A False Discovery Rate Oriented Approach to Parallel Sequential Change Detection Problems

Jie Chen, Wenyi Zhang, Senior Member, IEEE, and H. Vincent Poor, Fellow, IEEE

Abstract—The problem of sequentially detecting changes in parallel data streams is formulated and investigated. Each data stream may have its own change point at which the underlying probability distribution of its data changes, and the decision maker needs to declare, sequentially, which data streams have passed their change points. With a large number of parallel data streams, the error metric is the false discovery rate (FDR), which is the expected ratio of the number of falsely declared data streams to the total number of declared data streams. A data stream is falsely declared if the detected change point is ahead of its actual change point. Decision procedures which are guaranteed to control the FDR level are developed, and it is also shown that the average decision delays (ADD) of these decision procedures do not grow with the number of data streams. Numerical simulations and case studies are conducted to corroborate the analytical results, and to illustrate the utility of the decision procedures.

Index Terms—average decision delay, false discovery rate, large-scale inference, multiple change detection, multiple hypothesis testing

I. INTRODUCTION

As a fundamental breakthrough beyond the classical fixed-sample-size (FSS) binary hypothesis testing, sequential hypothesis testing, in which the decision maker is allowed to sequentially observe the data and make his/her decision at a randomly selected time, has been extensively studied since the seminal work by Wald [3]. For independent and identically distributed (i.i.d.) data, compared with the FSS Neyman-Pearson test [4], the sequential probability ratio test (SPRT) is superior in terms of the average sample size required for achieving the same error performance.

An important topic in the theory of sequential analysis is change detection. Suppose that the decision maker is monitoring a stochastic system, in which at some unknown time there is an abrupt change of the underlying probability distribution of the system state. Change detection concerns about methods for sequentially detecting the occurrence of the change point, and the goal is to minimize the delay between the actual change point and the declared time, subject to certain constraint on the risk of false alarms [5] [6]. In the Bayesian formulation, a prior probability distribution of the change point is imposed and exploited when designing the decision procedure; see, e.g., [7]. In the non-Bayesian (or minimax) formulation, there is no prior knowledge about the change point and the design objective is to optimize the performance under the worst-case scenario; see, e.g., [8] where a decision procedure based on the cumulative sum (CUSUM) statistic [9] was shown to be optimal in an asymptotic sense, and [10] where the CUSUM decision procedure was further shown to be exactly optimal.

In basic formulations of change detection problems, the decision maker usually only needs to treat a single data stream. But in many modern applications, the rapid development of sensing technology allows for the generation of large-scale real-time streaming data [11], corresponding to a large number of parallel data streams. Detecting the change points in such parallel data streams is the situation considered in our work here.

With multiple data streams, a formulation different from ours is that an unknown subset of the data streams have a common change point. This setting is mainly motivated by surveillance applications, in which a system is monitored by multiple sensors and at an unknown time a disruption leads to a change in the observations of a subset of deployed sensors; see, e.g., [12], [13], [14], [15], and [16]. In contrast, our formulation considers the scenario where each data stream has its own change point, and furthermore the data streams as well as their change points are mutually independent. This setting can be suitable for systems whose different components are essentially independent.

The formulation that a decision maker sequentially detects independent change points for parallel independent data streams may seem uninteresting at the first glance, since one may wonder that this simply boils down to a number of separate single-stream change detection procedures. Such a formulation, however, becomes interesting when the number of data streams becomes large, and the choice of error metric then turns out to be a key factor.

Let us start with the formulation of making $K$ hypothesis tests given $K$ observations, each corresponding to one of the tests. For any given decision procedure, Table I categorizes the outcomes of the $K$ tests. A direct definition of the error metric is the familywise error rate (FWER) [17] [18], which is the probability of rejecting any null hypothesis, i.e., $\Pr[V \geq 1]$. A simple decision procedure that guarantees the FWER no greater than $\alpha$ is Bonferroni’s procedure, which essentially
performs $K$ separate hypothesis tests, each of which guarantees the probability of rejecting a null (i.e., Type I error probability) no greater than $\alpha/K$. This requirement is very stringent for detecting weak signals when $K$ is large, and decision procedures that control the FWER generally have very low detection power. This motivates the proposal of an alternative error metric called the false discovery rate (FDR) \[19\] \[20\], which is the expected proportion of falsely rejected hypotheses at all rejected hypotheses, i.e., \[\text{FDR} = \mathbb{E} \left[ \frac{V}{K} \right]. \]

Compared with FWER-oriented decision procedures, decision procedures that control the FDR are more powerful in identifying alternative hypotheses \[11\], at the cost of increased error rates. The most widely known FDR-oriented decision procedure is the Benjamini-Hochberg (BH) procedure \[20\].

Returning to our problem of sequential change detection over parallel data streams, a rejection corresponds to declaring a change point before the actual change point of the data stream. Thus the FWER is the probability of making an early change decision for at least one of the data streams. Similar to the multiple hypothesis testing scenario in the previous paragraph, to guarantee a prescribed level of the FWER, the change detection for each data stream is highly strict in declaring a change, in turn leading to excessively large decision delays when the number of data streams is large. To remedy this, we propose to use the FDR as the error metric, which quantifies the expected ratio between the number of false rejections and the total number of rejections. As will be shown in this work, FDR-oriented decision procedures substantially reduce the average decision delay (ADD) compared with FWER-oriented decision procedures. In fact, for FDR-oriented decision procedures, the ADD does not grow with the number of data streams, in sharp contrast with FWER-oriented decision procedures for which the ADD grows logarithmically with the number of data streams.

The main contributions of this work are as follows.

1. We formulate the problem of change detection for parallel data streams, extend the FDR error metric to this problem, and develop the corresponding decision procedure, which is motivated by the classical BH procedure.

2. For the developed FDR-oriented decision procedure, we establish a theoretical guarantee for the FDR. Furthermore, we obtain the asymptotic behavior of the ADD as the number of data streams grows large, which does not grow with the number of data streams; in contrast, the ADD of FWER-oriented decision procedures grows logarithmically with the number of data streams.

3. To cope with the scenario where the post-change statistics are not perfectly known, we extend the problem formulation of change detection to change detection and isolation, for parallel data streams, and develop the corresponding decision procedures and performance guarantees.

4. We test the developed decision procedures for simulated data sets as well as a real-world data set of stock price time series, to corroborate the analytical results, and to illustrate the utility of the decision procedures.

The remaining part of this paper is organized as follows. Section II describes the system model, formulates the multiple change detection problem, and introduces the key definitions. Section III presents the decision procedures that aim at controlling the FDR and the FWER, respectively. Section IV establishes the FDR guarantee and the asymptotic behavior of the ADD. Section V extends the problem formulation, decision procedures, as well as their analysis to the multiple change detection and isolation problem. Section VI presents the simulation results, and Section VII presents a case study where we apply the developed decision procedures to a real-world data set of stock price time series. Finally Section VIII concludes this paper.

### II. Problem Setup

Consider $K \geq 2$ parallel data streams:

\[
X^{(1)}_k, X^{(2)}_k, \ldots, X^{(k)}_k, X^{(K)}_k, \ldots \quad \text{for } k \in [K] = \{1, \ldots, K\}. \tag{1}
\]

Denote the data at time epoch $n$ by $X_n^{[K]} = \{X_n^{(1)}, X_n^{(2)}, \ldots, X_n^{(K)}\}$. For the $k$-th data stream, there exists a time epoch $t_k^{(i)} \geq 1$ called its change point. In particular, we allow $t_k^{(i)} = \infty$; that is, the underlying statistics of a data stream may never change, and thus a data stream may not have any (finite) change point.

For a measurable space $(\Omega, \mathcal{F})$, consisting of a sample space $\Omega$ and a $\sigma$-field $\mathcal{F}$ of events, consider a family $\{p_k^{(i)}, k \in [K]\}$ of probability measures to describe the prior distribution of $t_k^{(i)}$ and the distribution of $\{X_k^{(i)}, k \in [K], n = 1, 2, \ldots\}$. Under $p_k^{(i)}$, $t_k^{(i)}$ is a random variable with prior distribution $\pi_m = \text{Pr}(t_k^{(i)} = m)$; conditioned upon $t_k^{(i)}$, $X_1^{(k)}, X_2^{(k)}, \ldots, X_{t_k^{(i)}}^{(k)}$ are independent and identically distributed (i.i.d.) with (a pre-change) distribution $h_0$ and $X_{t_k^{(i)}}^{(k)}, X_{t_k^{(i)}+1}^{(k)}, \ldots$ are i.i.d. with another (post-change) distribution $h_1$. In Section V we will address detection and isolation problems where there are multiple possible post-change distributions. We assume mutual independence among the $K$ data streams, and that $h_0$ and $h_1$ are probability densities with respect to some measure $P$ on $(\Omega, \mathcal{F})$.

Define the Kullback-Leibler (KL) divergence as

\[
I = \int h_1(x) \log \left( \frac{h_1(x)}{h_0(x)} \right) p(dx), \tag{2}
\]
and the mean change point as
\[ t = \mathbb{E}[P(l(k)|\pi(k)) < \infty]. \] (3)

Suppose that a statistician sequentially observes the \( K \) parallel data streams and makes decision regarding whether changes have occurred for these data streams, up to some deadline \( \bar{N} \). To be concrete, the statistician keeps observing \( X_n^{[k]} \) starting from \( n = 1 \), sequentially, until at a certain time epoch \( T_1 \), declaring that a change has occurred for a certain data stream indexed by \( D_1 \). Subsequently, the statistician keeps observing \( X_n^{[k]} \) starting from \( n = T_1 + 1 \), with the \( D_1 \)-th data stream excluded in the observation, until at a certain time epoch \( T_2 \), declaring that a change has occurred for a certain data stream indexed by \( D_2 \)... Such a procedure is executed until either changes have been declared for all the \( K \) data streams, or the time epoch \( j \) has reached the deadline \( \bar{N} \).

We can formalize the above decision procedure as follows.

**Definition** A decision procedure is a multiple stopping rule \( \mathcal{R} \) consisting of a sequence of stopping times and decisions, as
\[ \mathcal{R} = (T_1, D_1, T_2, D_2, \ldots, T_K, D_K), \] (4)
where \( 1 \leq T_1 \leq T_2 \leq \ldots \leq T_K \leq \bar{N} \), and \( \{D_k, k \in [K]\} \) constitute a permutation of \([K]\). Regarding the stopping times and decisions, \( T_1 = n_1, D_1 = d_1 \) depends upon \( X_n^{[1]}, X_n^{[2]}, \ldots, X_n^{[K]} \), and, in a recursive way, for general \( k \in [K] \), \( T_k = n_k, D_k = d_k \) depends upon \( X_n^{[k]} \) for \( n = 1, 2, \ldots, T_1; \ldots; X_n^{[k]} \) for \( n = T_{k-1} + 1, 2, \ldots, T_k \).

We can also define the stopping time for every data stream.

**Definition** For \( k \in [K] \), denote by \( T^{(k)} \) the stopping time for the \( k \)-th data stream, so that \( T^{(k)} = T_l \) for some \( l \in [K] \) such that \( D_l = k \).

We then define the false discovery rate (FDR) and the familywise error rate (FWER) of a decision procedure \( \mathcal{R} \) as follows.

**Definition** Denote by \( \bar{R} \) the number of change points declared by \( \mathcal{R} \), i.e., the number of elements in \( \{T^{(k)}, k \in [K]\} \) satisfying \( T^{(k)} < \bar{N} \), and by \( \bar{V} \) the number of falsely declared change points, i.e., the number of elements in \( \{T^{(k)}, k \in [K]\} \) satisfying \( T^{(k)} < \bar{N} \) and \( T^{(k)} < t^{(k)} \). The FDR of \( \mathcal{R} \) is
\[ \text{FDR} = \mathbb{E}[\bar{V} / \bar{R} \lor 1]. \] (5)

where \( a \lor b \) represents the maximum between \( a \) and \( b \), and \( \mathbb{E} \) represents the expectation with respect to \( \{P(l(k), k \in [K]\} \). The FWER of \( \mathcal{R} \) is
\[ \text{FWER} = \mathbb{P}[\bar{V} \geq 1]. \] (6)

Besides the FDR and the FWER, we are also interested in the average decision delay (ADD) of a decision procedure. This is the average number of additional samples before declaring a change point after that change point, normalized by the number of data streams with a finite change point. We denote by \( K_n \) the number of data streams without a finite change point, which is a binomial random variable \( \text{Bin}(K, \pi(\infty)) \), with probability mass function \( P(K_n = k_\infty) = \binom{K}{k_\infty} \pi(\infty)^{k_\infty} (1 - \pi(\infty))^{K-k_\infty} \). Formally, we have the following definition.

**Definition** For a decision procedure, the ADD is defined as
\[ \text{ADD} = \sum_{k=0}^{K} P(K_n = k_\infty) \text{ADD}_{k_\infty}, \] (7)
where \( \text{ADD}_{k_\infty} \) is the conditional ADD upon \( K_n = k_\infty \), which can be written as
\[ \text{ADD}_{k_\infty} = \frac{1}{K - k_\infty} \sum_{k=1}^{K-k_\infty} \sum_{i=1}^{K} \mathbb{E}[T^{(k)} - l^{(i)} | T^{(k)} \geq l^{(i)}]. \] (8)
for \( k_\infty = 1, \ldots, K - 1 \), and \( \text{ADD}_K = 0 \). Note that since all the data streams are i.i.d., it loses no generality to consider the case where the data streams of indices from 1 to \( K - k_\infty \) have finite change points. Using induction, we can also write \( \text{ADD}_{k_\infty} \) in the following form:
\[ \text{ADD}_{k_\infty} = \frac{1}{K - k_\infty} \sum_{k=1}^{K-k_\infty} \sum_{i=1}^{K} \mathbb{E}[T^{(k)} - l^{(i)} | T^{(k)} \geq l^{(i)}], \] (9)
which is usually more amenable to analysis in the sequel.

**III. Decision Procedures**

For the \( k \)-th data stream, \( k \in [K] \), we define a sequence of statistics
\[ G^{(k)} = \sum_{m=1}^{\infty} \pi_m \prod_{i=m}^{\infty} P(l^{(i)}), \quad n = 1, 2, \ldots, \] (10)
where \( L^{(k)} = h_1(X^{(k)}_t) / h_0(X^{(k)}_t) \), and for \( m > n \), set \( \prod_{i=m}^{n} P(l^{(i)}) = 1 \). So \( G^{(k)} \) is the average likelihood ratio between the hypotheses that the change occurs at \( t^{(k)} = m < \infty \) and that the change never occurs \( (t^{(k)} = \infty) \). Setting \( G^{(0)} = 1 \), we have a recursive relationship for \( G^{(k)} \) as
\[ G^{(k)} = G^{(k-1)} + \pi_{k+1} (1 - L^{(k)}), \] (11)
where \( \pi_{k+1} = \mathbb{P}[t^{(k)} \geq n + 1] \).

First we describe a decision procedure called MD-FDR that aims at controlling the FDR. Fix a parameter \( \alpha \in (0, 1) \), and set an array of thresholds as
\[ Q_i = \frac{K}{\alpha i}, \quad Q_K \leq Q_{K-1} \leq \cdots \leq Q_1. \] (12)

For convenience of description, we divide the operation of a decision procedure into multiple stages. At the beginning, the time epoch \( n = 1 \), the decision procedure starts with the first stage, and whenever at least a stopping time is reached and an associated decision for a data stream is declared, the current stage ends and a new stage begins from the next time epoch.

We use \( I_q \) to denote the indices of active data streams (i.e., the data streams that have not been stopped) at the beginning of the \( q \)-th stage, and use \( n_q \) to denote the time epoch at the end of the \( q \)-th stage. Initially we set \( I_1 = [K] \) and \( n_0 = 0 \). We describe the decision procedure MD-FDR in Algorithm 1, which is also illustrated in Figure 1.
Algorithm 1 MD-FDR and MD-Hochberg

For MD-FDR, set thresholds as (12); for MD-Hochberg, set thresholds as (15).

1) Initially set \( q = 1, I_q = [K] \), \( n = 1 \), and \( n_0 = 0 \).

2) Sample the active data streams \( \{X_n^{(k)}\}_{k \in I_n,q > n_0} \), and update the statistics \( \{G_n^{(k)}\}_{k \in I_n,q > n_0} \) using (11).

3) Sort the statistics in ascending order as \( \{G_n^{(i)}\} \), where \( i(n, l) \) denotes the index of the \( l \)-th ordered statistic at time epoch \( n \).

4) Repeat 2)-3) with \( n \) increased by 1 each time, until \( n \) equals \( n_q = \overline{N} \wedge \min \{ n > n_{q-1} \mid G_n^{(i(n,l))} \geq Q_{K-l+1}, \forall l \in [I_q]\} \).

A) If \( n_q < \overline{N} \), stop the \( q \)-th stage and declare change points at \( n_q \) for data streams that satisfy (16).

Update \( I_{q+1} \) to exclude the indices of these declared data streams. If \( |I_{q+1}| = 0 \), then stop; otherwise, set \( q = q + 1 \), and go to 2) to begin the next stage.

B) Otherwise, \( n_q = \overline{N} \), declare that all the active data streams in \( I_q \) have no change points, and stop.

Algorithm 2 MD-Bonferroni

1) Initially set \( q = 1, I_q = [K] \), \( n = 1 \), and \( n_0 = 0 \).

2) Sample the active data streams \( \{X_n^{(k)}\}_{k \in I_n,q > n_0} \), and update the statistics \( \{G_n^{(k)}\}_{k \in I_n,q > n_0} \) using (11).

3) Repeat 2) with \( n \) increased by 1 each time, until \( n \) equals \( n_q = \overline{N} \wedge \min \{ n > n_{q-1} \mid G_n^{(k)} \geq Q_{3k \in I_q} \} \).

A) If \( n_q < \overline{N} \), stop the \( q \)-th stage and declare change points at \( n_q \) for data streams that satisfy (16).

Update \( I_{q+1} \) to exclude the indices of these declared data streams. If \( |I_{q+1}| = 0 \), then stop; otherwise, set \( q = q + 1 \), and go to 2) to begin the next stage.

B) Otherwise, \( n_q = \overline{N} \), declare that all the active data streams in \( I_q \) have no change points, and stop.

For comparison, we also consider two decision procedures that control the FWER. Fix \( \alpha \in (0,1) \). The decision procedure based on Bonferroni’s procedure, denoted as MD-Bonferroni, has a single threshold \( Q = K/\alpha \), and is described in Algorithm 2. The decision procedure based on Hochberg’s procedure [22], denoted as MD-Hochberg, is similar with MD-FDR, but has a different array of thresholds as \( Q_i = \frac{K - s + 1}{\alpha}, \quad Q_K \leq Q_{K-1} \leq \ldots \leq Q_1 \).

The detailed description of MD-Hochberg is in Algorithm 1.

IV. Analysis of Decision Procedures

In this section, we analyze the statistical properties of the proposed decision procedure MD-FDR. On one hand, we show that MD-FDR controls the FDR. On the other hand, we compare the ADDs achieved by MD-FDR and by FWER-based decision procedures.

A. False Discovery Rate

The following result, Theorem 1, states that the proposed decision procedure MD-FDR controls FDR under our problem formulation.

Theorem 1. Suppose that \( 0 < I < \infty \), \( t < \infty \), and that \( \sum_{m=1}^{\infty} \pi_m |\log \pi_m| - \pi_\infty \log \pi_\infty | < \infty \).

For a given \( 0 < \alpha < 1 \), the decision procedure MD-FDR in Section III satisfies

\[
\text{FDR} \leq \alpha + \frac{\overline{N}}{N} \left( Kt + I^{-1} (K \log K - K \log \alpha) + o(K) \right),
\]

where \( o(K)/K \to 0 \) as \( K \to \infty \). Hence if we scale \( \overline{N} \) with \( K \) such that \( (K \log K)/\overline{N} \to 0 \) as \( K \to \infty \), we have

\[
\text{FDR} \leq \alpha.
\]

Proof. We prove Theorem 1 in two parts. In the first part, we consider a decision procedure similar to MD-FDR but without the deadline \( \overline{N} \), for which we prove that such a decision procedure controls the FDR at the level \( \alpha \). In the second part, we impose the deadline \( \overline{N} \) and estimate the extra cost upon the FDR.

Part 1:

As we remove the deadline \( \overline{N} \) in MD-FDR (i.e., setting \( \overline{N} = \infty \)), we adjust the definition of the FDR as

\[
\text{FDR}_\infty = \mathbb{E}_\tilde{X} \left[ \frac{V}{R \vee 1} \right],
\]

where \( R \) is the number of change points declared, i.e., the number of elements of \( [K] \) such that \( T^{(k)} < \infty \), and \( V \) is the number of falsely declared change points, i.e. the number of elements of \( [K] \) such that \( T^{(k)} < \delta^{(k)} \).
Let $K$ be the number of data streams satisfying $t^{(i)} < \infty$. For $k \in [K]$ and $s \in [K]$ define events
\[ W_{k,s} = \left\{ G_n^{(k)} \geq Q_s, \exists n < t^{(i)} \right\}. \]
(20)
For these events, we have
\[ P_\pi(W_{k,s}) = P_\pi(G_n^{(k)} \geq Q_s, \exists n < t^{(i)}) \leq \frac{str}{K}. \]
(21)

In order to prove (21), consider for each data stream, say, the $k$th one, a stopping time with respect to the observations $X_1^{(k)}, X_2^{(k)}, \ldots, X_{n-1}^{(k)}, X_n^{(k)}, \ldots$ as follows:
\[ T^{(k)} = \min \left\{ n \geq 1 \left| G_n^{(k)} \geq A \right. \right\}, \quad A > 1. \]
(22)
The following result is standard and can be found in several references, e.g., [23, Lem. 3.1].

**Lemma 2.** For any $A > 1$, $P_\pi(T^{(k)} < \infty) \leq 1/A$.

Note that the event $\{T^{(k)} < m\}$ is measurable with respect to the $\sigma$-field $F_{m-1}$, which implies
\[ P_m(T^{(k)} < m) = P_\pi(T^{(k)} < m). \]
(23)
Also note that
\[ \sup_{m \geq 1} P_\pi(T^{(k)} < m) = P_\pi(T^{(k)} < \infty). \]
(24)
Therefore, using Lemma 2, we have
\[ P_\pi(T^{(k)} < t^{(i)}) = \sum_{m=1}^{\infty} \pi_m P_\pi(T^{(k)} < m) \leq P_\pi(T^{(k)} < \infty) \leq \frac{1}{A}. \]
(25)

Hence, by setting $A = Q_s = K/(str)$ when defining $T^{(k)}$, we have
\[ P_\pi(W_{k,s}) = P_\pi(T^{(k)} < t^{(i)}) \leq \frac{1}{Q_s} = \frac{str}{K}. \]
(26)

Thus proving (21).

For $u \in \{0, \ldots, K\}$ and $v \in \{1, \ldots, K - u\}$, define
\[ \Omega_u = \{[\omega] \subseteq \{1, \ldots, K - u\} : |\omega| = v\} \]
\[ V_{v,u} = \{v \text{ data streams with indices in } \omega \text{ stopped early,} \}
\text{i.e., } T^{(k)} < t^{(i)}, k \in \omega; \]
\[ \text{and } u \text{ data streams correctly stopped} \}
\[ V_{v,u} = \bigcup_{\omega \in \Omega_u} V_{v,u}. \]

Note that the sets $V_{v,u}$ in the union $V_{v,u}$ are mutually disjoint. We claim that
\[ P_\pi(W_{k,v,u} \cap V_{v,u}) \geq 1 \left| k \in \omega \right| P_\pi(V_{v,u}). \]
(27)
The following proof of (27) is modified from an argument in [24]. If $k \notin \omega$, this inequality obviously holds. If $k \in \omega$, what we need to show is that $V_{v,u} \subseteq W_{k,v,u}$. For any outcome in $V_{v,u}$, the $k$-th data stream is stopped early, at some stage $q$. By (14) in the decision procedure MD-FDR, $k = i(n_q, l)$ for some $l \geq l_q$, so
\[ G_n^{(k)} \geq G_n^{(i(n_q, l))} \geq Q_{k-l_q+1}. \]
(28)
Since $K - l_q + 1$ is the number of stopped data streams until stage $q$, this value is no greater than the total number of stopped data streams in $V_{v,u, v + u}$. Hence
\[ G_n^{(k)} \geq Q_{k-l_q+1} \geq Q_{v+u}, \]
(29)
by noting that the thresholds $\{Q_s\}$ are decreasing with $s$. This thus shows that $V_{v,u} \subseteq W_{k,v,u}$ for any $k \in \omega$.

With (27) established, we now follow the argument of [25], with a few modifications suitable for our problem formulation:
\[ \sum_{k=1}^{K-u} P_\pi(W_{k,v,u} \cap V_{v,u}) = \sum_{k=1}^{K-u} P_\pi(W_{k,v,u} \cap V_{v,u}) \]
(30)
\[ \geq \sum_{k=1}^{K-u} \sum_{\omega \in \Omega_u} 1 \left| k \in \omega \right| P_\pi(V_{v,u}) \]
(31)
\[ = \sum_{\omega \in \Omega_u} \sum_{k=1}^{K-u} 1 \left| k \in \omega \right| P_\pi(V_{v,u}) \]
(32)
Here, (30) is because that the sets $\{V_{v,u}\}$ in the union $V_{v,u}$ are mutually disjoint, (31) is due to (27), and (32) is due to the definition of $\Omega_u$.

Using (32), we have that conditioned upon $K_1$, the corresponding (conditional) FDR$_{\omega}$ is
\[ K_1\text{-FDR}_{\omega} = \sum_{u=0}^{K_1} \sum_{v=1}^{K_u} \frac{v}{v+u} \sum_{k=1}^{K-u} P_\pi(W_{k,v,u} \cap V_{v,u}) \]
\[ \leq \sum_{u=0}^{K_1} \sum_{v=1}^{K_u} \frac{v}{v+u} \sum_{k=1}^{K-u} P_\pi(W_{k,v,u} \cap V_{v,u}) \]
\[ = \sum_{u=0}^{K_1} \sum_{v=1}^{K_u} \frac{v}{v+u} \sum_{k=1}^{K-u} P_\pi(W_{k,v,u} \cap V_{v,u}). \]
(33)

Define $U_{v,u,k}$ to be the event that, conditioned upon the $k$-th data stream being stopped early, some other $v - 1$ data streams are stopped early and some $u$ data streams correctly stopped. So $W_{k,v,u} \cap V_{v,u} = W_{k,v,u} \cap U_{v,u,k}$. Define $U_{k} = \bigcup_{v=1}^{K_u} U_{v,u,k}$. For any $k$, $U_{k, ..., U_{k, K}}$ partition the sample space. Then starting from (33), we have
\[ K_1\text{-FDR}_{\omega} \leq \sum_{u=0}^{K_1} \sum_{v=1}^{K_u} \frac{v}{v+u} \sum_{k=1}^{K-u} P_\pi(W_{k,v,u} \cap U_{v,u,k}) \]
(34)
Since the data streams are mutually independent, the event of the $k$th data stream being stopped early and the event of $U_{v,u,k}$
are independent, and we have

\[
K_1 \cdot \text{FDR}_{\infty} \leq \sum_{k=1}^{K-o} \left( \sum_{j=1}^{K} \frac{1}{s} P_\pi(W_{k,j}) P_\pi(U_{j,k}) \right) \\
\leq \sum_{k=1}^{K-o} \left( \sum_{j=1}^{K} \frac{1}{s} P_\pi(U_{j,k}) \right) \\
\leq \frac{\alpha}{K} \sum_{k=1}^{K} \sum_{j=1}^{K} P_\pi(U_{j,k}) \\
= \frac{\alpha}{K} K = \alpha,
\]

where (35) is due to (26). Because the bound (36) holds for any \(K_1\), we have \(\text{FDR}_{\infty} \leq \alpha\).

**Part 2:**

Now we turn to the decision procedure \(\text{MD-FDR}\) with a finite \(N\) and analyze the relationship between \(\text{FDR}_{\infty}\) in (19) and \(\text{FDR}\) in (5). With \(N\) imposed, the number of change points declared, \(\hat{R}\), is clearly no greater than \(R\). Denote \(R = \hat{R} + \Delta\), where \(\Delta\) is the size of \(\{T^{(k)} | N \leq T^{(k)} < \infty\}\), which can be further divided into two parts. The first part has a size of \(\Delta_1 = \{T^{(k)} | N \leq T^{(k)} < \infty\}\), which is the reduction of the number of false alarm events, and the second part has a size of \(\Delta_2 = \{T^{(k)} | \infty \leq T^{(k)} < \infty, N = \infty\}\). We have

\[
\text{FDR} = \mathbb{E}_\pi \left[ \frac{\sum \Delta_2 = 0}{\sum \Delta_2 \geq 1} P_\pi(\Delta_2 = 0) + \mathbb{E}_\pi \left[ \frac{\sum \Delta_2 \geq 1}{\sum \Delta_2 \geq 1} P_\pi(\Delta_2 = 1) \right] \right]
\]

where \(\pi(\cdot)\) means that this is the conditional probability \(P_\pi(T^{(k)} < \infty)\), and the last inequality is due to that \(\Delta_2\) satisfies \(T^{(k)} \leq \infty\).

Then we choose \(N\) to control \(\pi(\Delta_2 \geq 1)\). Let us write

\[
P_\pi(\Delta_2 \geq 1) = 1 - P_\pi(\Delta_2 = 0) \\
= 1 - \left[ 1 - P_\pi(\Delta_2 \geq 1) \right]^{K} \\
\geq 1 - \log \frac{K}{\alpha} \left[ 1 + o(1) \right].
\]

Following the argument in [23, Thm. 4.2], when \(\sum_{k=0}^{\infty} |\pi_{\alpha}| - \pi_\alpha \log \pi_\alpha < \infty\) holds, for every \(k \in [K]\), we have

\[
\mathbb{E}_\pi \left[ \left( T^{(k)} - \pi^{(k)} \right) | T^{(k)} \geq \pi^{(k)} \right] \geq \frac{1}{1 - \log \frac{K}{\alpha} + o(1)}.
\]

where \(o(1) \to 0\) as \(K \to \infty\). Applying Markov's inequality, we have

\[
P_\pi(T^{(k)} \geq \infty | T^{(k)} \geq \pi^{(k)}) \leq \frac{\mathbb{E}_\pi \left[ T^{(k)} | T^{(k)} \geq \pi^{(k)} \right]}{N} \leq \frac{t + I^{-1} \log \frac{K}{\alpha} + o(1)}{N}. \tag{39}
\]

Back to (38) and (37), we have

\[
\text{FDR} \leq \alpha + 1 - \left( 1 - t + I^{-1} \log \frac{K}{\alpha} + o(1) \right) \frac{K}{N} \leq \frac{\alpha + 1 - K t + I^{-1} (K \log K - K \log \alpha + o(K))}{N}. \tag{40}
\]

This completes the proof of Theorem 1. \(\square\)

In our problem formulation, we consider the simple case where all the parallel data streams are mutually independent. Following the approach in [25], it is possible to provide some degree of FDR control even when the data streams are somewhat correlated, e.g., satisfying the so-called positive regression dependence on a subset. A systematic investigation along that direction is an interesting topic for future research.

**B. Average Decision Delay**

For the three decision procedures considered in Section III, we have the following result about the behaviors of their ADDs.

**Theorem 3.** For the three decision procedures such that the FDR (for \(\text{MD-FDR}\)) and the FWER (for \(\text{MD-Bonferroni}\) and \(\text{MD-Hochberg}\)) are controlled by \(\alpha > 0\), if \(\pi_\alpha > 0\) holds, as \(K\) grows without bound, we have:

\[
\text{MD-FDR} : \quad \text{ADD} = O(1); \tag{41}
\]

\[
\text{MD-Bonferroni} : \quad \text{ADD} \sim \frac{\log K}{I}; \tag{42}
\]

\[
\text{MD-Hochberg} : \quad \text{ADD} \sim \frac{\log K}{I}. \tag{43}
\]

**Proof.** First, consider MD-FDR which controls the FDR. According to [23, Thm. 4.2], if the \(4\)th data stream is stopped by threshold \(Q_4 = K/(\alpha)\), we have

\[
\mathbb{E}_\pi \left[ \left( T^{(k)} - \pi^{(k)} \right) | T^{(k)} \geq \pi^{(k)} \right] \leq \frac{1}{I} \log \frac{K}{\alpha} + o(1); \tag{44}
\]

where \(o(1) \to 0\) as \(K \to \infty\). According to the definition of the ADD, we can derive its lower bound as follows:

\[
\text{ADD} = \sum_{k=0}^{K-1} \frac{P(K_\alpha = k_\alpha)}{K - k_\alpha} \frac{1}{(K - k_\alpha)} K^{k_\alpha} \sum_{k=1}^{K} \mathbb{E}_\pi \left[ T^{(k)} - \pi^{(k)} \right] \geq \frac{1}{I} \log \frac{K}{\alpha} - \sum_{k=0}^{K} \log \frac{K - k_\alpha}{\alpha} + o(1)
\]

\[
= \frac{1}{I} \log \frac{K}{\alpha} - \log (K - k_\alpha) + o(1)
\]

\[
= \frac{1}{I} \log \frac{K}{\alpha} - \log (K(1 - \pi_\alpha)) + o(1)
\]

\[
= \frac{1}{I} \log \pi_\alpha (1 + o(1)). \tag{45}
\]
where the inequality in the second line is due to (45) and that for a fixed $k_{\alpha_o}$, $Q_{k_{\alpha_o}}$ is the smallest threshold.

On the other hand, we can derive its upper bound as follows:

$$
ADD = \sum_{k_{\alpha_o}} P(K_{\alpha_o} = k_{\alpha_o}) \frac{1}{1 - k_{\alpha_o}} \sum_{k=1}^{K_{k_{\alpha_o}}} E \left[ T^{(k)} - \ell^{(k)} | T^{(k)} \geq \ell^{(k)} \right] \\
\leq \sum_{k_{\alpha_o}} P(K_{\alpha_o} = k_{\alpha_o}) \frac{1}{1 - k_{\alpha_o}} \sum_{k=1}^{K_{k_{\alpha_o}}} \log \frac{K_{k_{\alpha_o}}}{K_{\alpha_o}} + o(1) \\
= \frac{1}{T} \left( \log K + |\log \alpha| \right) (1 + o(1)),
$$

where the inequality in the second line is due to (45) and that the worst case of the thresholds occurs when the thresholds are ascending and different, and the equality in the last line is due to Stirling’s approximation. Combining the lower and the upper bounds we have $ADD = O(1)$ which does not grow with $K$.

Then, consider MD-Bonferroni and MD-Hochberg which control the FWER. We have for MD-Bonferroni,

$$
ADD = \sum_{k_{\alpha_o}} P(K_{\alpha_o} = k_{\alpha_o}) \frac{1}{1 - k_{\alpha_o}} \sum_{k=1}^{K_{k_{\alpha_o}}} E \left[ T^{(k)} - \ell^{(k)} | T^{(k)} \geq \ell^{(k)} \right] \\
\leq \sum_{k_{\alpha_o}} P(K_{\alpha_o} = k_{\alpha_o}) \frac{1}{1 - k_{\alpha_o}} \sum_{k=1}^{K_{k_{\alpha_o}}} \log \frac{K_{k_{\alpha_o}}}{K_{\alpha_o}} + o(1) \\
= \frac{1}{T} \left( \log K + |\log \alpha| \right) (1 + o(1)),
$$

where the equality in the second line is due to that MD-Bonferroni uses a single threshold $Q = K/\alpha$; and for MD-Hochberg,

$$
ADD = \sum_{k_{\alpha_o}} P(K_{\alpha_o} = k_{\alpha_o}) \frac{1}{1 - k_{\alpha_o}} \sum_{k=1}^{K_{k_{\alpha_o}}} E \left[ T^{(k)} - \ell^{(k)} | T^{(k)} \geq \ell^{(k)} \right] \\
\geq \sum_{k_{\alpha_o}} P(K_{\alpha_o} = k_{\alpha_o}) \frac{1}{1 - k_{\alpha_o}} \sum_{k=1}^{K_{k_{\alpha_o}}} \log \frac{k_{\alpha_o} + 1}{\alpha} + o(1) \\
= \frac{1}{T} \left( \sum_{k_{\alpha_o}} \sum_{k=1}^{K_{k_{\alpha_o}}} P(K_{\alpha_o} = k_{\alpha_o}) \log(k_{\alpha_o} + 1) - \log \alpha \right) + o(1) \\
= \frac{1}{T} \left( \log(K\pi_{\alpha_o} + 1) - \frac{1 - \pi_{\alpha_o}}{2K\pi_{\alpha_o}} - \log \alpha \right) + o(1) \\
= \frac{1}{T} \left( \log(K\pi_{\alpha_o} + 1) + |\log \alpha| \right) (1 + o(1)),
$$

where the inequality in the second line is similar to that in (46) and due to the thresholds in MD-Hochberg. □

From Theorem 3, we see that the ADD of MD-FDR does not increase with the number of data streams, and thus in this sense MD-FDR is a scalable decision procedure suitable for large scale problems. In contrast, the ADDs of MD-Bonferroni and MD-Hochberg scale at a logarithmic speed with the number of data streams. Such a discrepancy is due to that for procedures controlling the FDR, we allow a more relaxed decision criterion compared with procedures controlling the FWER.

V. EXTENSION TO DETECTION AND ISOLATION PROBLEMS

In many practical situations, although the pre-change distribution may be accurately obtained from data, the post-change distribution can be difficult to obtain. To handle such a situation, we allow the post-change distribution to be one of several possible ones, and thus are led to a problem formulation of detection and isolation.

A. Model and Problem Formulation

Consider a family $\{p_{i,j,k}, k \in [K], j \in [J]\}$ of probability measures on $(\Omega, \mathcal{F})$ that describes both the prior distribution of $p(i)$ and the distribution of $\{X_{n,k}, k \in [K], n = 1, 2,...\}$. Under $p_{i,j,k}, X^{(k)}_1, X^{(k)}_2, \ldots, X^{(k)}_{n_{k-1}}$ are i.i.d. with a (pre-change) distribution $h_{i,0}$, and $X^{(k)}_1, X^{(k)}_2, \ldots, X^{(k)}_{n_{k-1}}$ are i.i.d. with another (post-change) distribution $h_{i,1}$. $f_i([J]$ and further independent of $X^{(k)}_1, X^{(k)}_2, \ldots, X^{(k)}_{n_{k-1}}$.

For simplicity, we write $p_{i,j,k}$ for $p_{i,j,k}$. We also impose a prior probability distribution $p(j)$ on the change type $j$, and that $h_{0,j}$ and $h_{1,j}$ are probability densities with respect to some measure on $(\Omega, F)$.

Now the decision maker needs to decide, for each data stream, both its change point and its change type (i.e., the post-change distribution). So we extend the definition of a decision procedure in Section II as follows.

Definition A decision procedure for detection and isolation is a multiple stopping rule $R$ consisting of a sequence of stopping times and decisions, as

$$
R = \{T_1, D_1, J_1, T_2, D_2, J_2, \ldots, T_k, D_k, J_k\}
$$

where $[T_k, k \in [K]]$ and $[D_k, k \in [K]]$ are the same as those in (4) for a decision procedure for the basic problem formulation, and $J_k$ denotes the decided change type for the $D_k$-th data stream, $k \in [K]$.

We also define the stopping time and the decided change type for every data stream.

Definition For $k \in [K]$, denote by $T^{(k)}$ the stopping time for the $k$-th data stream, the same as that in Section II, and by $j^{(k)}$ the decided change type for the $k$-th data stream.

We then extend the definition of the FDR to incorporate the effect of deciding change types incorrectly.

Definition Denote by $\overline{R}$ is the number of change points declared by $R$, i.e., the number of elements in $\{T^{(k)}, k \in [K]\}$ satisfying $T^{(k)} < \overline{R}$, and by $\overline{V}$ the number of falsely declared change points or falsely classified change types, i.e., the number of elements in $\{T^{(k)}, k \in [K]\}$ satisfying $\{T^{(k)} < \ell^{(k)}, \overline{R} \}$ and $\{T^{(k)} < \ell^{(k)}, \overline{R} \}$ and $\overline{V}$. The FDR of $R$ is

$$
FDR = \frac{E_{\pi, \rho}}{\overline{R} + 1},
$$

where $E_{\pi, \rho}$ represents the expectation with respect to $p_{i,j,k}, p(j), k \in [K], j \in [J]$. For example, suppose we have ten data streams in total, indexed from 1 to 10, and data streams $\{1, 2, 4, 6, 7, 8\}$ have finite change points with change types $\{1, 2, 2, 1, 1, 1\}$. A
decision maker has declared seven stopping times for data streams 1, 2, 3, 4, 6, 7, 8, with change types 1, 2, 1, 2, 2, 1, 1. So data stream 3 is incorrectly stopped, and data stream 6 has its change type incorrectly decided. As a result, the empirical FDR for such an example is $\frac{7}{8}$. 

B. Decision Procedure

To construct a decision procedure controlling the FDR, for each data stream $k \in [K]$, we define a sequence of statistics according to [26, Thm. 6],

$$S_n^{(k)} = \max_{1 \leq j \leq J} \left\{ \max_{a_m \leq |a| \leq n} \sum_{x=1}^{n} Z_i^{(k)}(j, 0) - \max_{1 \leq j \leq J} \max_{a_m \leq |a| \leq n} \sum_{x=1}^{n} Z_i^{(k)}(g, 0) \right\},$$

(52) where $m_a$ is a window size we choose and

$$Z_i^{(k)}(j, 0) = \log \left( \frac{h_i(X_i^{(k)})}{h_0(X_i^{(k)})} \right).$$

Let

$$I_{jk} = \int \log \left( \frac{h_i(x)}{h_0(x)} \right) b_j(x) P(dx).$$

In the sequel, we assume that

$$I^* = \min_{1 \leq j \leq J} \min_{0 \leq \log \frac{a}{K} \leq \log \frac{a}{K}} I_{jk} > 0.$$ (53)

Furthermore, we assume that the distribution of the change point satisfies

$$\log \left( \sum_{t=1}^{\infty} \tau_t \right) = o \left( \log \frac{\alpha}{K} \right)$$

(54) as $K \to \infty$.

In the following, we fix a parameter $\alpha \in (0, 1)$. According to [26], we choose $m_a$ to satisfy

$$\liminf_{m_a} \frac{1}{\log \frac{\alpha}{K}} > 1 \min_{1 \leq j \leq J} \min_{0 \leq \log \frac{a}{K} \leq \log \frac{a}{K}} I_{jk}.$$ (55)

We set an array of threshold values

$$Q_k \leq Q_{K-1} \leq \ldots \leq Q_1,$$ (56)

according to

$$[2(J-1)m_a + 2J]m_a e^{-Q_1} = \frac{s}{K} \alpha.$$ (57)

We present the decision procedure, called MDI-FDR, in Algorithm 3.

For the stopping times $[T^{(k)}]$, $k \in [K]$, obtained from MDI-FDR, we have

$$\sup_{t \geq 1} P_{\rho \alpha} \{ t \leq T^{(k)} < t + m_a \} \leq \frac{2Jm_a}{e^{Q_1}},$$ (59)

$$\sup_{1 \leq t \leq y} P_{\rho \alpha} \{ r \leq T^{(k)} < r + m_a, \hat{J}^{(k)} \neq J^{(k)} \} \leq \frac{2(J-1)m_a}{e^{Q_1}},$$ (60)

where $(T^{(k)}, J^{(k)})$ is obtained by the decision procedure MDI-FDR. Then we have

$$\sum_{j=1}^{J} \sum_{p=1}^{\infty} \sum_{l \geq 2} \sum_{p=1}^{\infty} p(f^{(k)}) \pi_{\rho \alpha} P_{\rho \alpha} \{ T^{(k)} < \infty, \hat{J}^{(k)} \neq J^{(k)} \} \leq 2(J-1)m_a e^{-Q_1},$$

(61)

where the first inequality follows from an argument similar to the proof of [26, Thm. 7], and the second inequality is from (54) and (55).

**Algorithm 3 MDI-FDR**

1. Initially set $q = 1$, $i_q = [K]$, $n = 1$, and $n_0 = 0$.
2. Sample the active data streams $X_n^{(l)} | \alpha > n_0$, and update the statistics $S_n^{(l)}$ using (52).
3. Sort the statistics in ascending order as $S_n^{(l(n,0))}$, where $i(n, 0)$ denotes the index of the $i$-th ordered statistic at time epoch $n$.
4. Repeat 2)-3) with $n$ increased by 1 each time, until $n$ equals $n_q = \bar{N} \wedge \min \{ n > n_{q-1} \} S_n^{(l(n,0))} \geq Q_{K-1}, \forall [i_q] \}.$

a) If $n_q < \bar{N}$, stop the $q$-th stage and declare change points at $n_q$ for the following data streams:

$$X_{n_q}^{(l(n_q, q-1))}, X_{n_q}^{(l(n_{q+1}, q-1))}, \ldots, X_{n_q}^{(l(n_q, l_q))},$$

where

$$I_q = \min \{ l \in [I_q] | S_n^{(l(n, l_q))} \geq Q_{K-1} \}.$$ (63)

Meanwhile, declare change types as

$$\hat{J}^{(k)} = \arg \max_{1 \leq j \leq J} \left\{ \max_{n_q \leq |a| \leq n_0} \sum_{x=1}^{n} Z_i^{(k)}(j, 0) \right\},$$ (64)

for these declared data streams above. Update $l_{q+1}$ to exclude the indices of these declared data streams. If $|l_{q+1}| = 0$, then stop; otherwise, set $q = q + 1$, and go to 2) to begin the next stage.

b) Otherwise, $n_q = \bar{N}$, declare that all the active data streams in $I_q$ have no change points, and stop.

The following result provides a bound on the FDR of MDI-FDR.
Theorem 4. For a given $0 < \alpha < 1$, the decision procedure MDI-FDR satisfies

$$\text{FDR} \leq \alpha + \frac{Kt + (K \log K - K \log \alpha) / \Gamma + o(K)}{N},$$

(65)

where $o(K)/K \to 0$ as $K \to \infty$. Hence if we choose $N$ with $K$ such that $(K \log K)/N \to 0$ as $K \to \infty$, we have

$$\text{FDR} \leq \alpha.$$

(66)

Proof. The proof is similar to that of Theorem 1, and thus we only indicate the difference between the proofs. Instead of (20), we define $W_{k,t}$ as

$$W_{k,t} = \left\{ \begin{array}{l} S_{n}^{(k)} \geq Q_{n}, \exists n < t^{(k)}, \text{or} \ j^{(k)} \neq j^{(k)}, \exists n \geq t^{(k)} \end{array} \right\}.$$  

(67)

So according to (58), we have $P_{\pi_{\mathcal{P}}}(W_{k,t}) \leq \alpha / K$. We can then follow the proof of FDRcont in Theorem 1. Upon imposing $N$, we utilize the following result about the expected delay in detection and isolation problems:

Lemma 5. [26, Thm. 7] If $(T^{(k)}, J^{(k)})$ satisfies (58), then for every $1 \leq j \leq J$, 

$$\sum_{i=1}^{\infty} \pi_{\mathcal{P}_{i}} E_{\pi_{\mathcal{P}_{i}}, j} \left[T^{(k)} - t^{(k)} \mid T^{(k)} \geq t^{(k)} \right]$$

$$\sim \frac{(1 + o(1)) \log \alpha}{\min_{0 \leq l \leq j \neq n} I_{l}^{(n)}} \quad \text{as} \quad \alpha \to 0.$$  

(68)

Hence, we have for every $k \in [K]$,

$$E_{\pi_{\mathcal{P}}} \left[T^{(k)} - t^{(k)} \mid T^{(k)} \geq t^{(k)} \right]$$

$$= \sum_{j=1}^{\infty} \sum_{t=1}^{\infty} p_{t}^{(j)} \pi_{\mathcal{P}_{t}} E_{\pi_{\mathcal{P}_{t}}, j} \left[T^{(k)} - t^{(k)} \mid T^{(k)} \geq t^{(k)} \right]$$

$$\leq \frac{1}{T} \frac{K}{\alpha} (1 + o(1)),$$  

(69)

where $o(1) \to 0$ as $K \to \infty$ in our situation. The expectation $E_{\pi_{\mathcal{P}}, j}$ is taken with respect to $P_{\pi_{\mathcal{P}}, j}$. Then we use (69) to replace (39) in the proof of Theorem 1, and the remaining part readily follows.

We also study the ADD performance of MDI-FDR, which is defined in the same way as in Section II. Applying (69), we have the following result regarding the ADD.

Theorem 6. For the decision procedure MDI-FDR controlling the FDR by $\alpha$, as $K$ grows without bound, we have:

$$\text{ADD} \leq \frac{1}{T} \left( \log \alpha (1 - \pi_{\alpha}) + 1 \right) (1 + o(1)).$$  

(70)

The proof is similar to that of Theorem 3 and is thus omitted here.

We can see that the statistics in MDI-FDR do not exploit the knowledge of the distribution of change types. So we also propose a heuristic statistic with a Bayesian flavor as:

$$S_{n}^{(k)} = \max_{1 \leq j \leq J} \left\{ \frac{G_{n}^{(k)}(j, 0)}{\max_{1 \leq l \leq J} \{G_{n}^{(k)}(g, 0) \vee 1\}} \right\},$$  

(71)

where $G_{n}^{(k)}(j, 0)$ is a statistic similar to (10) as:

$$G_{n}^{(k)}(j, 0) = G_{n}^{(k)}(j, 0) \pi_{\mathcal{P}_{n}^{(k)}}(j, 0) + \gamma_{n+1} \left(1 - I_{n}^{(n)}(j, 0) \right).$$  

(72)

where $\gamma_{n+1} = \Pr\left(\beta^{(k)} \geq n + 1\right)$ and $L_{n}^{(k)}(j, 0) = h_{j}(X_{n}) / h_{0}(X_{n})$, $1 \leq j \leq J$. We use the thresholds $\{Q_{n}\}$ in (12). This thus leads to a heuristic decision procedure called MDI-Heuristic in Algorithm 4. Unfortunately we have not established FDR control for MDI-Heuristic, but our numerical experiments in Section VI-B suggest that this heuristic decision procedure performs well. A more thorough theoretic analysis of MDI-Heuristic is left for future research.

Algorithm 4 MDI-Heuristic

1) Initially set $q = 1$, $I_{q} = \{K\}$, $n = 1$, and $n_{0} = 0$.

2) Sample the active data streams $\{X_{n}^{(i)}\}_{i \in \mathcal{I}_{n}, n > n_{0}}$, and update the statistics $\{S_{n}^{(i)}\}_{i \in \mathcal{I}_{n}, n > n_{0}}$ using (71).

3) Sort the statistics in ascending order as $\{S_{n}^{(i(n,l))}\}_{l \in \mathcal{I}_{n}, n > n_{0}}$, where $i(n,l)$ denotes the index of the $l$-th ordered statistic at time epoch $n$.

4) Repeat 2)-3) with $n$ increased by 1 each time, until $n$ equals $n_{q} = N \wedge \min \left\{ n > n_{q-1} | S_{n}^{(i(n,l))} \geq Q_{K-\gamma_{q, l}}, \forall l \in [I_{q}] \right\}$.

(73)

4a) If $n_{q} < N$, stop the $q$-th stage and declare change points at $n_{q}$ for the following data streams:

$$X_{n_{q}^{(i(n,l))}}, X_{n_{q}^{(i(n,l+1))}}, \ldots, X_{n_{q}^{(i(n,l))}},$$

where

$$I_{q} = \min \{l \in [I_{q}] | S_{n}^{(i(n,l))} \geq Q_{K-\gamma_{q, l}} \}.$$  

(74)

Meanwhile, declare change types as

$$j^{(q)} = \arg \max_{1 \leq j \leq J} \left\{ \frac{G_{n}^{(q)}(j, 0)}{\max_{1 \leq l \leq J} \{G_{n}^{(q)}(g, 0) \vee 1\}} \right\}.$$  

(75)

for these declared data streams above. Update $I_{q+1}$ to exclude the indices of these declared data streams. If $\lfloor I_{q+1} \rfloor = 0$, then stop; otherwise set $q = q + 1$, and go to 2) to begin the next stage.

b) Otherwise, $n_{q} = N$, declare that all the active data streams in $I_{q}$ have no change points, and stop.

Finally, we also consider decision procedures that control theFWER, which is defined as $\Pr\left[\mathcal{V} \geq 1\right]$. Here we provide a decision procedure based on Bonferroni’s procedure, denoted as MDI-Bonferroni, as described in Algorithm 5.

In closing, we note that for the special case of $J = 1$ (i.e., there is only one possible type of post-change distribution), the problem formulation of detection and isolation reduces into that of detection only treated in Sections II-IV, and the decision procedures presented in this section are also applicable for the problem formulation therein. Indeed, in the case study of stock price time series in Section VII, we use MDI-FDR since its statistics (52) do not depend upon the prior distribution of change points, which are difficult to estimate for real-world data sets.
Algorithm 5 MDI-Bonferroni

1) Initially set \( q = 1 \), \( I_q = \{ k \} \), \( n = 1 \), and \( n_0 = 0 \).
2) Sample the active data streams \( \{ X_n^{(k)} \}_{k \in I_q, n > n_0} \), and update the statistics \( S_{n}^{(k)} \) \( \forall k \), \( \forall n \geq n_0 \) using (52).
3) Repeat 2) with \( n \) increased by 1 each time, until \( n \) equals \( n_q = \bar{N} \wedge \min \{ n > n_{q-1} | S_n^{(k)} \geq Q_1, \exists k \in I_q \} \). (76)

a) If \( n_q = \bar{N} \), stop the \( q \)-th stage and declare change points at \( n_q \) for data streams that satisfy (76). Meanwhile, declare change types as

\[
\hat{j}^{(k)} = \arg \max _{l \leq j \leq J} \left\{ \frac{1}{n_q - n_{q-l}} \sum _{i=1}^{n_q} z_{n_0}^{(k)}(j, 0) \right\},
\]

for these declared data streams above. Update \( I_{q+1} \) to exclude the indices of these declared data streams. If \( |I_{q+1}| = 0 \), then stop; otherwise, set \( q = q + 1 \), and go to 2) to begin the next stage.

b) Otherwise, \( n_q = \bar{N} \), declare that all the active data streams in \( I_q \) have no change points, and stop.

VI. SIMULATION RESULTS

In this section, we present the numerical results of Monte Carlo simulations designed to compare the proposed decision procedures controlling the FDR and the FWER. In order to make a fair comparison, we set the upper bounds of the FDR and the FWER the same, as \( \alpha = 0.1 \). The deadline \( \bar{N} \) is 2000.

A. Multiple Change Detection

Regarding the prior probability distribution of change points, we set the probability of that there is no finite change point as \( \pi_{\infty} = 0.2 \), and set the distribution of finite change points as a (conditional) geometric distribution with parameter \( p = 0.1 \). Table II displays the estimated FDR/FWER for K i.i.d. data streams, with pre/post-change probability distributions:

\[
h_0 \sim N(0, 1) \quad \text{v.s.} \quad h_1 \sim N(1, 1).
\]

We see that all the decision procedures well satisfy the upper bound \( \alpha \).

<table>
<thead>
<tr>
<th>( K )</th>
<th>Procedure</th>
<th>Estimated FDR/FWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>MD-FDR</td>
<td>0.0092</td>
</tr>
<tr>
<td></td>
<td>MD-Bonferroni</td>
<td>0.0153</td>
</tr>
<tr>
<td></td>
<td>MD-Hochberg</td>
<td>0.0275</td>
</tr>
<tr>
<td></td>
<td>MD-FDR</td>
<td>0.0061</td>
</tr>
<tr>
<td></td>
<td>MD-Bonferroni</td>
<td>0.0129</td>
</tr>
<tr>
<td></td>
<td>MD-Hochberg</td>
<td>0.0211</td>
</tr>
<tr>
<td>200</td>
<td>MD-FDR</td>
<td>0.0063</td>
</tr>
<tr>
<td></td>
<td>MD-Bonferroni</td>
<td>0.0135</td>
</tr>
<tr>
<td></td>
<td>MD-Hochberg</td>
<td>0.0238</td>
</tr>
<tr>
<td>500</td>
<td>MD-FDR</td>
<td>0.0062</td>
</tr>
<tr>
<td></td>
<td>MD-Bonferroni</td>
<td>0.0116</td>
</tr>
<tr>
<td></td>
<td>MD-Hochberg</td>
<td>0.0220</td>
</tr>
</tbody>
</table>

TABLE III

ERROR PERFORMANCE FOR CHANGE DETECTION WITH I.I.D. GAUSSIAN DATA STREAMS.

<table>
<thead>
<tr>
<th>( K )</th>
<th>Procedure</th>
<th>Estimated FDR/FWER</th>
<th>ADD</th>
<th>( \bar{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>MD-FDR</td>
<td>0.0114</td>
<td>3.49</td>
<td>9.13</td>
</tr>
<tr>
<td></td>
<td>MD-Bonferroni</td>
<td>0.0154</td>
<td>3.98</td>
<td>9.02</td>
</tr>
<tr>
<td>100</td>
<td>MD-FDR</td>
<td>0.0137</td>
<td>3.50</td>
<td>90.87</td>
</tr>
<tr>
<td></td>
<td>MD-Bonferroni</td>
<td>0.0137</td>
<td>4.68</td>
<td>89.65</td>
</tr>
<tr>
<td>500</td>
<td>MD-FDR</td>
<td>0.0136</td>
<td>3.54</td>
<td>455.53</td>
</tr>
<tr>
<td></td>
<td>MD-Bonferroni</td>
<td>0.0151</td>
<td>5.29</td>
<td>449.50</td>
</tr>
<tr>
<td>1000</td>
<td>MD-FDR</td>
<td>0.0146</td>
<td>3.54</td>
<td>912.00</td>
</tr>
</tbody>
</table>

Fig. 2 compares the ADD performance of the decision procedures. We see that the ADD is not visibly affected by the number of data streams \( K \) for MD-FDR, but tends to grow logarithmically with \( K \) for MD-Bonferroni and MD-Hochberg. This observation confirms the analytical result in Theorem 3, which indicates that FDR control decision procedures tend to be more scalable compared with FWER control decision procedures.

B. Multiple Change Detection and Isolation

We again consider the prior probability distribution of change points as a mixture of a singleton at infinity with probability \( \pi_{\infty} \) and a geometric distribution with parameter \( p = 0.05 \). We further set the post-change type as a uniform distribution on \([1, \ldots, J]\), and consider \( K \) i.i.d. data streams with pre/post-change probability distributions:

\[
h_0 \sim N(0, 1) \quad \text{v.s.} \quad h_1 \sim N(\mu_1, 1), \quad p(j) = 1/J.
\]

We run simulations for four configurations as follows:

<table>
<thead>
<tr>
<th>( P_j )</th>
<th>( J = 2, \pi_{\infty} = 0.1, [\mu_1, \mu_2] = [2, 6], )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( P_1 )</td>
</tr>
<tr>
<td>2</td>
<td>( P_2 )</td>
</tr>
<tr>
<td>3</td>
<td>( P_3 )</td>
</tr>
<tr>
<td>4</td>
<td>( P_4 )</td>
</tr>
</tbody>
</table>

We see that all the decision procedures well satisfy the upper bound \( \alpha \).

<table>
<thead>
<tr>
<th>( K )</th>
<th>Procedure</th>
<th>Estimated FDR/FWER</th>
<th>ADD</th>
<th>( \bar{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>MD-FDR</td>
<td>0.0114</td>
<td>3.49</td>
<td>9.13</td>
</tr>
<tr>
<td></td>
<td>MD-Bonferroni</td>
<td>0.0154</td>
<td>3.98</td>
<td>9.02</td>
</tr>
<tr>
<td>100</td>
<td>MD-FDR</td>
<td>0.0137</td>
<td>3.50</td>
<td>90.87</td>
</tr>
<tr>
<td></td>
<td>MD-Bonferroni</td>
<td>0.0137</td>
<td>4.68</td>
<td>89.65</td>
</tr>
<tr>
<td>500</td>
<td>MD-FDR</td>
<td>0.0136</td>
<td>3.54</td>
<td>455.53</td>
</tr>
<tr>
<td></td>
<td>MD-Bonferroni</td>
<td>0.0151</td>
<td>5.29</td>
<td>449.50</td>
</tr>
<tr>
<td>1000</td>
<td>MD-FDR</td>
<td>0.0146</td>
<td>3.54</td>
<td>912.00</td>
</tr>
</tbody>
</table>

Tables III through VI display the results, wherein we include \( \bar{R} \), the average number of change points declared.
Comparing Tables III and IV, we see that as the gap between \( \mu_1 \) and \( \mu_2 \) decreases, the FDR/FWER and the ADD increase, reflecting that the task of isolation of post-change type becomes challenging. Comparing Tables III and V, we see that a larger probability of infinite change point considerably worsens the FDR/FWER, whereas still controlled by the prescribed upper bound. From Table VI, we see that an increase in the number of post-change types decreases the FDR/FWER, but at the cost of increasing the ADD. This is because in the decision procedure the thresholds increase with \( J \) and the error probability is with respect to the worst case. This suggests that when \( J \) is large, it may be possible to set smaller thresholds to control the error rates at a similar level but with reduced decision delays.

In terms of the ADD, MDI-FDR outperforms MDI-Bonferroni. The performance shown in Fig. 3 exhibits a similar trend as that in Fig. 2: the ADD in MDI-FDR is not visibly affected by \( K \), but tends to grow logarithmically with \( K \) in MDI-Bonferroni.

We also conduct simulations for the heuristic decision procedure MDI-Heuristic, and compare its performance with MDI-FDR in Table VII. We see that both decision procedures control the FDR at a similar level, and that MDI-Heuristic achieves lower ADD compared with MDI-FDR, which is consistent with our intuition that MDI-Heuristic makes better use of the prior information.

### VII. A Case Study

We conduct a case study in which we apply the proposed decision procedures to online monitoring of abrupt price changes in the stock market.

#### Table IV

<table>
<thead>
<tr>
<th>( K )</th>
<th>Procedure</th>
<th>Estimated FDR/FWER</th>
<th>ADD</th>
<th>( % )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>MDI-FDR</td>
<td>0.0306</td>
<td>4.96</td>
<td>9.28</td>
</tr>
<tr>
<td></td>
<td>MDI-Bonferroni</td>
<td>0.0379</td>
<td>5.69</td>
<td>9.04</td>
</tr>
<tr>
<td>100</td>
<td>MDI-FDR</td>
<td>0.0319</td>
<td>5.06</td>
<td>92.38</td>
</tr>
<tr>
<td></td>
<td>MDI-Bonferroni</td>
<td>0.0392</td>
<td>6.89</td>
<td>89.68</td>
</tr>
<tr>
<td>500</td>
<td>MDI-Bonferroni</td>
<td>0.0343</td>
<td>7.77</td>
<td>449.52</td>
</tr>
<tr>
<td>1000</td>
<td>MDI-FDR</td>
<td>0.0300</td>
<td>5.07</td>
<td>926.56</td>
</tr>
<tr>
<td></td>
<td>MDI-Bonferroni</td>
<td>0.0320</td>
<td>8.17</td>
<td>900.07</td>
</tr>
</tbody>
</table>

#### Table V

<table>
<thead>
<tr>
<th>( K )</th>
<th>Procedure</th>
<th>Estimated FDR/FWER</th>
<th>ADD</th>
<th>( % )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>MDI-FDR</td>
<td>0.0835</td>
<td>3.59</td>
<td>4.39</td>
</tr>
<tr>
<td></td>
<td>MDI-Bonferroni</td>
<td>0.0917</td>
<td>3.88</td>
<td>4.09</td>
</tr>
<tr>
<td>100</td>
<td>MDI-FDR</td>
<td>0.0790</td>
<td>3.80</td>
<td>45.81</td>
</tr>
<tr>
<td></td>
<td>MDI-Bonferroni</td>
<td>0.0787</td>
<td>4.77</td>
<td>40.43</td>
</tr>
<tr>
<td>500</td>
<td>MDI-FDR</td>
<td>0.0788</td>
<td>3.79</td>
<td>217.31</td>
</tr>
<tr>
<td></td>
<td>MDI-Bonferroni</td>
<td>0.0760</td>
<td>5.27</td>
<td>200.27</td>
</tr>
<tr>
<td>1000</td>
<td>MDI-FDR</td>
<td>0.0790</td>
<td>3.80</td>
<td>435.56</td>
</tr>
<tr>
<td></td>
<td>MDI-Bonferroni</td>
<td>0.0800</td>
<td>5.54</td>
<td>401.40</td>
</tr>
</tbody>
</table>

#### Table VI

<table>
<thead>
<tr>
<th>( K )</th>
<th>Procedure</th>
<th>Estimated FDR/FWER</th>
<th>ADD</th>
<th>( % )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>MDI-FDR</td>
<td>0.0001</td>
<td>7.21</td>
<td>90.04</td>
</tr>
<tr>
<td></td>
<td>MDI-Bonferroni</td>
<td>0.0028</td>
<td>8.13</td>
<td>90.02</td>
</tr>
<tr>
<td>100</td>
<td>MDI-FDR</td>
<td>0.0028</td>
<td>7.31</td>
<td>90.44</td>
</tr>
<tr>
<td></td>
<td>MDI-Bonferroni</td>
<td>0.0035</td>
<td>9.70</td>
<td>90.20</td>
</tr>
<tr>
<td>500</td>
<td>MDI-FDR</td>
<td>0.0029</td>
<td>10.97</td>
<td>450.52</td>
</tr>
<tr>
<td>1000</td>
<td>MDI-FDR</td>
<td>0.0029</td>
<td>7.33</td>
<td>903.22</td>
</tr>
<tr>
<td></td>
<td>MDI-Bonferroni</td>
<td>0.0027</td>
<td>11.62</td>
<td>900.79</td>
</tr>
</tbody>
</table>

A. Data Description

The data we use is from the Machine Learning Data Set Repository (http://www.mldata.org). The data streams are the daily stock prices (at closing) of \( K = 100 \) different stocks, each of length \( N = 1813 \) during the period between January 3, 2000 and March 30, 2007. Apparently, these stocks may experience multiple phases with distinct statistical characteristics, and this makes detecting changes in stock price processes challenging.

Fig. 4 displays an example of a stock (NYSE: CVH) price process. An eye inspection reveals that there are many local maximum/minimum points, which may correspond to changes in the statistical characteristics. Fig. 5 displays the same stock price process, using the daily relative prices \( d_i = \log(X_i/X_{i-1}) \), \( i = 1, \ldots, N - 1 \). We use the daily relative prices as the data stream when running our decision procedures. As shown in Fig. 5, it is difficult to visually identify abrupt changes in the daily relative prices, except for two outliers which are indeed the result of stock splits.

B. Offline Change Detection

In order to evaluate the performance of our decision procedures, we need to provide a reference of the true change points of the stock prices. Since each data stream may consist of multiple change points, for simplicity, we use an offline change detection rule to estimate its first change point. For this, we use the so-called Brodsky-Darkhovsky (BD) statistic [27], which for the \( k \)-th stock is defined as

\[
Y_n^{(k)} = \sqrt{\frac{n(N-n)}{N^2}} \left[ 1 - \frac{1}{n} \sum_{i=1}^n d_i^{(k)} - \frac{1}{N-n} \sum_{i=n+1}^N d_i^{(k)} \right].
\]
where \( 1 \leq n \leq N - 1 \). From the structure of this statistic, we see that it basically computes the difference between the sample means before and after each time epoch, and we thus divide the \( k \)-th data stream into two segments at

\[
\hat{p}^{(k)} = \arg \max_{1 \leq l \leq N-1} |\tilde{y}_n^{(k)}| + 1. \tag{80}
\]

Iteratively computing the BD statistics (79) for each segment and further dividing each segment according to (80), we obtain a series of dividing time epochs. We estimate the first change point as the minimum dividing time epoch such that its corresponding BD statistic exceeds a prescribed threshold, chosen as 0.001 in our case study.

For example, Fig. 6 shows the behavior of \( |\tilde{y}_n^{(k)}| \) of NYSE: CVH daily relative prices. We see that its maximum value occurs at time epoch 242, consistent with our observation made in Fig. 4. Following the iterative procedure described in the previous paragraph, we estimate the first change point at time epoch 50, and the second at time epoch 138. Using the segment from 1 to 50 and the segment from 51 to 138, we get the pre-change and post-change sample mean and standard deviation as \((\hat{\mu}_0, \hat{\sigma}_0) = (2.56 \times 10^{-3}, 0.0373)\) and \((\hat{\mu}_1, \hat{\sigma}_1) = (0.0076, 0.0437)\), respectively.

Running the above procedure for the 100 data streams, we estimate their first change points as well as the pre-change/post-change sample means/standard deviations, while still finding that there are 32 data streams without any change point.

**C. Online Change Detection**

We now apply the proposed sequential decision procedures to these stock price data streams. For ease of implementation, we assume that the pre-change and the post-change daily relative prices follow Gaussian distributions, with mean and variance fitted using the experiments in the offline change detection in the previous subsection. Since it is difficult to model the prior distribution of the change point, we use the decision procedures MDI-FDR and MDI-Bonferroni developed in Section V, with \( J = 1 \).

We tabulate the performance results in Table VIII, wherein we use the number of false discoveries to represent the error performance, because the experiment is one-shot only. Two decision procedures incur a similar number of false discoveries. Fig. 7 displays the decision delays for all the data streams (excluding those without any change point). We make some interesting observations. There is a break in the red curve at the 11th stock, because MDI-FDR does not detect its change point. Besides that, MDI-FDR generally achieves pretty small decision delays. In contrast, MDI-Bonferroni incurs a very large decision delay at the 34th stock, because it has a very stringent stopping criterion due to the FWER control. We display the daily stock prices of the 34th stock in Fig. 8.

---

**TABLE VII**

Comparison of MDI-FDR and MDI-Heuristic for change detection and isolation with i.i.d. Gaussian data streams.

<table>
<thead>
<tr>
<th>( K )</th>
<th>Procedure</th>
<th>Estimated FDR</th>
<th>ADD</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>MDI-FDR</td>
<td>0.0256</td>
<td>5.97</td>
</tr>
<tr>
<td></td>
<td>MDI-Heuristic</td>
<td>0.0303</td>
<td>3.56</td>
</tr>
<tr>
<td>100</td>
<td>MDI-FDR</td>
<td>0.0217</td>
<td>6.06</td>
</tr>
<tr>
<td></td>
<td>MDI-Heuristic</td>
<td>0.0263</td>
<td>3.66</td>
</tr>
<tr>
<td>500</td>
<td>MDI-FDR</td>
<td>0.0224</td>
<td>6.07</td>
</tr>
<tr>
<td></td>
<td>MDI-Heuristic</td>
<td>0.0260</td>
<td>3.67</td>
</tr>
<tr>
<td>1000</td>
<td>MDI-FDR</td>
<td>0.0220</td>
<td>6.07</td>
</tr>
<tr>
<td></td>
<td>MDI-Heuristic</td>
<td>0.0260</td>
<td>3.67</td>
</tr>
</tbody>
</table>

Fig. 4. Daily stock prices for NYSE: CVH in the period between January 3, 2000 and March 30, 2007.

Fig. 5. Daily relative prices for NYSE: CVH in the period between January 3, 2000 and March 30, 2007.

Fig. 6. BD statistics of NYSE: CVH daily relative prices.

Fig. 7. Decision delays for all the data streams (excluding those without any change point). We use the number of false discoveries to represent the error performance, because the experiment is one-shot only.
To cope with the emerging situation of analyzing large-scale real-time streaming data, in this work, we examined the problem of sequentially detecting changes in parallel data streams, and brought the error metric of FDR into the problem formulation. Our proposed FDR-oriented decision procedure is a sequential variant of the classical BH procedure, where the \( p \)-value is replaced with a sequential detection statistic. Thanks to the characteristic of FDR-oriented decision procedures, when it comes to change detection problems, the average decision delay is significantly reduced compared with conventional FWER-oriented decision procedures. Our theoretical findings are also corroborated by numerical experiments and case studies using real-world data sets.

**References**


