

We call $T^b_{i,t}$ the detection statistics. GLR is a good detection procedure that performs well in detecting mean shifts [17]. Other statistics, such as the Shiryaev–Roberts statistic [38] and CUSUM statistic, can also play the role of the detection statistic. Here, we use GLR as it does not involve unknown parameters.

We expect that the changes from null to alternative can be detected well with GLR. However, the detection statistic $T^b_{i,t}$ is also large if the epidemic change happens completely in the window (see the bottom-left panel of Fig. 1), which causes the false rejection if we only use the detection statistic. Inspired by this fact, if we consider another statistic $T^e_{i,t}$ which can detect the change from alternative to null, the epidemic change in the window also makes $T^e_{i,t}$ large. We call $T^e_{i,t}$ the return statistic. A concrete implement of $T^e_{i,t}$ is

$$T^e_{i,t} = \max_{\tau\in\{0,\ldots,w-1\}} \left(\frac{\sum_{j=t-w}^{t-\tau-1} Z_{i,j}}{w-\tau} - \frac{\sum_{j=t-\tau}^{t} Z_{i,j}}{\tau+1} \right) \Big/ \left(\frac{1}{w-\tau} + \frac{1}{\tau+1} \right)^{1/2}, \tag{4}$$

which is the CUSUM statistic for one-sided changes in the batch setting [6]. The form of $T^e_{i,t}$ in (4) is not necessary, and other statistics playing the same role as (4) can be also considered. Then, we propose to penalize $T^b_{i,t}$ by subtracting scaled $T^e_{i,t}$, i.e.,

$$T_{i,t}(\theta) = T^b_{i,t} - \theta \times T^e_{i,t}, \tag{5}$$

where $\theta \in [0, \infty)$ is a tuning parameter used to control the amount of penalization. The proposed statistic $T_{i,t}$ can be expected to be positively large only when a change from null to alternative happens after setting proper penalization parameter θ and window size w. Thus, $T_{i,t}$ can be used to recover the states of $Z_{i,t}$ by a thresholding rule.

It is worth noting that the choice of θ is important and involves with a trade-off between the detection power and false detections. For small θ, it will yield a lot of false detections after an epidemic change; while for large θ, the detection power is low due to overlarge penalization. We will introduce a data-driven selection criterion for θ in Section 2.3.

2.2. Dynamic multiple testing

We target to make inference on $(\delta_{1,t},\ldots,\delta_{m,t})$ at each time t based on observations $\mathbb{Z}_{t-w},\ldots,\mathbb{Z}_t$ in the sliding window. To avoid stringent assumptions on data generating process, we consider the following m hypotheses at time t:

$$\mathbb{H}^0_{i,t}:\mu_{i,t-w}=\cdots=\mu_{i,t}=0,\quad \text{versus}\quad \mathbb{H}^1_{i,t}:\mu_{i,j}>0,\ \text{for some } j\in[t-w,t]. \tag{6}$$

Then, the true state to recover in this context is $\vartheta^w_{i,t}$, which satisfies $\vartheta^w_{i,t}=1$ if the alternative hypothesis in (6) is true and $\vartheta^w_{i,t}=0$ otherwise. Let $\mathcal{I}_{0;t}=\{i:\vartheta^w_{i,t}=0\}$ and $\mathcal{I}_{1;t}=\{i:\vartheta^w_{i,t}=1\}$ denote the true null set and alternative set at time t, respectively. By this modification from $\vartheta_{i,t}$ to $\vartheta^w_{i,t}$, we do not need to place restrictions on the data generating mechanism since the hypotheses are based on observations in the window, and thus expand the scope of the procedure developed below. Note that the alternatives should be distinguished since they include three types of changes and only one change is desired. So we term the discovery of changes with $\mu_{i,t}>0$ (see the top-right panel in Fig. 1) as power and the discovery of changes with $\mu_{i,t}=0$ (see the bottom panels in Fig. 1) as pseudo power, respectively. This partitions the alternative set $\mathcal{I}_{1;t}$ into $\mathcal{A}_{1;t}=\{i\in\mathcal{I}_{1;t}:\mu_{i,t}>0\}$ and $\mathcal{B}_{1;t}=\{i\in\mathcal{I}_{1;t}:\mu_{i,t}=0\}$. Our goal is to make discoveries of $\mathcal{A}_{1;t}$ at each time t. Note that we treat the discoveries in $\mathcal{B}_{1;t}$ as true positives due to the formulation in (6), but we hope data streams in $\mathcal{B}_{1;t}$ should be discovered as less as possible.

The penalized statistic $T_{i,t}(\theta)$ is expected to be positively large when a change from null to alternative happens in the window. Hence, we proceed to design a multiple testing procedure to determine the threshold for $T_{i,t}(\theta)$ and reject the hypotheses with their statistics greater than the threshold. The following empirical processes are defined:

$$\begin{aligned} &V_t(q,\theta)=\#\{i\in\mathcal{I}_{0;t}:T_{i,t}(\theta)\ge q\},\quad S_{1;t}(q,\theta)=\#\{i\in\mathcal{A}_{1;t}:T_{i,t}(\theta)\ge q\},\\ &R_t(q,\theta)=\#\{i\in[m]:T_{i,t}(\theta)\ge q\},\quad S_{2;t}(q,\theta)=\#\{i\in\mathcal{B}_{1;t}:T_{i,t}(\theta)\ge q\},\end{aligned} \tag{7}$$

where $[m]=\{1,\ldots,m\}$, $S_{1;t}(q,\theta)$ is the number of true rejections in $\mathcal{A}_{1;t}$ at the current time t, whilst $S_{2;t}(q,\theta)$ is used to count the number of true rejected data streams in $\mathcal{B}_{1;t}$. As a convention, $V_t(q,\theta)$ and $R_t(q,\theta)$ are the number of false rejections and the total number of rejections, respectively. The false discovery proportion with respect to the threshold q is defined as $\mathrm{FDP}_t(q,\theta)=V_t(q,\theta)/\{R_t(q,\theta)\vee 1\}$ with $a\vee b=\max\{a,b\}$. The false discovery rate is $\mathrm{FDR}_t(q,\theta)=\mathrm{E}\{\mathrm{FDP}_t(q,\theta)\}$.

To estimate and control FDR at time t, the unobserved term $V_t(q,\theta)$ is estimated with the penalized test statistics $\{T_{i,t}(\theta), i\in\{1,\ldots,m\}\}$. As discussed by [36], if the values of $T_{i,t}(\theta)$ satisfy the weak dependence assumption so that $\mathrm{Card}(\mathcal{I}_{0;t})^{-1}V_t(q,\theta)$ converges to a distribution function almost surely when the number of the streams goes to infinity, then a consistent estimate of $V_t(q,\theta)$ can be obtained. Hence, it is theoretically important to seek conditions on Σ under which the weak dependence assumption of $T_{i,t}(\theta)$ is satisfied. A sufficient condition is Condition C1 given in Appendix. Some commonly used covariance matrices, such as the autoregressive covariance structure or the banded covariance structure, satisfy Condition C1. Note that the correlation structure is not required to implement the proposed procedure; it is only used to justify the theoretical results.

Proposition 1. *Let $F_{0;t}(q,\theta)=\Pr\{T_{i,t}(\theta)\ge q\mid i\in\mathcal{I}_{0;t}\}$ be the true null distribution of $T_{i,t}(\theta)$ under $\mathbb{H}^0_{i,t}$, and $F^i_{1;t}(q,\theta)=\Pr\{T_{i,t}(\theta)\ge q\mid i\in\mathcal{A}_{1;t}\}$ and $F^i_{2;t}(q,\theta)=\Pr\{T_{i,t}(\theta)\ge q\mid i\in\mathcal{B}_{1;t}\}$ be the two types of alternative distribution for $i\in\mathcal{A}_{1;t}$ and $i\in\mathcal{B}_{1;t}$. Suppose that Conditions C1–C2 in the Appendix hold. For a fixed θ and window size w, we have*

(i) $\mathrm{Card}(\mathcal{I}_{0;t})^{-1}V_t(q,\theta)$ *converges to the true null distribution $F_{0;t}(q,\theta)$ almost surely.*
(ii) *If we further assume that*

$$\mathrm{Card}(\mathcal{A}_{1;t})^{-1}\sum_{i\in\mathcal{A}_{1;t}}F^i_{1;t}(q,\theta)\to F_{1;t}(q,\theta),\quad \mathrm{Card}(\mathcal{B}_{1;t})^{-1}\sum_{i\in\mathcal{B}_{1;t}}F^i_{2;t}(q,\theta)\to F_{2;t}(q,\theta),$$

then the empirical processes $S_{1;t}(q,\theta)$, $S_{2;t}(q,\theta)$ and $R_t(q,\theta)$ in (7) converge in the sense that $\mathrm{Card}(\mathcal{A}_{1;t})^{-1}S_{1;t}(q,\theta)\to F_{1;t}(q,\theta)$, $\mathrm{Card}(\mathcal{B}_{1;t})^{-1}S_{2;t}(q,\theta)\to F_{2;t}(q,\theta)$ and $\mathrm{Card}(\mathcal{I}_t)^{-1}R_t(q,\theta)\to F_t(q,\theta)$ almost surely, where $F_t(q,\theta)=\pi_{0;t}F_{0;t}(q,\theta)+\pi_{1;t}F_{1;t}(q,\theta)+\pi_{2;t}F_{2;t}(q,\theta)$ with $\pi_{0;t}=\lim_{m\to\infty}\mathrm{Card}(\mathcal{I}_{0;t})/m$, $\pi_{1;t}=\lim_{m\to\infty}\mathrm{Card}(\mathcal{A}_{1;t})/m$ and $\pi_{2;t}=\lim_{m\to\infty}\mathrm{Card}(\mathcal{B}_{1;t})/m$ being the asymptotic proportions of null, and two types of alternatives.

If we further assume that the observations at each time point, i.e., $\mathbb{Z}_t$, follows a multivariate normal distribution, the conclusions in (i) and (ii) continues to hold when Condition C1 is replaced by $\{m(m-1)\}^{-1}\sum\sum_{i_1\ne i_2}\sigma_{i_1,i_2}=O(m^{-\delta})$, for $\delta>0$.

To validate the convergence of the empirical process $R_t(q,\theta)$, in Proposition 1 (ii) we require the averaged alternative distribution functions converges as $m\to\infty$, which is easily satisfied if the data streams in the alternative set share the same temporal structure. In the case where the locations of change points and the signal strength are heterogeneous but are sampled from some prior distributions, the convergence also hold.

Proposition 1 indicates that the penalized test statistics satisfy the weak dependence assumptions in [36], and thus the standard estimation and controlling procedure [36] can be applied directly using $F_{0;t}(q,\theta)$. The estimation and controlling approaches are outlined as follows:

Estimation approach: A conservative estimate of the $\mathrm{FDR}_t(q,\theta)$ is given by

$$\widehat{\mathrm{FDR}}_{\lambda;t}(q,\theta)=\frac{\widehat{V}_t(q,\theta)}{R_t(q,\theta)\vee 1}=\frac{m\widehat{\pi}_{0;t}(\lambda,\theta)F_{0;t}(q,\theta)}{R_t(q,\theta)\vee 1}, \tag{8}$$

where $\widehat{\pi}_{0;t}(\lambda,\theta)=\sum_{i=1}^{m}\mathrm{I}(T_{i,t}(\theta)<\lambda)/[m\{1-F_{0;t}(\lambda,\theta)\}]$ is an upper-biased estimate of $\pi_{0;t}$. The tuning parameter λ is chosen such that a negligible number of hypotheses have penalized test statistics of less than λ in the alternative set. See Section 4 for more discussions.

Controlling approach: Given a significance level α, the data-driven threshold is

$$\widehat{q}_{\alpha;t}(\theta)=\inf_q\left\{q:\widehat{\mathrm{FDR}}_{\lambda;t}(q,\theta)\le\alpha\right\}. \tag{9}$$

Hypotheses with a value of $T_{i,t}(\theta)$ exceeding $\widehat{q}_{\alpha;t}(\theta)$ are rejected.

We refer to this procedure as penalization-assisted dynamic detection (PADD).

2.3. A selection criterion for the penalization coefficient

We discuss the effect of the penalization coefficient θ on the FDR and detection power of the PADD procedure and propose a criterion for selection of θ by setting a fixed percentage of power loss. This can be achieved by connecting the frequentist version of FDR to a Bayesian rationale in terms of a two-group random mixture model, which follows directly from the Bayes theorem. According to [11], the Bayesian FDR is defined as the posterior probability of a null hypothesis being true if the test statistic is within the rejection region:

$$\mathrm{Fdr}_t(q,\theta)=\Pr\{H_{i,t}^0\mid T_{i,t}(\theta)\ge q\}=\frac{\pi_{0;t}F_{0;t}(q,\theta)}{F_t(q,\theta)}.$$

Based on the results in Proposition 1, the frequentist FDR converges to the Bayesian FDR [36].

To express the detection power when the Bayesian FDR is controlled at α, we note that (9) is asymptotically equivalent to:

$$q_{\alpha;t}(\theta)=\inf\{q:k(\theta)\times\mathrm{Fdr}_t(q,\theta)\le\alpha\}, \tag{10}$$

where $k(\theta)=1+(1-\pi_{0;t})\{1-F_{12;t}(\lambda,\theta)\}/[\pi_{0;t}\{1-F_{0;t}(\lambda,\theta)\}]$. Accordingly, the overall detection power for both types of alternatives is evaluated by $\Pr\{T_{i,t}(\theta)\ge q_{\alpha;t}(\theta)\mid i\in\mathcal{A}_{1;t}\cup\mathcal{B}_{1;t}\}\equiv\{\pi_{1;t}F_{1;t}(q_{\alpha;t}(\theta),\theta)+\pi_{2;t}F_{2;t}(q_{\alpha;t}(\theta),\theta)\}/\{\pi_{1;t}+\pi_{2;t}\}$, denoted as $F_{12;t}(q_{\alpha;t}(\theta),\theta)$. We can see that the overall power $F_{12;t}(q_{\alpha;t}(\theta),\theta)$ consists of the power $F_{1;t}(q_{\alpha;t}(\theta),\theta)$ and the pseudo power $F_{2;t}(q_{\alpha;t}(\theta),\theta)$. Intuitively, a small θ leads to a high power, but also to a high pseudo power.

We use a toy example to illustrate the power, pseudo-power, and overall power. We generate $m=2000$ data streams over a period of 200 time points, 10% of which are experiencing epidemic changes with $\mu_{i,t}=0.4$, from time $t=111$ to the present time $t=200$, and another 10% of which have already changed from a signal period $[b_i,e_i]=[121,170]$ with $\mu_{i,t}=0.2$ to the null state for thirty time units. The left panel of Fig. 2 shows the FDR and three types of detection powers evaluated with the data-driven threshold $\widehat{q}_{\alpha;t}(\theta)$, where $\alpha=0.2$, $w=100$, and $\lambda=0$. We observe that the FDR levels can be well controlled and that both types of detection powers decrease smoothly when the penalization coefficient θ is increased. This motivates us to select θ by setting a fixed percentage of power loss relative to the case in which no penalization is applied. More specifically, given a fixed percentage of power loss β, θ is selected by

$$\theta_{0;t}=\sup\left\{0\le\theta<\infty;\ \frac{F_{12;t}(q_{\alpha;t}(\theta),\theta)}{F_{12;t}(q_{\alpha;t}(0),0)}\ge 1-\beta\right\}. \tag{11}$$

To estimate $\theta_{0;t}$, the detection power functions must be estimated. By (10), the power function is proportional to the marginal distribution if the Bayesian FDR is strictly controlled at the level α, that is,

$$F_t(q_\alpha(\theta),\theta)=\frac{1-\pi_{0;t}}{1-\alpha/k(\theta)}F_{12;t}(q_\alpha(\theta),\theta).$$

In the cases in which $\pi_{0;t}\approx 1$ or $F_{12;t}(\lambda,\theta)\approx 1$, the term $k(\theta)$ is approximately 1. As a result, the criterion (11) can be replaced by

$$\theta_{0;t}=\sup\left\{0\le\theta<\infty;\ \frac{F_t(q_{\alpha;t}(\theta),\theta)}{F_t(q_{\alpha;t}(0),0)}\ge 1-\beta\right\}. \tag{12}$$

With the data-driven threshold $\widehat{q}_{\alpha;t}(\theta)$ in (9) and the empirical estimate of the marginal distribution, a data-driven choice of θ in (12) is given by

$$\widehat{\theta}_t=\sup\left\{0\le\theta<\infty;\ \frac{R_t(\widehat{q}_{\alpha;t}(\theta),\theta)/m}{R_t(\widehat{q}_{\alpha;t}(0),0)/m}\ge 1-\beta\right\}. \tag{13}$$

As a side note, the upper boundary H in the modified CUSUM [18] plays a similar role to θ, reducing the pseudo-power with a certain loss of the desired power. For the ith data stream, they employ the modified CUSUM in (2) with parameters

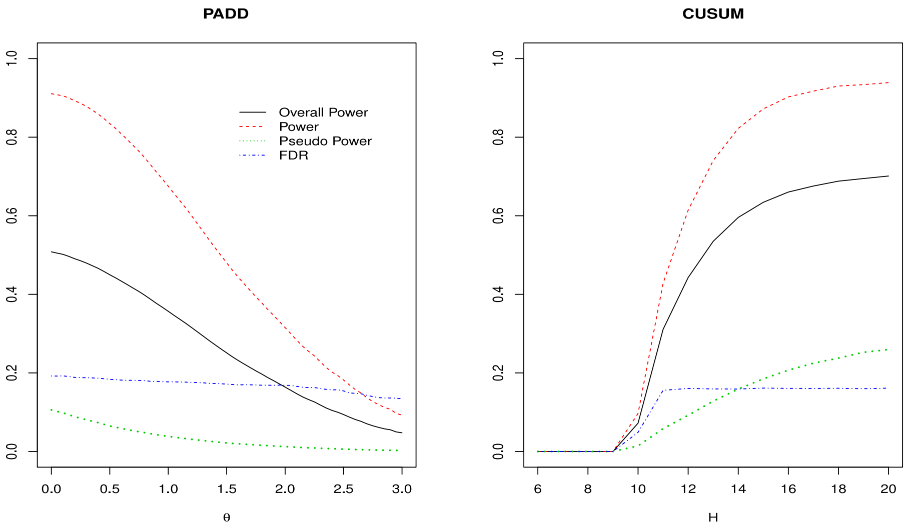

Fig. 2. The curves of the FDR and three types of detection power. Left panel: the PADD with different values of θ; Right panel: the modified CUSUM with different upper boundaries H.

(k_i, H_i). Then they adopt the Benjamini–Hochberg procedure to control the FDR at each time t. The right panel of Fig. 2 depicts the FDR and three types of detection power for the modified CUSUM with various values of $H_i = H$ under the same setting as in the left panel. The parameter k_i is chosen as 0.15 for all streams. A small value of H yields conservative FDRs and little power. This can be understood by noting that good control of the FDR for the modified CUSUM needs the p-values corresponding to $S_{i,t}$ under the null to follow an approximately uniform distribution. However, this requirement would be violated when H is not sufficiently large, because the null distribution of $S_{i,t}$ has a point mass at H. How large the H should be depends on $k_i, \alpha, \pi_{0;t}$, and the alternative distribution. We also observe that the power and pseudo-power exhibit similar patterns as the H increases, which demonstrates the importance of choosing a suitable H. As opposed to using a fixed H in the modified CUSUM, the θ in PADD is determined dynamically based on a fixed percentage of power loss. One of our main contributions is to theoretically establish the FDR control with a data-driven choice of θ, which will be highlighted in the next section.

3. Theoretical properties

We first show that the estimation approach can conservatively estimate the FDR uniformly for all thresholds q and θ. Technical conditions are given in the Appendix. Throughout this section, we assume that λ and w are fixed.

Theorem 1. *Suppose Conditions C1–C3 in the Appendix hold. Then, for each time $t > w$,*

$$\lim_{m\to\infty} \inf_{q\leq\bar{q}} \inf_{0\leq\theta\leq\bar{\theta}} \left\{\widehat{\mathrm{FDR}}_{\lambda;t}(q,\theta) - \mathrm{FDR}_t(q,\theta)\right\} \geq 0, \quad \textit{almost surely}, \tag{14}$$

where $\bar{q} \in (-\infty, \infty)$ and $\bar{\theta} \in (0, \infty)$ are two sufficiently large constants.

Because the support set of θ contains infinite choices of value, careful investigation is required to control the FDR of the proposed PADD procedure using the data-driven threshold $\widehat{\theta}_t$. If $\widehat{\theta}_t$ converges to its true value, say $\theta_{0;t}$, the possible range of θ in (13) will be reduced to a local interval around the true value $\theta_{0;t}$ when m is sufficiently large, thus allowing the FDR to be controlled.

Proposition 2. *Under Conditions C1–C5 in the Appendix, $\widehat{\theta}_t \to \theta_{0;t}$ almost surely.*

By Theorem 1 and Proposition 2, we can show that the PADD method with the data-driven threshold $\widehat{q}_{\alpha;t}(\widehat{\theta}_t)$ is able to control the FDR.

Theorem 2. *Suppose that Conditions C1–C5 hold. Then $\limsup_{m\to\infty}\left\{\mathrm{FDR}_t(\widehat{q}_{\alpha;t}(\widehat{\theta}_t), \widehat{\theta}_t)\right\} \leq \alpha$, for each time $t > w$.*

We next turn to a theoretical justification of the observation that the PADD procedure achieves a lower pseudo power than the standard GLR method. Let $q_{\alpha;t}(\theta)$ denote the threshold such that $\mathrm{Fdr}_t(q,\theta) = \alpha$. Then, the pseudo-power function can be formulated by $F_{2;t}(q_{\alpha;t}(\theta), \theta) = \beta' F_{0;t}(q_{\alpha;t}(\theta), \theta)$, with $\beta' = (1/\alpha - 1)\pi_{0;t}/(1 - \pi_{0;t})$. Our goal is to quantify the pseudo-power that can be reduced by combining the statistic $T^e_{i,t}$ with $T^b_{i,t}$. Using similar techniques in [8], the ratio of the power of the proposed procedure to the method using the GLR ($\theta = 0$) can be derived as

$$\frac{F_{2;t}(q_{\alpha;t}(\theta), \theta)}{F_{2;t}(q_{\alpha;t}(0), 0)} = \frac{F_{0;t}(q_{\alpha;t}(\theta), \theta)}{F_{0;t}(q_{\alpha;t}(0), 0)} = 1 + \frac{\frac{\partial}{\partial\theta}\left\{F_{0;t}(q_{\alpha;t}(\theta), \theta)\right\}\big|_{\theta=0}\theta}{F_{0;t}(q_{\alpha;t}(0), 0)} + O\left(\theta^2\right) = 1 + \Delta(\theta) + O\left(\theta^2\right).$$

In our asymptotic analysis on $\mathcal{B}_{1;t}$, we assume that $e - b \to \infty$, $t - e \to \infty$ and $0 < \lim_{t\to\infty}(e - b + 1)/(t - b + 1) \to \gamma < 1$ as $t \to \infty$. In addition, we consider the case with $w \le (e - b + 1)$ and suppose the change magnitude μ satisfies $\lim_{t\to\infty}\sqrt{t - b + 1}\mu/(\log\log w)^{1/2} \to \infty$. Derivations in the Appendix yield that the ratio of power reduced is approximated by

$$1 + \Delta(\theta) = 1 - \frac{C \times \left\{\sqrt{\gamma(1-\gamma)}\sqrt{t - b + 1}\mu - (2\log\log w)^{1/2}\right\}}{\left[\left\{\frac{\partial F_{2;t}(q,\theta)}{\partial q} - \beta'\frac{\partial F_{0;t}(q,\theta)}{\partial q}\right\}\Big|_{q=q_{\alpha;t}(\theta)}\right]\Big|_{\theta=0}} \times \frac{\theta}{F_{0;t}(q_{\alpha;t}(0), 0)} + O(\theta^2), \tag{15}$$

where C is a positive constant independent of θ. A similar argument by [14] yields that

$$\left[\left\{\frac{\partial F_{2;t}(q, \theta)}{\partial q} - \beta'\frac{\partial F_{0;t}(q, \theta)}{\partial q}\right\}\Big|_{q=q_{\alpha;t}(\theta)}\right]\Big|_{\theta=0} < 0.$$

Combining these, we have $1 + \Delta(\theta) < 1$. In other words, there is a greater probability that the proposed procedure will not make a discovery at time t for $i \in \mathcal{B}_{1;t}$, which results in a lower pseudo-power.

As illustrated in Section 2.3, both power and pseudo power are expected to be decreased as θ increases. So we sacrifice some detection power to reduce the pseudo power, where the power loss is determined by the user specified parameter β. As the upper bound H in (2) [13], there exists a trade-off in the PADD procedure between detecting changes from null to alternative and not detecting changes from alternative to null. But the trade-off is determined by a meaningful parameter β, which is the percentage of power loss and can be flexibly specified by the user.

4. Practical guidelines

4.1. Choice of the tuning parameter λ

The selection of λ to estimate the null proportion is well studied in the multiple testing literature. Benjamini and Hochberg [2] suggested a simple data-driven approach for the selection of λ that achieves a bias–variance trade-off. The effect of the choice of λ on the variance part vanishes as the number of tests increases. Thus, one prefers to choose a small λ to mitigate the bias effect. To this end, $\lambda = 0$ is recommended in our PADD procedure to guarantee that no bias occurs with a high probability, say $F_{12;t}(\lambda, \theta) \approx 1$.

4.2. Unknown initial parameters

We assumed $\mu_{i,0}$ and $\sigma_{i,i}$ are known for $i \in \{1, \ldots, m\}$ in the previous sections. When the initial mean parameters are unknown, invariant transformations [3,29] can be used. Yakir [43] then explored more general cases which include both initial mean and variance parameters are unknown. Formally, for each data stream at time t, we consider a set of invariant statistics $(X_{i,1}, \ldots, X_{i,t}) = \varphi(Z_{i,1}, \ldots, Z_{i,t})$ for some transformation φ, the initial parameters after which are known. Then, the penalized statistic (5) is computed with $X_{i,1}, \ldots, X_{i,t}$ instead of $Z_{i,1}, \ldots, Z_{i,t}$, and the PADD procedure can be still applied. Concretely, for unknown $\mu_{i,0}$ and known $\sigma_{i,i}$, an invariant function [29] is $X_{i,1} = 0, X_{i,2} = Z_{i,2} - Z_{i,1}, X_{i,3} = Z_{i,3} - Z_{i,1}, \ldots$, whose initial mean is zero. For both unknown $\mu_{i,0}$ and $\sigma_{i,i}$, suppose one can obtain a learning sample $Z_{i,-n}, Z_{i,-n+1}, \ldots, Z_{i,-1}$, $n \ge 2$, of prechange observations. Then, the invariant statistics from [43] are $X_{i,t} = (Z_{i,t} - \bar{Z}_{-n})/s_{-n}$, $t \in \{1, 2, \ldots\}$, where $\bar{Z}_{-n} = \sum_{i=1}^n Z_{-i}/n$ and $s^2_{-n} = \sum_{i=1}^n (Z_{-i} - \bar{Z}_{-n})^2/(n-1)$.

4.3. Robust penalization-assisted dynamic detection

The statistics $T^b_{i,t}$ in (3) and $T^e_{i,t}$ in (4) are close to the likelihood ratio forms based on normally distributed data. So good performance can be expected for PADD when underlying data follow the (multivariate) normal distribution. Indeed, utilizing a nonparametric transformation technique [23,27], PADD can be extended to a robust procedure against heavy-tailed or skewed distributions. For the ith data stream, we consider the sequential ranks $\{r^t_{i,j}\}^t_{j=t-w}$ at current time t, where

$$r^t_{i,j} = \frac{1}{j - (t - w) + 2}\sum_{k=t-w}^{j}\mathbb{I}(Z_{i,k} \le Z_{i,j}),\ j \in [t - w, t]. \tag{16}$$

If there is no change, $r^t_{i,j}$ are independent and asymptotically Uniform[0, 1] distributed. Then, the distribution of $\Phi^{-1}(r^t_{i,j})$ tends to that of Φ^{-1}(Uniform[0, 1]), which is rightly $\Phi(t)$, where $\Phi(t)$ is the cumulative distribution function of the standard normal random variable. Hence, after replacing $\{Z_{i,j}\}^t_{j=t-w}$ with $\{\Phi^{-1}(r^t_{i,j})\}^t_{j=t-w}$, the robust surveillance statistic is

$$T^r_{i,t}(\theta) = T^{r,b}_{i,t} - \theta \times T^{r,e}_{i,t},$$

where $\theta \in [0, \infty)$ is a tuning parameter playing the same role as in (5), and

$$T^{r,b}_{i,t} = \max_{\tau \in \{0,\ldots,w\}} \frac{\sum_{j=t-\tau}^{t} \Phi^{-1}(r^t_{i,j})}{\sqrt{\tau+1}},$$
$$T^{r,e}_{i,t} = \max_{\tau \in \{0,\ldots,w-1\}} \left(\frac{\sum_{j=t-w}^{t-\tau-1} \Phi^{-1}(r^t_{i,j})}{w-\tau} - \frac{\sum_{j=t-\tau}^{t} \Phi^{-1}(r^t_{i,j})}{\tau+1} \right) \Big/ \left(\frac{1}{w-\tau} + \frac{1}{\tau+1} \right)^{1/2}.$$

We expect the asymptotic behavior of $T^r_{i,t}(\theta)$ is similar to that of $T_{i,t}(\theta)$, due to the asymptotic property of sequential ranks. Therefore, the PADD procedure in Section 2.2 and the data-driven selection criterion for θ in Section 2.3 are both applicable to $\{T^r_{i,t}(\theta)\}_{i=1}^m$. For the sake of compactness, we omit the statement on making discoveries and selecting penalty parameter θ. We abbreviate this robust PADD procedure as R-PADD, which also enjoys the benefit of not being affected by the unknown initial parameters.

4.4. Two-sided changes

In practice, we are often concerned with both positive and negative shifts; that is, the alternative hypothesis in (6) is $\mathbb{H}^1_{i,t} : \mu_{i,j} \neq 0$, for some $j \in [t-w, t]$. As a convention, the lower-sided statistics can be similarly defined as

$$L^b_{i,t} = \max_{\tau \in \{0,\ldots,w\}} -\frac{\sum_{j=t-\tau}^{t} Z_{i,j}}{\sqrt{\tau+1}}, \quad L^e_{i,t} = \max_{\tau \in \{0,\ldots,w-1\}} -\left(\frac{\sum_{j=t-w}^{t-\tau-1} Z_{i,j}}{w-\tau} - \frac{\sum_{j=t-\tau}^{t} Z_{i,j}}{\tau+1} \right) \Big/ \left(\frac{1}{w-\tau} + \frac{1}{\tau+1} \right)^{1/2}.$$

Accordingly, $\tilde{T}_{i,t}(\theta) = \max\{T_{i,t}(\theta), L^b_{i,t} - \theta L^e_{i,t}\}$ can be used to replace $T_{i,t}(\theta)$ in the definition of the proposed procedure. An extension to two-sided changes for the robust statistic $T^r_{i,t}(\theta)$ can be derived similarly.

4.5. Choice of the window size w

We specify a constant window size w in the null hypotheses (6). If the window size w matches the length of a signal period, as illustrated by Fig. 1, the proposed procedure will be effective. However, the length of the alternative periods depends on the specific stream and time point. The $T^b_{i,t}$ and $T^e_{i,t}$ with an excessively small w would fail to accumulate sufficient evidence after the changes, whilst an excessively large w may result in contamination of the most recent information. Accordingly, the detection power would be degraded to a certain degree. Ideally, w should be determined in an adaptive manner.

Let $w_{i,t}$ be the window size used for the ith stream at time t. We allow the window size to grow as time proceeds but reset the window size to one if there is evidence that the stream has returned to the null. The change-point test statistic $T^e_{i,t}$ can be used to detect such a transition. That is, w is selected as

$$w_{i,t} = \begin{cases} w_{i,t-1} + 1, & \text{if } T^e_{i,t} < c, \\ 1, & \text{if } T^e_{i,t} \geq c, \end{cases}$$

where the cutoff c should be set high to guarantee that resetting does not occur frequently. The numerical results show that selection of c as the upper 0.5% quantile of the null distribution of $T^e_{i,t}$ suffices for most practical uses.

5. Numerical studies

5.1. Competitors

Before getting into simulations, we introduce some competitors of the proposed PADD procedure.

- CUSUMs — We consider the aforementioned modified CUSUM chart (see [13] and (2)). For all data streams, we pick the same tuning parameters, i.e., $k_i \equiv k$ and $H_i \equiv H$. Two combinations of parameters are examined, $(k, H) = (0.25, 20)$ (CUSUM1) and $(k, H) = (0.5, 15)$ (CUSUM2).
- GLRs — We compare with the GLR statistics, which has the forms (3) for abrupt changes (GLR-A) and $G_{i,[t-w,t]} = \max_{t-w \leq \tau_2 \leq t} \max_{t-w-1 \leq \tau_1 < \tau_2} \left(\sum_{j=\tau_1+1}^{\tau_2} Z_{i,j} \right) / \sqrt{\tau_2 - \tau_1}$ for epidemic changes (GLR-E).
- SRs — We also consider the Shiryaev–Roberts type procedures (e.g., [16,38,39]), which take the forms $J^t_{i,t}$ for abrupt changes (SR-A) and $J_{i,[t-w,t]} = (w+1)^{-1} \max_{t-w \leq \tau_2 \leq t} J^t_{i,\tau_2}$ for epidemic changes (SR-E), where $J^t_{i,\tau_2} = \sum_{k=t-w}^{\tau_2} \exp\left\{ \sum_{j=k}^{\tau_2} \left(Z_{i,j}\mu_{i,j} - \mu^2_{i,j}/2 \right) \right\}$. Since μ_i is unknown, we use the sample mean $\widehat{\mu}^k_{i,j}$ to substitute $\mu_{i,j}$ as formulated in [38], where $\widehat{\mu}^k_{i,j} := (j-k)^{-1} \sum_{\ell=k}^{j-1} Z_{i,\ell}$ is the sample mean based on $\{Z_{i,k}, \ldots, Z_{i,j-1}\}$ and $\widehat{\mu}^k_{i,k} = 0$.
- Shewart — The classic Shewart chart [28,32] based procedure (Shewart) is also compared.

In summary, we have eight procedures including the proposed procedure, PADD, CUSUM1, CUSUM2, GLR-A, GLR-E, SR-A, SR-E, and Shewart.

Like R-PADD, for the procedures based on GLRs or SRs, we can substitute $\{Z_{i,j}\}_{j=t-w}^{t}$ by $\{\Phi^{-1}(r_{i,j})\}_{j=t-w}^{t}$ using sequential ranks (16), which yields robust versions of GLR- and SR-type statistics. We term them as R-GLR-A, R-GLR-E, R-SR-A and R-SR-E, respectively. The CUSUM and Shewart are not based on a sliding window $[t-w,t]$, so the technique of sequential ranks is not applicable.

For all procedures, we simulate the null distribution in (8) with standard normal random variables. Then, for FDR control, we apply Eq. (8)–(9) to each procedure mentioned above to decide the rejection threshold. The window sizes of the mentioned procedures except CUSUMs and Shewart are set as $w = 200$ and the parameter λ used to estimate the true null proportion $\pi_{0;t}$ is fixed at the median of the simulated null distribution for all procedures including PADD for fairness. We also run the proposed PADD method with a data-driven choice of window size, and the results are similar to those with a carefully chosen w that depends on the simulation designs.

5.2. Simulation results

In this section, we evaluate the performance of our proposed procedures via simulations. We study the accuracy of FDR control and then compare the power and pseudo power of the PADD method with those of other competitors. All results in this section were obtained with 500 replications. The R and Matlab codes are available upon request.

We first simulate $\mathbb{Z}_j$ with $m = 500$ dimensions from a multivariate normal distribution with a mean vector $\mu_j = (\mu_{1,j},\ldots,\mu_{m,j})^\top$ and a common covariance matrix $\Sigma = \left(\rho^{|i_1-i_2|}\right)_{m\times m}$ sequentially for $j \in \{1,\ldots,T\}$, with $T = 2000$ and $\rho = 0.5$. A proportion π_1 of data streams experience multiple periodic mean changes, while the other data streams keeps the mean level $\mu_{i,j} = 0$, throughout the monitoring process. The data stream in the alternative set obeys the recurrent one-sided epidemic change model (1) with $\mu_{i,0} = 0$, i.e.,

$$\begin{cases} \mu_{i,t} = 0, & t \notin [b_{i,k}, e_{i,k}], \\ \mu_{i,t} > 0, & t \in [b_{i,k}, e_{i,k}], \end{cases}$$

which alternates between the null state and alternative state. Each data stream in the alternative set follows the process below to be generated. For generating change periods of the ith data stream, the location of the first change-point $b_{i,1}$ is randomly chosen from the set $\{11, 21, \ldots, 101\}$. Then, the ith data stream experiences a period of mean changes with length $L_{i,1}$, and this change period is $[b_{i,1}, e_{i,1}]$, where $e_{i,1} = b_{i,1} + L_{i,1} - 1$. After that, a null period ($\mu_{i,t} = 0$) with length L_{null} is observed, so the new change period begins at $b_{i,2} = e_{i,1} + L_{\text{null}}$. By repeating this generating process, we decide the change periods $[b_{i,k}, e_{i,k}]$ for the ith data stream. In our simulations, we set $L_{\text{null}} = 200$. For the mean level $\mu_{i,t}$ in $[b_{i,k}, e_{i,k}]$ and the length of change period $L_{i,k}$, we explore two scenarios:

Scenario I: Consider homogeneous change periods $[b_{i,k}, e_{i,k}]$ with length $L_{i,k} = e_{i,k} - b_{i,k} + 1$ and signal $\mu_{i,j} = \delta_{i,k}$ for $j \in [b_{i,k}, e_{i,k}]$. For each change period $[b_{i,k}, e_{i,k}]$, parameters $(L_{i,k}, \delta_{i,k})$ are randomly selected from $\{(100, 1), (200, 5), (300, 0.2)\}$.

Scenario II: Consider mixed regimes of change periods $[b_{i,k}, e_{i,k}]$ with length $L_{i,k} = e_{i,k} - b_{i,k} + 1$. For each change period $[b_{i,k}, e_{i,k}]$, the period length $L_{i,k}$ is randomly sampled from $\{50, 100, \ldots, 500\}$ and the signal magnitudes $\{\mu_{i,j}\}_{j=b_{i,k}}^{e_{i,k}}$ randomly adopt one of the following four regimes: (i) constant alternative with $\mu_{i,j} = 0.5$; (ii) relative constant alternative with $\mu_{i,j} = 5/L_{i,k}$; (iii) linear drift alternative with $\mu_{i,j} = 2/(e_{i,k} + 1 - j)$ and (iv) arbitrary drift alternative with $\mu_{i,j}$ being randomly sampled from $\{0.2, 0.5, 1\}$.

These complex setups are used to assess the efficiency and robustness of the proposed testing procedure. The FDR level α and the power loss percentage β in PADD are both set at 0.2.

We first confirm that the proposed method can substantially reduce the pseudo-power, whilst the FDR can be controlled at the significance level. Table 1 shows the mean and standard deviation of the empirical FDR, power and pseudo-power across $T = 2000$ time points for all procedures. The FDRs of the PADD procedure can be controlled at the nominal level $\alpha = 0.2$ within acceptable ranges. Compared to GLRs, SRs and CUSUMs, the PADD procedure is able to considerably reduce the pseudo-power in this recurrent scheme. This is consistent with our theoretical analysis and demonstrates the validity of this method of selecting the penalization coefficient θ. The Shewart provides both low power and pseudo power in the two scenarios where signals are not large.

Except the metrics (FDR, Power, and Pseudo-power), we also concern about two types of delays: (i) the detection delay, that is, the minimum number of time units that the procedure takes to detect a signal within each alternative period, and (ii) the return delay, that is, the minimum number of time units it requires to not declare the signal after experiencing each alternative period. The detection delay is the same as in the literature on sequential change detection, but the return delay is a new metric to be noticed in this recurrent scheme. We do not concern the average run length to a false alarm, because we introduce the notion of FDR to characterize the type-I error. In simulations, if there is no detection in the alternative period, the detection delay are simply set as the length of that specific alternative period plus one; whilst if there is no return (i.e., declaring no signal) between two alternative periods, the return delay are simply set as the length of that null period plus one. We record the means of two delays amongst all change periods in each replication for each procedure.

Table 1
Mean and standard deviation (in parentheses) of empirical FDR, power and pseudo-power (%) across $T = 2000$ time points for different π_1 with nominal size $\alpha = 20\%$ in the setting of multivariate normal distribution. PADD is the procedure we proposed in Section 2 and its competitors (CUSUM1, CUSUM2, GLR-A, SR-A, GLR-E, SR-E, Shewart) are introduced in details in Section 5.1.

	$\pi_1 = 0.2$			$\pi_1 = 0.1$		
	FDR	Power	Pseudo-power	FDR	Power	Pseudo-power
Scenario I						
PADD	$19.9_{(0.79)}$	$61.5_{(8.61)}$	$37.7_{(16.1)}$	$21.6_{(0.90)}$	$55.2_{(8.06)}$	$35.8_{(16.5)}$
CUSUM1	$18.9_{(2.44)}$	$58.1_{(10.0)}$	$61.9_{(17.8)}$	$19.0_{(2.59)}$	$53.0_{(9.99)}$	$53.0_{(16.5)}$
CUSUM2	$18.5_{(1.57)}$	$37.8_{(8.84)}$	$59.9_{(20.6)}$	$18.3_{(1.62)}$	$34.4_{(8.15)}$	$47.8_{(17.9)}$
GLR-A	$19.3_{(0.42)}$	$67.8_{(9.48)}$	$55.1_{(21.9)}$	$19.9_{(0.55)}$	$60.5_{(8.91)}$	$51.0_{(22.0)}$
SR-A	$18.8_{(0.56)}$	$62.5_{(9.41)}$	$49.2_{(21.9)}$	$19.2_{(0.62)}$	$56.1_{(8.99)}$	$46.0_{(21.9)}$
GLR-E	$18.7_{(0.47)}$	$51.7_{(8.25)}$	$62.3_{(22.2)}$	$18.8_{(0.48)}$	$45.8_{(8.01)}$	$58.8_{(22.5)}$
SR-E	$18.8_{(0.55)}$	$53.0_{(8.52)}$	$61.7_{(22.2)}$	$19.1_{(0.55)}$	$47.3_{(8.27)}$	$58.4_{(22.5)}$
Shewart	$17.0_{(1.63)}$	$0.4_{(0.13)}$	$0.0_{(0.06)}$	$17.9_{(1.69)}$	$0.4_{(0.12)}$	$0.0_{(0.05)}$
Scenario II						
PADD	$19.3_{(0.76)}$	$63.0_{(11.4)}$	$27.4_{(5.51)}$	$21.3_{(0.82)}$	$60.3_{(11.4)}$	$26.4_{(5.30)}$
CUSUM1	$18.7_{(2.56)}$	$65.1_{(11.4)}$	$66.9_{(12.1)}$	$19.0_{(2.75)}$	$62.9_{(11.6)}$	$58.2_{(11.0)}$
CUSUM2	$18.7_{(1.43)}$	$54.4_{(9.72)}$	$66.6_{(12.6)}$	$19.0_{(1.59)}$	$51.1_{(9.66)}$	$51.0_{(11.1)}$
GLR-A	$18.7_{(0.45)}$	$65.1_{(11.3)}$	$53.2_{(9.78)}$	$19.7_{(0.66)}$	$63.1_{(11.5)}$	$49.7_{(9.35)}$
SR-A	$18.3_{(0.57)}$	$63.4_{(11.5)}$	$47.8_{(8.97)}$	$19.2_{(0.73)}$	$61.6_{(11.7)}$	$44.7_{(8.58)}$
GLR-E	$18.2_{(0.48)}$	$60.0_{(12.0)}$	$62.2_{(11.3)}$	$18.8_{(0.81)}$	$57.6_{(12.2)}$	$59.4_{(11.3)}$
SR-E	$18.4_{(0.40)}$	$60.2_{(12.0)}$	$61.4_{(11.2)}$	$19.1_{(0.61)}$	$58.0_{(12.2)}$	$58.7_{(11.2)}$
Shewart	$16.2_{(1.47)}$	$1.2_{(0.43)}$	$0.0_{(0.07)}$	$17.5_{(1.57)}$	$1.0_{(0.33)}$	$0.0_{(0.05)}$

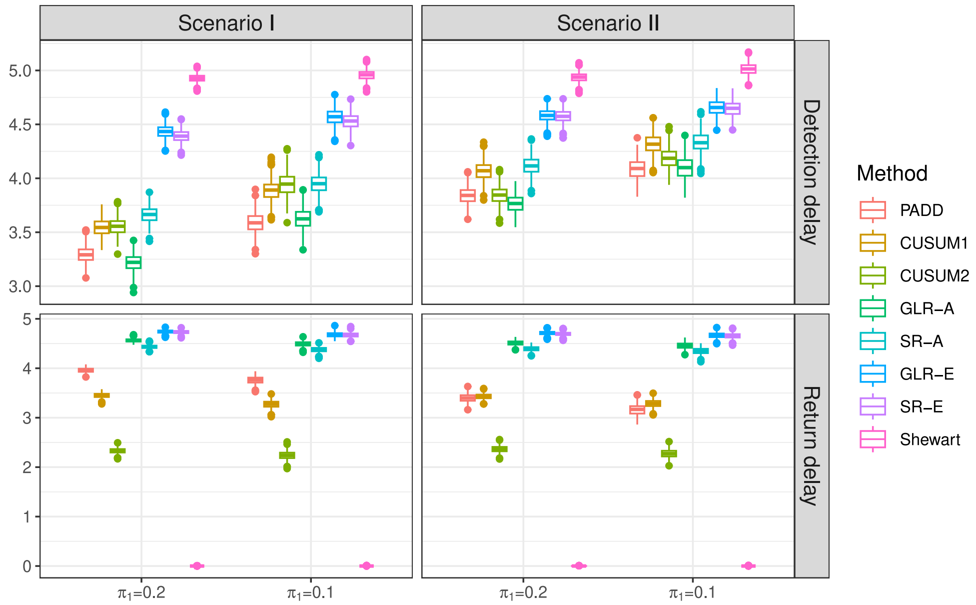

Fig. 3. Boxplots of the detection delay and return delay in log-scale based on 500 replications for the PADD, GLR, SR, Shewart, and modified CUSUM schemes in the setting of multivariate normal distribution.

Fig. 3 summarizes the detection delays and return delays into boxplots. Our PADD method is slightly inferior to CUSUMs on the return delay but provides lower detection delays than CUSUMs. The GLRs provide faster detections but slightly slower returns than SRs. The PADD outperforms the GLRs and SRs by a significant margin in both scenarios from the viewpoint of return delay. It is surprising that the PADD performs comparable with or even better than GLRs and SRs under these two scenarios in terms of detection delay. This may be partly due to the fact that the statistic $T_{i,t}^{e}$ is negative directly after the change-point $b_{i,k}$ with a high probability in this scenario, giving $T_{i,t}(\theta)$ a "boost" compared to the GLR statistic (3).

We also test the robust procedures introduced in Sections 4.3 and 5.1. We simulate $\mathbb{Z}_j$ with the same settings as before except from the standardized multivariate t-distribution with degrees of freedom three. The results on FDR, power, and pseudo power in this setting is listed in Table 2. The results on detection and return delays are shown in Fig. 4. The results are similar to those in the setting of multivariate normal distribution, so we omit the statements.

Table 2
Mean and standard deviation (in parentheses) of empirical FDR, power and pseudo-power (%) across $T = 2000$ time points for different π_1 with nominal size $\alpha = 20\%$ in the setting of multivariate t-distribution. R-PADD and its competitors (R-GLR-E, R-SR-E, R-GLR-A, R-SR-A) are the robust variant of PADD and competitors introduced in Section 5.1 with the nonparametric transformation technique in Section 4.3.

	$\pi_1 = 0.2$			$\pi_1 = 0.1$		
	FDR	Power	Pseudo-power	FDR	Power	Pseudo-power
Scenario I						
R-PADD	$17.9_{(1.35)}$	$37.0_{(11.00)}$	$4.3_{(5.06)}$	$20.9_{(1.49)}$	$34.0_{(10.22)}$	$4.4_{(4.82)}$
R-GLR-E	$17.3_{(1.40)}$	$37.1_{(11.1)}$	$17.3_{(16.9)}$	$19.2_{(1.61)}$	$34.0_{(10.3)}$	$16.7_{(15.9)}$
R-SR-E	$18.9_{(1.08)}$	$39.7_{(10.7)}$	$57.3_{(19.3)}$	$19.7_{(1.16)}$	$35.5_{(10.1)}$	$53.7_{(19.4)}$
R-GLR-A	$16.7_{(1.28)}$	$42.0_{(12.7)}$	$9.0_{(9.35)}$	$18.0_{(1.24)}$	$38.4_{(11.7)}$	$8.4_{(8.44)}$
R-SR-A	$18.6_{(1.00)}$	$44.0_{(11.9)}$	$60.7_{(19.6)}$	$19.4_{(1.12)}$	$39.5_{(11.1)}$	$56.5_{(19.6)}$
Scenario II						
R-PADD	$17.6_{(1.21)}$	$36.3_{(7.92)}$	$2.7_{(2.50)}$	$20.9_{(1.34)}$	$34.1_{(7.52)}$	$2.7_{(2.34)}$
R-GLR-E	$17.0_{(1.03)}$	$37.8_{(8.86)}$	$18.1_{(9.63)}$	$19.1_{(1.10)}$	$34.9_{(8.29)}$	$16.4_{(8.66)}$
R-SR-E	$18.4_{(1.13)}$	$39.4_{(9.30)}$	$57.3_{(11.1)}$	$19.6_{(1.14)}$	$36.1_{(8.69)}$	$53.3_{(10.8)}$
R-GLR-A	$16.7_{(1.18)}$	$41.7_{(9.33)}$	$7.3_{(5.68)}$	$18.0_{(1.21)}$	$39.1_{(8.90)}$	$6.3_{(4.49)}$
R-SR-A	$18.1_{(1.02)}$	$42.3_{(9.55)}$	$57.3_{(11.4)}$	$19.2_{(1.17)}$	$39.5_{(9.13)}$	$53.5_{(11.7)}$

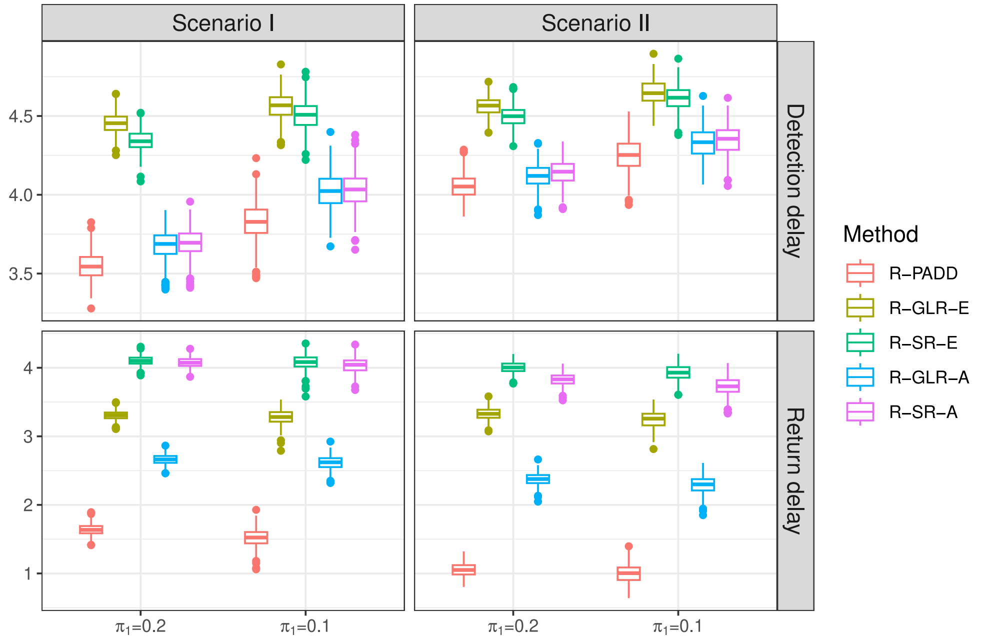

Fig. 4. Boxplots of the detection delay and return delay in log-scale based on 500 replications for the R-PADD, R-GLR and R-SR in the setting of multivariate t-distribution.

In conclusion, the PADD could provide considerable power as those procedures in literature while reducing the "false" discoveries (i.e., pseudo-power). From the viewpoints of detection/return delay, PADD can keep the ability of fast detection and have a small return delay. Some additional simulation results, including the case with independent data streams and the case when $\alpha = \beta = 0.1$, are provided in the Supplementary Material. Similar conclusions can be drawn.

5.3. Fund selection

As an illustrative example, we consider the case of an investor selecting skilled funds in the Chinese financial market. This dataset records all daily net values of the open-end equity funds $Y_{i,t}$ from October 2015 to December 2016 and consists of a total of 304 daily records (excluding holidays) on $m = 1149$ funds. The data from the 2015 calendar year, 61 days in total, are used to estimate the parameters, and the rest are used for testing. We compute the logarithmic function of the returns for each fund: $\tilde{Z}_{i,t} = \log(Y_{i,t}/Y_{i,t-1})$. Because fund performance is heavily affected by stock market volatility, amongst many other factors, we use the median of $\{\tilde{Z}_{1,t}, \ldots, \tilde{Z}_{m,t}\}$, denoted as $\widehat{\eta}_t$, and obtain the standardized observation $Z_{i,t} = (\tilde{Z}_{i,t} - \widehat{\eta}_t)/\widehat{\sigma}_i$, where $\widehat{\sigma}_i^2$ is the sample variance computed by the training sample. The main focus of this section is to

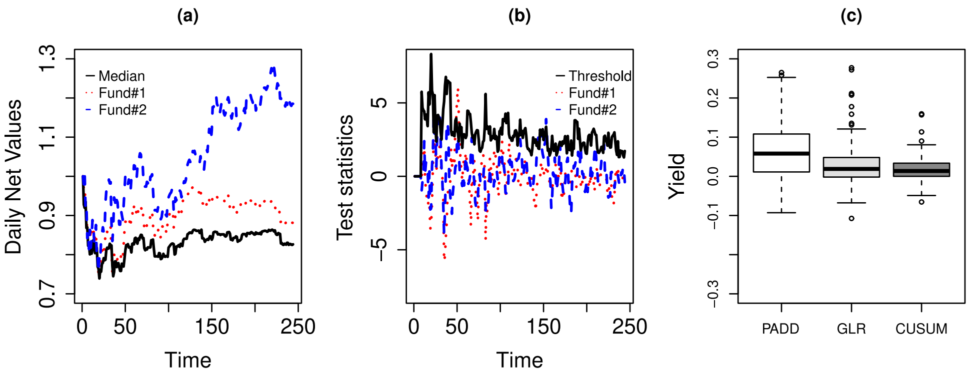

Fig. 5. (a) Daily net values of two selected funds in the dataset along with the median values of all of the funds; (b) Test statistics of PADD for the two selected funds along with the threshold value $\widehat{q}_{\alpha,t}(\widehat{\theta}_t)$; (c) Boxplot of the yields for the trading strategies based on PADD, GLR-A and CUSUM.

identify the states of funds, $\mu_{i,t} > 0$ or not, in model (1). Fig. 5(a) illustrates the original observations of the two selected funds along with the median values of all of the funds.

We treat the first 30 days of 2016 as a "warm-up" phase, meaning that we only compute the test statistics without triggering any signals during this period. From the simulations above, GLR-A provides better performance amongst GLRs, SRs and Shewart. Thus, we apply the PADD with a data-driven choice of the window size as in Section 4.5, GLR-A and CUSUM (with parameters $(k_i, H_i) = (0.25, 20)$) to identify skill funds in the remaining days with $\alpha = 0.2$ and $\beta = 0.2$. The results for other choices of α are similar. Fig. 5(b) presents the PADD statistics for the two selected funds considered in Fig. 5(a) along with the dynamic threshold values. The three methods make 952, 1814 and 1013 discoveries, respectively, over a total period of 213 days.

The true change-points are unknown, so it is impossible to compare the three methods in terms of their FDR, power, detection delay, and return delay as in the simulation study. However, we can evaluate their performance from an investor's perspective. For simplicity, the trading strategy is known as the "automatic investment plan", in which a predetermined amount of money is invested daily in funds identified with the testing methods. We hold each selected fund until the end and compute the average returns. Fig. 5(c) presents the boxplots of the returns for all three methods. The trading strategy based on the PADD produces more positive returns than that based on the CUSUM, which suggests that the reference values in the CUSUM may affect its performance to a degree. The PADD also performs better than the GLR in terms of average yields, whilst its variation is relatively large. This may be partly due to the data-driven choices of the penalization parameter θ and window size w in our method.

6. Concluding remarks

The proposed PADD framework makes dynamic discoveries for high-dimensional data streams while guaranteeing the FDR at each time point controlled. Some relevant directions can be considered for future research.

We permit some covariance structures across data streams for the proposed procedure, but as one referee pointed out, we construct the test statistic for each data stream separately. How to integrate the high-dimensional covariance to make the existence of covariance a blessing rather than a curse is still a hot topic and some developments made in multiple testing field [7,41] may be considered as a starting point.

The relaxation on the temporal dependence is challenging, especially in this setting of high-dimensional data streams. An intuitive idea is to block data temporally. Then, the temporal structure can be retained within each block and the dependence between blocks are expected to be small, so that the methodology in this article can be borrowed.

Last but not least, in practice, data streams often contain group structures. For example, in fMRI analysis, often, of interest are the regions of interest, which are the anatomical division of the brain, and each contains a lot of voxels in fMRI images. Thus, the analysis after grouping is necessary. A direct extension of the proposed procedure to such settings is to use multivariate GLR and CUSUM to formulate the PADD statistic. But both new theoretical and practical problems need to be solved in this complex setting.

CRediT authorship contribution statement

Lilun Du: Conceptualization, Methodology, Theory, Writing, Supervision. **Mengtao Wen:** Methodology, Writing, Simulations, Data analysis.

Acknowledgments

The authors are grateful to the referees, Associate Editor, and Editor for comments that have significantly improved the article. The authors thank Changliang Zou from Nankai University for some initial discussions. Lilun Du's research is supported by Hong Kong RGC-GRF 16302620. Mengtao Wen's research was supported by the China National Key R&D Program (Grant Nos. 2019YFC1908502, 2022YFA1003703, 2022YFA1003802, 2022YFA1003803) and NNSF of China Grants (Nos. 11925106, 12231011,11931001 and 11971247).

Appendix A

We provide the conditions used in the main text, the detailed proofs of Propositions 1, 2, Theorems 1, and 2 in this Appendix.

A.1. Conditions

For notational simplicity, we re-define the empirical processes in (7) as follows:

$$\widehat{F}_{0;t}(q,\theta)=\text{Card}(\mathcal{I}_{0;t})^{-1}V_t(q,\theta),\qquad \widehat{F}_{1;t}(q,\theta)=\text{Card}(\mathcal{A}_{1;t})^{-1}S_{1;t}(q,\theta),$$
$$\widehat{F}_{2;t}(q,\theta)=\text{Card}(\mathcal{B}_{1;t})^{-1}S_{2;t}(q,\theta),\qquad \widehat{F}_{t}(q,\theta)=\text{Card}(\mathcal{I}_{t})^{-1}R_t(q,\theta).$$

We require the following technical conditions.

- C1 The $\Sigma=(\sigma_{i_1,i_2})$ satisfies $\max_{i\in\{1,\ldots,p\}}\sum_{j=1}^{p}I(\sigma_{i,j}\neq 0)=o(m^{\delta})$, for some $\delta<1$.
- C2 $F_{0;t}(q,\theta)$, $F_{1;t}(q,\theta)$, $F_{2;t}(q,\theta)$, and $F_t(q,\theta)$ are continuously differentiable with respect to q and θ. In addition, their second-order derivatives with respect to q are uniformly bounded.
- C3 Let $p_{\alpha;t}(\theta)$ be the α lower percentile of the true null distribution $F_{0;t}(q,\theta)$. The estimator $\widehat{F}_{0;t}(p_{\alpha;t}(\theta),\theta)$ satisfies the Lipschitz continuity condition with respective to θ as follows: $\sup_{\alpha\in\mathbb{Q}}\sup_m|\widehat{F}_{0;t}(p_{\alpha;t}(\theta),\theta)-\widehat{F}_{0;t}(p_{\alpha;t}(\theta'),\theta')|\leq C|\theta-\theta'|$, where C is a constant that does not depend on α and m, and $\mathbb{Q}$ is the set of rational numbers. Moreover, $\widehat{F}_{1;t}(q,\theta)$, $\widehat{F}_{2;t}(q,\theta)$, and $\widehat{F}_{t}(q,\theta)$ satisfy the Lipschitz continuity condition.
- C4 Assume that $\left.\frac{\partial \text{Fdr}_t(q,\theta)}{\partial q}\right|_{q=q_{\alpha;t}(\theta)}\neq 0$ for each θ.
- C5 There exists some $\delta>0$ such that $F_t(q_{\alpha;t}(\theta'),\theta')>F_t(q_{\alpha;t}(\theta_{0;t}),\theta_{0;t})$ for $\theta'\in(\theta_{0;t}-\delta,\theta_{0;t})$.

Remark 1. Condition C1 is closely related to its variant assumed under multivariate normal case that the average of the correlation coefficients converges to zero at a polynomial rate. Similar conditions are popular in the literature, including that in [12]. If the correlation matrix contains many non-zero entries, this condition may not hold; a certain degree of sparseness of Σ is needed. Under the normality condition, the joint distribution function of $(T^b_{i,t},T^e_{i,t})$ is continuously differentiable of all orders. Because $T_{i,t}(\theta)$ is simply a linear combination of $T^b_{i,t}$ and $T^e_{i,t}$, $F_{0;t}(q,\theta)$, $F^i_{1;t}(q,\theta)$ and $F^i_{2;t}(q,\theta)$ are continuously differentiable of all orders with respect to q and θ. As a result, Condition C2 holds under homoscedasticity that $[b_{i,k},e_{i,k}]$ are the same across the data streams, and the signal strength remains constant within and across the alternative periods. In fact, Condition C2 also holds when the change points and the signal magnitude are sampled from some prior distributions. Condition C3 is directly borrowed from [8] to verify the uniform consistency of the empirical processes in (7) with respect to θ. Condition C4 is a technical one to ensure that the true FDR curve is monotone around the significance level α; see also Theorem 5 of [36]. This condition is weaker than the monotone likelihood ratio condition in [37]. Condition C5 implies that the overall power function deceases with respect to θ, so that a fixed percentage of power loss in the penalization method (13) can be achieved. We need this condition to guarantee the almost surely uniform convergence of the empirical distributions.

We also give a simplified scenario where Condition C2 and C4–C5 can be verified in the following proposition.

Proposition 3. *Consider the setting, where $[b,e]$ is the homogeneous change period across data streams in the alternative set. Assume $e-b\to\infty$, $t-e\to\infty$, and $0<\lim_{t\to\infty}(e-b+1)/(t-b+1)\to\gamma<1$ as $t\to\infty$. In addition, we consider the case with $w\leq(e-b+1)$, $\lambda=+\infty$ and suppose the change magnitude μ satisfies $\lim_{t\to\infty}\sqrt{t-b+1}\mu/(\log\log w)^{1/2}\to\infty$. Then, Condition C2 and C4–C5 hold.*

Appendix B. Proofs of theorems and propositions

Proof of Proposition 1. To prove Proposition 1, we first supply a strong law of large number for weakly dependent data from [20].

Lemma 1. *Let $\{x_i\}_{i=1}^{\infty}$ be a sequence of real-valued random variables such that $|x_i|\leq 1$ and $\sum_{m\geq 1}\frac{1}{m}E\{\frac{1}{m}\sum_{i\leq m}x_i\}^2<\infty$. Then $\lim_{m\to\infty}\frac{1}{m}\sum_{i\leq m}x_i=0$ almost surely.*

As the proofs of (i) and (ii) of Proposition 1 are similar, we only prove (i). We first prove the results under the modified Condition C1 and normality assumption.

For ease of exposition, we define a function $\mathbb{F}(.)$ that inputs the data $\{Z_{i,j}, j \in \{t-w,\ldots,t\}\}$ and outputs the test statistic $T_{i,t}(\theta)$, that is

$$\mathbb{F}(Z_{i,j}, j \in \{t-w,\ldots,t\}) = T_{i,t}^b - \theta \times T_{i,t}^e = \max_{\tau\in\{0,\ldots,w\}} \frac{\sum_{j=t-\tau}^{t} Z_{i,j}}{\sqrt{\tau+1}} - \theta \times \max_{\tau\in\{0,\ldots,w-1\}} \frac{\sum_{j=t-w}^{t-\tau-1} Z_{i,j}/(w-\tau) - \sum_{j=t-\tau}^{t} Z_{i,j}/(\tau+1)}{\sqrt{1/(w-\tau)+1/(\tau+1)}}.$$

The expectation of $\mathrm{Card}(\mathcal{I}_{0;t})^{-1}V_t(q,\theta)$ can be derived as

$$\mathrm{E}\{\mathrm{Card}(\mathcal{I}_{0;t})^{-1}V_t(q,\theta)\} = \Pr\{T_{i,t}(\theta) \ge q \mid i \in \mathcal{I}_{0;t}\} = \Pr\{\mathbb{F}(Z_{i,j}, j \in \{t-w,\ldots,t\}) \ge q \mid i \in \mathcal{I}_{0;t}\} = F_{0;t}(q,\theta).$$

To apply Lemma 1, we set $x_i = \mathrm{I}(T_{i,t}(\theta) \ge q) - F_{0;t}(q,\theta)$. Then, the conclusion (i) holds if we can show that

$$\mathrm{Var}\{m_{0;t}^{-1}V_t(q,\theta)\} = O(m_{0;t}^{-\delta}), \quad \text{for some } \delta > 0, \tag{B.1}$$

where $m_{0;t} = \mathrm{Card}(\mathcal{I}_{0;t})$. The variance can be further expressed as

$$\mathrm{Var}\{m_{0;t}^{-1}V_t(q,\theta)\} = m_{0;t}^{-2}\sum_{i\in\mathcal{I}_{0;t}} \mathrm{Var}\{\mathrm{I}(T_{i,t}(\theta)\ge q)\} + m_{0;t}^{-2}\sum_{i_1\ne i_2} \mathrm{Cov}\{\mathrm{I}(T_{i_1,t}(\theta)\ge q), \mathrm{I}(T_{i_2,t}(\theta)\ge q) \mid i_1 \in \mathcal{I}_{0;t}, i_2 \in \mathcal{I}_{0;t}\}.$$

The variance term is simply bounded by $O(m_{0;t}^{-1})$ due to the fact that $\mathrm{Var}\{\mathrm{I}(T_{i,t}(\theta)\ge q)\} \le 1/4$. The covariance term can be expanded as

$$\begin{aligned} &\mathrm{Cov}\{\mathrm{I}(T_{i_1,t}(\theta)\ge q), \mathrm{I}(T_{i_2,t}(\theta)\ge q) \mid i_1 \in \mathcal{I}_{0;t}, i_2 \in \mathcal{I}_{0;t}\} \\ &\quad = \Pr\{T_{i_1,t}(\theta)\ge q, T_{i_2,t}(\theta)\ge q \mid i_1 \in \mathcal{I}_{0;t}, i_2 \in \mathcal{I}_{0;t}\} - \{F_{0;t}(q,\theta)\}^2. \end{aligned} \tag{B.2}$$

By the bivariate normality of $\{Z_{i_1,j}, Z_{i_2,j}\}$, we decompose the i_1-th and i_2-th datastreams within the window $[t-w,t]$ as $Z_{i_1,j} = \sqrt{\rho_{i_1,i_2}}C_j + \sqrt{1-\rho_{i_1,i_2}}U_j$ and $Z_{i_2,j} = \sqrt{\rho_{i_1,i_2}}C_j + \sqrt{1-\rho_{i_1,i_2}}V_j$, for $j \in \{t-w,\ldots,t\}$; the case that $\rho_{i_1,i_2} \le 0$ can be discussed similarly. Here $\{U_j, V_j, C_j, j \in \{t-w,\ldots,t\}\}$ are independent standard normal random variables. Hence, the first term of (B.2) can be derived as

$$\begin{aligned} &\Pr\{T_{i_1,t}(\theta)\ge q, T_{i_2,t}(\theta)\ge q \mid i_1 \in \mathcal{I}_{0;t}, i_2 \in \mathcal{I}_{0;t}\} \\ &\quad = \Pr\{\mathbb{F}(Z_{i_1,j}, j \in \{t-w,\ldots,t\}) \ge q, \mathbb{F}(Z_{i_2,j}, j \in \{t-w,\ldots,t\}) \ge q, \mid i_1 \in \mathcal{I}_{0;t}, i_2 \in \mathcal{I}_{0;t}\} \\ &\quad = \Pr\{\mathbb{F}(U_j, j \in \{t-w,\ldots,t\}) \ge \bar{q}(\sqrt{\rho_{i_1,i_2}}), \mathbb{F}(V_j, j \in \{t-w,\ldots,t\}) \ge \bar{q}(\sqrt{\rho_{i_1,i_2}}) \mid i_1 \in \mathcal{I}_{0;t}, i_2 \in \mathcal{I}_{0;t}\} \\ &\quad = \int\cdots\int \{F_{0;t}(\bar{q}(\sqrt{\rho_{i_1,i_2}}),\theta)\}^2 dC_{t-w}\ldots dC_t, \end{aligned} \tag{B.3}$$

where $\bar{q}(\sqrt{\rho_{i_1,i_2}}) = \frac{q-\sqrt{\rho_{i_1,i_2}}\mathbb{F}(C_j, j\in\{t-w,\ldots,t\})}{\sqrt{1-\rho_{i_1,i_2}}}$. Note that when $\rho_{i_1,i_2} \to 0$, $\bar{q}(\sqrt{\rho_{i_1,i_2}}) \to q$ and $F_{0;t}(\bar{q}(\sqrt{\rho_{i_1,i_2}}),\theta) \to F_{0;t}(q,\theta)$. Hence, applying Taylor expansion to $F_{0;t}(\bar{q}(\sqrt{\rho_{i_1,i_2}}),\theta)$ with respect to $\sqrt{\rho_{i_1,i_2}}$ yields that

$$\begin{aligned} F_{0;t}(\bar{q}(\sqrt{\rho_{i_1,i_2}}),\theta) &= F_{0;t}(q,\theta) + \frac{\partial F_{0;t}}{\partial \bar{q}}\frac{\partial \bar{q}}{\partial\sqrt{\rho_{i_1,i_2}}}\Big|_{\sqrt{\rho_{i_1,i_2}}=0}\sqrt{\rho_{i_1,i_2}} \\ &\quad + \frac{1}{2}\{\frac{\partial^2 F_{0;t}}{\partial \bar{q}^2}\{\frac{\partial \bar{q}}{\partial\sqrt{\rho_{i_1,i_2}}}\}^2 + \frac{\partial F_{0;t}}{\partial \bar{q}}\frac{\partial^2 \bar{q}}{\partial\sqrt{\rho_{i_1,i_2}}^2}\}\Big|_{\sqrt{\rho_{i_1,i_2}}=0}\{\sqrt{\rho_{i_1,i_2}}\}^2 + o(|\rho_{i_1,i_2}|), \end{aligned}$$

where $\frac{\partial \bar{q}}{\partial\sqrt{\rho_{i_1,i_2}}}\Big|_{\sqrt{\rho_{i_1,i_2}}=0} = -\mathbb{F}(C_j, j \in \{t-w,\ldots,t\})$ and $\frac{\partial^2 \bar{q}}{\partial\sqrt{\rho_{i_1,i_2}}^2}\Big|_{\sqrt{\rho_{i_1,i_2}}=0} = q$. By the fact that $\int\ldots\int \mathbb{F}(C_j, j \in \{t-w,\ldots,t\})dC_{t-w}\ldots dC_t = 0$, the term in (B.3) can be bounded by

$$\{F_{0;t}(q,\theta)\}^2 + \{\frac{\partial F_{0;t}}{\partial \bar{q}}\}^2\Big|_{\sqrt{\rho_{i_1,i_2}}=0} \times Moment_2 \times |\rho_{i_1,i_2}| + F_{0;t}(q,\theta) \times \{\frac{\partial^2 F_{0;t}}{\partial \bar{q}^2} \times Moment_2 + \frac{\partial F_{0;t}}{\partial \bar{q}} \times q\}\Big|_{\sqrt{\rho_{i_1,i_2}}=0}|\rho_{i_1,i_2}| + o(|\rho_{i_1,i_2}|), \tag{B.4}$$

where $Moment_2 = \int\ldots\int\{\mathbb{F}(C_j, j \in \{t-w,\ldots,t\})\}^2 dC_{t-w}\ldots dC_t$ is the second moment of $\mathbb{F}(C_j, j \in \{t-w,\ldots,t\})$. By Condition C2, the coefficients in (B.4) are bounded and the bound does not depend on i_1 and i_2. Consequently, the covariance in (B.2) is bounded by $O(|\rho_{i_1,i_2}|)$. This together with the modified Condition C1 implies that (B.1) is satisfied. The proof of (i) is completed.

Without normality assumption, it is easy to verify that (B.1) holds under Condition C1. □

Proof of Theorem 1. Before we prove Theorem 1, we strengthen the conclusion in Proposition 1 as in the following lemma.

Lemma 2. *Suppose the conditions in Proposition 1 and conditions C1–C3 in the Appendix hold. Then,*

$$\sup_q \sup_{0\le\theta\le\bar\theta} |\widehat{F}_{0;t}(q,\theta) - F_{0;t}(q,\theta)| \xrightarrow{a.s.} 0, \quad \sup_q \sup_{0\le\theta\le\bar\theta} |\widehat{F}_{1;t}(q,\theta) - F_{1;t}(q,\theta)| \xrightarrow{a.s.} 0,$$

$$\sup_q \sup_{0\le\theta\le\bar\theta} |\widehat{F}_{2;t}(q,\theta) - F_{2;t}(q,\theta)| \xrightarrow{a.s.} 0, \quad \sup_q \sup_{0\le\theta\le\bar\theta} |\widehat{F}_{t}(q,\theta) - F_{t}(q,\theta)| \xrightarrow{a.s.} 0,$$

where $0 < \bar\theta < \infty$.

Proof. Without loss of generality, let $\bar\theta = 1$. We extend the standard techniques used in *Clivenko–Cantelli* theorem [10] to prove this lemma. Let $q_{j/k}(\theta)$ be the j/k-th lower percentile of $F_{0;t}(q,\theta)$, for $j \in \{0,\dots,k\}$. We further partition the domain of θ into L equally lengths of intervals $\cup_{\ell=1}^{L}[\theta_{\ell-1},\theta_\ell]$ such that $|\theta_{\ell-1}-\theta_\ell| \le 1/Ck$, where C is the constant in Condition C3. By Proposition 1, there exists a sufficiently large m_k such that when $m > m_k$,

$$\left|\widehat{F}_{0;t}(q_{j/k}(\theta_\ell),\theta_\ell) - F_{0;t}(q_{j/k}(\theta_\ell),\theta_\ell)\right| < 1/k,$$

for $j \in \{0,\dots,k\}$ and $\ell \in \{0,\dots,L\}$. Then for any pair of value $(q,\theta) \in [q_{(j-1)/k}(\theta), q_{j/k}(\theta)] \times [\theta_{\ell-1},\theta_\ell]$, $\widehat{F}_{0;t}(q,\theta)$ can be upper bounded by

$$\begin{aligned}\widehat{F}_{0;t}(q,\theta) &\le \widehat{F}_{0;t}(q_{j/k}(\theta),\theta) \le \widehat{F}_{0;t}(q_{j/k}(\theta_{\ell-1}),\theta_{\ell-1}) + 1/k \le F_{0;t}(q_{j/k}(\theta_{\ell-1}),\theta_{\ell-1}) + 2/k \\ &= F_{0;t}(q_{(j-1)/k}(\theta_{\ell-1}),\theta_{\ell-1}) + 3/k = F_{0;t}(q_{(j-1)/k}(\theta),\theta) + 3/k \le F_{0;t}(q,\theta) + 3/k.\end{aligned}$$

Similarly, we can have $\widehat{F}_{0;t}(q,\theta) \ge F_{0;t}(q,\theta) - 3/k$. Hence, $\sup_q \sup_\theta |\widehat{F}_{0;t}(q,\theta) - F_{0;t}(q,\theta)| \le 3/k$ when $m > m_k$. This shows that $\widehat{F}_{0;t}(q,\theta)$ uniformly converges to $F_{0;t}(q,\theta)$ almost surely. Similar results for $\widehat{F}_{1;t}(q,\theta)$, $\widehat{F}_{2;t}(q,\theta)$, and $\widehat{F}_t(q,\theta)$ can be obtained. □

Now we use Lemma 2 to prove Theorem 1. Some calculations yield that

$$\begin{aligned}\lim_{m\to\infty} \inf_{q\le\bar q} \inf_{0\le\theta\le\bar\theta} \left\{\widehat{\mathrm{FDR}}_{\lambda;t}(q,\theta) - \mathrm{FDP}_t(q,\theta)\right\} &\ge \lim_{m\to\infty} \inf_{q\le\bar q} \inf_{0\le\theta\le\bar\theta} \left\{\widehat{\mathrm{FDR}}_{\lambda;t}(q,\theta) - \frac{m\pi_{0;t}F_{0;t}(q,\theta)}{R_t(q,\theta)\vee 1}\right\} \\ &\quad - \lim_{m\to\infty} \sup_{q\le\bar q} \sup_{0\le\theta\le\bar\theta} \left|\frac{m\pi_{0;t}F_{0;t}(q,\theta)}{R_t(q,\theta)\vee 1} - \mathrm{FDP}_t(q,\theta)\right|.\end{aligned} \tag{B.5}$$

By Lemma 2, we have $\lim_m \inf_{0\le\theta\le\bar\theta}\{\widehat{\pi}_{0;t}(\lambda,\theta) - \pi_{0;t}\} \ge 0$ almost surely. Hence, the first term in (B.5) is nonnegative asymptotically. The second term can be derived as

$$\begin{aligned}\lim_{m\to\infty} \sup_{q\le\bar q} \sup_{0\le\theta\le\bar\theta} \left|\frac{m\pi_{0;t}F_{0;t}(q,\theta)}{R_t(q,\theta)\vee 1} - \frac{V_t(q,\theta)}{R_t(q,\theta)\vee 1}\right| &\le \lim_{m\to\infty} \sup_{0\le\theta\le\bar\theta} \frac{m}{R_t(\bar q,\theta)} \times \lim_{m\to\infty} \sup_{q\le\bar q} \sup_{0\le\theta\le\bar\theta} \left|V_t(q,\theta)/m - \pi_{0;t}F_{0;t}(q,\theta)\right| \\ &= \frac{1}{\inf_{0\le\theta\le\bar\theta} F_t(\bar q,\theta)} \times \lim_{m\to\infty} \sup_{q\le\bar q} \sup_{0\le\theta\le\bar\theta} \left|V_t(q,\theta)/m - \pi_{0;t}F_{0;t}(q,\theta)\right|.\end{aligned}$$

By normality assumption, $F_{0;t}(q,\theta)$ is continuously differentiable of all orders with respect to q and θ, implying that $\inf_{0\le\theta\le\bar\theta} F_{0;t}(\bar q,\theta) > 0$. Therefore,

$$\inf_{0\le\theta\le\bar\theta} F_t(\bar q,\theta) \ge \inf_{0\le\theta\le\bar\theta} \pi_{0;t}F_{0;t}(\bar q,\theta) > 0. \tag{B.6}$$

According to Lemma 2, the second term converges to zero uniformly for q and θ. Combining the above results, (B.5) is nonnegative asymptotically. To prove (14), it suffices to show that

$$\lim_{m\to\infty} \sup_{q\le\bar q} \sup_{0\le\theta\le\bar\theta} \left|\mathrm{FDP}_t(q,\theta) - \mathrm{FDR}_t(q,\theta)\right| = 0 \quad a.s. \tag{B.7}$$

(B.7) holds if we can verify the following two convergence results, that is,

$$\lim_{m\to\infty} \sup_{q\le\bar q} \sup_{0\le\theta\le\bar\theta} \left|\mathrm{FDP}_t(q,\theta) - \mathrm{Fdr}_t(q,\theta)\right| = 0 \text{ a.s.}, \quad \lim_{m\to\infty} \sup_{q\le\bar q} \sup_{0\le\theta\le\bar\theta} \left|\mathrm{Fdr}_t(q,\theta) - \mathrm{FDR}_t(q,\theta)\right| = 0 \text{ a.s.}$$

The proof of 2 can be derived in a similar way as in [36]. The proof of Theorem 1 is hence completed. □

Proof of Proposition 2. We first provide the following lemma regarding the consistency of the data-driven threshold $\widehat{q}_{\alpha;t}(\theta)$.

Lemma 3. *Assume Conditions C1–C4 hold. Then, we have*

$$\sup_{0\le\theta\le\bar\theta} \left|\widehat{q}_{\alpha;t}(\theta) - q_{\alpha;t}(\theta)\right| \xrightarrow{a.s.} 0.$$

Proof. For any fixed $\delta_1 > 0$, let $\check{q}(\theta)$ be any curve that $\check{q}(\theta) < q_{\alpha;t}(\theta) - \delta_1$. Then, $\widehat{\mathrm{FDR}}_{\lambda;t}(\check{q}(\theta), \theta)$ can be lower bounded by

$$\begin{aligned}\frac{m\widehat{\pi}_{0;t}(\lambda,\theta)F_{0;t}(\check{q}(\theta),\theta)}{R_t(\check{q}(\theta),\theta)\vee 1} &\geq \frac{k(\theta)\pi_{0;t}F_{0;t}(\check{q}(\theta),\theta) - |\widehat{\pi}_{0;t}(\lambda,\theta) - k(\theta)\pi_{0;t}|F_{0;t}(\check{q}(\theta),\theta)}{F_t(\check{q}(\theta),\theta) + |R_t(\check{q}(\theta),\theta)/m - F_t(\check{q}(\theta),\theta)|} \\ &= \frac{k(\theta)\pi_{0;t}F_{0;t}(\check{q}(\theta),\theta)/F_t(\check{q}(\theta),\theta) - |\widehat{\pi}_{0;t}(\lambda,\theta) - k(\theta)\pi_{0;t}|F_{0;t}(\check{q}(\theta),\theta)/F_t(\check{q}(\theta),\theta)}{1 + |R_t(\check{q}(\theta),\theta)/m - F_t(\check{q}(\theta),\theta)|/F_t(\check{q}(\theta),\theta)} \\ &\geq \frac{k(\theta)\pi_{0;t}F_{0;t}(\check{q}(\theta),\theta)/F_t(\check{q}(\theta),\theta) - \epsilon_1}{1+\epsilon_2} = \frac{k(\theta)\mathrm{Fdr}_t(\check{q}(\theta),\theta) - \epsilon_1}{1+\epsilon_2},\end{aligned}$$

where

$$\begin{aligned}\epsilon_1 &= \sup_{q\leq q_{\alpha;t}(\theta)-\delta_1}\ \sup_{0\leq\theta\leq\bar{\theta}} |\widehat{\pi}_{0;t}(\lambda,\theta) - k(\theta)\pi_{0;t}|F_{0;t}(q,\theta)/F_t(q,\theta), \\ \epsilon_2 &= \sup_{q\leq q_{\alpha;t}(\theta)-\delta_1}\ \sup_{0\leq\theta\leq\bar{\theta}} |R_t(q,\theta)/m - F_t(q,\theta)|/F_t(q,\theta).\end{aligned}$$

By (10), Condition C2, and implicit function theorem, $q_{\alpha;t}(\theta)$ is continuously differentiable with respect to θ. Moreover, it simply holds that $q_{\alpha;t}(\theta) < \infty$ for each θ; otherwise the true FDR cannot be achieved at α level. As a result, $\sup_{0\leq\theta\leq\bar{\theta}} q_{\alpha;t}(\theta) < \infty$. This together with Lemma 2 and (B.6) yields that $\epsilon_1 \xrightarrow{a.s.} 0$ and $\epsilon_2 \xrightarrow{a.s.} 0$. Note that for each θ, $k(\theta)\mathrm{Fdr}_t(\check{q}(\theta),\theta) > \alpha$; otherwise it contradicts the fact that $q_{\alpha;t}(\theta)$ is the infimum in the definition of (10). By Condition C2, one can have $\inf_{0\leq\theta\leq\bar{\theta}} k(\theta)\mathrm{Fdr}_t(\check{q}(\theta),\theta) > \alpha$. Hence, for a sufficiently large $M_1(\delta_1)$, when $m > M_1(\delta_1)$, it follows that $\inf_{0\leq\theta\leq\bar{\theta}} \widehat{\mathrm{FDR}}_{\lambda;t}(\check{q}(\theta),\theta) > \alpha$ almost surely. This implies that $\widehat{q}_{\alpha;t}(\theta) \geq q_{\alpha;t}(\theta) - \delta_1$ for all θ.

We now prove a similar inequality on the other side. By Condition C2, the partial derivative

$$s(q,\theta) = \frac{\partial \mathrm{Fdr}_t(q,\theta)}{\partial q}$$

is continuously differentiable with respect to q and θ. It follows that $s(q_{\alpha;t}(\theta),\theta)$ must be negative according to Condition C4; otherwise, $q_{\alpha;t}(\theta)$ cannot be the true infimum for all q that $k(\theta)\mathrm{Fdr}_t(q,\theta) \leq \alpha$. By continuity, there exists a sufficiently small $\delta_1 > 0$ such that $\inf_{|q-q_{\alpha;t}(\theta)|\leq\delta_1} \inf_{0\leq\theta\leq\bar{\theta}} |s(q,\theta)| > 0$. For $\tilde{q}(\theta) \in [q_{\alpha;t}(\theta) + \delta_1/2, q_{\alpha;t}(\theta) + \delta_1]$, Taylor expansion leads to

$$k(\theta)\mathrm{Fdr}_t(\tilde{q}(\theta),\theta) = k(\theta)\mathrm{Fdr}_t(q_{\alpha;t}(\theta),\theta) + \{\tilde{q}(\theta) - q_{\alpha;t}(\theta)\}k(\theta)s(q',\theta) \leq \alpha - \delta_1/2 \inf_{|q-q_{\alpha;t}(\theta)|\leq\delta_1}\ \inf_{0\leq\theta\leq\bar{\theta}} \{k(\theta)|s(q,\theta)|\} < \alpha.$$

Consequently, $\widehat{\mathrm{FDR}}_{\lambda;t}(\tilde{q}(\theta),\theta)$ is upper bounded as

$$\begin{aligned}\widehat{\mathrm{FDR}}_{\lambda;t}(\tilde{q}(\theta),\theta) = \frac{\widehat{\pi}_{0;t}(\lambda,\theta)F_{0;t}(\tilde{q}(\theta),\theta)}{\{R_t(\tilde{q}(\theta),\theta)\vee 1\}/m} &\leq \frac{k(\theta)\pi_{0;t}F_{0;t}(\tilde{q}(\theta),\theta) + |\widehat{\pi}_{0;t}(\lambda,\theta) - k(\theta)\pi_{0;t}|F_{0;t}(\tilde{q}(\theta),\theta)}{F_t(\tilde{q}(\theta),\theta) - |R_t(\tilde{q}(\theta),\theta)/m - F_t(\tilde{q}(\theta),\theta)|} \\ &= \frac{k(\theta)\mathrm{Fdr}_t(\tilde{q}(\theta),\theta) + \epsilon_3}{1-\epsilon_4} < \frac{\alpha+\epsilon_3}{1-\epsilon_4},\end{aligned}$$

where

$$\begin{aligned}\epsilon_3 &= \sup_{q\leq q_{\alpha}(\theta)+\delta_1}\ \sup_{0\leq\theta\leq\bar{\theta}} |\widehat{\pi}_{0;t}(\lambda,\theta) - k(\theta)\pi_{0;t}|F_{0;t}(q,\theta)/F_t(q,\theta), \\ \epsilon_4 &= \sup_{q\leq q_{\alpha}(\theta)+\delta_1}\ \sup_{0\leq\theta\leq\bar{\theta}} |R_t(q,\theta)/m - F_t(q,\theta)|/F_t(q,\theta).\end{aligned}$$

Analogously, one can show that $\epsilon_3 \xrightarrow{a.s.} 0$ and $\epsilon_4 \xrightarrow{a.s.} 0$. As a result, there exists a sufficiently large $M_2(\delta_1)$ such that when $m > M_2(\delta_1)$, $\sup_{0\leq\theta\leq\bar{\theta}} \widehat{\mathrm{FDR}}_{\lambda;t}(\tilde{q}(\theta),\theta) < \alpha$ for $\tilde{q}(\theta) \in [q_{\alpha;t}(\theta) + \delta_1/2, q_{\alpha;t}(\theta) + \delta_1]$. By the selection rule of $\widehat{q}_{\alpha;t}(\theta)$, $\widehat{q}_{\alpha;t}(\theta) < q_{\alpha;t}(\theta) + \delta_1$ for all θ when $m > M_2(\delta_1)$. Combining this and the previous result, we obtain that $\sup_{0\leq\theta\leq\bar{\theta}} |\widehat{q}_{\alpha;t}(\theta) - q_{\alpha;t}(\theta)| \xrightarrow{a.s.} 0$. □

Now we proceed to prove Proposition 2. We first show that $\widehat{F}_t(\widehat{q}_{\alpha;t}(\theta),\theta)$ converges to $F_t(q_{\alpha;t}(\theta),\theta)$ almost surely. Based on Condition C2 and the results from Proposition 1, and Lemma 3, we have

$$\begin{aligned}|\widehat{F}_t(\widehat{q}_{\alpha;t}(\theta),\theta) - F_t(q_{\alpha;t}(\theta),\theta)| &\leq |\widehat{F}_t(\widehat{q}_{\alpha;t}(\theta),\theta) - F_t(\widehat{q}_{\alpha;t}(\theta),\theta)| + |F_t(\widehat{q}_{\alpha;t}(\theta),\theta) - F_t(q_{\alpha;t}(\theta),\theta)| \\ &\leq \sup_q |\widehat{F}_t(q,\theta) - F_t(q,\theta)| + O(|\widehat{q}_{\alpha;t}(\theta) - q_{\alpha;t}(\theta)|) \xrightarrow{a.s.} 0.\end{aligned} \tag{B.8}$$

The consistency of $\widehat{\theta}_t$ can be obtained by (B.8) and Theorem 5 in [36] as follows. By definition of $\theta_{0;t}$, for each $\theta' > \theta_{0;t}$, there exists some $\varepsilon > 0$ such that $F_t(q_{\alpha;t}(\theta'),\theta')/F_t(q_{\alpha;t}(0),0) = 1 - \beta - \varepsilon$. Thus, by (B.8), we can take m sufficient large that

$$\left|\widehat{F}_t(\widehat{q}_{\alpha;t}(\theta'),\theta')/\widehat{F}_t(\widehat{q}_{\alpha;t}(0),0) - F_t(q_{\alpha;t}(\theta'),\theta')/F_t(q_{\alpha;t}(0),0)\right| < \varepsilon/2,$$

and thus $\widehat{F}_t(\widehat{q}_{\alpha;t}(\theta'),\theta')/\widehat{F}_t(\widehat{q}_{\alpha;t}(0),0) < 1-\beta$ eventually with probability one. Hence, $\limsup_{m\to\infty}\widehat{\theta}_t \le \theta_{0;t}$ almost surely. By Condition C5, there is a neighborhood of size $\delta > 0$, such that, for $\theta' \in [\theta_{0;t}-\delta,\theta_{0;t})$, we have $F_t(q_{\alpha;t}(\theta'),\theta')/F_t(q_{\alpha;t}(0),0) > F_t(q_{\alpha;t}(\theta_{0;t}),\theta_{0;t})/F_t(q_{\alpha;t}(0),0)$. By a similar argument to that for the previous one, we have that, for any θ' in this neighborhood, $\widehat{F}_t(\widehat{q}_{\alpha;t}(\theta'),\theta')/\widehat{F}_t(\widehat{q}_{\alpha;t}(0),0) > 1-\beta$ eventually with probability one. Thus $\liminf_{m\to\infty}\widehat{\theta}_t \ge \theta_{0;t}$ almost surly. Putting these together, we have the results. □

Proof of Theorem 2. To justify Theorem 2, we provide Lemma 4.

Lemma 4. *Under Conditions C1–C5, we have for any q,*

$$\lim_{m\to\infty}|\widehat{F}_{0;t}(q,\widehat{\theta}_t)-F_{0;t}(q,\theta_{0;t})| \xrightarrow{a.s.} 0,\quad \lim_{m\to\infty}|\widehat{F}_t(q,\widehat{\theta}_t)-F_t(q,\theta_{0;t})| \xrightarrow{a.s.} 0.$$

Proof. We prove the consistency of $\widehat{F}_{0;t}(q,\widehat{\theta}_t)$, while the result for $\widehat{F}_t(q,\widehat{\theta}_t)$ is similar. By derivation,

$$|\widehat{F}_{0;t}(q,\widehat{\theta}_t)-F_{0;t}(q,\theta_{0;t})| \le |\widehat{F}_{0;t}(q,\widehat{\theta}_t)-F_{0;t}(q,\widehat{\theta}_t)| + |F_{0;t}(q,\widehat{\theta}_t)-F_{0;t}(q,\theta_{0;t})|.$$

According to Proposition 2 and Condition C2, the second term is asymptotically negligible. For the first term, it is upper bounded by $\sup_{0\le\theta\le\bar{\theta}}|\widehat{F}_{0;t}(q,\theta)-F_{0;t}(q,\theta)|$, which converges to zero by Lemma 2. This completes the proof of Lemma 4. □

Now we prove Theorem 2. By Lemma 4, $\widehat{\mathrm{FDR}}_{\lambda;t}(q,\widehat{\theta}_t)$ converges to $k(\theta)\mathrm{Fdr}_t(q,\theta_{0;t})$ almost surely. By Condition C2, There exists a $q' < \infty$, such that $\varepsilon = \alpha - k(\theta)\mathrm{Fdr}_t(q',\theta_{0;t}) > 0$. We can take a sufficiently large m such that $|\widehat{\mathrm{FDR}}_{\lambda;t}(q',\widehat{\theta}_t)-k(\theta)\mathrm{Fdr}_t(q',\theta_{0;t})| \le \varepsilon/2$, which implies that $\widehat{\mathrm{FDR}}_{\lambda;t}(q',\widehat{\theta}_t) < \alpha$ and $\widehat{q}_{\alpha;t}(\widehat{\theta}_t) < q'$. Therefore, $\limsup_{m\to\infty}\widehat{q}_{\alpha;t}(\widehat{\theta}_t) \le q'$ with probability one. By Theorem 1, we have

$$\liminf_{m\to\infty}\left[\widehat{\mathrm{FDR}}_{\lambda;t}(\widehat{q}_{\alpha;t}(\widehat{\theta}_t),\widehat{\theta}_t)-\mathrm{FDP}_t(\widehat{q}_{\alpha;t}(\widehat{\theta}_t),\widehat{\theta}_t)\right] \ge \lim_{m\to\infty}\inf_{q\le 2q'}\inf_{0\le\theta\le 3}\left[\widehat{\mathrm{FDR}}_{\lambda;t}(q,\theta)-\mathrm{FDP}_t(q,\theta)\right] \ge 0.$$

As $\widehat{\mathrm{FDR}}_{\lambda;t}(\widehat{q}_{\alpha;t}(\widehat{\theta}_t)) \le \alpha$, it follows that $\limsup_{m\to\infty}\mathrm{FDP}_t(\widehat{q}_{\alpha;t}(\widehat{\theta}_t),\widehat{\theta}_t) \le \alpha$ with probability one. By Fatou's lemma,

$$\limsup_{m\to\infty} E\left\{\mathrm{FDP}_t(\widehat{q}_{\alpha;t}(\widehat{\theta}_t),\widehat{\theta}_t)\right\} \le E\left\{\limsup_{m\to\infty}\mathrm{FDP}_t(\widehat{q}_{\alpha;t}(\widehat{\theta}_t),\widehat{\theta}_t)\right\} \le \alpha.$$

This completes the proof of Theorem 2. □

Derivation of $\Delta(\theta)$ in (15):

$$\Delta(\theta) = \frac{\partial}{\partial\theta}\left\{F_{0;t}(q_{\alpha;t}(\theta),\theta)\right\}\Big|_{\theta=0} \times \frac{\theta}{F_{0;t}(q_{\alpha;t}(0),0)} = \left[\left\{\frac{\partial F_{0;t}(q,\theta)}{\partial q}\frac{\partial q_{\alpha;t}(\theta)}{\partial\theta} + \frac{\partial F_{0;t}(q,\theta)}{\partial\theta}\right\}\Big|_{q=q_{\alpha;t}(\theta)}\right]\Big|_{\theta=0} \times \frac{\theta}{F_{0;t}(q_{\alpha;t}(0),0)}. \tag{B.9}$$

Derivation similar to (A.6) in [8] yields that

$$\frac{\partial q_{\alpha;t}(\theta)}{\partial\theta} = \left\{\frac{\beta'\frac{\partial F_{0;t}(q,\theta)}{\partial\theta} - \frac{\partial F_{2;t}(q,\theta)}{\partial\theta}}{\frac{\partial F_{2;t}(q,\theta)}{\partial q} - \beta'\frac{\partial F_{0;t}(q,\theta)}{\partial q}}\right\}\Bigg|_{q=q_{\alpha;t}(\theta)}. \tag{B.10}$$

Plugging (B.10) into (B.9), $\Delta(\theta)$ can be expressed explicitly as

$$\Delta(\theta) = \frac{\left[\left\{\frac{\partial F_{0;t}(q,\theta)}{\partial\theta}\frac{\partial F_{2;t}(q,\theta)}{\partial q} - \frac{\partial F_{0;t}(t,\theta)}{\partial q}\frac{\partial F_{2;t}(q,\theta)}{\partial\theta}\right\}\Big|_{q=q_{\alpha;t}(\theta)}\right]\Big|_{\theta=0}}{\left[\left\{\frac{\partial F_{2;t}(q,\theta)}{\partial q} - \beta'\frac{\partial F_{0;t}(q,\theta)}{\partial q}\right\}\Big|_{q=q_{\alpha;t}(\theta)}\right]\Big|_{\theta=0}} \times \frac{\theta}{F_{0;t}(q_{\alpha;t}(0),0)}.$$

By a modification of Theorem 1.3.1 in [6], we have the asymptotic null distribution of $T^b_{i,t}$

$$\lim_{m\to\infty}\Pr\left\{A(\log w)T^b_{i,t} \le q + D(\log w)\right\} = \exp(-e^{-q})$$

for any q, where $A(x) = (2\log x)^{1/2}$, $D(x) = 2\log x + 1/2\log\log x - 1/2\log\pi$. By Theorem A.2.4.2 in [6], we have $T^e_{i,t}/\{A(\log w)\} \to 1$ in probability. Thus, the $F_{0;t}(q,\theta)$ with small θ can be approximated as

$$1-\exp\left[-e^{-\left\{A(\log w)q+\theta A^2(\log w)-D(\log w)\right\}}\right]. \tag{B.11}$$

By Lemma 1.5.1 in [6], we can verify that under the alternative hypothesis $\mathcal{B}_{1;t}$,

$$T^b_{i,t}-\sqrt{t-b+1}\sqrt{\gamma}\mu \xrightarrow{\mathcal{L}} N(0,1),\quad T^e_{i,t}-\sqrt{t-b+1}\sqrt{\gamma(1-\gamma)}\mu \xrightarrow{\mathcal{L}} N(0,1).$$

Accordingly, $T_{i,t}^{e} = \sqrt{t-b+1}\sqrt{\gamma(1-\gamma)}\mu\{1+o_p(1)\}$ by the condition that

$$\lim_{t\to\infty} \sqrt{t-b+1}\mu/(\log\log w)^{1/2} \to \infty.$$

Hence, the $F_{2;t}(q,\theta)$ with a small value of θ can be approximated as

$$1-\Phi\left[\left\{\theta\sqrt{\gamma(1-\gamma)}-\sqrt{\gamma}\right\}\sqrt{t-b+1}\mu+q\right]. \tag{B.12}$$

The result in (15) can be immediately obtained by taking partial derivatives of (B.11) and (B.12) with respect to q and θ. □

Proof of Proposition 3. Since the considered setting is the same as that for power analysis, we have, by (B.11) and (B.12), $F_{1;t}(q,\theta)=0$ and

$$F_t(q,\theta) = 1-\pi_{0;t}\exp\left[-e^{-\left\{A(\log w)q+\theta A^2(\log w)-D(\log w)\right\}}\right]-\pi_{12;t}\Phi\left[\left\{\theta\sqrt{\gamma(1-\gamma)}-\sqrt{\gamma}\right\}\sqrt{t-b+1}\mu+q\right].$$

Thus, Condition C2 is easy to verify by taking derivative. Also, by calculations, we have

$$\begin{aligned}\frac{\partial \mathrm{Fdr}_t(q,\theta)}{\partial q} &= \frac{\pi_{0;t}\pi_{12;t}}{\{F_t(q,\theta)\}^2}\left\{\frac{\partial F_{0;t}(q,\theta)}{\partial q}F_{2;t}(q,\theta)-F_{0;t}\frac{\partial F_{2;t}(q,\theta)}{\partial q}\right\}\\ &= \frac{\pi_{0;t}\pi_{12;t}}{\{F_t(q,\theta)\}^2}\left(\phi\left[\left\{\theta\sqrt{\gamma(1-\gamma)}-\sqrt{\gamma}\right\}\sqrt{t-b+1}\mu+q\right]F_{0;t}(q,\theta)\right.\\ &\quad\left.-\{1-F_{0;t}(q,\theta)\}e^{-\left\{A(\log w)q+\theta A^2(\log w)-D(\log w)\right\}}A(\log w)F_{12;t}(q,\theta)\right)\neq 0.\end{aligned}$$

Thus, Condition C4 is verified. For small θ, $\partial \mathrm{Fdr}_t(q,\theta)/\partial q<0$, and we calculate

$$\frac{\partial F_{0;t}(q_{\alpha;t}(\theta),\theta)}{\partial\theta} = -\{1-F_{0;t}(q,\theta)\}\exp\left[-\left\{A(\log w)q+\theta A^2(\log w)-D(\log w)\right\}\right]\left\{A(\log w)\frac{\partial q_{\alpha;t}(\theta)}{\partial\theta}+A^2(\log w)\right\},$$

$$\frac{\partial F_{12;t}(q_{\alpha;t}(\theta),\theta)}{\partial\theta} = -\phi\left[\left\{\theta\sqrt{\gamma(1-\gamma)}-\sqrt{\gamma}\right\}\sqrt{t-b+1}\mu+q\right]\left\{\sqrt{\gamma(1-\gamma)}\sqrt{t-b+1}\mu+\frac{\partial q_{\alpha;t}(\theta)}{\partial\theta}\right\}.$$

By the definition of (10), $\lambda=+\infty$ and $\partial \mathrm{Fdr}_t(q,\theta)/\partial q<0$, we can get $\partial q_{\alpha;t}(\theta)/\partial\theta>0$. Therefore, Condition C5 is verified. □

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2023.105224.

References

[1] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B Stat. Methodol. 57 (1) (1995) 289–300.
[2] Y. Benjamini, Y. Hochberg, On the adaptive control of the false discovery rate in multiple testing with independent statistics, J. Educ. Behav. Stat. 25 (1) (2000) 60–83.
[3] R.L. Brown, J. Durbin, J.M. Evans, Techniques for testing the constancy of regression relationships over time, J. R. Stat. Soc. Ser. B Stat. Methodol. 37 (1975) 149–192.
[4] H.P. Chan, Optimal sequential detection in multi-stream data, Ann. Statist. 45 (6) (2017) 2736–2763.
[5] Y. Chen, T. Wang, R.J. Samworth, High-dimensional, multiscale online changepoint detection, J. R. Stat. Soc. Ser. B Stat. Methodol. 84 (1) (2022) 234–266.
[6] M. Csörgő, L. Horváth, Limit Theorems in Change-Point Analysis, in: Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 1997, p. xvi+414, With a foreword by David Kendall.
[7] L. Du, X. Guo, W. Sun, C. Zou, False discovery rate control under general dependence by symmetrized data aggregation, J. Amer. Statist. Assoc. 118 (541) (2023) 607–621.
[8] L. Du, C. Zhang, Single-index modulated multiple testing, Ann. Statist. 42 (4) (2014) 30–79.
[9] L. Du, C. Zou, On-line control of false discovery rates for multiple datastreams, J. Statist. Plann. Inference 194 (2018) 1–14.
[10] R. Durrett, Probability—Theory and Examples, in: Cambridge Series in Statistical and Probabilistic Mathematics, vol. 49, Cambridge University Press, Cambridge, 2019, p. xii+419, Fifth edition of [MR1068527].
[11] B. Efron, R. Tibshirani, Empirical bayes methods and false discovery rates for microarrays, Genet. Epidemiol. 23 (1) (2002) 70–86.
[12] J. Fan, X. Han, W. Gu, Estimating false discovery proportion under arbitrary covariance dependence, J. Amer. Statist. Assoc. 107 (499) (2012) 1019–1035.
[13] A. Gandy, F.D.-H. Lau, Non-restarting cumulative sum charts and control of the false discovery rate, Biometrika 100 (1) (2013) 261–268.
[14] C. Genovese, L. Wasserman, Operating characteristics and extensions of the false discovery rate procedure, J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (3) (2002) 499–517.
[15] O.A. Grigg, D.J. Spiegelhalter, H.E. Jones, Local and marginal control charts applied to methicillin resistant *Staphylococcus aureus bacteraemia* reports in UK acute national health service trusts, J. R. Stat. Soc. Ser. A 172 (1) (2009) 49–66.

[16] G. Gurevich, A. Vexler, Retrospective change point detection: from parametric to distribution free policies, Comm. Statist. Simulation Comput. 39 (5) (2010) 899–920.
[17] D. Han, F. Tsung, A generalized EWMA control chart and its comparison with the optimal EWMA, CUSUM and GLR schemes, Ann. Statist. 32 (1) (2004) 316–339.
[18] F.D.-H. Lau, A. Gandy, Optimality of non-restarting CUSUM charts, Sequential Anal. 32 (4) (2013) 458–468.
[19] Y. Li, F. Tsung, False discovery rate-adjusted charting schemes for multistage process monitoring and fault identification, Technometrics 51 (2) (2009) 186–205.
[20] R. Lyons, Strong laws of large numbers for weakly correlated random variables, Michigan Math. J. 35 (3) (1988) 353–359.
[21] O.-A. Maillard, Sequential change-point detection: Laplace concentration of scan statistics and non-asymptotic delay bounds, in: A. Garivier, S. Kale (Eds.), Proceedings of the 30th International Conference on Algorithmic Learning Theory, in: Proceedings of Machine Learning Research, vol. 98, PMLR, 2019, pp. 610–632.
[22] C. Marshall, N. Best, A. Bottle, P. Aylin, Statistical issues in the prospective monitoring of health outcomes across multiple units, J. R. Stat. Soc. Ser. A 167 (3) (2004) 541–559.
[23] D. McDonald, A cusum procedure based on sequential ranks, Nav. Res. Logist. 37 (5) (1990) 627–646.
[24] Y. Mei, Efficient scalable schemes for monitoring a large number of data streams, Biometrika 97 (2) (2010) 419–433.
[25] E.S. Page, Continuous inspection schemes, Biometrika 41 (1954) 100–115.
[26] M. Pollak, Average run lengths of an optimal method of detecting a change in distribution, Ann. Statist. 15 (2) (1987) 749–779.
[27] M. Pollak, A robust changepoint detection method, Sequential Anal. 29 (2) (2010) 146–161.
[28] M. Pollak, A.M. Krieger, Shewhart revisited, Sequential Anal. 32 (2) (2013) 230–242.
[29] M. Pollak, D. Siegmund, Sequential detection of a change in a normal mean when the initial value is unknown, Ann. Statist. 19 (1) (1991) 394–416.
[30] P. Qiu, D. Xiang, Univariate dynamic screening system: an approach for identifying individuals with irregular longitudinal behavior, Technometrics 56 (2) (2014) 248–260.
[31] H. Ren, C. Zou, N. Chen, R. Li, Large-scale datastreams surveillance via pattern-oriented-sampling, J. Amer. Statist. Assoc. 117 (538) (2022) 794–808.
[32] W. Shewhart, Economic Control of Quality of Manufactured Product, in: Bell Telephone Laboratories series, American Society for Quality Control, 1931.
[33] D. Siegmund, E.S. Venkatraman, Using the generalized likelihood ratio statistic for sequential detection of a change-point, Ann. Statist. 23 (1) (1995) 255–271.
[34] Y.S. Soh, V. Chandrasekaran, High-dimensional change-point estimation: combining filtering with convex optimization, Appl. Comput. Harmon. Anal. 43 (1) (2017) 122–147.
[35] D. Spiegelhalter, C. Sherlaw-Johnson, M. Bardsley, I. Blunt, C. Wood, O. Grigg, Statistical methods for healthcare regulation: rating, screening and surveillance, J. R. Stat. Soc. Ser. A 175 (1) (2012) 1–47.
[36] J.D. Storey, J.E. Taylor, D. Siegmund, Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach, J. R. Stat. Soc. Ser. B Stat. Methodol. 66 (1) (2004) 187–205.
[37] W. Sun, T.T. Cai, Oracle and adaptive compound decision rules for false discovery rate control, J. Amer. Statist. Assoc. 102 (479) (2007) 901–912.
[38] A. Vexler, Guaranteed testing for epidemic changes of a linear regression model, J. Statist. Plann. Inference 136 (9) (2006) 3101–3120.
[39] A. Vexler, C. Wu, An optimal retrospective change point detection policy, Scand. J. Stat. 36 (3) (2009) 542–558.
[40] J. Wang, L. Du, C. Zou, Z. Wu, Dynamic statistical inference in massive datastreams, 2021, arXiv preprint arXiv:2111.01339.
[41] M. Wen, G. Wang, C. Zou, Z. Wang, Activation discovery with FDR control: Application to FMRI data, Statist. Sinica (2022) in press.
[42] Y. Xie, D. Siegmund, Sequential multi-sensor change-point detection, in: 2013 Information Theory and Applications Workshop, ITA, IEEE, 2013, pp. 1–20.
[43] B. Yakir, On the average run length to false alarm in surveillance problems which possess an invariance structure, Ann. Statist. 26 (3) (1998) 1198–1214.
[44] Y. Yu, O.H.M. Padilla, D. Wang, A. Rinaldo, A note on online change point detection, 2020, arXiv preprint arXiv:2006.03283.
[45] C. Zou, Z. Wang, W. Jiang, X. Zi, An efficient online monitoring method for high-dimensional data streams, Technometrics 57 (3) (2015) 374–387.