

High-dimensional change point detection for vector moving average model

Ruiqi Chen¹, Xiangchao Li¹, and Qing Yang² ✉

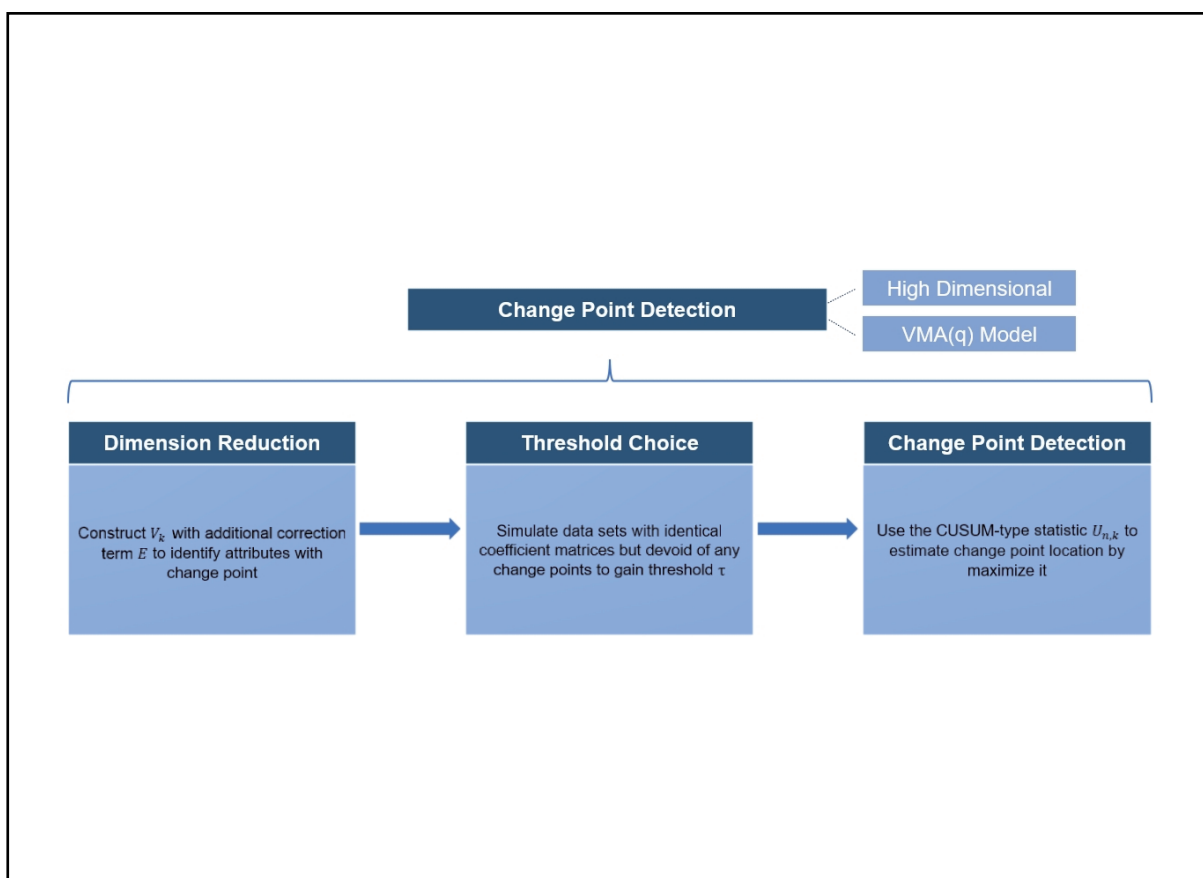
¹School of Management, University of Science and Technology of China, Hefei 230026, China;

²International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Qing Yang, E-mail: yangq@ustc.edu.cn

© 2026 The Author(s). This is an open access article under the CC BY-NC-ND 4.0 license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Graphical abstract



Research framework.

Public summary

- The q th-order vector moving average model is introduced as a dependent data model for change point detection.
- The dimension reduction statistics have been optimized to achieve better change point detection performance.
- Simulations, numerical comparisons and real data analysis about comparative genomic hybridization, and processed inventory investment are presented to illustrate the performance of the proposed method.

High-dimensional change point detection for vector moving average model

Ruiqi Chen¹, Xiangchao Li¹, and Qing Yang² 

¹School of Management, University of Science and Technology of China, Hefei 230026, China;

²International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230026, China

 Correspondence: Qing Yang, E-mail: yangq@ustc.edu.cn

© 2026 The Author(s). This is an open access article under the CC BY-NC-ND 4.0 license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Cite This: JUSTC, 2026, 56(X): (13pp)



Read Online



Supporting Information

Abstract: This paper investigates single change point detection for the q th-order vector moving average (VMA) model. We perform dimension reduction on the raw data by calculating the weighted sum of the squared differences between the mean vector estimators for each attribute. An additional rectification is further introduced to ensure the statistical properties of the expectation. The change point position is subsequently estimated via the reduced data. Specifically, we construct a CUSUM-type statistic that discards high-order moment terms. The feasibility of the dimension reduction step and the performance of our change point detection approach are validated through various simulated data experiments and real-data analysis.

Keywords: change point detection; dimension reduction; vector moving average model

CLC number: O212.4

Document code: A

2020 Mathematics Subject Classification: 62M10

1 Introduction

The topic of change point analysis is a prevalent research area in contemporary statistical studies. It was initially introduced in the study of quality control^[1,2] and has thereafter been proven useful in bioinformatics^[3], econometrics and finance^[4–6], climate and natural environment science^[7], computer science^[8], clinical neuroscience^[9], and many other fields.

Many researchers have studied the problem of change point detection from different perspectives, including methods for constructing tests for the existence of change point hypotheses and estimating the position of change points. Aue et al.^[10] and Jandhyala et al.^[11] provided a detailed introduction and summary of these contents. A key issue in identifying structural breaks in multivariate data is detecting changes in a sequence of means of dependent data. A classic method for change point detection of the mean vector under the assumption of independent data involves applying different variants of the CUSUM statistic^[12–14]. Many other methods have also been proposed, including pruned exact linear time (PELT)^[15] and Binary Segmentation^[16]. In regard to dependent data, numerous methods for this topic have been suggested and studied, such as the Bayesian method^[17], the double CUSUM statistic^[18], minimum contrast estimators^[19], and the likelihood ratio test method^[20]. Pang et al.^[21] discussed the asymptotic inferences for the AR(1) model with a single change point stationary and nearly non-stationary cases. Galeano and Peña^[22] investigated the solution to covariance change detection in multivariate dependent time series. Jirak^[23] conducted a change point detection test on d dependent data under the

condition that both d and sample size n tend to infinity. Cho and Fryzlewicz^[24] also proposed the WCM.gSa method for dependent cases. High dimensionality has emerged as a key focal concern among scholars in recent years. Cho and Fryzlewicz^[25] introduced sparsified binary segmentation to address this issue, while Wang and Samworth^[26] developed the sparse projection method for high-dimensional time series data. Furthermore, numerous studies have investigated the problem of change point detection in vector autoregressive (VAR) models^[27–29].

This article aims to detect the change point position in the mean vector for the q th-order vector moving average (VMA(q)) model, assuming that there is a single change point. A VMA(q) model is defined as follows:

$$\mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t - \mathbf{B}_1 \boldsymbol{\varepsilon}_{t-1} - \cdots - \mathbf{B}_q \boldsymbol{\varepsilon}_{t-q}, \quad (1)$$

where $\mathbf{B}_1, \dots, \mathbf{B}_q$ are $p \times p$ coefficient matrices, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,p})^T$, $i = 1 - q, \dots, n$, are i.i.d. centered random vectors with the covariance matrix $\mathbf{I}_p$, and the mean vector $\boldsymbol{\mu}$ is nonrandom. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a sequence of observations following Model (1), with an unknown location k_0 such that $\mathbb{E}\mathbf{x}_t = \boldsymbol{\mu}_1$ for $t \leq k_0$ and $\mathbb{E}\mathbf{x}_t = \boldsymbol{\mu}_2$ for $t > k_0$. We state that a change point occurs if $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, and our objective is to estimate k_0 . Our main contribution in this work is to introduce a novel change point detection method for high-dimensional data following the VMA(q) model, as detailed in Section 2. Motivated by the dimension reduction technique^[30], our proposed method initially performs dimension reduction on the original data by retaining only those components whose

weighted mean of squared differences exceeds a given threshold. We subsequently construct a CUSUM-type statistic based on the reduced components to estimate the location of the change point.

There are relatively few works that extend change point detection methods to the VMA model, particularly in high-dimensional scenarios. In this study, we aim to address both the high-dimensional and dependent aspects of the data simultaneously, with a specific focus on the VMA model. Through simulation experiments, we observed that existing dimension reduction methods are not highly effective when dealing with the VMA model. To overcome this limitation, we propose appropriate rectification of the dimension reduction statistic V_k to mitigate the autocorrelation effects inherent in the model.

The remainder of the paper is organized as follows. In Section 2, we introduce the change point detection method, including both the dimension reduction step and the detection step. Additionally, we provide detailed information on threshold selection. In Section 3, simulations and numerical comparisons are presented to illustrate the performance of our proposed method. Furthermore, two real data applications are conducted in Section 3. A brief summary is provided to conclude the paper in Section 4. Supplementary numerical results are provided in the Appendix.

Notations. We present some notations to be used throughout the paper. The matrix I_p represents the $p \times p$ identity matrix. For a p -dimensional vector, $\|\cdot\|$ is used to denote its L_2 norm. The determinant of a square matrix is represented by $\det(\cdot)$, and $\circ$ is used to denote the Hadamard product.

2 Methodology

Let $\mu_1 = (\mu_{1,1}, \dots, \mu_{1,p})^T$ and $\mu_2 = (\mu_{2,1}, \dots, \mu_{2,p})^T$ be nonrandom vectors and $\mu_1 \neq \mu_2$. Suppose that

$$\mathbf{x}_t = \mathbf{v}_t + \boldsymbol{\varepsilon}_t - \mathbf{B}_1 \boldsymbol{\varepsilon}_{t-1} - \dots - \mathbf{B}_q \boldsymbol{\varepsilon}_{t-q}, \quad \mathbf{v}_t = \begin{cases} \mu_1, & 1 \leq t \leq k_0; \\ \mu_2, & k_0 + 1 \leq t \leq n, \end{cases} \quad (2)$$

are p -dimensional observations, where the parameter k_0 is the true change point position. For the coefficient matrices $\mathbf{B}_1, \dots, \mathbf{B}_q$, we assume that $\det(\mathbf{I}_p - \mathbf{B}_1 z - \mathbf{B}_2 z^2 - \dots - \mathbf{B}_q z^q) \neq 0$ for all $|z| \leq 1$ to ensure that the process is invertible. We aim to estimate the true position k_0 . Our methodology can be divided into the following two steps:

Step 1. We perform the dimension reduction procedure to address the problem of high dimensionality. Denote

$$\begin{aligned} \hat{\mu}_1^k &= (\hat{\mu}_{1,1}, \dots, \hat{\mu}_{1,p})^T = \frac{1}{k} \sum_{i=1}^k \mathbf{x}_i, \\ \hat{\mu}_{k+1}^n &= (\hat{\mu}_{k+1,1}, \dots, \hat{\mu}_{k+1,p})^T = \frac{1}{n-k} \sum_{i=k+1}^n \mathbf{x}_i. \end{aligned} \quad (3)$$

Our dimension reduction method only retains the attributes that make the following statistics sufficiently large:

$$\sum_{k=2}^{n-2} k(n-k) (\hat{\mu}_{1,a}^k - \hat{\mu}_{k+1,a}^n)^2, \quad a \in \{1, \dots, p\}.$$

Step 2. We estimate the location of the change point of the

data after the dimension reduction. We use $\tilde{\mu}_1^k, \tilde{\mu}_{k+1}^n$ to represent the corresponding versions of $\hat{\mu}_1^k, \hat{\mu}_{k+1}^n$ after dimension reduction. The change point is estimated by finding the point that maximizes the statistic

$$\frac{1}{n^4} \|(n-k)k(\tilde{\mu}_1^k - \tilde{\mu}_{k+1}^n)\|^2.$$

We provide a comprehensive explanation of these two steps in the next sections.

2.1 Step 1: dimension reduction

Define

$$\begin{aligned} E(k, n, q, \mathbf{B}_1, \dots, \mathbf{B}_q) &= \frac{2}{k(k-1)} \sum_{m=1}^{\min(q,k)} (k-m) \mathbf{H}(m) + \\ &\quad \frac{2}{(n-k)(n-k-1)} \sum_{m=1}^{\min(q,n-k)} (n-k-m) \mathbf{H}(m) - \\ &\quad \frac{2}{k(n-k)} \sum_{m=1}^{\min(q,k,n-k)} m \mathbf{H}(m), \end{aligned}$$

where $\mathbf{H}(m)$ is a vector composed of the diagonal elements of the autocovariance matrix $\boldsymbol{\Gamma}_m = \mathbb{E}[(\mathbf{x}_t - \mathbf{v}_t)(\mathbf{x}_{t+m} - \mathbf{v}_{t+m})^T]$. We introduce the following statistic:

$$\begin{aligned} V_k &= (V_{k,a})_{1 \leq a \leq p} = \\ &\quad \frac{1}{k(k-1)} \sum_{i \neq j \leq k} \mathbf{x}_i \circ \mathbf{x}_j + \\ &\quad \frac{1}{(n-k)(n-k-1)} \sum_{i \neq j > k} \mathbf{x}_i \circ \mathbf{x}_j - \\ &\quad \frac{2}{k(n-k)} \sum_{i=1}^k \sum_{j=k+1}^n \mathbf{x}_i \circ \mathbf{x}_j - E(k, n, q, \mathbf{B}_1, \dots, \mathbf{B}_q). \end{aligned} \quad (4)$$

The motivation behind the construction of the statistic V_k is to approximate

$$\begin{aligned} V_k^* &= \frac{1}{k^2} \sum_{i,j=1}^k \mathbf{x}_i \circ \mathbf{x}_j + \frac{1}{(n-k)^2} \sum_{i,j=k+1}^n \mathbf{x}_i \circ \mathbf{x}_j - \\ &\quad \frac{2}{k(n-k)} \sum_{i=1}^k \sum_{j=k+1}^n \mathbf{x}_i \circ \mathbf{x}_j, \end{aligned}$$

where $V_k^* = ((\hat{\mu}_{1,a}^k - \hat{\mu}_{k+1,a}^n)^2)_{1 \leq a \leq p}$. Clearly, V_k^* is an estimator of $(\mu_1 - \mu_2)^2$ when $k = k_0$ and can be used to measure the difference between the mean vectors. We mention that there are two main differences between V_k and V_k^* . First, unlike V_k^* , V_k excludes the terms corresponding to $i = j$ to eliminate the influence of $\mathbf{x}_i \circ \mathbf{x}_i$'s. Second, an additional correction term $E(k, n, q, \mathbf{B}_1, \dots, \mathbf{B}_q)$ is introduced to account for the correlations in the VMA(q) model. The theoretical basis for this can be observed through Theorem 1 below.

However, the exact position of k_0 is unknown in reality, so we analyze the properties of the statistic for all positions $k \in \{1, \dots, n\}$.

Theorem 1. Assume that the observations $\mathbf{x}_1, \dots, \mathbf{x}_n$ follow Model (2). The expectation of $V_{k,a}$ is given by

$$\mathbb{E}V_{k,a} = \begin{cases} (\mu_{1,a} - \mu_{2,a})^2 & \text{if } k = k_0; \\ \frac{k_0(k_0-1)}{k(k-1)}(\mu_{1,a} - \mu_{2,a})^2 & \text{if } k > k_0; \\ \frac{(n-k_0)(n-k_0-1)}{(n-k)(n-k-1)}(\mu_{1,a} - \mu_{2,a})^2 & \text{if } k < k_0. \end{cases} \quad (5)$$

Proof. Let $x_{i,a}$ and $E(k, n, q, \mathbf{B}_1, \dots, \mathbf{B}_q)_a$ be the a th element of $\mathbf{x}_i$ and the a th element of $E(k, n, q, \mathbf{B}_1, \dots, \mathbf{B}_q)$, respectively. Moreover, $\mathbf{H}(m)_a = (\mathbf{F}_m)_{a,a}$. Then, $\mathbb{E}V_{k,a}$ can be written as

$$\begin{aligned} \mathbb{E}V_{k,a} &= \frac{1}{k(k-1)} \sum_{i \neq j=1}^k \mathbb{E}(x_{i,a}x_{j,a}) + \\ &\quad \frac{1}{(n-k)(n-k-1)} \sum_{i \neq j=k+1}^n \mathbb{E}(x_{i,a}x_{j,a}) - \\ &\quad \frac{2}{k(n-k)} \sum_{i=1}^k \sum_{j=k+1}^n \mathbb{E}(x_{i,a}x_{j,a}) - E(k, n, q, \mathbf{B}_1, \dots, \mathbf{B}_q)_a. \end{aligned}$$

Since

$$\begin{aligned} \mathbb{E}(x_{i,a}x_{j,a}) &= \\ v_{i,a}v_{j,a} + \mathbb{E} \left[\left[\varepsilon_{i,a} - \sum_{t=1}^q \sum_{l=1}^p (B_t)_{a,l} \varepsilon_{i-t,l} \right] \left[\varepsilon_{j,a} - \sum_{t=1}^q \sum_{l=1}^p (B_t)_{a,l} \varepsilon_{j-t,l} \right] \right] &= \\ v_{i,a}v_{j,a} - (B_{|i-j|})_{a,a} + (B_{|i-j|+1} \mathbf{B}_1^T)_{a,a} + (B_{|i-j|+2} \mathbf{B}_2^T)_{a,a} + \dots + (B_q \mathbf{B}_{q-|i-j|}^T)_{a,a} &= \\ v_{i,a}v_{j,a} + H(|i-j|)_a, \end{aligned}$$

and

$$\begin{aligned} \frac{1}{k(k-1)} \sum_{i \neq j=1}^k H(|i-j|)_a + \frac{1}{(n-k)(n-k-1)} \sum_{i \neq j=k+1}^n H(|i-j|)_a &= \\ \frac{2}{k(n-k)} \sum_{i=1}^k \sum_{j=k+1}^n H(|i-j|)_a + E(k, n, q, \mathbf{B}_1, \dots, \mathbf{B}_q)_a, \end{aligned}$$

we have

$$\begin{aligned} \mathbb{E}V_{k,a} &= \frac{1}{k(k-1)} \sum_{i \neq j=1}^k v_{i,a}v_{j,a} + \frac{1}{(n-k)(n-k-1)} \sum_{i \neq j=k+1}^n v_{i,a}v_{j,a} - \\ &\quad \frac{2}{k(n-k)} \sum_{i=1}^k \sum_{j=k+1}^n v_{i,a}v_{j,a}. \end{aligned}$$

When $k = k_0$, the expectation of $V_{k,a}$ is

$$\begin{aligned} \mathbb{E}V_{k_0,a} &= \frac{1}{k_0(k_0-1)} k_0(k_0-1) \mu_{1,a}^2 + \\ &\quad \frac{1}{(n-k_0)(n-k_0-1)} (n-k_0)(n-k_0-1) \mu_{2,a}^2 - \\ &\quad \frac{2}{k_0(n-k_0)} k_0(n-k_0) \mu_{1,a} \mu_{2,a} = \\ &\quad (\mu_{1,a} - \mu_{2,a})^2. \end{aligned}$$

When $k > k_0$,

$$\begin{aligned} \mathbb{E}V_{k,a} &= \\ \frac{1}{k(k-1)} \left[\sum_{i \neq j=1}^{k_0} v_{i,a}v_{j,a} + \sum_{i=1}^{k_0} \sum_{j=k_0+1}^k v_{i,a}v_{j,a} + \sum_{i=k_0+1}^k \sum_{j=1}^{k_0} v_{i,a}v_{j,a} + \sum_{i \neq j=k_0+1}^k v_{i,a}v_{j,a} \right] &+ \\ \frac{1}{(n-k)(n-k-1)} \sum_{i \neq j=k+1}^n v_{i,a}v_{j,a} - \\ \frac{2}{k(n-k)} \left[\sum_{i=1}^{k_0} \sum_{j=k+1}^n v_{i,a}v_{j,a} + \sum_{i=k_0+1}^k \sum_{j=k+1}^n v_{i,a}v_{j,a} \right] &= \\ \frac{1}{k(k-1)} [k_0(k_0-1) \mu_{1,a}^2 + 2k_0(k-k_0) \mu_{1,a} \mu_{2,a} + (k-k_0)(k-k_0-1) \mu_{2,a}^2] &+ \\ \frac{1}{(n-k)(n-k-1)} (n-k)(n-k-1) \mu_{2,a}^2 - \\ \frac{2}{k(n-k)} [k_0(n-k) \mu_{1,a} \mu_{2,a} + (k-k_0)(n-k) \mu_{2,a}^2] &= \\ \frac{k_0(k_0-1)}{k(k-1)} (\mu_{1,a} - \mu_{2,a})^2. \end{aligned}$$

Similarly, it can be verified that when $k < k_0$,

$$\mathbb{E}V_{k,a} = \frac{(n-k_0)(n-k_0-1)}{(n-k)(n-k-1)} (\mu_{1,a} - \mu_{2,a})^2.$$

Combining the results above, we have proven (5).

According to Theorem 1, the expectation of $V_{k,a}$ reaches the maximum value $(\mu_{1,a} - \mu_{2,a})^2$ at $k = k_0$. The coefficients $\frac{k_0(k_0-1)}{k(k-1)}$ and $\frac{(n-k_0)(n-k_0-1)}{(n-k)(n-k-1)}$ before $(\mu_{1,a} - \mu_{2,a})^2$ only depend on k . For any fixed k , a larger $V_{k,a}$ implies a larger difference between $\mu_{1,a}$ and $\mu_{2,a}$. To identify the significant attributes, we introduce the weighted sum

$$\mathbf{D} = \frac{1}{n-3} \sum_{k=2}^{n-2} \frac{k(n-k)}{n} V_k. \quad (6)$$

The introduction of this weight is motivated by the observation that the variances of V_k differ when k takes different values, resulting in varying proportions of the estimated mean difference of V_k calculated at different positions k . By assigning weights to V_k corresponding to their respective k values, we can effectively balance the impact caused by disparate V_k variances and ensure consistency in estimating the impact across all positions. After constructing the vector $\mathbf{D}$ via this approach, we retain attributes with a $\mathbf{D}$ value that exceeds a predefined threshold to achieve dimension reduction.

2.1.1 Choice of threshold

In this section, our objective is to determine a threshold τ to filter out elements in $\mathbf{D}$ whose values exceed the threshold τ ; the remaining elements serve as the attributes retained during our dimension reduction step. To approximate the distribution properties in the absence of a change point, we employ a bootstrap method. Define

$$\begin{aligned} \mathbf{T} &= (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \in \mathbb{R}^{p \times n}, \\ \mathbf{Z} &= (\mathbf{x}_2 - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_2, \dots, \mathbf{x}_n - \mathbf{x}_{n-1}) \in \mathbb{R}^{p \times (n-1)}. \end{aligned} \quad (7)$$

Then, except for $\mathbf{x}_{k_0+1} - \mathbf{x}_{k_0}$, each column of the matrix $\mathbf{Z}$ satisfies

$$\begin{aligned} \mathbf{x}_t - \mathbf{x}_{t-1} &= \varepsilon_t - (B_1 + I) \varepsilon_{t-1} - (B_2 - B_1) \varepsilon_{t-2} - \dots - \\ &\quad (B_q - B_{q-1}) \varepsilon_{t-q} + B_q \varepsilon_{t-q-1}. \end{aligned}$$

Recalling the definition of VMA, we observe that these

columns can be considered new vector moving average model with order $q+1$ and coefficient matrices

$$B_1^* = B_1 + I, B_2^* = B_2 - B_1, \dots, B_q^* = B_q - B_{q-1}, B_{q+1}^* = -B_q,$$

or equivalently,

$$\begin{aligned} B_1 &= B_1^* - I, B_2 = B_2^* + B_1^* - I, \\ B_3 &= B_3^* + B_2^* + B_1^* - I, \dots, B_q = \sum_{i=1}^q B_i^* - I. \end{aligned} \quad (8)$$

Given that the matrix $\mathbf{Z}$ contains only one special column with a nonzero mean, its influence on model parameter estimation can be disregarded when the sample size n is sufficiently large. Therefore, we calculate the corresponding sample covariance matrices to approximate the autocovariance matrices $\mathbf{\Gamma}_0, \mathbf{\Gamma}_1, \dots, \mathbf{\Gamma}_{q+1}$. Since the covariance matrix of the model is related only to the coefficient matrices $B_1^*, \dots, B_{q+1}^*$ and the variance of white noise, we can estimate the coefficient matrices $B_1^*, \dots, B_{q+1}^*$ from the corresponding sample covariance matrices. Then, the estimates of $B_1, \dots, B_q$ can be directly derived from (8), denoted as $\hat{B}_1, \dots, \hat{B}_q$. To determine the threshold τ , we generate a new data matrix $\mathbf{Y} = (\mathbf{y}_{1-q}, \dots, \mathbf{y}_n) = (\mathbf{y}_{i,j}) \in \mathbb{R}^{p \times (n+q)}$, where $y_{i,j}$'s are i.i.d. standard normal random variables. Define

$$\mathbf{x}_t^* = \mathbf{y}_t - \hat{B}_1 \mathbf{y}_{t-1} - \hat{B}_2 \mathbf{y}_{t-2} - \dots - \hat{B}_q \mathbf{y}_{t-q}.$$

We substitute the original data $\mathbf{x}_1, \dots, \mathbf{x}_n$ with $\mathbf{x}_1^*, \dots, \mathbf{x}_n^*$ in the calculation of $\mathbf{D}$, aiming to simulate datasets with identical coefficient matrices but devoid of any change points. Denote the new vector by

$$\mathbf{D}^* = (D_a^* : 1 \leq a \leq p).$$

The threshold τ is defined as the maximum element of $\mathbf{D}^*$, i.e.,

$$\tau = \max_{1 \leq a \leq p} D_a^*.$$

We can subsequently identify all significant attributes whose corresponding entries in $\mathbf{D}$ exceed τ . These identified attributes will be exclusively considered in Step 2.

2.2 Step 2: change point detection after dimension reduction

In what follows, we use $\tilde{\mathbf{x}}_t$ to denote the vectors after dimension reduction. The change point estimation is conducted via the following CUSUM-type statistic:

$$U_{n,k} = \frac{1}{n^4} \sum_{(t \neq l)=1}^k \sum_{(j \neq l)=k+1}^n \sum_{i=1}^n (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)^T (\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}}_l), \quad (9)$$

and the change point location is estimated as

$$\hat{k} = \arg \max_{1 \leq k \leq n} U_{n,k}. \quad (10)$$

The statistic $U_{n,k}$ is motivated by the CUSUM statistic, which is frequently used in the change point literature. It takes the form

$$U_{n,k}^* = \frac{1}{n^4} \sum_{i,l=1}^k \sum_{j,l=k+1}^n (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)^T (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_l) = \frac{1}{n^4} \| (n-k)k(\tilde{\boldsymbol{\mu}}_1^k - \tilde{\boldsymbol{\mu}}_{k+1}^n) \|^2,$$

where

$$\tilde{\boldsymbol{\mu}}_1^k = \frac{1}{k} \sum_{i=1}^k \tilde{\mathbf{x}}_i, \quad \tilde{\boldsymbol{\mu}}_{k+1}^n = \frac{1}{n-k} \sum_{i=k+1}^n \tilde{\mathbf{x}}_i.$$

Note that the difference

$$\begin{aligned} U_{n,k}^* - U_{n,k} &= \frac{1}{n^4} \sum_{i=1}^k \sum_{j=k+1}^n \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|^2 + \\ &\quad \frac{1}{n^4} \sum_{i=1}^k \sum_{(j \neq l)=k+1}^n \sum_{l=k+1}^n (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)^T (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_l) + \\ &\quad \frac{1}{n^4} \sum_{(t \neq l)=1}^k \sum_{j=k+1}^n \sum_{l=k+1}^n (\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}}_j)^T (\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}}_l) \end{aligned} \quad (11)$$

is of a smaller order than $U_{n,k}$ when k is far from 1 and n . (11) that the statistic $U_{n,k}$ does not consist of the terms corresponding to $i=t$ and $j=l$, thereby eliminating the influence of $\tilde{\mathbf{x}}_i^T \tilde{\mathbf{x}}_i$. Consequently, we avoid estimating high-order moments, leading to an enhancement in the finite sample performance of the procedure.

3 Numerical studies

3.1 Simulations

In this section, we study the numerical performance of the proposed method in simulated settings. For each $p \in \{100, 200, 300, 400\}$, we generate a sequence of $n = 200$ observations from the model

$$\mathbf{x}_t = \begin{cases} \boldsymbol{\mu}_1 + \boldsymbol{\varepsilon}_t - \mathbf{B}_1 \boldsymbol{\varepsilon}_{t-1}, & \text{if } 1 \leq t < 100; \\ \boldsymbol{\mu}_2 + \boldsymbol{\varepsilon}_t - \mathbf{B}_1 \boldsymbol{\varepsilon}_{t-1}, & \text{if } 100 \leq t \leq 200, \end{cases}$$

where $\boldsymbol{\mu}_1 = \mathbf{0}$, and $\boldsymbol{\varepsilon}_t$'s are i.i.d. Gaussian random vectors with mean zero and population covariance matrix $\mathbf{I}_p$. We choose $\boldsymbol{\mu}_2 = (0.5, 0, \dots, 0)^T$ or $\boldsymbol{\mu}_2 = (1, 0, \dots, 0)^T$ so that $\|\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\| = 0.5$ or 1, and we use “dmu” to refer to $\|\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\|$ in the figures. The matrix $\mathbf{B}_1$ is $0.8\mathbf{I}_p$ or a random matrix satisfying the aforementioned invertibility condition of the VMA model. Let $\hat{r} = \hat{k}/n$ be the estimate of the fraction of the change point. Since the true change point is located at 100, the true fraction $k_0/n = 0.5$. The mean value (mean), absolute deviation (AD), standard deviation (SD), and mean square error (MSE) of $\hat{k}/n$ are calculated for each parameter setting on the basis of 200 replications. These results are recorded in Table 1, where the column “stationary” corresponds to the scenario where $\mathbf{B}_1$ is a random matrix. It is evident that the estimated fractions closely approximate the true value of 0.5 while exhibiting small values for AD, sd, and MSE.

Fig. 1 records the threshold τ and the statistic $\max\{D(-1)\} := \max_{a \in \{2, \dots, p\}} D_a$, and Fig. 2 compares τ with $D(1) := D_1$. Under our settings, Figs. 1 and 2 display the performance of D statistic in dimension reduction for attributes without changes and with significant changes, respectively. Regardless of the changes in p and $\|\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\|$, $D(1) > \tau > \max\{D(-1)\}$ always holds. This observation implies that the

Table 1. Mean, AD, sd, and MSE of the change point estimator under different parameter settings.

p	$\ \mu_2 - \mu_1\ $	B_1	Mean	AD	SD	MSE
100	1	$0.8I_p$	0.4958	0.0042	0.0089	0.000096
100	0.5	$0.8I_p$	0.4915	0.0085	0.0179	0.000390
200	1	$0.8I_p$	0.4960	0.0040	0.0091	0.000099
200	0.5	$0.8I_p$	0.4918	0.0082	0.0157	0.000312
300	1	$0.8I_p$	0.4952	0.0048	0.0094	0.000110
300	0.5	$0.8I_p$	0.4927	0.0073	0.0158	0.000302
400	1	$0.8I_p$	0.4963	0.0037	0.0080	0.000077
400	0.5	$0.8I_p$	0.4923	0.0077	0.0175	0.000365
100	1	Stationary	0.4954	0.0046	0.0107	0.000136
100	0.5	Stationary	0.4830	0.0170	0.0371	0.001659
200	1	Stationary	0.4948	0.0052	0.0120	0.000171
200	0.5	Stationary	0.4860	0.0140	0.0415	0.001907
300	1	Stationary	0.4938	0.0062	0.0131	0.000208
300	0.5	Stationary	0.4826	0.0174	0.0430	0.002141
400	1	Stationary	0.4950	0.0050	0.0117	0.000161
400	0.5	Stationary	0.4804	0.0196	0.0481	0.002689

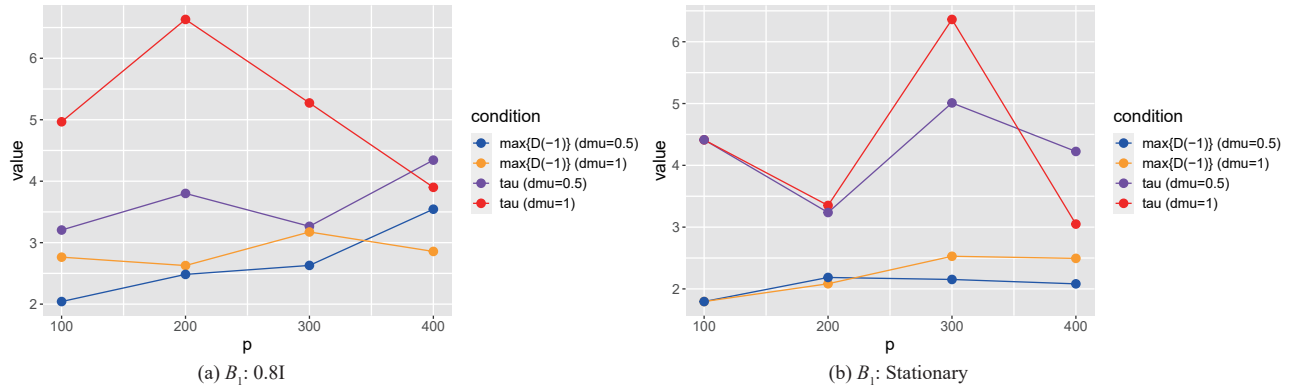


Fig. 1. Performance of statistics on attributes without change points.

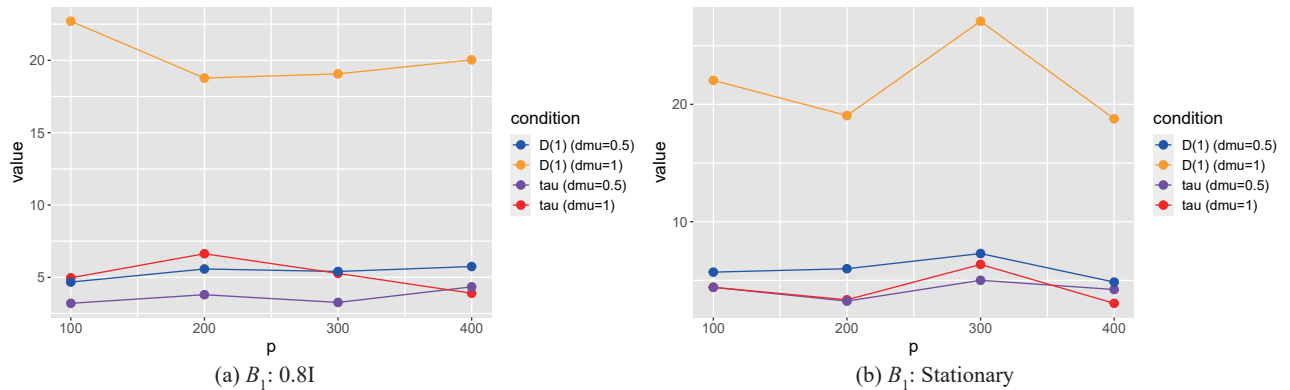


Fig. 2. Performance of statistics on attributes with change points.

dimension reduction step effectively identifies the dimension containing the change point as well as those without any change points. This finding is further supported by Fig. 3, which displays the scatter plots for all attributes when $B_1 = 0.8I_p$ and $p = 200$. Scatter plots under other settings can

be found in the Appendix.

The histograms of the estimated change point positions when $B_1 = 0.8I_p$ and $p = 100$ are visualized in Fig. 4 for $\|\mu_2 - \mu_1\| = 0.5$ and $\|\mu_2 - \mu_1\| = 1$, respectively. In Fig. 5, the dimension p is increased to 200. As expected, Figs. 4 and 5

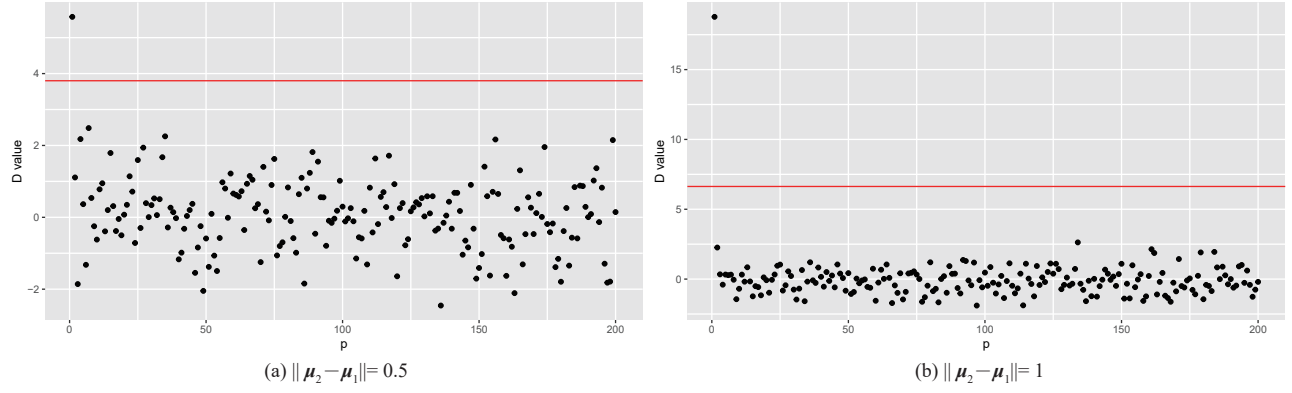


Fig. 3. Dimension reduction results. $B_1 = 0.8I_p$, $p = 200$. The red line represents the threshold value.

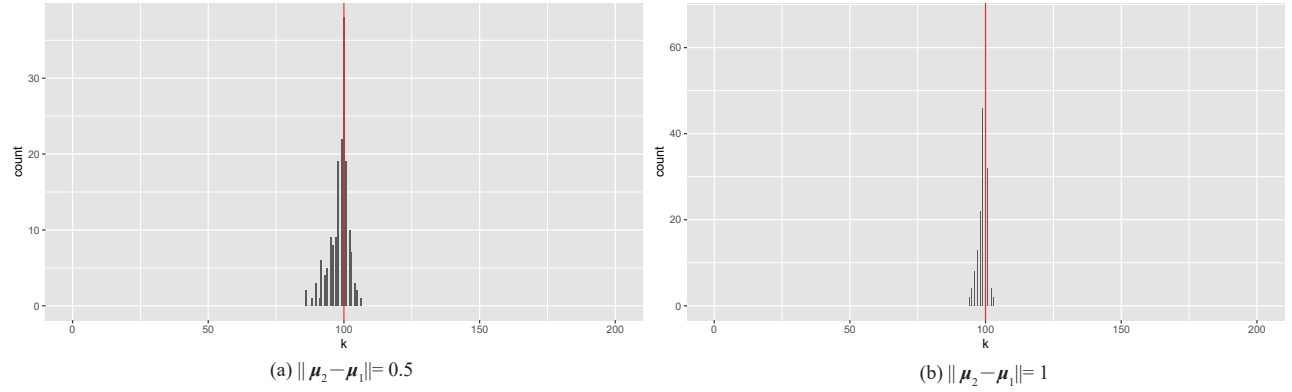


Fig. 4. Histograms of the estimated change point positions. The red line represents the true position. The coefficient matrix $B_1 = 0.8I_p$, and the dimension $p = 100$.

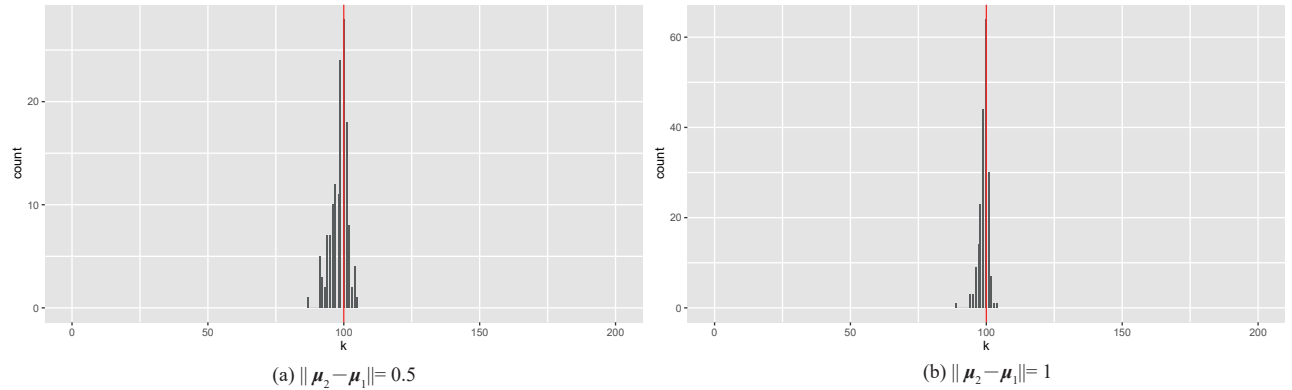


Fig. 5. Histograms of the estimated change point positions. The red line represents the true position. The coefficient matrix $B_1 = 0.8I_p$, and the dimension $p = 200$.

show that the larger the difference, the smaller the variance of the estimated change point positions. Moreover, the two figures and Table 1 illustrate that the proposed method is stable with respect to the dimension, which is due to the dimension reduction step that correctly identifies the components containing significant changes and discards those without changes. Additional figures depicting the performance of dimension reduction and change point detection under different parameter settings are presented in the Appendix.

3.2 Comparison with other change point detection methods

This section aims to compare the performance of the pro-

posed method with that of several existing methods under the settings detailed in Section 3.1. We first compare the proposed method with the sparsified binary segmentation (SBS) method^[25]. The simulated results when $\|\mu_2 - \mu_1\| = 1$ and $B_1 = 0.8I_p$ are recorded in Table 2. As can be observed, the proposed method yields smaller AD, sd, and MSE. The additional simulation results for SBS under alternative choices of $\|\mu_2 - \mu_1\|$ and B_1 are documented in Table 3. When compared with our results in Table 1, it can be observed that SBS exhibits worse performance when $\|\mu_2 - \mu_1\| = 0.5$, resulting in a significantly reduced success rate of detecting change points when the magnitude of change is small.

Next, we compare the proposed method with the

Table 2. Comparison between the proposed change point estimator (estimate) and the SBS change point estimator.

p	Estimate				SBS			
	Mean	AD	SD	MSE	Mean	AD	SD	MSE
100	0.4958	0.0042	0.0089	0.000096	0.5152	0.0152	0.0598	0.003792
200	0.4960	0.0040	0.0091	0.000099	0.5458	0.0458	0.0978	0.011603
300	0.4952	0.0048	0.0094	0.000110	0.5628	0.0628	0.1085	0.015660
400	0.4963	0.0037	0.0080	0.000077	0.5814	0.0814	0.1175	0.020350

Table 3. Proportion of successful detection of change points via the SBS method.

p	$\ \mu_2 - \mu_1\ $	B_1	Success rate of detection
100	0.5	$0.8I_p$	0.000
200	0.5	$0.8I_p$	0.000
300	0.5	$0.8I_p$	0.000
400	0.5	$0.8I_p$	0.000
100	1	Stationary	0.980
200	1	Stationary	0.965
300	1	Stationary	0.960
400	1	Stationary	0.945
100	0.5	Stationary	0.075
200	0.5	Stationary	0.070
300	0.5	Stationary	0.055
400	0.5	Stationary	0.025

WCM.gSa method^[24]. Note that the WCM.gSa method is limited to one-dimensional data. Therefore, when applying WCM.gSa in our study, we perform a preliminary step by considering only the first attribute of the data, effectively reducing it to a one-dimensional dataset containing a change point. In contrast, our methods are applied using the full p -dimensional data. The results are presented in Table 4 and indicate that despite this special treatment for WCM.gSa, our estimator demonstrates comparable performance to the WCM.gSa method in terms of sd and MSE, albeit with a slightly larger AD. Importantly, our approach possesses the ability to address high-dimensional change point problems and thus enhances its universal applicability.

Finally, we compare our method with the sparse projection approach^[26]. This method is designed to handle high-dimensional data and performs well when the data is dependent. The results are given in Table 5. We observe that the sparse projection estimator performs better than ours in terms of AD when $\|\mu_1 - \mu_2\| = 1$; however, this advantage diminishes and even disappears as the mean difference $\|\mu_1 - \mu_2\|$ decreases to 0.5. In contrast, the proposed method is more stable than the sparse projection estimator, as evident from the smaller sd and MSE.

3.3 Real data analysis

3.3.1 CGH data

In this section, we apply the proposed method to comparative genomic hybridization (CGH) profiles. CGH is a technique used to determine the number of copies of genes in a DNA profile. The number of copies of specific genes can vary between tumor cells and normal cells; thus, change points

exist. The dataset utilized in this section is publicly available at http://microarrays.curie.fr/publications/oncologie_moleculaire/bladder_TCM/. It comprises $p = 57$ bladder tumor samples, where each profile represents a sequence of 2385 relative quantities of DNA that describe the copy number variation. Our focus lies on the 16th chromosome, which exhibits a single change point. Following the suggestion of Enikeeva^[31], we remove vectors containing outliers to obtain a final dataset consisting of $n = 66$ observations on the 16th chromosome. In Fig. 6, we present a graphical representation of this specific chromosomal segment. As recommended by Refs. [31, 32], we calculate the scaled norm of the mean multiplied by the sign of the highest profile mean within each segment and depict it as a black piecewise constant line to visually represent both the magnitude and direction of change in the mean vector. Applying the proposed methodology to this dataset, we determine that the estimated location of the change point is 32, which aligns precisely with the findings from Ref. [31].

3.3.2 Processed inventory investment data

In this section, we apply the proposed method to the quarterly inventory investment data released by the Canadian National Bureau of Statistics from 1961 to 2023. Given that inventory levels are periodically reset to a predetermined optimal level, this implies a temporal constraint on the impact of inventory investment. Once this time period elapses, autocorrelation diminishes to zero, which is an inherent characteristic of the moving average process. Consequently, the behavior of inventory investment can be effectively modeled via a moving average process. We specifically consider market values from

Table 4. Comparison between the proposed change point estimator (estimate) and the WCM.gSa method (WCM.gSa is limited to the first dimension of the p -dimensional data).

p	$\ \mu_2 - \mu_1\ $	B_1	Estimate				WCM.gSa			
			Mean	AD	SD	MSE	Mean	AD	SD	MSE
100	1	$0.8I_p$	0.4958	0.0042	0.0089	0.000096	0.4999	0.0001	0.007026	0.000049
100	0.5	$0.8I_p$	0.4915	0.0085	0.0179	0.000390	0.5018	0.0018	0.013781	0.000192
200	1	$0.8I_p$	0.4960	0.0040	0.0091	0.000099	0.4990	0.0010	0.007469	0.000057
200	0.5	$0.8I_p$	0.4918	0.0082	0.0157	0.000312	0.5012	0.0012	0.023705	0.000561
300	1	$0.8I_p$	0.4952	0.0048	0.0094	0.000110	0.5005	0.0005	0.006661	0.000044
300	0.5	$0.8I_p$	0.4927	0.0073	0.0158	0.000302	0.5041	0.0041	0.033720	0.001148
400	1	$0.8I_p$	0.4963	0.0037	0.0080	0.000077	0.4999	0.0001	0.007493	0.000056
400	0.5	$0.8I_p$	0.4923	0.0077	0.0175	0.000365	0.5031	0.0031	0.048692	0.002369
100	1	Stationary	0.4954	0.0046	0.0107	0.000136	0.4987	0.0013	0.009437	0.000090
100	0.5	Stationary	0.4830	0.0170	0.0371	0.001659	0.4999	0.0001	0.049415	0.002430
200	1	Stationary	0.4948	0.0052	0.0120	0.000171	0.5013	0.0013	0.009606	0.000094
200	0.5	Stationary	0.4860	0.0140	0.0415	0.001907	0.5030	0.0030	0.044479	0.001977
300	1	Stationary	0.4938	0.0062	0.0131	0.000208	0.4989	0.0011	0.010401	0.000109
300	0.5	Stationary	0.4826	0.0174	0.0430	0.002141	0.5030	0.0030	0.050000	0.002496
400	1	Stationary	0.4950	0.0050	0.0117	0.000161	0.4958	0.0042	0.033884	0.001160
400	0.5	Stationary	0.4804	0.0196	0.0481	0.002689	0.5010	0.0010	0.057448	0.003285

Table 5. Comparison between the proposed change point estimator (estimate) and the sparse projection estimator.

p	$\ \mu_2 - \mu_1\ $	B_1	Estimate				Sparse projection			
			Mean	AD	SD	MSE	Mean	AD	SD	MSE
100	1	$0.8I_p$	0.4958	0.0042	0.0089	0.000096	0.5003	0.0003	0.006607	0.000044
100	0.5	$0.8I_p$	0.4915	0.0085	0.0179	0.000390	0.4994	0.0006	0.062318	0.003865
200	1	$0.8I_p$	0.4960	0.0040	0.0091	0.000099	0.5003	0.0003	0.007794	0.000061
200	0.5	$0.8I_p$	0.4918	0.0082	0.0157	0.000312	0.4877	0.0123	0.079063	0.006371
300	1	$0.8I_p$	0.4952	0.0048	0.0094	0.000110	0.4994	0.0006	0.006846	0.000047
300	0.5	$0.8I_p$	0.4927	0.0073	0.0158	0.000302	0.4856	0.0144	0.140367	0.019811
400	1	$0.8I_p$	0.4963	0.0037	0.0080	0.000077	0.5003	0.0003	0.007368	0.000054
400	0.5	$0.8I_p$	0.4923	0.0077	0.0175	0.000365	0.4891	0.0109	0.175659	0.030820
100	1	Stationary	0.4954	0.0046	0.0107	0.000136	0.4987	0.0013	0.009408	0.000090
100	0.5	Stationary	0.4830	0.0170	0.0371	0.001659	0.4993	0.0007	0.090523	0.008154
200	1	Stationary	0.4948	0.0052	0.0120	0.000171	0.4989	0.0011	0.008356	0.000071
200	0.5	Stationary	0.4860	0.0140	0.0415	0.001907	0.4994	0.0006	0.091787	0.008383
300	1	Stationary	0.4938	0.0062	0.0131	0.000208	0.4990	0.0010	0.010398	0.000109
300	0.5	Stationary	0.4826	0.0174	0.0430	0.002141	0.5133	0.0133	0.149541	0.022428
400	1	Stationary	0.4950	0.0050	0.0117	0.000161	0.4995	0.0005	0.010765	0.000116
400	0.5	Stationary	0.4804	0.0196	0.0481	0.002689	0.5191	0.0191	0.164890	0.027416

Q1 1986 to Q4 2010, excluding three non-stationary attributes and scaling the data accordingly, as shown in Fig. 7b. Additionally, we manually introduce a change point in the “wholesale” attribute at position 50 with a magnitude of 3 units, as shown in Fig. 7a. By applying our proposed methodology to this dataset, we accurately detect the change point at position 50.

4 Conclusions

We reduce dimensionality via a weighted sum of squared differences among attribute mean vector estimators, introduce a rectification to preserve their expectation properties, and estimate the change point on the reduced data using a CUSUM-

type statistic that discards higher-order moment terms.

Our research focuses primarily on investigating change point detection in the context of a single change point. To extend our methodology to encompass multiple change points, a preliminary approach would involve employing the widely-used binary segmentation method^[16], which is commonly employed for detecting multiple change points. After completing a change point detection process and obtaining an estimated value for the actual position of the change point via our proposed method, we partition the original sequence into two segments on the basis of this estimated position and subsequently perform separate change point detection steps on these subintervals. This iterative process continues until all estimated change points no longer exhibit statistically signi-

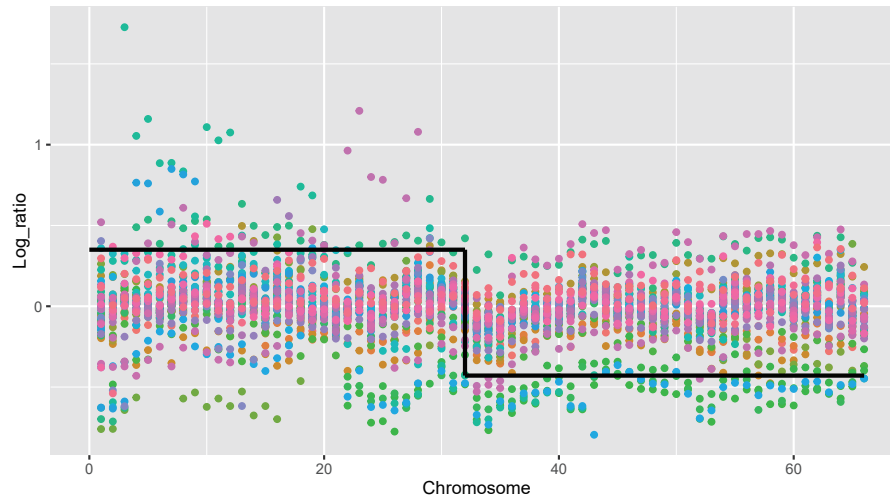


Fig. 6. 57 CGH profiles on the 16th chromosome segment.

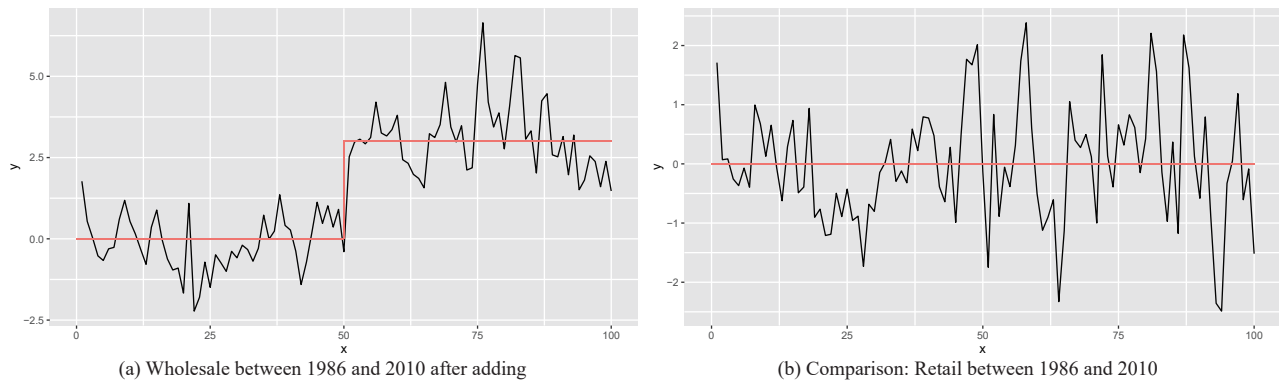


Fig. 7. Comparison between real data and data with a change point.

ficance. By leveraging the binary segmentation method as a foundation, the proposed method can be extended to incorporate WBS^[33], WBS2^[34], and other methods suitable for detecting multiple change points. The detailed exploration of multiple change point detection in the VAR model is left for future research.

Acknowledgements

This work was supported by the National Key R&D Program of China (2022YFA1008000) and the National Natural Science Foundation of China (12571297).

Biographies

Ruiqi Chen is a master's graduate at the School of Management of the University of Science and Technology of China. His research mainly focuses on change point detection.

Qing Yang is an Associate Professor at the Department of Statistics and Finance, School of Management, University of Science and Technology of China. She received her Ph.D. degree from Nanyang Technological University, Singapore. Her research mainly focuses on high-dimensional statistical inference and random matrix theory.

References

[1] Wald A. Sequential tests of statistical hypotheses. *The Annals of*

Mathematical Statistics, **1945**, 16 (2): 117–186.
 [2] Page E S. Continuous inspection schemes. *Biometrika*, **1954**, 41: 100–115.
 [3] Olshen A B, Venkatraman E S, Lucito R, et al. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, **2004**, 5 (4): 557–572.
 [4] Hansen B E. The new econometrics of structural change: Dating breaks in US labor productivity. *Journal of Economic Perspectives*, **2001**, 15 (4): 117–128.
 [5] Andreou E, Ghysels E. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics*, **2002**, 17 (5): 579–600.
 [6] Bai J, Perron P. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, **2003**, 18 (1): 1–22.
 [7] Reeves J, Chen J, Wang X L, et al. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, **2007**, 46 (6): 900–915.
 [8] Ranganathan A. PLISS: labeling places using online changepoint detection. *Autonomous Robots*, **2012**, 32 (4): 351–368.
 [9] Younes L, Albert M, Miller M I, et al. Inferring changepoint times of medial temporal lobe morphometric change in preclinical Alzheimer's disease. *NeuroImage: Clinical*, **2014**, 5: 178–187.
 [10] Aue A, Horváth L. Structural breaks in time series. *Journal of Time Series Analysis*, **2013**, 34 (1): 1–16.
 [11] Jandhyala V, Fotopoulos S, MacNeill I, et al. Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, **2013**, 34 (4): 423–446.
 [12] Chu C J, Stinchcombe M, White H. Monitoring structural change. *Econometrica*, **1996**, 64 (5): 1045–1065.
 [13] Horváth L, Kokoszka P, Steinebach J. Testing for changes in multivariate dependent observations with an application to temperature changes. *Journal of Multivariate Analysis*, **1999**, 68 (1):

- 96–119.
- [14] Horváth L, Hušková M. Change-point detection in panel data. *Journal of Time Series Analysis*, **2012**, 33 (4): 631–648.
- [15] Wambui G D, Waititu G A, Wanjoya A. The power of the pruned exact linear time (PELT) test in multiple changepoint detection. *American Journal of Theoretical and Applied Statistics*, **2015**, 4 (6): 581–586.
- [16] Vostrikova L Y. Detecting “disorder” in multidimensional random processes. In: *Doklady Akademii Nauk*. Moscow: Russian Academy of Sciences, **1981**, 259: 270–274.
- [17] Ray B K, Tsay R S. Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis*, **2002**, 23 (6): 687–705.
- [18] Cho H. Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics*, **2016**, 10 (2): 2000–2038.
- [19] Lavielle M. Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and Their Applications*, **1999**, 83 (1): 79–102.
- [20] Berkes I, Horváth L, Ling S, et al. Testing for structural change of AR model to threshold AR model. *Journal of Time Series Analysis*, **2011**, 32 (5): 547–565.
- [21] Pang T, Zhang D, Chong T T. Asymptotic inferences for an AR(1) model with a change point: Stationary and nearly nonstationary cases. *Journal of Time Series Analysis*, **2014**, 35 (2): 133–150.
- [22] Galeano P, Peña D. Covariance changes detection in multivariate time series. *Journal of Statistical Planning and Inference*, **2007**, 137 (1): 194–211.
- [23] Jirak M. Uniform change point tests in high dimension. *The Annals of Statistics*, **2015**, 43 (6): 2451–2483.
- [24] Cho H, Fryzlewicz P. Multiple change point detection under serial dependence: Wild contrast maximisation and gappy Schwarz algorithm. *Journal of Time Series Analysis*, **2024**, 45 (3): 479–494.
- [25] Cho H, Fryzlewicz P. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2015**, 77 (2): 475–507.
- [26] Wang T, Samworth R J. High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2018**, 80 (1): 57–83.
- [27] Bai P, Safikhani A, Michailidis G. Multiple change point detection in reduced rank high dimensional vector autoregressive models. *Journal of the American Statistical Association*, **2023**, 118 (544): 2776–2792.
- [28] Gombay E. Change detection in autoregressive time series. *Journal of Multivariate Analysis*, **2008**, 99 (3): 451–464.
- [29] Safikhani A, Shojaie A. Joint structural break detection and parameter estimation in high-dimensional nonstationary VAR models. *Journal of the American Statistical Association*, **2022**, 117 (537): 251–264.
- [30] Dette H, Pan G, Yang Q. Estimating a change point in a sequence of very high-dimensional covariance matrices. *Journal of the American Statistical Association*, **2022**, 117 (537): 444–454.
- [31] Enikeeva F, Harchaoui Z. High-dimensional change-point detection under sparse alternatives. *Annals of Statistics*, **2019**, 47 (4): 2051–2079.
- [32] Stransky N, Vallot C, Reyat F, et al. Regional copy number-independent deregulation of transcription in cancer. *Nature Genetics*, **2006**, 38 (12): 1386–1396.
- [33] Fryzlewicz P. Wild binary segmentation for multiple change-point detection. *Annals of Statistics*, **2014**, 42 (6): 2243–2281.
- [34] Fryzlewicz P. Detecting possibly frequent change-points: Wild Binary Segmentation 2 and steepest-drop model selection. *Journal of the Korean Statistical Society*, **2020**, 49 (4): 1027–1070.

Appendix

This appendix presents additional performance results for dimension reduction and change point detection that were not included in Section 3.1, as shown in Figs. A1–A12. The model settings remain consistent with those described in the main text: For each $p \in \{100, 200, 300, 400\}$, a sequence of $n = 200$ observations is generated from the VMA model, with $\mu_1 = \mathbf{0}$. We choose

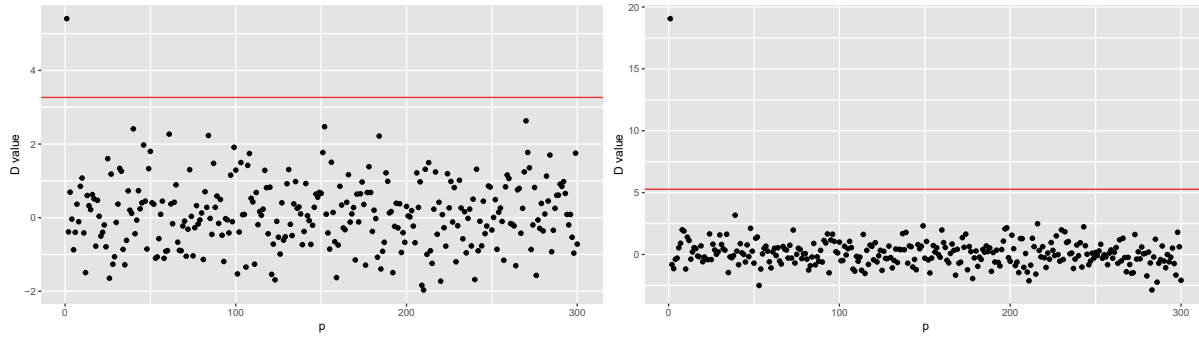


Fig. A1. Dimension reduction results ($B_1 = 0.8I_p$, $p = 300$).

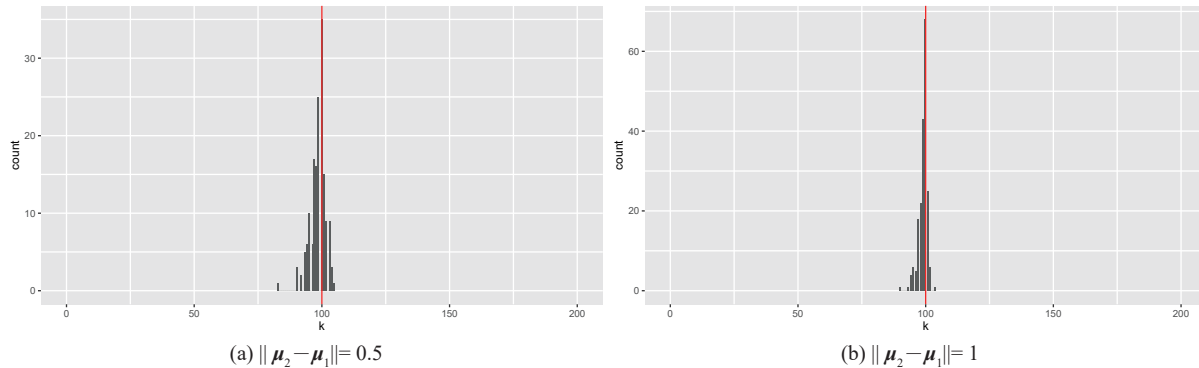


Fig. A2. Change point detection results ($B_1 = 0.8I_p$, $p = 300$).

$\mu_2 = (0.5, 0, \dots, 0)^T$ or $\mu_2 = (1, 0, \dots, 0)^T$. And B_1 is $0.8I_p$ or a random matrix satisfies the invertibility condition of the VMA model, and $k_0/n = 0.5$. We conduct 200 simulations for each setting.

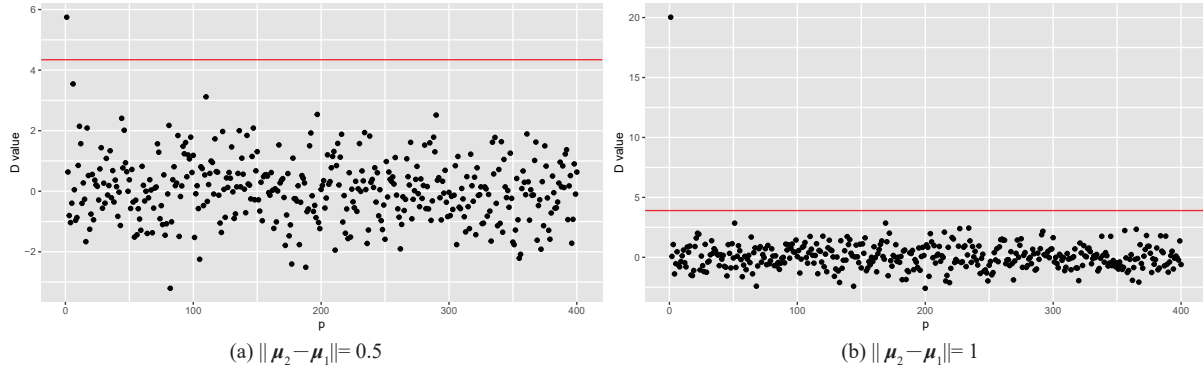


Fig. A3. Dimension reduction results ($B_1 = 0.8I$, $p = 400$). The red line represents the threshold value.

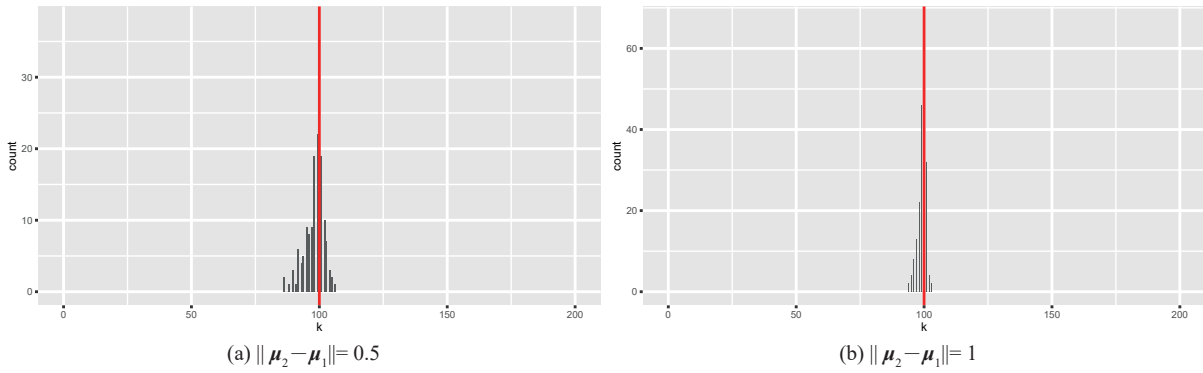


Fig. A4. Change point detection results ($B_1 = 0.8I_p$, $p = 400$).

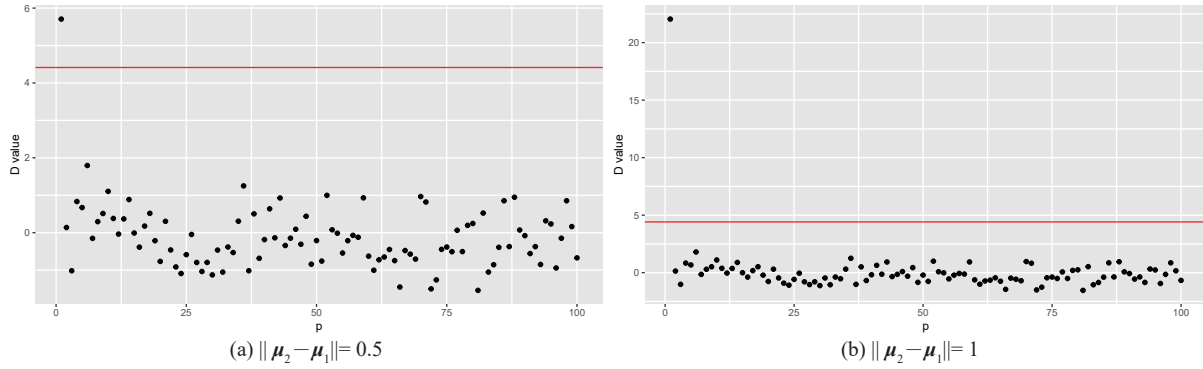


Fig. A5. Dimension reduction results (stationary B_1 , $p = 100$). The red line represents the threshold value.

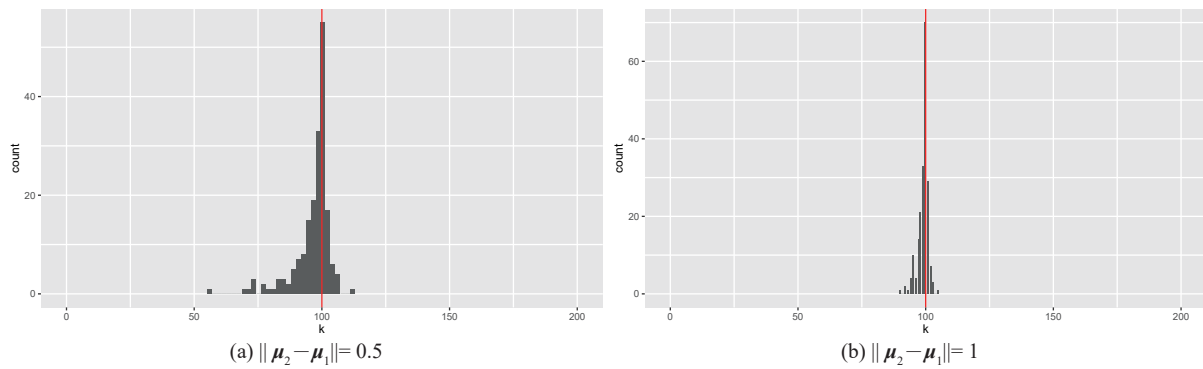


Fig. A6. Change point detection results (stationary B_1 , $p = 100$).

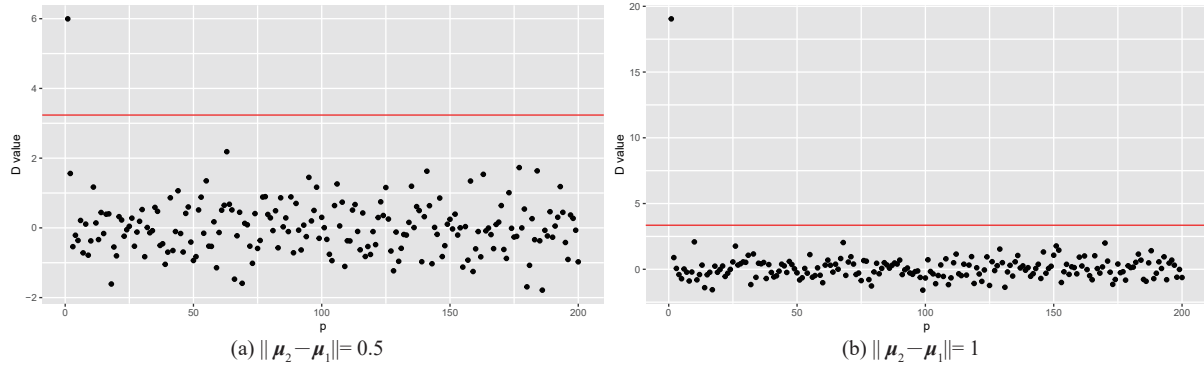


Fig. A7. Dimension reduction results (stationary B_1 , $p = 200$). The red line represents the threshold value.

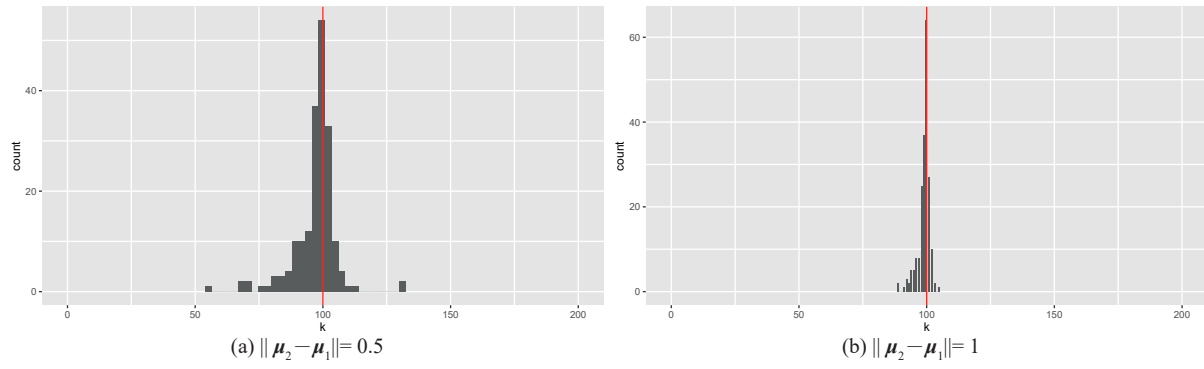


Fig. A8. Change point detection results (stationary B_1 , $p = 200$).

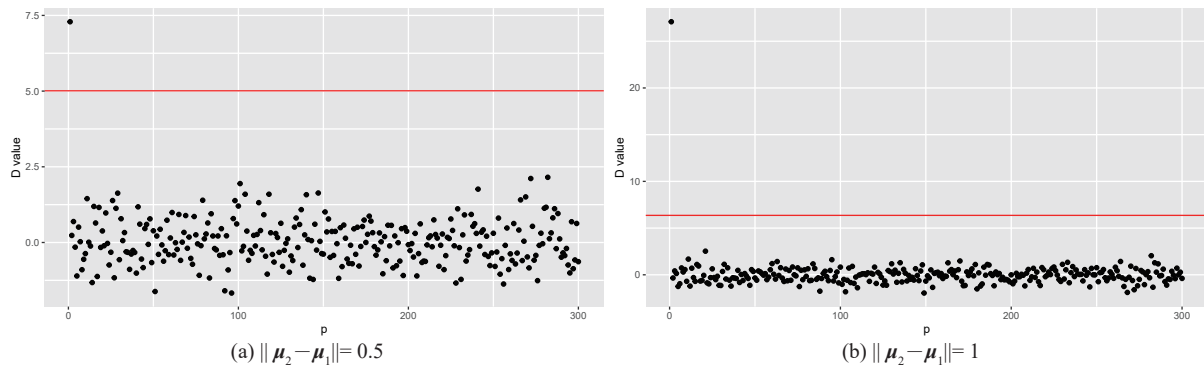


Fig. A9. Dimension reduction results (stationary B_1 , $p = 300$).

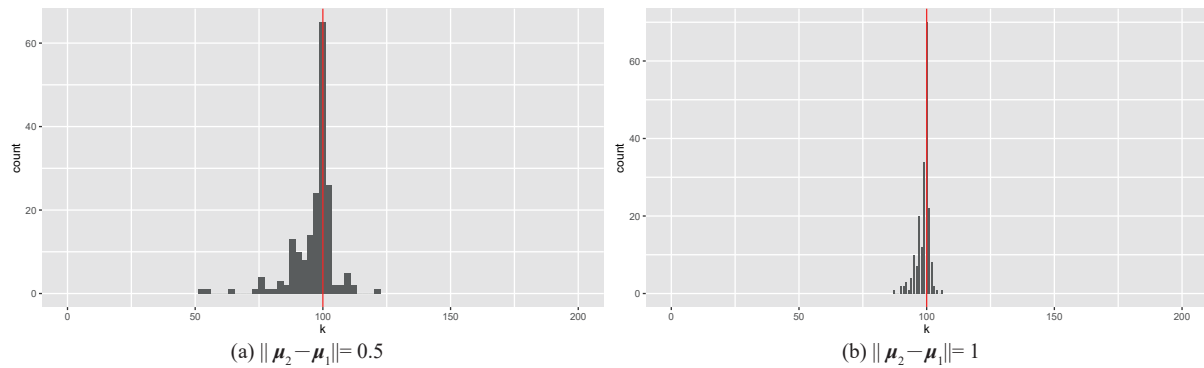


Fig. A10. Change point detection results (stationary B_1 , $p = 300$).

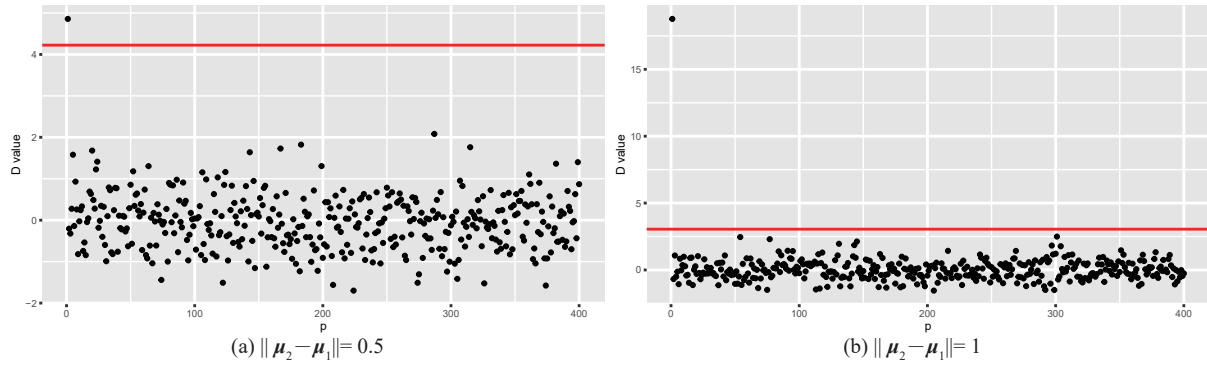


Fig. A11. Dimension reduction results (stationary B_1 , $p = 400$).

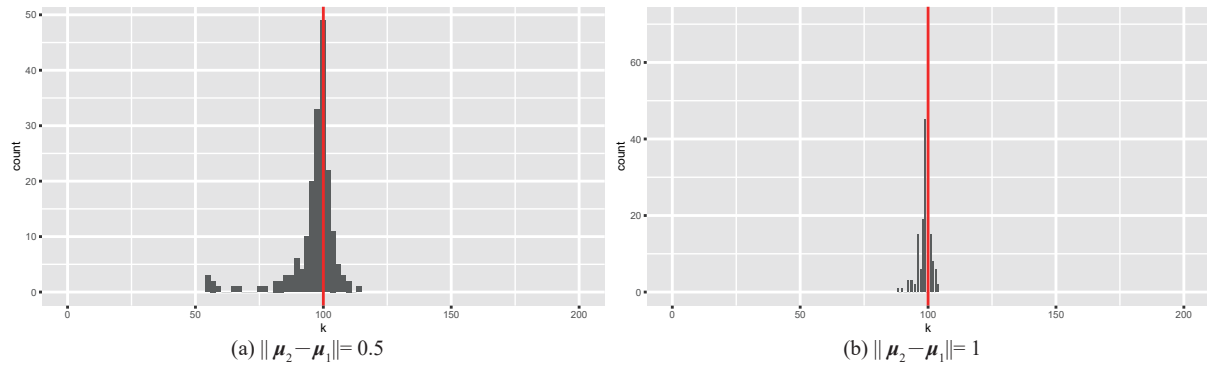


Fig. A12. Change point detection results (stationary B_1 , $p = 400$).