Finite sample change point inference and identification for high-dimensional mean vectors

Mengjia Yu | Xiaohui Chen

Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Correspondence
Mengjia Yu, Department of Statistics, University of Illinois at Urbana-Champaign, 725 S. Wright Street, Champaign, IL 61820, USA.
Email: myu17@illinois.edu

Funding information
NSF, Grant/Award Number: DMS-1404891 and DMS-1752614; UIUC Research Board Awards, Grant/Award Number: RB17092 and RB18099; Simons Fellowship

Abstract
Cumulative sum (CUSUM) statistics are widely used in the change point inference and identification. For the problem of testing for existence of a change point in an independent sample generated from the mean-shift model, we introduce a Gaussian multiplier bootstrap to calibrate critical values of the CUSUM test statistics in high dimensions. The proposed bootstrap CUSUM test is fully data dependent and it has strong theoretical guarantees under arbitrary dependence structures and mild moment conditions. Specifically, we show that with a boundary removal parameter the bootstrap CUSUM test enjoys the uniform validity in size under the null and it achieves the minimax separation rate under the sparse alternatives when the dimension $p$ can be larger than the sample size $n$.

Once a change point is detected, we estimate the change point location by maximising the $\ell^\infty$-norm of the generalised CUSUM statistics at two different weighting scales corresponding to covariance stationary and non-stationary CUSUM statistics. For both estimators, we derive their rates of convergence and show that dimension impacts the rates only through logarithmic factors, which implies that consistency of the CUSUM estimators is possible when $p$ is much larger than $n$. In the presence of multiple change points, we propose a principled bootstrap-assisted binary segmentation (BABS) algorithm to dynamically adjust the change point detection rule and recursively estimate their locations. We derive its rate of convergence under suitable signal separation and strength conditions.

The results derived in this paper are non-asymptotic and we provide extensive simulation studies to assess the finite sample performance. The empirical evidence shows an encouraging agreement with our theoretical results.
1 | INTRODUCTION

This paper studies the problems of change point inference and identification for mean vectors of high-dimensional data in finite samples. High-dimensional data are now ubiquitous in many scientific and engineering fields and data heterogeneity is the rule rather than the exception. A central problem of studying data heterogeneity is to detect structural breaks in the underlying data generation process. Perhaps the two most fundamental questions for abrupt changes are: (i) is there a change point in data? (ii) if so, when does the change occur? In this work, we consider change point detection and identification for temporally independent data with cross-sectional dependence. Specifically, let \( X_n = \{X_1, \ldots, X_n\} \) be a sequence of independent random vectors in \( \mathbb{R}^p \) generated from the mean-shift model:

\[
X_i: = (X_{i1}, \ldots, X_{ip})' = \mu + \delta_n 1(i > m) + \xi_i, \quad i = 1, \ldots, n, \tag{1}
\]

where \( \mu \in \mathbb{R}^p \) is the population mean parameter, \( \delta_n \in \mathbb{R}^p \) is the mean-shift signal parameter, \( m \) is the change point location, and \( \xi_1, \ldots, \xi_n \) are (temporally) independent and identically distributed (i.i.d.) mean-zero noise random vectors in \( \mathbb{R}^p \) with common distribution function \( F \). Let \( \Sigma = \text{Cov}(\xi_1) \) be the unknown noise covariance matrix that is not necessarily diagonal, and thus we allow cross-sectional (sometimes also referred as spatial) dependence among the components \( X_{i1}, \ldots, X_{ip} \) for each \( i = 1, \ldots, n \). Under the mean-shift model, if \( \delta_n = 0 \) or \( m = n \), then \( X_1, \ldots, X_n \) form a sample of i.i.d. random vectors and no change point occurs. In this paper, our first goal is to test for whether or not there is a change point in the mean vectors \( \mu_i = \mathbb{E}(X_i) \), that is, to test for

\[
H_0: \delta_n = 0 \quad \text{and} \quad H_1: \delta_n \neq 0 \quad \text{and there exists an} \ m \in (1, \ldots, n - 1), \tag{2}
\]

where the alternative hypothesis \( H_1 \) is parameterised by the change point signal \( \delta_n \) and location \( m \). If a change point is detected in the mean vectors (i.e. \( H_0 \) is rejected), then our second goal is to estimate the change point location \( m \).

For i.i.d. Gaussian noise \( \xi_i \sim N(0, \Sigma) \), the log-ratio of the maximised likelihoods between \( H_1 \) with a change point at \( s = 1, \ldots, n - 1 \) and \( H_0 \) without change point is given by

\[
\log(A_s) = Z_n(s)^\top \Sigma^{-1} Z_n(s) / 2, \tag{3}
\]

where

\[
Z_n(s) = \sqrt{\frac{s(n-s)}{n}} \left( \frac{1}{s} \sum_{i=1}^{s} X_i - \frac{1}{n-s} \sum_{i=s+1}^{n} X_i \right) \tag{4}
\]

is a sequence of the normalised mean differences before and after \( s \). Then \( H_0 \) is rejected if \( \max_{1 \leq s \leq n} \log(A_s) \) is larger than a critical value. In literature, \( \{Z_n(s)\}_{s=1}^{n-1} \) are often called the cumulative sum (CUSUM) statistics (Csörgő & Horváth, 1997). Note that the log-ratio statistics of the maximised likelihoods in

KEYWORDS
binary segmentation, bootstrap, CUSUM, change point analysis, Gaussian approximation, high-dimensional data
Equation (3) require the knowledge or an estimate of the unknown covariance matrix $\Sigma$. In the high-dimensional setting where $p$ is larger (or even much larger) than $n$, estimation of $\Sigma$ itself becomes a challenging problem. Spectral norm consistency of $\Sigma$ (or the inverse $\Sigma^{-1}$) is possible under additional structural assumptions (such as sparsity or low-rankness) on the covariance matrix (Barigozzi et al., 2018; Bickel & Levina, 2008a,b; Cai & Zhou, 2012; Cai et al., 2010; Chen et al., 2013), which may not hold in practical applications. In contrast, tests based on the CUSUM statistics in Equation (4) do not involve $\Sigma$ and they are more robust to the misspecification on covariance structures. Therefore, this motivates us to study the problems of change point testing and estimation based on the high-dimensional CUSUM statistics.

To build a decision rule for change-point detection, we need to cautiously aggregate the (dependent) random vectors $Z_n(s), s = 1, \ldots, n − 1$. Enikeeva and Harchaoui (2019) considers the change point detection on mean vectors under the mean-shift model (1) with i.i.d. $\xi_i \sim N(0, \sigma^2 I_p)$. They propose the linear and scan statistics based on the $\ell^2$-norm aggregation of the CUSUM statistics and derive the change point detection boundary. Jirak (2015) considers the $\ell^\infty$-norm aggregation of the CUSUM statistics and establishes a Gumbel limiting distribution under $H_0$. Jirak (2015) also considers the bootstrap approximations to improve the rate of convergence. Wang and Samworth (2018) considers the estimation in the high-dimensional mean vectors in reduced dimensions by sparse projections and they derive the rate of convergence for estimating the change point location. In the three aforementioned papers, strong structural assumptions (i.e. cross-sectional sparsity in the sense that the components of $X_i$ are independent or weakly dependent) are imposed to substantially reduce the intrinsic complexity. Cho and Fryzlewicz (2015) relaxes the sparsity assumption and consider the estimation problem in the (marginal) variances of high-dimensional time series under a multiplicative process in $\mathbb{R}^p$. However, even for $\rho$ components of CUSUM statistics and it may be viewed as a data-driven alternative for selecting the threshold in Cho and Fryzlewicz (2015).

In this paper, besides some mild moment conditions, we do not make any assumption on the cross-sectional dependence structure of the underlying data distribution. We consider the multivariate CUSUM statistics (4) in the $\ell^\infty$-norm aggregated form:

$$T_n = \max_{s \leq s \leq n - \frac{1}{2}} |Z_n(s)| = \max_{s \leq s \leq n - \frac{1}{2}} \max_{1 \leq j \leq \rho} |Z_{nj}(s)| = \max_{s \leq s \leq n - \frac{1}{2}} \max_{1 \leq j \leq \rho} |Z_{nj}(s)|,$$

where $s \in [1, n/2]$ is a user-specified boundary removal parameter. Removing boundary points is necessary in change point detection since the distributions of $|Z_n(s)|$ that are closer to endpoints are more difficult to approximate because of fewer data points. Then $H_0$ is rejected if $T_n$ is larger than a critical value such as the $(1 - \alpha)$ quantile of $T_n$. Under $H_0$, $\{Z_n(s)\}_{s=1}^{n-1}$ is a centered and covariance stationary process in $\mathbb{R}^p$ (i.e. $\mathbb{E}[Z_n(s)] = 0$ and $\text{Cov}(Z_n(s)) = \Sigma$). To approximate the distribution of $T_n$, extreme value theory is a common technique to derive the Gumbel-type limiting distributions (Leadbetter et al., 1983; Resnick, 1987). However, even for $p = 1$, the convergence rate of maxima of the CUSUM process $\{Z_n(s)\}_{s=1}^{n-1}$ is known to be slow (Hall, 1991; Resnick, 1987).

### 1.1 Our contributions

To overcome the fundamental difficulty in calibrating the distribution of $T_n$, we consider the bootstrap approximation to the finite sample distribution of $T_n$ without referring to a weak limit of $\{Z_n(s)\}_{s=1}^{n-1}$. 
In Section 2, we propose a Gaussian multiplier bootstrap tailored to the CUSUM test statistics in Equation (4). The proposed test is fully data dependent and requires no tuning parameter (except for a pre-specified boundary removal parameter $\delta$). This is in contrast with the thresholding-aggregation method in Cho and Fryzlewicz (2015), which requires further procedure to select threshold that is not easy to justify. We will show in Section 3.1 that the proposed test is a uniformly valid inferential procedure under $H_0$ where $p$ can grow sub-exponentially fast in $n$ and no explicit condition on the dependence structure among $\{X_{ij}\}_{j=1}^p$ is needed. This is in contrast with Enikeeva and Harchaoui (2019), Jirak (2015) and Wang and Samworth (2018) where the components are assumed to be either independent or weakly dependent, and with Cho and Fryzlewicz (2015) and Cho (2016) where the dimension can only grow polynomially fast in sample size. Moreover, we will show that, under a mild signal strength condition, our bootstrap CUSUM test is consistent in the sense that the sum of type I and type II errors is asymptotically vanishing (Giné & Nickl, 2015, Chapter 6.2). In addition, the requirement on the signal strength can achieve the minimax separation rate derived in Enikeeva and Harchaoui (2019) under the sparse alternative (i.e. the change occurs only in a few number of components).

If a change point is detected, then we estimate its location by maximising the $\ell_\infty$-norm of the generalised CUSUM statistics (8) at two different weighting scales. The first estimator is based on the covariance stationary CUSUM statistics in Equation (4). In Section 3.2, we show that it is consistent in estimating the location at the parametric rate $n^{-1/2}$ (up to logarithmic factors) for sub-exponential observations. The second estimator is a non-stationary CUSUM statistics assigning less weights on the boundary data points. In this case, we show that it achieves the best possible rate of convergence on the order $n^{-1}$ (up to logarithmic factors) under stronger side conditions. In both cases, dimension impacts the rate of convergence only through the logarithmic factors. Thus, consistency of the CUSUM location estimators can be achieved when $p$ grows sub-exponentially fast in $n$.

Our bootstrap change point inference can be naturally extended to handle multiple change points via the generic binary segmentation (BS) technique. Once a change point is claimed by our bootstrap test and located by our estimator, BS continues the same testing and estimation procedure on the segments before and after the change until no further change point can be detected by the bootstrap test (cf. Algorithm 1 in Section 2.3). Thus, the bootstrap CUSUM test can dynamically adjust detection rule during the iterations. We derive the rate of convergence of this bootstrap-assisted binary segmentation (BABS) for recursively estimating the multiple change points under suitable signal separation and strength conditions.

1.2 | Literature review

CUSUM statistics (Page, 1955) are originally introduced in the sequential testing problems to distinguish between the in-control hypothesis $\delta_n = 0$ and the out-control mean-shift hypothesis for a given $\delta_n \neq 0$ in model (1), aiming to minimise the expected average run length (Chernoff & Zacks, 1964; Hinkley, 1970; Lai, 2001; Lorden, 1971; Page, 1955; Qiu, 2014; Wald, 1945; Wald & Wolfowitz, 1948; Woodall & Ncube, 1985). The current paper uses CUSUM statistics for fixed sampled size tests, as in many other statistical change point testing and estimation works (Aue et al., 2009; Bai, 1997; Berkes et al., 2009; Bhattacharya, 1987; Brodsky & Darkhovsky, 1993; Carlstein, 1988; Csörgő & Horváth, 1997; Frick et al., 2014; Fryzlewicz, 2014; Garreau & Arlot, 2018; Harchaoui & Lévy-Leduc, 2010; Hariz et al., 2007; Harlé et al., 2016; Kokoszka & Leipus, 2000; Loader, 1996; Ombao et al., 2005; Yao, 1987; Zhang et al., 2010). There is a recent surge of literature on change point analysis for high-dimensional data: change point detection is considered in (Enikeeva & Harchaoui, 2019; Jirak, 2015), estimation of the number and locations of change points is considered in (Cho,
2016; Cho & Fryzlewicz, 2015; Jirak, 2015; Wang & Samworth, 2018), and bootstrap inference is considered in (Cho, 2016) (without giving rigorous statistical guarantees).

Finite sample approximations to the distribution of maxima for sums of independent mean-zero random vectors in high dimensions are studied in Chernozhukov et al. (2013, 2017). We highlight that validity of our bootstrap CUSUM test for the change point does not (at least directly) follow the Gaussian and bootstrap approximation results in Chernozhukov et al. (2013, 2017). The reason is that, in the change-point detection context, the extreme-value type test statistic \( T_n \) defined in Equation (5) is the maximum of a sequence of dependent random vectors \( Z_n(s), s = s, \ldots, n - s \). Therefore, the distributional approximation results developed in Chernozhukov et al. (2013, 2017) require considerable modifications tailored to the change point analysis. A main technical innovation of this work is that the CUSUM statistics are affine transformations of the independent data points in an augmented space so that we can make use of the high-dimensional Gaussian and bootstrap approximations without overpaying the price of the increased dimensionality in the embedded larger space.

### 1.3 | Organisation

The rest of this paper is organised as follows. The bootstrap change point test, change point location estimators and extension to multiple change points algorithm are described in Section 2. In Section 3, we derive the size validity and power properties of the bootstrap test, the rate of convergence for two change point location estimators and the consistency of BABS. Sections 4 and 5 provide extensive simulation results and two real data examples, respectively. Discussions on detailed comparisons with literature, proofs of the main results, and additional simulation results are given in the Supplementary Material (SM).

### 1.4 | Notation

For \( q > 0 \) and a generic vector \( x \in \mathbb{R}^d \), we denote \( |x|_q = \left( \sum_{i=1}^p |x_i|^q \right)^{1/q} \) for the \( \ell^q \) norm of \( x \) and we write \( |x| = |x|_2 \). For a random variable \( X \), denote \( \|X\|_q = (\mathbb{E}|X|^q)^{1/q} \). For \( \beta > 0 \), let \( \psi_\beta(x) = \exp(x^\beta) - 1 \) be a function defined on \( [0, \infty) \) and \( L_{\psi_\beta} \) be the collection of all real-valued random variables \( X \) such that \( \mathbb{E}[\psi_\beta(|X|/C)] < \infty \) for some \( C > 0 \). For \( X \in L_{\psi_\beta} \) define \( \|X\|_{\psi_\beta} = \inf\{C > 0 : \mathbb{E}[\psi_\beta(|X|/C)] \leq 1\} \). Then, for \( \beta \in [1, \infty) \), \( \| \cdot \|_{\psi_\beta} \) is an Orlicz norm and \( (L_{\psi_\beta}, \| \cdot \|_{\psi_\beta}) \) is a Banach space (Ledoux & Talagrand, 1991). For \( \beta \in (0, 1) \), \( \| \cdot \|_{\psi_\beta} \) is a quasi-norm, that is, there exists a constant \( C(\beta) > 0 \) such that \( \|X + Y\|_{\psi_\beta} \leq C(\beta) (\|X\|_{\psi_\beta} + \|Y\|_{\psi_\beta}) \) holds for all \( X, Y \in L_{\psi_\beta} \) (Adamczak, 2008). Let \( \rho(X, Y) = \sup_{t \in \mathbb{R}} |\mathbb{P}(X \leq t) - \mathbb{P}(Y \leq t)| \) be the Kolmogorov distance between two random variables \( X \) and \( Y \). We shall use \( C_1, C_2, \ldots \) and \( K_1, K_2, \ldots \) to denote positive and finite constants that may have different values. We assume \( n \geq 4 \) and \( p \geq 3 \).

### 2 | METHODOLOGY

#### 2.1 | Bootstrap CUSUM test

We first introduce a bootstrap procedure to approximate the distribution of \( T_n \). Let \( e_1, \ldots, e_n \) be i.i.d. \( N(0, 1) \) random variables independent of \( X_1^s \). Let \( \bar{X}_{s}^- = s^{-1} \sum_{i=1}^{s} X_i \) and \( \bar{X}_{s}^+ = (n - s)^{-1} \sum_{i=s+1}^{n} X_i \) be the left and right sample averages at \( s \). Define
Then the bootstrap test statistic is defined as
\[ T^*_n = \max_{1 \leq s \leq n - 1} |Z^*_n(s)|_\infty, \] (7)
and the \((1 - \alpha)\) conditional quantile of \(T^*_n\) given \(X^n_1\), defined as
\[ q_{T^*_n | X^n_1}(1 - \alpha) = \inf \{ t \in \mathbb{R} : P(T^*_n \leq t | X^n_1) \geq 1 - \alpha \}, \]
is used as critical values of the bootstrap test to approximate the quantiles of \(T^*_n\). In particular, for any \(\alpha \in (0, 1)\), we reject \(H_0\) if \(T^*_n > q_{T^*_n | X^n_1}(1 - \alpha)\).

Note that the Gaussian multiplier bootstrap test statistic \(T^*_n\) and its conditional quantile \(q_{T^*_n | X^n_1}(1 - \alpha)\) are 
\textit{computable} since we can draw Monte Carlo samples by simulating the multiplier random variables \(e_1, \ldots, e_n\) to approximate the distribution of \(T^*_n\).

\textbf{Remark 1} (Comments on centering terms in the bootstrap CUSUM statistics) We may also consider the following version of bootstrap CUSUM statistics
\[ \tilde{Z}^*_n(s) = \sqrt{\frac{n - s}{ns}} \sum_{i=1}^{s} e_i X_i - \sqrt{\frac{s}{n(n - s)}} \sum_{i=s+1}^{n} e_i X_i \]
without left and right centering by \(\bar{X}_s\) and \(\bar{X}_s^+\). It can be shown that the bootstrap CUSUM tests based on \(Z^*_n(s)\) and \(\tilde{Z}^*_n(s)\) have the same rate of convergence in the size and power analysis (Theorem 3.1, Corollary 3.2, and Theorem 3.3). Since \(\text{Cov}(Z^*_n(s) | X^n_1) = \text{Cov}(\tilde{Z}^*_n(s) | X^n_1)\) as a matrix inequality, \(\tilde{Z}^*_n(s)\) incurs a larger (conditional) covariance matrix than \(Z^*_n(s)\) and it is recommended to use \(Z^*_n(s)\) rather than \(\tilde{Z}^*_n(s)\).

\textbf{Remark 2} (Generalisation to time series: a block multiplier bootstrap CUSUM test) Since the CUSUM test statistics \(Z_n(s)\) in Equation (4) can be re-written as a block sum, the Gaussian multiplier bootstrap CUSUM test statistics in Equations (6) and (7) can be modified to a block version to accommodate the temporal dependence for time series data. Let \(M, B\) be positive integers such that \(n = MB\). We divide the sample \(X^n_i\) into \(B\) blocks of size \(M\). In particular, for \(b = 1, \ldots, B\), let \(L_b = \{(b - 1)M + 1, \ldots, bM\}\) be the \(b\)-th block indices. Then, for \(s = 1, \ldots, n - 1\), we can write \(Z_n(s)\) in Equation (4) as
\[ Z_n(s) = \sqrt{\frac{n - s}{ns}} \sum_{b=1}^{B} \sum_{i \in L_b} X_i 1(i \leq s) - \sqrt{\frac{s}{n(n - s)}} \sum_{b=1}^{B} \sum_{i \in L_b} X_i 1(i > s). \]

For any \(\alpha \in (0, 1)\), we reject \(H_0\) if the test statistic \(T_n = \max_{1 \leq s \leq n - s} |Z_n(s)|_\infty\) is larger than a critical value. Since the distributions of \(Z_n(s)\) under \(H_0\) and \(H_1\) for dependent error processes are different from the i.i.d. errors, we need to accommodate the dependence in calibrating the distributions of the test statistic \(T_n\). The idea is to modify the Gaussian multiplier bootstrap \(Z^*_n(s)\) in Equation (6) and the bootstrap CUSUM test statistic \(T^*_n\) in Equation (7) to their block versions. Specifically, to approximate the (finite sample) distribution of \(T^*_n\), we use a \textit{block Gaussian multiplier bootstrap} tailored to the CUSUM statistics setting. Let \(e_1, \ldots, e_B\) be i.i.d. standard Gaussian random variables. Define
\[ Z^*_n(b) = \sqrt{\frac{n - s}{ns}} \sum_{b=1}^{B} e_b V^-_b(s) - \sqrt{\frac{s}{n(n - s)}} \sum_{b=1}^{B} e_b V^+_b(s), \]
where \(V^-_b(s)\) and \(V^+_b(s)\) are block Gaussian multiplier random variables.
where \( V_b^-(s) = \sum_{i \in L_b} (X_i - \bar{X}_s^-) \mathbf{1}(i \leq s) \) and \( V_b^+(s) = \sum_{i \in L_b} (X_i - \bar{X}_s^+) \mathbf{1}(i > s) \). Then, the distribution of \( T_n \) is approximated by its bootstrap analog given by \( T_{n}^B = \max_{s \leq s \leq n-s} |Z_{n}^b(s)|_{\infty} \) and we reject \( H_0 \) if \( T_n > q_{r_1}^{\alpha} |X_1^r| (1 - \alpha) \), where \( q_{r_1}^{\alpha} |X_1^r| (1 - \alpha) \) is the \((1 - \alpha)\) conditional quantile of \( T_{n}^B \) given \( X_i^r \). Note that if the block size \( M = 1 \) (i.e. \( B = n \)), then \( Z^b_{n}(s) = Z^{*}_{n}(s) \). Thus, the bootstrap CUSUM test statistic for independent sequences is a special case of the block CUSUM test statistic. Generally, larger \( M \) is needed for stronger temporal dependence, while this would reduce the effective sample size. Some empirical performance of the block bootstrap CUSUM test is assessed in the SM Section D.3.

### 2.2 Estimating the change point location under the alternative hypothesis

If a change point is detected in the mean vectors (i.e. \( H_0 \) is rejected), then our next goal is to identify the change point location \( m \). Specifically, we estimate \( t_m = m/n \), \( m = 1, \ldots, n \), where the data \( X_i^n \) are observed at evenly spaced time points and their index variables are normalised to \([0, 1]\). We consider the change point location estimator based on the generalised CUSUM statistics (Hariz et al., 2007):

\[
Z_{\theta,n}(s) = \left[ \frac{s(n-s)}{n} \right]^{1-\theta} \left( \frac{1}{s} \sum_{i=1}^{s} X_i - \frac{1}{n-s} \sum_{i=s+1}^{n} X_i \right),
\]

where \( 0 \leq \theta < 1 \) is a weighting parameter. Obviously, the CUSUM statistics \( Z_n(s) \) in Equation (4) is a special case of \( \theta = 1/2 \), that is, \( Z_n(s) = Z_{1/2,n}(s) \). Then, we estimate \( m \) by

\[
\hat{m}_\theta = \arg \max_{1 \leq s < n} |Z_{\theta,n}(s)|_{\infty}.
\]

and we use \( t_{\hat{m}_\theta} = \hat{m}_\theta / n \) to estimate \( t_m \). It is seen that, for smaller values of \( \theta \), \( Z_{\theta,n}(s) \) assigns less weights on the boundary data points. Therefore, if the true change point location is bounded away from the two endpoints, we expect that \( t_{\hat{m}_\theta} \) with a smaller weighting parameter can achieve better rate of convergence. For example, if \( t_m \in (0, 1) \) is fixed and \( p = 1 \), then it is known that the \( \{Z_{\theta,n}(s)\}_{s=1}^{n-1} \) converges weakly to a functional of the Weiner process and the corresponding maximiser \( \hat{m}_\theta \) achieves the best possible rate of convergence of the order \( n^{-1} \) (Bai, 1997; Hariz et al., 2007). Instead of considering the whole family of the generalised CUSUM statistics indexed by \( \theta \in [0, 1] \), we consider two important cases of \( \theta = 1/2 \) (covariance stationary) and \( \theta = 0 \) (non-stationary) in this paper. For \( \theta = 1/2 \), \( Z_{1/2,n}(s) \) is related to the proposed bootstrap CUSUM statistics \( Z^*(s) \) in Equation (6) and the log-ratio statistics in Equation (3) under normality with \( \Sigma = \sigma^2 \text{Id}_p \). For \( \theta = 0 \), \( Z_{0,n}(s) \) is related to the parametric bootstrap in Jirak (2015).

**Remark 3** (Comments on the boundary removal) It should be noted that we must remove boundary points to approximate the distribution of \( T_{n}^* \). If the boundary points are included in the maxima \( T_n \) and \( T^*_n \), then the conditional distribution of \( T^*_n \) (given \( X_1^r \)) does not provide an accurate approximation to the distribution of \( T_{n}^* \). Theorems 3.1 and 3.3 provide the precise rate of convergence that characterises the boundary removal parameter \( \xi \) to ensure the consistency (in terms of the sum of type I and type II errors) of the bootstrap CUSUM test. Moreover, the estimation problem in Equation (9) does not exclude the endpoints outside \([\xi, n-\xi]\). However, in practice, if the existence of a change point is not known as a priori and it is decided by a test, then the boundary
restriction is implicitly imposed for both testing and estimation in empirical applications (Bai, 1997). Further discussions on the theoretical choice of \( \tilde{\gamma} \) can be found in Remark 4.

2.3 Bootstrap-assisted binary segmentation (BABS) for multiple change points

Suppose there are \( \nu \) change points \( m_0 = 1 < m_1 < \ldots < m_\nu < m_{\nu+1} = n \) and consider the following multiple mean-shifts model:

\[
X_i = \mu + \sum_{k=1}^{\nu} \delta_n^{(k)} 1(i > m_k) + \xi_i, \quad i = 1, \ldots, n, \tag{10}
\]

where \( \delta_n^{(k)} \in \mathbb{R}^p \) are non-zero mean-shift vectors and \( \xi_i \) are again i.i.d. mean-zero random vectors in \( \mathbb{R}^p \). Without loss of generality, we may assume \( \mu = 0 \) and \( \delta_n^{(0)} = \delta_n^{(\nu+1)} = 0 \) such that the mean vectors \( \mu_i = \mathbb{E}[X_i] \) are piecewise constant \( \mu_{1+m_k} = \ldots = \mu_{m_{k+1}} = \sum_{l=0}^{k} \delta_n^{(l)} \). Given a beginning time point \( b \) and an ending time point \( e \), we can compute the CUSUM statistics on the initial data segment \( \{X_i\}_{i=b}^{e} \):

\[
Z_{n,b,e}(s) = \sqrt{\frac{(s-b+1)(e-s)}{e-b+1}} \left( \frac{1}{s-b+1} \sum_{i=b}^{s} X_i - \frac{1}{e-s} \sum_{i=s+1}^{e} X_i \right).
\]

Note that the normalisation in \( Z_{n,b,e}(s) \) corresponds to the case \( \theta = 1/2 \) in Equation (8). It can be shown that the maximiser of \( |\mathbb{E}Z_{n,b,e}(s)|_\infty \) s = b, \ldots, e always occurs at one of the change points \( \{ m_k, k = 1, \ldots, \nu \} \cap \{ b, e \} \) (cf. Lemma B.6 in the SM Section B). Therefore, under multiple change points model (10), we can use \( Z_{n,b,e}(s) \) to locate one shift in the interval [b,e]. If our bootstrap CUSUM test (calculated based on \( \{ X_i \}_{i=b}^{e} \)) rejects \( H_0 \) at the significance level \( \alpha \), then \( \hat{m}_b^e = \arg \max_{s=b, \ldots, e} |Z_{n,b,e}(s)|_\infty \) is marked as a change point. Thus, we may recursively apply the binary segmentation to search along the two directions \( [b, \hat{m}_b^e] \) and \( [\hat{m}_b^e + 1, e] \) until no further change point would be detected by the subsequent bootstrap tests. The pseudo-code for our bootstrap-assisted binary segmentation algorithm for multiple change points detection and estimation, referred as BABS(\( \alpha, b, e \)), is summarised in Algorithms 1.

**Algorithm 1. BABS(\( \alpha, b, e \))**

1: \( \text{if } e - b + 1 < 2\tilde{\gamma} \text{ then} \)
2: \( \quad \text{STOP} \)
3: \( \text{else} \)
4: \( \quad \hat{m}_b^e = \arg \max_{s=b, \ldots, e} |Z_{n,b,e}(s)|_\infty \)
5: \( \text{if } \text{our bootstrap CUSUM test concludes the existence of a change in [b,e] then} \)
6: \( \quad \text{add } \hat{m}_b^e \text{ to the set of estimated change-points;} \)
7: \( \quad \text{BABS(\( \alpha, b, \hat{m}_b^e \));} \)
8: \( \quad \text{BABS(\( \alpha, \hat{m}_b^e + 1, e \));} \)
9: \( \text{else} \)
10: \( \quad \text{STOP} \)
11: \( \quad \text{end if} \)
12: \( \text{end if} \)
13: \( \text{return estimated change points.} \)
THEORETICAL RESULTS

3.1 Size and power of the bootstrap CUSUM test: single change point

Denote \( P_0(\cdot) \) and \( P_1(\cdot) \) as the probability computed under \( H_0 \) and \( H_1 \), respectively. Our first main result (cf. Theorem 3.1) is to establish finite sample bounds for the (random) Kolmogorov distance between \( T_n \) and \( T_n^* : \rho^*(T_n, T_n^*) = \sup_{t \in \mathbb{R}} |P_0(T_n \leq t) - P_0(T_n^* \leq t|X_1^n)| \). From this, we can derive the asymptotic bootstrap validity for certain high-dimensional scaling limit of \((n, p)\). Particularly, with \( \rho^*(T_n, T_n^*) = o_P(1) \), we can show that type I error of the test is asymptotically controlled at the exact nominal level \( \alpha \in (0, 1) \), that is, \( P_0(T_n > q_{T_n^*} |X_1^n(1 - \alpha)) \rightarrow \alpha \) (cf. Corollary 3.2). Let \( \underline{b}, \overline{b}, q > 0 \). We make the assumptions:

1. \( \text{Var}(\xi_{ij}) \geq \underline{b} \) for all \( j = 1, \ldots, p \).
2. \( \mathbb{E}[|\xi_{ij}|^{2 + \ell}] \leq \overline{b}^{\ell} \) for \( \ell = 1, 2 \) and for all \( i = 1, \ldots, n \) and \( j = 1, \ldots, p \).
3. \( \|\xi_{ij}\|_{\psi_1} \leq \overline{b} \) for all \( i = 1, \ldots, n \) and \( j = 1, \ldots, p \).
4. \( \mathbb{E}[\max_{1 \leq j \leq p} (|\xi_{ij}| / \overline{b})^q] \leq 1 \) for all \( i = 1, \ldots, n \).

Condition (A) is a non-degeneracy assumption. Condition (B) is a mild moment growth condition. Without loss of generality, we may take \( \overline{b} \geq 1 \). Conditions (C) and (D) impose sub-exponential and uniform polynomial moment requirements on the observations, respectively. Define \( \sigma_{1,n} = \left( \frac{\log^4(np)}{s} \right)^{1/6} \) and \( \sigma_{2,n} = \left( \frac{\sigma^2 \log^4(np)}{s} \right)^{1/3} \).

**Theorem 3.1** (Main result I: bounds on the Kolmogorov distance between \( T_n \) and \( T_n^* \) under \( H_0 \)) Suppose \( H_0 \) is true and assume (A) and (B) hold. Let \( \gamma \in (0, e^{-1}) \) and suppose that \( \log(\gamma^{-1}) \leq K \log(np) \) for some constant \( K > 0 \).

(i) If (C) holds, then there exists a constant \( C > 0 \) only depending on \( \overline{b}, \overline{b}, K \) such that

\[
P(\rho^*(T_n, T_n^*) \leq C \sigma_{1,n}) \geq 1 - \gamma.
\]

(ii) If (D) holds, then there exists a constant \( C > 0 \) only depending on \( \underline{b}, \overline{b}, K, q \) such that

\[
P(\rho^*(T_n, T_n^*) \leq C (\sigma_{1,n} + \sigma_{2,n})) \geq 1 - \gamma.
\]

Based on Theorem 3.1, we have the uniform size validity of the bootstrap CUSUM test.

**Corollary 3.2** (Uniform size validity of the bootstrap CUSUM test) Suppose \( H_0 \) is true and assume (A) and (B) hold. Let \( \gamma \in (0, e^{-1}) \) and suppose that \( \log(\gamma^{-1}) \leq K \log(np) \) for some constant \( K > 0 \).

(i) If (C) holds, then there exists a constant \( C > 0 \) only depending on \( \underline{b}, \overline{b}, K \) such that

\[
\sup_{\alpha \in (0, 1)} |P_0(T_n \leq q_{T_n^*} |X_1^n(\alpha)) - \alpha| \leq C \sigma_{1,n} + \gamma.
\]

Consequently, if \( \log^4(np) = o(s) \), then \( \sup_{\alpha \in (0, 1)} |P_0(T_n \leq q_{T_n^*} |X_1^n(\alpha)) - \alpha| \) uniformly in \( \alpha \in (0, 1) \) as \( n \rightarrow \infty \).

(ii) If (D) holds, then there exists a constant \( C > 0 \) only depending on \( \underline{b}, \overline{b}, K, q \) such that

\[
\sup_{\alpha \in (0, 1)} |P_0(T_n \leq q_{T_n^*} |X_1^n(\alpha)) - \alpha| \leq C (\sigma_{1,n} + \sigma_{2,n}) + \gamma.
\]
Consequently, if \( \max \{ \log^2(np), n^{2/3} \log^3 + \epsilon(np) \} = o(\frac{1}{n}) \) for some \( \epsilon > 0 \), then \( P_0(T_n \leq q_T^* | \xi_n^* (\alpha)) \to \alpha \) uniformly in \( \alpha \in (0, 1) \) as \( n \to \infty \).

Remark 4 (Choice of boundary removal parameter) There is a trade-off for the choice of boundary removal parameter: the larger \( \gamma \), the smaller of the error bounds and the more data points are removed from the change point detection (so that the regime allowed by the bootstrap CUSUM test is smaller). In theory, the lower bound of \( \gamma \) is given in Corollary 3.2 for size validity of the bootstrap CUSUM test. Specifically, if the data distribution has sub-exponential tail (i.e. Condition (C) holds), then we need \( \gamma \gg \log^2(np) \) for \( \omega_{1,n} = o(1) \); if the data distribution has polynomial tail (i.e. Condition (D) holds) with \( q > 0 \), then we need \( \gamma \gg \max \{ \log^2(np), n^{2/3} \log^3 + \epsilon(np) \} \) for \( \omega_{1,n} + \omega_{2,n} = o(1) \). This implies that we can choose \( \gamma = c_1n^{\gamma_2} \) for some constants \( c_1 > 0, 1 \geq c_2 > 0 \) in either sub-exponential or polynomial case.

As a leading example, we consider a fixed normalised true change point location \( t_m = m / n \in (0, 1) \). Then we may choose \( \gamma = c_0n \) for some small constant \( c_0 > 0 \) in order to include the true change point in the interval \( [\underline{\delta}, n - \underline{\delta}] \). For this setup, asymptotic size validity of the bootstrap CUSUM test is obtained if \( \rho = O(e^{\rho'}) \) for some \( 1/\rho > c > 0 \) under the sub-exponential moment condition on the observations.

Our second main result is to analyse the power of the bootstrap CUSUM test. We are mainly interested in characterising the change point signal strength (quantified by the location \( t_m \)) such that \( H_0 \) and \( H_1 \) can be (asymptotically) separated by our bootstrap CUSUM test. Without loss of generality, we may assume that \( |\delta_n|_{\infty} \leq 1 \).

**Theorem 3.3** (Main result II: power of the bootstrap CUSUM test under \( H_1 \)) Suppose \( H_1 \) is true with a change point \( m \in [\underline{\delta}, n - \underline{\delta}] \) and assume (A) and (B) hold. Let \( \zeta \in (0, 1/2) \) and \( \gamma \in (0, e^{-1}) \) such that \( \log(\gamma^{-1}) \leq K \log(np) \) for some constant \( K > 0 \).

(i) If (C) holds and

\[
|\delta_n|_{\infty} \geq C_1 \sqrt{\frac{\log(\zeta^{-1}) \log(np) + \log(np / n)}{n(1 - t_m)}} \tag{15}
\]

for some large enough constant \( C_1 := C_1(\underline{b}, \overline{b}, K) > 0 \), then there exists a constant \( C_2 := C_2(\underline{b}, \overline{b}, K) > 0 \) such that

\[
P_1(T_n > q_T^* | \xi_n^* (\alpha)) \geq 1 - \gamma - C_2 \omega_{1,n} - 2\zeta. \tag{16}
\]

(ii) If (D) holds and \( |\delta_n|_{\infty} \) obeys Equation (15) for some large enough constant \( C_1 := C_1(\underline{b}, \overline{b}, K, q) > 0 \), then there exists a constant \( C_2 := C_2(\underline{b}, \overline{b}, K, q) > 0 \) such that

\[
P_1(T_n > q_T^* | \xi_n^* (\alpha)) \geq 1 - \gamma - C_2 \{ \omega_{1,n} + \omega_{2,n} \} - 2\zeta. \tag{17}
\]

Remark 5 (Rate-optimality on the change point detection for sparse alternatives) For i.i.d. Gaussian errors \( \xi_j \sim N(0, I_{d_\psi}) \) in the mean-shift model (1), the change point detection boundary is characterised in Enikeeva and Harchaoui (2019). Suppose a change \( a > 0 \) occurs in the first
Remark 6 (Monotonicity of power in the signal strength) Inspecting the proof of Theorem 3.3, we see that the type II error of the bootstrap CUSUM test is bounded by a probability depending on the change point signal strength $|\delta_n|_\infty$ and location $m$ (cf. Equation (44) in the SM Section B). Specifically,

$$\text{Type II error} \leq \Pr(\hat{T}_n \geq \hat{\Delta} - q_{T^{*}_n} \mid X_1^n(1 - \alpha)),$$

where $\hat{\Delta} = \sqrt{\log(p)} / nt_m(1 - t_m) \mid \delta_n \mid_\infty$, $\hat{T}_n = \max_{s \leq n \leq \infty} |Z^s_n(s)|$, and $Z^s_n(s)$ are the CUSUM statistics computed on the $\xi_1^n$ noise random vectors. Since the distribution of $\hat{T}_n$ does not depend on $\delta_n$ and the conditional quantile $q_{T^{*}_n} \mid X_1^n(1 - \alpha)$ is bounded by $O(\sqrt{\log(np)})$ with high probability under $H_0$, the power of the bootstrap CUSUM test is lower bounded by a quantity that is non-decreasing in $|\delta_n|_\infty$. Simulation examples in Section 4 confirm our theoretical observation. In addition, since $t_m(1 - t_m)$ is maximised at $t_m = 1 / 2$, a change point near the middle is easier to detect than it is near the boundary.

3.2 Rate of convergence of the change point location estimator

Our third main result is concerned with the rate of convergence of the change point location estimator $t_{\hat{m}_\theta}$, where $\hat{m}_\theta$ is defined through Equations (9) and (8). We first consider the case of $\theta = 1 / 2$ corresponding to the covariance stationary CUSUM statistics.

Theorem 3.4 (Main result III: rate of convergence for change point location estimator: $\theta = 1 / 2$) Suppose that (B) holds and $H_1$ is true. Suppose that $\log(\gamma^{-1}) \leq K \log(np)$ for some constant $K > 0$.

(i) If (C) holds, then there exists a constant $C := C(\theta, K) > 0$ such that

$$\Pr(\mid t_{\hat{m}_{1/2}} - t_m \mid \leq C \log^2(np) / \sqrt{\log(n)} \mid \delta_n \mid_\infty) \geq 1 - \gamma. \tag{18}$$
(ii) If (D) holds with \( q > 2 \), then there exists a constant \( C : \equiv C(\delta, K, q) > 0 \) such that
\[
\Pr \left( |t_{\hat{m}_{1/2}} - t_m| \leq \frac{C n^{1/q} (\log (np) + \gamma^{-1/q})}{\sqrt{n t_m (1 - t_m)} |\delta|_\infty} \right) \geq 1 - \gamma. \tag{19}
\]

Note that the non-degeneracy Condition (A) is not needed in estimating the change point location. Consider a fixed \( t_m \in (0, 1) \) as a leading example in Remark 4. Theorem 3.4 guarantees consistency of \( t_{\hat{m}_{1/2}} \) if the signal strength satisfying: i) \( \delta \gg n^{-1/2} \log^2(np) \) in the sub-exponential moment case; ii) \( \delta \gg n^{-1/2 + 1/q} \log(np) \) in the polynomial moment case. From Part (i) of Theorem 3.4, it should also be noted that the change point location estimator \( t_{\hat{m}_{1/2}} \) does not attain the optimal rate of convergence. Consider the setup where \( t_m \in (0, 1) \), \( p = 1 \), and \( |\delta_n| = c \) is a constant signal. Then the rate of convergence in Equation (18) reads \( O(\log^2(n) / \sqrt{n}) \); that is, up to a logarithmic factor, the change point estimator has the rate of convergence \( n^{-1/2} \). In such setup, however, it is known that the best possible rate of convergence for estimating the change point location is \( n^{-1} \) (Hariz et al., 2007), which is achieved by maximising \( |Z_{0n}(s)| \). Therefore, it is interesting to study the impact of dimensionality on the rate in the case of \( \theta = 0 \) when the true change point \( t_m \in (0, 1) \) is fixed. This is the content of the next theorem. Denote \( \hat{\delta}_n = \min_{j \in S} |\delta_j| \).

**Theorem 3.5** (Main result IV: rate of convergence for change point location estimator: \( \theta = 0 \)) Suppose that (B) holds and \( H_1 \) is true with a change point \( m \) satisfying \( c_1 \leq t_m \leq c_2 \) for some constants \( c_1, c_2 \in (0, 1) \). Suppose that \( \log^3(np) \leq Kn \) and \( \log(\gamma^{-1}) \leq K \log(np) \) for some constant \( K > 0 \).

(i) If (C) holds, then there exists a constant \( C : \equiv C(\delta, K, c_1, c_2) > 0 \) such that
\[
\Pr \left( |t_{\hat{m}_0} - t_m| \leq \frac{C \log^4(np)}{n \hat{\delta}_n^2} \right) \geq 1 - \gamma. \tag{20}
\]

(ii) If (D) holds for some \( q \geq 2 \), then there exists a constant \( C : \equiv C(\delta, K, q, c_1, c_2) > 0 \) such that
\[
\Pr \left( |t_{\hat{m}_0} - t_m| \leq \frac{C \log(np)}{n \hat{\delta}_n^2 \max \left\{ 1, \frac{n^{2/q} \log(np)}{\gamma^{2/q}} \right\}} \right) \geq 1 - \gamma. \tag{21}
\]

Based on Theorem 3.5, we see that the dimension impacts the optimal rate of convergence for estimating the change point location only on the logarithmic scale. Compared with Theorem 3.4, we see that faster convergence of \( t_{\hat{m}_0} \) than that of \( t_{\hat{m}_{1/2}} \) is possible when \( t_m \in (0, 1) \) is fixed and the dimension is allowed to grow sub-exponentially fast in the sample size. Moreover, \( t_{\hat{m}_{1/2}} \) is more robust to estimate the change point when its location is near the boundary, that is, \( t_m \to 0 \) and \( t_m \to 1 \) are allowed to maintain the consistency in Theorem 3.4; see our simulation result in Section 4 for numeric comparisons.

### 3.3 Rate of convergence of BABS

Under the multiple mean-shifts model (10), we consider the testing problem for \( H_0 \) against the alternative hypothesis with multiple change points
\[
H'_1 : \delta_n^{(k)} \neq 0 \text{ for some } 1 = m_0 < m_1 < \ldots < m_v < m_{v+1} = n \text{ and } v \geq 1. \tag{22}
\]
The following Lemma 3.6 controls the power of our bootstrap CUSUM test based on $T_n^*$ in Equation (7) in presence of multiple change points. This is the initial step of the BABS (Algorithm 1) and the power control is crucial for deriving the overall rate of convergence of recursively estimating the multiple change points. Denote $\delta_n = \min_{k=1,...,v} |\delta_n^{(k)}|_{\infty}$.

**Lemma 3.6 (Power of the bootstrap CUSUM test under $H_1'$)** Suppose $H_1'$ is true under the multiple mean-shift model (10) with $\{m_k\}_{k=1}^v \subset [s, n-s]$. Assume (A), (B) and $\min_{k=0,...,v} |m_{k+1} - m_k| \geq D_v$ for some $D_v > s$. Let $\zeta \in (0, 1/2)$ and $\gamma \in (0, e^{-1})$ such that $\log(\gamma^{-1}) \leq K \log(np)$ for some constant $K > 0$.

(i) If (C) holds and

$$\max_{k=1,...,v} \frac{D_v^2 \delta_n}{\sqrt{n^2 t_m (1 - t_m)}} \geq C_1 \left[ \log(\zeta^{-1}) \log(np) + \nu^2 \log(np/\alpha) \right]$$  \hspace{1cm} (23)

for some large enough constant $C_1 := C_1(\bar{b}, b, K) > 0$, then there exists a constant $C_2 := C_2(\bar{b}, b, K) > 0$ such that Equation (16) holds.

(ii) If (D) holds and $|\delta_n|_{\infty}$ obeys (23) for some large enough constant $C_1 := C_1(\bar{b}, b, K, q) > 0$, then there exists a constant $C_2 := C_2(\bar{b}, b, K, q) > 0$ such that Equation (17) holds.

**Remark 7** (Comments on the signal strength under multiple change points alternative) The signal strength on the LHS of Equation (23) depends on the smallest mean shift $\delta_n$ in $c^{\infty}$-norm and change point locations that are closest to boundary, which is the most difficult situation for CUSUM statistic to detect mean change. If $\nu = 1$, then $D_v = \min \{m, n - m\}$ and $m(n - m) / n \leq D_v \leq 2m(n - m) / n$. Thus, the LHS of Equation (23) has the same order as $n^{1/2}(t_m(1 - t_m))^{3/2} |\delta_n|_{\infty}$ which is stronger than the requirement of lower bound $n^{1/2}(t_m(1 - t_m))^{1/2} |\delta_n|_{\infty}$ in Theorem 3.3. This extra cost comes from handling the possible mean shift cancellation in analysing the general case of multiple change points. If the single change point is bounded from boundaries (i.e. $t_m$ can be treated as a constant), then Lemma 3.6 gives the same lower bound (15) as in Theorem 3.3.

Now we turn to the bootstrap-assisted binary segmentation algorithm BABS($\alpha, b, e$). We make the following assumptions in addition to (A)-(D).

1. $\min_{k=0,...,v} |m_{k+1} - m_k| \geq D_v$, where $D_v \geq n^{\Theta}$ for some $\Theta \leq 1$.
2. $\min_{k=1,...,v} \min_{e \in D_k} |\delta_n^{(k)}| \geq -\delta_n$, where $D_k = \{1 \leq j \leq p: \delta_n^{(j)} \neq 0\}$ and $\delta_n \geq n^{-\omega}$ for some $\omega \geq 0$.
3. $\Theta - \frac{\omega}{2} > \frac{3}{4}$.
4. $n^{3/2 \Theta - \omega} > C \max \{\log^2(np), \sqrt{\log(\zeta^{-1}) \log(np) + \nu^2 \log(np/\alpha)}\}$, where $\zeta, K$ are constants defined in Lemma 3.6. Here, $C > 0$ is a constant depending only on $\bar{b}, b, K$ under (C) and on $\bar{b}, b, K, q$ under (D).
5. $e_n < s < D_v$, where $e_n$ is defined in Equation (24) below.

Assumptions (a)-(c) are standard signal separation and strength requirements in estimating the multiple change point locations via binary segmentation, see for example, Theorem 1 in Fryzlewicz (2014). Assumption (d) ensures that the bootstrap CUSUM test is able to consistently detect the
mean-shift signals (cf. Lemma 3.6). Under Assumption d), the expected signal size \( |\mathbb{E}Z_{n,b,c}(s)|_\infty \) dominates the random behaviour of \( Z_{n,b,c}(s) \), thus, obeying Equation (23) with large probability. Assumption (e) is a minimal condition on the boundary removal parameter \( \delta \), which is smaller than the separation distance between any consecutive change points and larger than the rate of convergence \( e_n \) for consistently estimating all change point locations. Note that the signal strength requirement in estimation depends on \( \min_{j \in \mathcal{D}_k} |\delta_{nj}^{(k)}| \) in assumption (b), which is typically stronger than \( \max_{1 \leq j \leq p} |\delta_{nj}^{(k)}| \) used in the testing problem.

**Theorem 3.7**  (Main result V: rate of convergence of BABS) Let \( \hat{\nu} \) denote the number of change points and \( \hat{m}_1 < \ldots < \hat{m}_\nu \) the change point locations estimated from BABS(\( \alpha, 1, n \)). Assume (A), (B) and (a)–(e) hold. Let \( \gamma \in (0, e^{-1}) \) such that \( \log(\gamma^{-1}) \leq \text{Klog}(D, p) \leq \text{Klog}(np) \) and \( \zeta \) is defined as in Theorem 3.3. Define

\[
\epsilon_n = \begin{cases} 
\frac{n^2\log^4(np)}{D^2 \delta_n^2}, & \text{if (C) holds} \\
\frac{n^2(\log^2(np) + \gamma^{-2/4})}{D^2 \delta_n^2}, & \text{if (D) holds}
\end{cases}
\]

(i) If (C) holds, then there exist constants \( C_0 = C_0(\bar{b}, \underline{b}, K), C_0' = C_0'(\alpha, \bar{b}, K) \) such that

\[
\mathbb{P}(S_n) \geq 1 - 2\gamma - \nu(\gamma + 2\zeta + C_0 e_{1,n}) - (\nu + 1)\alpha,
\]

where \( S_n = \{ \hat{\nu} = \nu \text{ and } \max_{k=1,...,\nu} |\hat{m}_k - m_k| \leq C_0' \epsilon_n \} \).

(ii) If (D) holds, then there exist constants \( C_0 = C_0(\bar{b}, \underline{b}, K, q), C_0' = C_0'(\alpha, \bar{b}, K, q) \) such that

\[
\mathbb{P}(S_n) \geq 1 - 2\gamma - \nu(\gamma + 2\zeta + C_0(\sigma_{1,n} + \sigma_{2,n})) - (\nu + 1)\alpha.
\]

Theorem 3.7 reveals an interesting size-power trade-off of the BABS algorithm in multiple change point detection and estimation. For smaller \( \alpha \), stronger signal strength and larger change point separation are needed to fulfill Assumption d) required by Theorem 3.7. For larger \( \alpha \), the bootstrap CUSUM test used in BABS tends to reject more \( H_0 \) and consequently the BABS is likely to over-estimate the number of change points under \( H_1' \). This is reflected by the multiple testing term \( (\nu + 1)\alpha \) in the lower bound of \( \mathbb{P}(S_n) \) in Theorem 3.7. Moreover, under \( H_1' \), both type I error (quantified by \( \alpha \)) and type II error (quantified by \( \zeta \)), together with their Bonferroni type multiple testing adjustment (quantified by \( \nu \)), affect consistency of the BABS algorithm.

Binary segmentation was also considered in Cho and Fryzlewicz (2015) and Cho (2016), both of which are consistent under their own conditions with different rate of convergence \( e_{n,p} \). Our rate of \( e_{n} \) is similar to that in Cho and Fryzlewicz (2015) up to a logarithmic factor when \( \Theta \in (3/4, 1] \) and it is sharper than that in Cho (2016) for sparse alternative. Comparisons on explicit rates are given in Remark 11 in the SM.

4 | SIMULATION STUDIES

In this section, we perform extensive simulation studies to investigate the size and power of the proposed bootstrap change point test, the estimation error of the change point location(s), as well as empirical performance of BABS. In all setups, 200 bootstrap samples (if necessary) are drawn for each simulation.
4.1 Setup

We generate i.i.d. \( \xi_i \) in the mean-shift model (1) from three distributions.

1. Multivariate Gaussian distribution: \( \xi_i \sim N(0, V) \).
2. Multivariate elliptical \( t \)-distribution with degree of freedom \( \nu \): \( \xi_i \sim t_\nu(V) \) with the probability density function (Muirhead, 1982, Chapter 1)

\[
\begin{align*}
  f(x; \nu, V) &= \frac{\Gamma((\nu + p)/2)}{\Gamma(\nu/2)(v\pi)^{p/2} \det(V)^{1/2}} \left( 1 + \frac{x^\top V^{-1}x}{\nu} \right)^{-(\nu+p)/2}.
\end{align*}
\]

The covariance matrix of \( \xi_i \) is \( \Sigma = \nu / (\nu - 2)V \). In our simulation, we use \( \nu = 6 \).
3. Contaminated Gaussian: \( \xi_i \sim \text{ctm-Gaussian}(\epsilon, \nu, V) \) with density

\[
\begin{align*}
  f(x; \epsilon, \nu, V) &= (1 - \epsilon) \frac{\exp\left(-\frac{x^\top V^{-1}x}{2}\right)}{(2\pi)^{p/2} \det(V)^{1/2}} + \epsilon \frac{\exp\left(-\frac{x^\top V^{-1}x}{2\nu^2}\right)}{(2\pi\nu^2)^{p/2} \det(V)^{1/2}}.
\end{align*}
\]

The covariance matrix of \( \xi_i \) is \( \Sigma = \{ (1 - \epsilon) + \epsilon\nu^2 \} V \). We will fix \( \epsilon = 0.2 \) and \( \nu = 2 \).

We consider three cross-sectional dependence structures of \( V \) for each distribution.

1. Independent: \( V = \text{Id}_p \), where \( \text{Id}_p \) is the \( p \times p \) identity matrix.
2. Strongly dependent (compound symmetry): \( V = 0.8J + 0.2\text{Id}_p \) where \( J \) is the \( p \times p \) matrix containing all ones.
3. Moderately dependent (autoregressive): \( V_{ij} = 0.8^{i-j} \).

4.2 Simulation results for single change point model

4.2.1 Size of the bootstrap CUSUM test

We fix the sample size \( n = 500 \) and vary the dimension \( p = 10, 300, 600 \). For the bootstrap CUSUM test, we set the boundary removal parameter \( s = 30, 40 \). For a significance level \( \alpha \in (0, 1) \), we denote \( \hat{R}(\alpha) \) as the proportion of empirically rejected null hypothesis in 1000 simulations.

Under \( H_0 \), the upper half of Table 1 reports the uniform error-in-size \( \sup_{\alpha \in (0,1)} | \hat{R}(\alpha) - \alpha | \), a quantity that reflects the Kolmogorov distance between distributions of \( T_n \) and its bootstrap analog \( T^*_n \), the smaller uniform error-in-size, the closer \( \rho^*(T_n, T^*_n) \). Each column corresponds to a combination of noise distributions and cross-sectional dependence structures. The lower half of Table 1 shows the empirical type I error \( \hat{R}(\alpha) \) at the significance level \( \alpha = 0.05 \). We can draw several conclusions for our bootstrap CUSUM test by comparing results under different choices of boundary removal parameter, distribution family, and cross-sectional dependence structure. First, in most cases the uniform error-in-size of \( \xi = 40 \) are smaller than those of \( \xi = 30 \), meaning that the greater the \( \xi \), the better the approximation under \( H_0 \). Moreover, \( \hat{R}(0.05) \) is generally close to the nominal size 0.05 for \( \xi = 40 \). Next, uniform errors-in-size is usually smaller for the Gaussian distribution than that of \( t_6 \) or ctm-Gaussian cases. Lower Table 1 delivers a similar message that Gaussianity helps to control \( \hat{R}(0.05) \). Finally, our method is robust to the cross-sectional dependence structure. In many cases, stronger dependence
(II > III > I) is more beneficial for reducing the approximation errors. In summary, size can be better controlled if $g$ is large, data are Gaussian distributed, and strong cross-sectional dependence exists. As a visualisation of the accuracy for size control, Figure 1 displays three example curves of $\hat{R}(\alpha)$ for our proposed test where $p = 600$, $g = 40$. The rejection rate $\hat{R}(\alpha)$ follows closely along the diagonal line in dash (i.e. the line of $\hat{R}(\alpha) = \alpha$).

A thorough comparison between our bootstrap CUSUM test and two benchmark methods (i.e. bootstrap log-ratio of maximised likelihood test and the oracle test with known covariance matrix) can be found in the SM Section D.1. In SM Section D.1, we also compare our method with the test statistics in Jirak (2015) (denoted as $B_n$) and Enikeeva and Harchaoui (2019) (denoted as $\psi$) under the setting $n = 500$, $p = 600$, $g = 40$.

**Table 1** Uniform error-in-size, $\sup_{\alpha \in [0,1]} \left| \hat{R}(\alpha) - \alpha \right|$, and empirical type I error with nominal level 0.05, $\hat{R}(0.05)$, for our bootstrap CUSUM test under $H_0$, where $p = 10, 300, 600$, $g = 30, 40$ and data are simulated from all combinations of distribution families and covariance dependence structures.

|          | Gaussian | $t_6$       | ctm-Gaussian |
|----------|----------|-------------|--------------|
|          | I        | II          | III          | I            | II          | III          | I            | II          | III          |
| $p = 10$ | $g = 30$ | 0.034       | 0.036        | 0.041        | 0.048        | 0.041        | 0.039        | 0.036        | 0.042        | 0.021        |
|          | $g = 40$ | 0.042       | 0.034        | 0.037        | 0.043        | 0.037        | 0.033        | 0.041        | 0.042        | 0.043        |
| $p = 300$| $g = 30$ | 0.054       | 0.051        | 0.050        | 0.085        | 0.036        | 0.049        | 0.115        | 0.025        | 0.065        |
|          | $g = 40$ | 0.046       | 0.026        | 0.035        | 0.058        | 0.030        | 0.040        | 0.057        | 0.032        | 0.055        |
| $p = 600$| $g = 30$ | 0.051       | 0.035        | 0.048        | 0.122        | 0.044        | 0.088        | 0.103        | 0.030        | 0.096        |
|          | $g = 40$ | 0.060       | 0.055        | 0.046        | 0.083        | 0.038        | 0.087        | 0.079        | 0.026        | 0.057        |
| $\hat{R}(0.05)$ | | | | | | | | | |
| $p = 10$ | $g = 30$ | 0.051       | 0.052        | 0.051        | 0.044        | 0.056        | 0.034        | 0.043        | 0.056        | 0.052        |
|          | $g = 40$ | 0.046       | 0.055        | 0.052        | 0.045        | 0.050        | 0.048        | 0.054        | 0.053        | 0.040        |
| $p = 300$| $g = 30$ | 0.026       | 0.054        | 0.039        | 0.018        | 0.034        | 0.018        | 0.021        | 0.043        | 0.027        |
|          | $g = 40$ | 0.045       | 0.043        | 0.044        | 0.024        | 0.046        | 0.036        | 0.026        | 0.056        | 0.034        |
| $p = 600$| $g = 30$ | 0.026       | 0.060        | 0.027        | 0.010        | 0.034        | 0.020        | 0.010        | 0.053        | 0.019        |
|          | $g = 40$ | 0.031       | 0.038        | 0.036        | 0.020        | 0.044        | 0.016        | 0.015        | 0.042        | 0.027        |

**Figure 1** Empirical rejection rate, $\hat{R}(\alpha)$, in selected data generating schemes under $H_0$. (Left) Gaussian distribution with Covariance I; (Middle) $t_6$ distribution with Covariance II; (Right) ctm-Gaussian distribution with Covariance III. Parameters: $n = 500$, $p = 600$, $g = 40$.
4.2.2 | Power of the bootstrap CUSUM test

Under $H_1 : \mu_1 = \cdots = \mu_m \neq \mu_{m+1} = \cdots = \mu_p$, we consider the single change point location $m$ at 
{50, 150, 250} (i.e. $t_m = m/n = 1/10, 3/10, 5/10$ for $n = 500$). Denote $k$ as the number of components that have change points, that is, $\delta_{n,1} = \cdots = \delta_{n,k} \neq 0$. Two types of the mean-shift signal are considered: $k=1$ for sparse signal and $k=50$ for dense signal. Due to the space limit, we only present the sparse alternative case of $k=1$ in this section, and results of dense signal for $\psi$ can be found in the SM Section D.4. To analyse the power under $H_1$, we fix $n = 500, p = 600, \xi = 40$ and the significance level $\alpha = 0.05$.

We first investigate the impact of change point location and distribution to our test. Figure 2(a) shows the empirical power curves v.s. the signal strength $|\delta_n|_\infty = |\delta_{n,1}|$. In all cases, the powers monotonically increase and eventually reach 1 as $|\delta_n|_\infty$ gets large enough. Comparing the three curves corresponding to ctm-Gaussian distribution (I) in Figure 2(a), we observe that change points closer to boundaries are harder to detect at the same signal strength. When we narrow down to the four curves corresponding to $t_m = 1/10$, we can see the distributional influence: Gaussianity and cross-sectional dependence is helpful to improve the power. More simulation results of powers can be found in Table 7 in the SM Section D.4.

Next, we compare our method with $B_n$ of Jirak (2015) and $\psi$ of Enikeeva and Harchaoui (2019). Figure 2(b) shows power trends of $B_n$ under $t$-distributed data for $k = 1$ when $t_m = 1/2, 1/10$. The test $B_n$ performs better for central change point ($t_m = 1/2$) than boundary change point ($t_m = 1/10$). Compared to Figure 2(a), we see that boundary change point brings more challenge to $B_n$ than to our test because our powers increase faster than $B_n$ under the same setup. Table 8 in the SM Section D.4 gives a more detailed power report of $B_n$ in all scenarios under sparse $H_1$. We note that although $B_n$ returns slightly higher power than ours at $t_m = 1/2$, it is computed with true long-run variance and it tends to over reject $H_0$ (i.e. size distortion). Figure 2(c) displays power trends of the $\psi$ ($\xi = 1$, no boundary removal) and $\psi$-improved (boundary removal with $\xi = 40$) for Gaussian distributed data at $t_m = 1/2$. Neither $\psi$ test has valid power curve when the independent covariance assumption is violated. Furthermore, unreported results (due to the space concern) show invalid power curves in all other non-Gaussian distributed data. It is unsurprising since $\psi$ suffers from serious size distortion (cf. Table 6 in the SM Section D.4). We refer to Table 9 in the SM Section D.4 for complete power reports of $\psi$ in Gaussian scenarios with independent components under both sparse and dense $H_1$.

![Figure 2](image-url)  
**Figure 2** Power curves under sparse alternative: (a) $T_n^*$ (our) under selected distributions, covariances and $t_m = 1/10, 3/10, 5/10$; (b) $B_n$ (Jirak, 2015) under $t_k$ distribution, three covariances and $t_m = 1/2, 1/10$; (c) original and improved $\psi$ (i.e. $\xi = 1$ and $\xi = 40$) (Enikeeva & Harchaoui, 2019) under Gaussian distribution, three covariances and $t_m = 1/2$. Parameters: $n = 500, p = 600$ [Colour figure can be viewed at wileyonlinelibrary.com]
Now we examine the performance of our location estimators under sparse alternative where \( t_m = 1/10, 3/10, 5/10 \). The performance measure is the root-mean-square error (RMSE). Figure 3 shows comparison between two estimators across different change point locations and signal sizes \( |\delta_n|_{\infty} \). First, Figure 3(a) illustrates that boundary change points (such as \( t_m = 1/10 \)) are harder to estimate as the RMSEs are uniformly larger than that for \( t_m = 5/10 \). Second, as implied by Theorems 3.4 and 3.5, RMSEs of the estimator with \( t_m = 1/2 \) are smaller than that of the estimator with \( t_m = 1/10 \) because \( Z_{\theta,n}(s) \) assigns less weights to the boundary points for smaller values of \( \theta \). This is also empirically confirmed by Figure 3(b): \( t_m = 1/2 \) is more in favour of boundary points when \( |\delta_n|_{\infty} = 0 \), and \( t_m = 5/10 \) slightly leans to the center when change (\( |\delta_n|_{\infty} = 0.842 \)) is not in the middle of sequence.

Next, we compare our estimators with Wang and Samworth (2018) and Cho and Fryzlewicz (2015). In Wang and Samworth (2018), a projection based estimator \( \text{Inspect} \) is proposed. Theoretical analysis of this algorithm requires the data follow Gaussian distribution. In Cho and Fryzlewicz (2015), the proposed SBS estimator is the maximiser of threshold \( \ell^{-1} \)-aggregated CUSUM statistics after thresholding. This method is sensitive to threshold tuning parameters selected by bootstrap. Both approaches...
allow multiple change points. For now, we first compare with their single change point versions (see R packages \textit{InspectChangepoint} and \textit{hdbinseg}). We also include a truncated version of our location estimator \( \hat{\theta}_y = \arg\max_{\frac{s}{n} \leq s < \frac{n - 1}{n}} |Z_{\theta_n}(s)|_\infty \) (cf. Remark 3) for fair comparison. Both \( k = 1 \), 50 are considered, where \( k \) represents signal density (\( n^{-1} = \ldots = n^{-1} \)).

Figure 4(a) compares our non-truncated \( \hat{\theta}_y \) with \textit{Inspect} that has no boundary removal, where \( t_m = 3 / 10 \) and data are \( t_6 \) distributed with Covariance II. The RMSEs of our estimator is uniformly smaller than \textit{Inspect} if \( k = 1 \) or data do not have an isotropic Gaussian distribution. Figure 4(b) shows non-monotone RMSEs of \textit{SBS} returned from \textit{hdbinseg} when \( s = 40, t_m = 5 / 10, k = 1 \) and data are from ctm-Gaussian with Covariance II. We also note that the empirical detection rate of change point drops from 91% to 72% when \( |\delta_n|_\infty \) grows from 0.84 to 2. Although the \textit{SBS} works well for \( t_m = 5 / 10, k = 50 \) or \( t_m = 1 / 10, k = 1 \) (cf. SM Section D.2), it means that large CUSUM values under sparse \( H_1 \) may lead to unreasonable selection of their threshold. The full RMSEs of our method with sparse signal and selected RMSEs of Wang and Samworth (2018) and Cho and Fryzlewicz (2015) are reported in Tables 10, 11 and 12 in the SM Section D.2.

### 4.3 | Multiple change points estimation using BABS

In the multiple change-point scenario, we first consider the \( k \)-th component of \( \delta_n^{(k)} \) to have the same mean shift, that is, \( \delta_{n,1}^{(1)} = \delta_{n,2}^{(2)} = \ldots = \delta_{n,v}^{(v)} = \delta \neq 0 \). Since change point estimation can be viewed as a special case of clustering, the accuracy (consistency) can be measured by the Adjusted Rand Index (ARI) (Hubert & Arabie, 1985; Rand, 1971). We also report the average ARI over 500 simulations. The bootstrap resampling is \( B = 200 \).

First, we provide simulation results for BABS using \( t_6 \) distribution with Covariance III as an illustrative example. We consider \( n = 1000, p = 1200, \xi = 40, \alpha = 0.05 \) and the two change points \( (m_1, m_2) = (300, 600) \). From Table 2 (left), we see that when signal is small (e.g. \( \delta = 0.317 \)), BABS cannot locate mean shift accurately. As signal gets larger, both the number and the locations of change points can be estimated more consistently, and ARI is also increasing to 1 (i.e. perfect estimation). We further add one more change point, \( (m_1, m_2, m_3) = (300, 600, 800) \). Table 2 (right) shows that the

| \( (m_1, m_2) = (300, 600) \) | \( (m_1, m_2, m_3) = (300, 600, 800) \) |
|---|---|
| \( \delta \) | 0 | 0.317 | 0.733 | 1.282 | 2.004 | 0 | 0.317 | 0.733 | 1.282 | 2.004 |
| Estimated number of change points | 0 | 497 | 378 | 1 | 0 | 0 | 491 | 360 | 0 | 0 | 0 |
| | 1 | 3 | 117 | 13 | 0 | 0 | 9 | 133 | 5 | 0 | 0 |
| | 2 | 0 | 5 | 458 | 464 | 470 | 0 | 7 | 141 | 0 | 0 |
| | 3 | 0 | 0 | 26 | 35 | 30 | 0 | 0 | 328 | 455 | 474 |
| | 4 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 25 | 40 | 25 |
| | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 5 | 1 |
| Sum | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 |
| ARI | 0.994 | 0.128 | 0.935 | 0.978 | 0.989 | 0.982 | 0.106 | 0.871 | 0.973 | 0.989 |
estimation is slightly worse under the same signal size in the left. This is because the effective sample size cuts down after each binary segmentation and one more multiple testing adjustment is needed in the stopping criterion of BABS. Nonetheless, our algorithm eventually detects all change points consistently when signal is large enough (e.g. $\delta = 2$). Figures 11 and 12 in the SM Section D.4 visualize this observation for the two cases, respectively.

Next, we experiment the setup in Section 5.3 of Wang and Samworth (2018) to compare BABS with Inspect. Consider $n = 2000$, $p = 200$, $(m_1, m_2, m_3) = (500, 1000, 1500)$ and $(|\delta_n^{(1)}|_2, |\delta_n^{(2)}|_2, |\delta_n^{(3)}|_2) = (\delta, 2\delta, 3\delta)$ for signal strength $\delta \in \{0.4, 0.6\}$. Table 3 below and Figure 13 in the SM Section D.4 summarise estimation performance of our BABS and the Inspect. If $k = 40$, then our method to detect. However, when $k = 1$ such that $\delta_n^{(i)}$, $i = 1, 2, 3$ is sparse (with the $\ell^2$-norm unchanged), then our algorithm shows superiority in terms of both $\widehat{V}$ and ARI. Again, our BABS algorithm has advantage when the $\ell^\infty$-norm of signal is bounded below and likewise when data are non-Gaussian or cross-sectionally dependent as shown in Sections 4.2.2 and 4.2.3. For comparison with Cho and Fryzlewicz (2015), Cho (2016), Jirak (2015), we refer to Section 5.3 of Wang and Samworth (2018) for a comprehensive simulation study.

5 | REAL DATA APPLICATIONS

5.1 | Array CGH data

The microarray dataset aCGH from the ecp R package in James and Matteson (2015) consists of $p = 43$ individuals with bladder tumours. There are $n = 2215$ log-intensity-ratio fluorescent measurements of DNA segments that share almost identical change points because the individuals have the same disease. We set $B = 1000$, $\alpha = 0.05$, $s = 60$. Our BABS finds 27 change points in the copy-number that are marked as red vertical dashed lines in Figure 5. Compared to Inspect, which identifies
FIGURE 5  Change point estimation in the log-intensity-ratio fluorescent measurements of aCGH data (the first 10 patients displayed) estimated by BABS using whole aCGH data. Parameters: $n = 2215$, $p = 43$, $B = 1000$, $\alpha = 0.05$, $\delta = 60$ [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4  Quantiles of bootstrapped CUSUM test statistic in the stock return data on log-scale for different block sizes

| $q_{0.90}$ | $q_{0.95}$ | $q_{0.99}$ |
|---------|---------|---------|
| M = 1   | M = 2   | M = 5   | M = 10  |
| 2.350   | 3.380   | 5.219   | 6.981   |
| 2.470   | 3.860   | 5.464   | 7.327   |
| 2.782   | 4.580   | 5.746   | 8.861   |

FIGURE 6  Change points detected by extension of BABS using the block CUSUM bootstrap test and two different location estimators $\hat{m}_{1/2}$ and $\hat{m}_0$ in the stock return data on log-scale. Parameters: $\delta = 25$, $B = 400$, $M = 5$ and $\alpha = 0.05$ [Colour figure can be viewed at wileyonlinelibrary.com]
254 change points using the default threshold, our discovery is more reasonable and stable under the existence of outliers (e.g. the segment (1724, 1836) or (1965, 2044)).

5.2 Stock return data

We run our block bootstrap CUSUM test and location estimators to stock return data that is available on https://finance.yahoo.com. The dataset (read through the R package BatchGetSymbols) contains daily closing prices of \( p = 440 \) stocks from the S&P500 index during the trading days between August 27, 2007 to August 24, 2009 (\( n = 503 \) time points). The daily closing prices are transformed to log scales due to their multiplicative nature. For this dataset, the CUSUM test statistic \( T_n = 38.699 \).

We set \( \xi = \lfloor 0.05n \rfloor = 25 \), bootstrap repeats \( B = 200 \) and the block sizes \( M = 1, 2, 5, 10 \) for our block bootstrap calibration. Table 4 shows the (conditional) quantiles of our block bootstrap CUSUM test statistic. Compared with the critical value 8.861 corresponding to the 99%-quantile for \( M = 10 \), we reject \( H_0 \). In addition, \( \hat{m}_{1/2} = 265 \) corresponds to September 12, 2008, the last trading day before Lehman Brothers Holdings Inc. declared bankruptcy on September 15, 2008. Figure 14 in the SM Section D.4 plots the top 5 stocks in this financial crash.

A binary segmentation extension based on the block bootstrap CUSUM test is considered as well. We implement this extension with \( \hat{m}_{1/2} \) and \( \hat{m}_0 \) separately, whose estimations are shown in Figure 6 on top and bottom, respectively. We set \( \xi = 25, B = 400, M = 5 \) and \( \alpha = 0.05 \) for both scenarios. Overall, the two estimators share common time points on detection when mean-shift signals are large enough. There are seven overlapping change points identified by both algorithms, and the one using \( \hat{m}_{1/2} \) additionally locates June 12 and July 21 in 2008 but misses October 17, 2008. That is, in the interval between April 21 and September 12 in 2008, \( \hat{m}_{1/2} \) is sensitive to change points on the boundary points (i.e. June 12 and July 21, 2008). However, \( \hat{m}_0 \) is sensitive to the middle change point, namely October 17, 2008, in the interval between September 12 and November 26 in 2008. This exactly reflects our observation in the simulation.

We would like to make two comments for this example. First, the estimator \( \hat{m}_{1/2} \) is compatible with our test statistic \( T_n \) using stationary weight \( \theta = 1/2 \), while \( \hat{m}_0 \) is also a reasonable choice since stock prices are likely to be integrated. Therefore, we evaluate both of them in this stock prices dataset. Second, neither of the two algorithms identifies change point between January 13, 2009 and June 11, 2009, even though there seems to be fluctuations in the mean return. One possible reason is that there exists non-synchronous change points (e.g. around time index at 380), which are not estimable. However, this is a common issue in multiple change-point analysis and it is necessary to make some minimal separation or spacing assumptions, cf. Fryzlewicz (2014), Cho and Fryzlewicz (2015), Cho (2016), Barigozzi et al. (2018).

ACKNOWLEDGEMENTS

The authors thank two anonymous referees, an Associate Editor and the Co-Editor Simon Wood for their careful and constructive comments that greatly improved the quality of this paper. This work was completed in part with the high-performance computing resource provided by the Illinois Campus Cluster Program at UIUC. X. Chen’s research was supported in part by NSF DMS-1404891, NSF CAREER Award DMS-1752614, UIUC Research Board Awards (RB17092, RB18099), and a Simons Fellowship. X. Chen acknowledges that part of this work was carried out at the MIT Institute for Data, System, and Society (IDSS).

REFERENCES

Adamczak, R. (2008) A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. Electronic Journal of Probability, 13(34), 1000–1034.
Aue, A., Hörmann, S., Horváth, L. & Reimherr, M. (2009) Break detection in the covariance structure of multivariate time series models. *Annals of Statistics*, 37(6B), 4046–4087.

Bai, J. (1997) Estimation of a change point in multiple regression models. *The Review of Economics and Statistics*, 79(4), 551–563.

Barigozzi, M., Cho, H. & Fryzlewicz, P. (2018) Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics*, 206(1), 187–225.

Berkes, I., Gabrys, R., Horváth, L. & Kokoszka, P. (2009) Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B*, 71(5), 475–507.

Bhattacharya, P.K. (1987) Maximum likelihood estimation of a change-point in the distribution of independent random variables: General multiparameter case. *Journal of Multivariate Analysis*, 23(2), 183–208.

Bickel, P.J. & Levina, E. (2008a) Covariance regularization by thresholding. *Annals of Statistics*, 36(6), 2577–2604.

Bickel, P.J. & Levina, E. (2008b) Regularized estimation of large covariance matrices. *Annals of Statistics*, 36(1), 199–227.

Brodsky, B.E. & Darkhovsky, B.S. (1993) *Nonparametric methods in change point problems*. Netherlands: Springer.

Cai, T.T. & Zhou, H.H. (2012) Optimal rates of convergence for sparse covariance matrix estimation. *Annals of Statistics*, 40(5), 2389–2420.

Cai, T.T., Zhang, C.-H. & Zhou, H.H. (2010) Optimal rates of convergence for covariance matrix estimation. *Annals of Statistics*, 38(4), 2118–2144.

Carlstein, E. (1988) Nonparametric change-point estimation. *Annals of Statistics*, 14(1), 188–197.

Chen, X., Xu, M. & Wu, W.B. (2013) Covariance and precision matrix estimation for high-dimensional time series. *Annals of Statistics*, 41(6), 2994–3021.

Chernoff, H. & Zacks, S. (1964) Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics*, 35(3), 999–1018.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2013) Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics*, 41(6), 2786–2819.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2017) Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, 45(4), 2309–2352.

Cho, H. (2016) Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics*, 10(2), 2000–2038.

Cho, H. & Fryzlewicz, P. (2015) Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B*, 77(2), 475–507.

Csörgő, M. & Horváth, L. (1997) *Limit theorems in change-point analysis*. New York: Wiley.

Enikeeva, F. & Harchaoui, Z. (2019) High-dimensional change-point detection under sparse alternatives. *Annals of Statistics*, 47(4), 2051–2079.

Frick, K., Munk, A. & Sieling, H. (2014) Multiscale change point inference. *Journal of the Royal Statistical Society: Series B*, 76(3), 495–580.

Fryzlewicz, P. (2014) Wild binary segmentation for multiple change-point detection. *Annals of Statistics*, 42(6), 2243–2281.

Garreau, D. & Arlot, S. (2018) Consistent change-point detection with kernels. *Electronic Journal of Statistics*, 12(2), 4440–4486.

Giné, E. & Nickl, R. (2015) *Mathematical foundations of infinite-dimensional statistical models*. Cambridge: Cambridge University Press.

Hall, P. (1991) On convergence rates of suprema. *Probability Theory and Related Fields*, 89, 447–455.

Harchaoui, Z. & Lévy-Leduc, C. (2010) Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, 105(492), 1480–1493.

Hariz, S.B., Wylie, J.J. & Zhang, Q. (2007) Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences. *Annals of Statistics*, 35(4), 1802–1826.

Harlé, F., Chatelain, F., Gouy-Pailler, C. & Achard, S. (2016) Bayesian model for multiple change-points detection in multivariate time series. *IEEE Transactions Signal Processing*, 64(16), 4351–4362.

Hinkley, D.V. (1970) Inference about the change-point in a sequence of random variables. *Biometrika*, 57(1), 1–17.

Hubert, L. & Arabie, P. (1985) Comparing partitions. *Journal of Classification*, 2, 193–218.

James, N.A. & Matteson, D.S. (2015) ecp: An R package for nonparametric multiple change point analysis of multivariate data. *Journal of Statistical Software*, 62(7), 1–25.
Jirak, M. (2015) Uniform change point tests in high dimension. *Annals of Statistics*, 43(6), 2451–2483.

Kokoszka, P. & Leipus, R. (2000) Change-point estimation in ARCH models. *Bernoulli*, 6(3), 513–539.

Lai, T.L. (2001) Sequential analysis: Some classical problems and new challenges. *Statistica Sinica*, 11(2), 303–351.

Leadbetter, M.R., Lindgren, G. & Rootzén, H. (1983) *Extremes and related properties of random sequences and processes*. New York: Springer-Verlag.

Ledoux, M. & Talagrand, M. (1991) *Probability in Banach spaces: Isoperimetry and processes*. New York: Springer-Verlag.

Loader, C.R. (1996) Change point estimation using nonparametric regression. *Annals of Statistics*, 24(4), 1667–1678.

Lorden, G. (1971) Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 42(6), 1897–1908.

Muirhead, R.J. (1982) *Aspects of multivariate statistical theory*. New Jersey: John Wiley & Sons, Inc.

Ombao, H., von Sachs, R. & Guo, W. (2005) SLEX analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, 100(470), 519–531.

Page, E.S. (1955) A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3), 523–527.

Qiu, P. (2014) *Introduction to statistical process control*. Boca Raton: CRC Press.

Rand, W.M. (1971) Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336), 846–850.

Resnick, S.I. (1987) *Extreme values, regular variation and point processes*. New York: Springer-Verlag.

Wald, A. (1945) Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, 16(2), 117–186.

Wald, A. & Wolfowitz, J. (1948) Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 19(3), 326–339.

Wang, T. & Samworth, R.J. (2018) High dimensional change point estimation via sparse projection. *Journal of Royal Statistical Society: Series B*, 80(1), 57–83.

Wang, D., Yu, Y. & Rinaldo, A. (2017) Optimal covariance change point localization in high dimension. *Preprint arXiv: 1712.09912*.

Woodall, W. & Ncube, M. (1985) Multivariate CUSUM quality-control procedures. *Technometrics*, 27(3), 285–292.

Yao, Y.-C. (1987) Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Annals of Statistics*, 15(3), 1321–1328.

Zhang, N.R., Siegmund, D.O., Ji, H. & Li, J.Z. (2010) Detecting simultaneous changepoints in multiple sequences. *Biometrika*, 97(3), 631–645.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

---

**How to cite this article:** Yu M, Chen X. Finite sample change point inference and identification for high-dimensional mean vectors. *J R Stat Soc Series B*. 2021;83:247–270. [https://doi.org/10.1111/rssb.12406](https://doi.org/10.1111/rssb.12406)