Robust Autocovariance Change Point Analysis for Heavy-tailed Time Series

by

Kaiwen Han

(Under the Direction of Yuan Ke)

Abstract

This dissertation studies autocovariance change point problems in heavy-tailed and high-dimensional time series under both offline and online scenarios. In the first project, we consider the offline multiple autocovariance change point detection problems in high-dimensional and heavy-tailed time series. First, we introduce an element-wise truncated autocovariance estimator for high dimensional and nonstationary time series. Next, we introduce a moving sum statistic and a binary segmentation segmentation algorithm to detect the number and locations of change points. Detection consistency is guaranteed under mild moments, dependence and signal-to-noise ratio conditions. Simulation study demonstrates superior performance of the proposed approach. The second project of this dissertation involves autocovariance change point problems in a online manner. Besides the element-wise truncated estimator, we introduce a spectrum-wise truncated estimator of which the nonasymptotic property is provided. Next, we construct CUSUM-type statistics with the two estimators to run through the data sequence as new observation arrives concurrently and detect the change point as soon as it occurs. We show that, under mild moments,

dependence and signal-to-noise ratio conditions, with appropriate threshold of certain order, false alarm rate of the scheme can be controlled, and detection delay is upper bounded with high probability. We introduce a more efficient algorithm with linear computational cost, which preserve the same theoretical guarantees as before. In addition, we study the delay and provide a minimax lower bound from independent Gaussian setting. The proposed online approach is evaluated by two experiments in the end. The last project is concerned about application of proposed change point detection approach to pandemic time series data. We conduct retrospective analysis with offline method and identify several change points which might relate to several important events over the course of Covid-19 pandemic. Besides that, we apply the online method in the monitoring case and provide a hybrid model with the assistance of change points to improve the forecasting. Results shows the proposed model moderately improves the accuracy and dramatically boosts the efficiency.

INDEX WORDS: autocovariance change point, offline detection, online detection, heavy-tailed time series

Robust Autocovariance Change Point Analysis for Heavy-tailed Time Series

by

Kaiwen Han

B.S., University of Minnesota Duluth, 2018

A Dissertation Submitted to the Graduate Faculty of the

University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Doctor of Philosophy

Athens, Georgia

2023

Robust Autocovariance Change Point Analysis for Heavy-tailed Time Series

by

Kaiwen Han

Major Professor: Yuan Ke

Committee: Pengsheng Ji

Liang Liu

Ting Zhang

Electronic Version Approved:

Ron Walcott

Vice Provost for Graduate Education and Dean of the Graduate School

The University of Georgia

May 2023

Acknowledgments

My research work could not have been done without help and support of several people. I would like to express my sincere gratitude to my advisor Professor Yuan Ke for his continuous support and patience throughout my Ph.D. journey. He has set an excellent example for me with his dedication and knowledge. It is fortunate for me to have him as my advisor. Many thanks to fellow graduate students who provided constructive comments to my work. Last but not least, I would like to thank the faculty members in the Department of Statistics. Their instruction has guided me in my first several years of course study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Over the past decade, time series analysis has become increasingly important and been widely applied in various aspects of science and industry. Time series is a sequence of observations taken in time order, measuring the behavior of the subject over time. Due to some internal or external events, the property of the series may abruptly change. Change point detection (CPD) is a common task in time series analysis and signal processing. It is to identify the times when the underlying structure of the observed data changes. The fundamental assumption of CPD is that the observed process is stationary between the change points (Truong et al., 2018).

The discovered change point may signal valuable retrospective information for researchers to analyze or raise alarm on the monitored system for timely reaction. CPD methods have extensive applications in various domains. Some examples are shown as follow.

Finance. CPD has been widely applied in financial market time series such as stock and foreign exchange market (Lynch and Mestel, 2019; Liu et al., 2010; Takayasu, 2015).

Climate Analysis. CPD methods are applied to find discontinuities in the climatological records such as temperature and precipitation (Reeves et al., 2007; Ducré-Robitaille et al., 2003; Itoh and Kurths, 2010).

Speech Recognition. Change points are detected for online environment learning in automatic speech recognition and audio segmentation, which is an essential preprocessing step in audio processing applications (Rybach et al., 2009; Harchaoui et al., 2009; Chowdhury et al., 2012).

Medical Analysis. Online medical condition monitoring utilize CPD to evaluate the patient's health. For examples, heart rate and electroencephalogram (EEG) monitoring (Ping et al., 2006; Gao et al., 2018).

Brain image analysis Functional magnetic resonance imaging (fMRI) is now a well-established technique for studying the brain. CPD can help infer connectivities in the brain and model the onset times and durations of underlying psychological activity when they are unknown(Lindquist et al., 2007; "Detecting changes in the covariance structure of functional time series with application to fMRI data", 2021).

In general, change point detection methods can be divided into two branches: 'online' and 'offline'. *Offline* methods focus on identifying the location and the number of change points in a retrospective manner when the entire series is available. The change points fond often contain meaningful information and represent some events of interest, which are valuable for practitioners to interpret and investigate. On the other hand, *Online* techniques are applied in a streaming setting where they run concurrently with each data point acquired. The objective is to detect the change point as soon as it occurs for prompt reaction while keeping the chance of falsely triggering alerts as small as possible.

In the next two sections, we will briefly introduce a number of popular offline and online CPD methods in the last few decades.

1.2 Introduction to offline change point detection

In this section, we review important works in the literature regarding change point detection in the offline setting. Consider a $\mathbb{R}^d$-valued random process $\{\boldsymbol{Y}_t\}_{t=1}^T$ where $d \geq 1$. We say it is piecewise stationary if the statistical property of the series change at some unknown times $t_0 < t_1 < t_2 < \ldots t_K < t_{K+1}$, where $t_0 = 1$ and $t_{K+1} = T$. Consequently, those K change points will divide the sequence into $K + 1$ segments, with the ith segment $Y_{(t_{i-1}+1):t_i}$. The offline change point detection consists in estimating the locations t_i, where $i = 1, 2, \ldots, K$, and the number of change points K, if it is unknown.

In the context of model selection, the task of identifying multiple change points can be viewed as determining the best segmentation $\mathcal{T}$ according to some criterion $V(\mathcal{T})$ that needs to be minimized (Truong et al., 2018). Here the offline detection problem can be framed as the following discrete optimization problem

$$\min_{\mathcal{T}} V(\mathcal{T}, Y) + P(\mathcal{T}), \tag{1.2.1}$$

where $P(\mathcal{T})$ can regarded as a penalty on the number of change points $|\mathcal{T}| = K$, measuring the complexity of a segmentation, and it equals zero when K is known. $V(\mathcal{T})$ can be defined as the sum of costs for a particular segmentation

$$V(\mathcal{T}, y) = \sum_{i=0}^{K} \mathcal{C}(Y_{(t_{i-1}+1):t_i}),$$

where K is the number of change points, and $\mathcal{C}(\cdot)$ is a cost function measuring the goodness-of-fit of each segment.

In terms of cost function, negative log likelihood is a common choice, especially for the distributional change in the normally distributed data (Horvath, 1993; Pein et al., 2017; Keshavarz et al., 2018). When detecting the change in mean, it is ubiquitous to use the quadratic loss function (Sen and Srivastava, 1975; Yao and Au., 1989; Lavielle and Moulines, 2000). It is equivalent to minimise the negative log likelihood when assuming $i.i.d$ Gaussian with unchanged variance. Due to the fact that squared-error loss suffers from outliers, Fearnhead and Rigaill, 2019 adapt the idea of robust estimation and propose to use absolute error loss, Huber loss and bi-weight loss to estimate the change points in the presence of outliers.

The cumulative sum (CUSUM) has been widely adopted in offline change point detection. In general it can also be viewed as a way to construct the objective function $V(\mathcal{T}, Y)$ in 1.2.1 although it is not often connected with cost function. Many efforts have been made on this to deal with both first and second order change. Inclan and Tiao, 1994 use cumulative sums of squares to identify multiple changes in variance in the sequence of independent variables. J. Chen and Gupta, 1997 construct a CUSUM-type statistic based on Schwarz information criterion and study the variance change in the sequence of independent Gaussian variables. Cho, 2016a propose the double CUSUM statistic which aggregates the signals in multivariate series to detect mean changes in panel data. Jirak, 2015 introduced coordinate-wise CUSUM statistic for testing mean in high dimensional dependent time series data. Zhang and Lavitas, 2018 propose an unsupervised self-normalized test statistic for testing changes in mean and and other quantites such as median. They use a function of CUSUM process as self-normalizer to avoid the direct estimation of asymptotic variance.

Moreover, non-parametric methods have appeared in the literature in the sense that they do not impose any parametric assumptions on the distribution of the series. Zou et al., 2014) develop a non-parametric multiple change point detection procedure based on the empirical cumulative distribution. Lung-Yut-Fong et al., 2015 propose rank-based homogeneity test statistic to compare multiple samples and apply it on the change point detection problem. S. Li et al., 2015a adopt the idea of B-test and construct two computationally efficient kernel-based M-statistic using the maximum mean discrepancy to compare the difference between the right window and re-sampled left window. Shi et al., 2017 consider the change point in distribution based on the minimal spanning trees and design a Bayesian-type running statistic with the shortest Hamiltonian path.

Generally speaking, there are two major approaches to solve the minimization problem 1.2.1, exact search and approximate search. Exact search aims to find the optimal solution of problem 1.2.1. A naive example is to find the exact solution by exhaustively searching all possible segmentations. Thanks to the dynamic programming (Bellman & Dreyfus, 1962), several exact search algorithms have been designed over the past decade. Jackson et al., 2005 introduce the Optimal Partitioning (OP) which first considers the cost for the last segment according to the last change point and then minimizes the cost of the segmentation given the last change point. The time complexity of this recursion is or order $\mathcal{O}(T^2)$. Killick et al., 2012 propose the Pruned Exact Linear Time (PELT) method which increase the computational efficiency by introducing a pruning step within the OP algorithm. The expected CPU cost of PELT is shown to be bounded above by LT for some constant $L < \infty$ under certain assumptions while the optimality of the solution remains unchanged. Functional Pruning Optimal partitioning (FPOP) (Maidstone et al., 2017) is another updated version of OP, which uses functional pruning to reduce the computational cost and is shown to prune more and therefore works more efficiently than PELT.

When the computational complexity of exact search methods for detecting change points is too high for a particular application, approximate methods can be used instead.

Binary segmentation (BinSeg) (Scott & Knott, 1974) is arguably the most popular search algorithm in the change point literature. The idea is to iteratively apply single change point detection method on sub-sequences. BinSeg first searches for the change point with the largest signal. The series is then split into two sub-sequences, and the search is repeated on the new sequences. This procedure will continue until certain stopping criterion is met. The computational cost of BinSeg is of order $\mathcal{O}(T \log T)$. Thanks to its low complexity and well-established theory, BinSeg has been widely adopted in the detection procedure (Bai, 1997; Olshen et al., 2004; Niu and Zhang, 2012; Cho, 2016a; Jirak, 2015 ; M. Yu and Chen, 2021; Wang et al., 2020). However, BinSeg may not always be as accurate or reliable as optimal methods. Jandhyala et al., 2013 argues that the change points tends to be inaccurately identified in the presence of short spacing between them. The following approach can mitigate this issue.

A well-known extension to BinSeg is Wild Binary Segmentation algorithm (WBS) (Fryzlewicz, 2014). In short, the single change point detection in each iteration is performed on a number of random sub-intervals of which the starting and ending indices are drawn (independently with replacement) uniformly from the the parent interval. WBS is shown to be effective when the change points are close and jump magnitude of the signal is small.

In the context of the machine learning, one possible way to approach the offline CPD is to frame it as a binary classification problem, where one class represents all possible change points and the other class includes all sequences in-between the change points. Some interpretable supervised models such as support vector machines and logistic regression (KD et al., 2015) can help address this problem if informative features are available. However, such a learning problem will become more complicated if a

large number of possible types of change point exist and may also suffer from class imbalance problem (Cook & Krishnan, 2015). From a different perspective, the problem of change point detection can be considered as a clustering problem with a unknown number of clusters, such that observations within clusters are identically distributed. One famous clustering approach used for change point detection combines sliding window and bottom up methods into an algorithm called SWAB (Sliding Window and Bottom-up) (Keogh et al., 2001).

1.3 Introduction to online change point detection

Most of the change point literature is centered around the offline or *posterior* analysis on the observed data sequence, as introduced in the preceding section. Despite its long history, online change point detection algorithms has grown in popularity over the last decade and has been applied in many situations due to the dramatic development in technology. Online or sequential change point analysis involves monitoring the sequence as each data point arrives successively and conducting quickest detection once the underlying statistical structure of the monitored sequence has changed. A false alarm occurs when a change point is identified before it actually happens. The goal of sequential change point detection is to detect the change point as soon as it occurs while keeping the probability of raising a false alarm small. In this section, we review important works in online methods.

The earliest research on online or sequential change point detection can trace back to Shewhart's control chart (Shewart, 1931) and Page's cumulative sum or CUSUM (Page, 1954). They were originally designed from the standpoint of quality control. Typically, at a production line, people tend to observe the output and assume that a particular feature fluctuates within a specific control limit. However, there

are instances where this feature unexpectedly starts to fluctuate outside of that control limit, possibly due to a malfunction of the production equipment.

CUSUM is arguably the most prominent and established statistic in the online change point literature and many variants or extensions have been developed. Suppose we are monitoring a sequence of $i.i.d$ univariate Gaussian variables $\{X_t\}$ with mean μ and unit variance and see if the mean changes or not. We define the partial sum $S(s,t) = \sum_{i=s}^{t}(x_i - \mu)$. Based on that, Kirch and Weber, 2018 pointed out several commonly used statistics.

$$\begin{aligned}
\text{CUSUM} \quad & \frac{1}{\sqrt{n}} S(0,t) \\
\text{Page's CUSUM} \quad & \frac{1}{\sqrt{n}} max_{1 \le s \le t-1} S(s,t) \\
\text{MOSUM} \quad & \frac{1}{\sqrt{w}} S(t-w,t) \\
\text{mMOSUM} \quad & \frac{1}{\sqrt{\lfloor kt \rfloor}} S(t-\lfloor kt \rfloor, t)
\end{aligned}$$

MOSUM was first considered in Eiauer and Hackl, 1978 and studied in Horváth et al., 2008, Aue et al., 2012 and Avanesov and Buzun, 2021. Its modified version mMOSUM was proposed by Z. Chen and Tian, 2010. The scaler is used to standardize the variance of $S()$. Usually we monitor the statistic value of data stream and declare a change point if it exceeds some pre-determined threshold. The threshold is often a function of the length of observed sequence so far.

Mean change is a fundamental topic that researchers focus heavily on. Mei, 2010 focuses on mean change in multiple data streams by looking at the sum of local CUSUM statistics. Aue et al., 2012 develop

the theory the MOSUM statistic for monitoring univariate mean change and focus on the limiting distribution for the delay time. Y. Yu et al., 2020 study the CUSUM-type statistic for univariate mean change of which the detection delay is nearly minimax optimal while controlling the false alarm rate. Romano et al., 2021 propose a efficient online algorithm using functional pruning CUSUM statistics of which the expected time complexity is $\mathcal{O}(\log T)$. Gösmann et al., 2022 consider the weighted CUSUM-type statistic to learn mean change in high-dimensional time series with temporal and spatial dependence. Y. Chen et al., 2022 study the change in mean of high-dimensional Gaussian data by utlizing likelihood ratio test and aggregating statistics across different scales and coordinates.

In contrast to the extensive research on mean change, second-order change is uncommonly studied in the literature. Choi et al., 2008 propose non-parametric spectral-based methods to identify change in the autocorrelation structure by looking at the change in the Fourier or wavelet-based spectrum. Avanesov and Buzun, 2021 construct a MOSUM statistic for testing change in covariance matrix and propose a non-standard bootstrap scheme for selecting threshold. L. Li and Li, 2019 propose a stopping rule for change in covariance structure of M-dependent time series and derive explicit expression for detection delay.

When monitoring a data sequence, which characteristics has shifted is in fact unknown to us. This will bring issues if we are not specifically aware of how to deal with the data at hand. Fortunately, many approaches have been developed that can monitor more than one characteristic or general distributional change. Kawahara and Sugiyama, 2009 provided a detection algorithm based on direct density-ratio estimation that can be computed very efficiently in an online manner. S. Li et al., 2015b constructed a B-statistic based on maximum mean discrepancy that can identify change points in various settings. H. Chen and Zhang, 2015 developed graph-based methods to measure the similarity between the left and

right windows around the candidate change points. H. Chen, 2019 propose a two sample test based on k-nearest-neighborhoods. They both can be applied to sequences of multivariate observations. Bayesian Online Change Point Detection (BOCPD) proposed by Adams and MacKay, 2007 utilize a recursive message passing algorithm to evaluate the posterior probability of run length given the data sequence. Saatchi et al., 2010 use the BOCPD algorithm in Gaussian process and allow the algorithm to learn the hyper parameters which have to be manually chosen before.

Other than the common time series, sequential change point detection has been adopted in many other models in the recent year of study. Chu et al., 1996 develop the CUSUM statistic of recursive residuals and the parameter fluctuation to detect change of parameters in linear regression. Marangoni-Simonsen and Xie, 2015 develop three algorithms for identifying the emergence of a community in large networks. Keshavarz et al., 2020 focus on the shift in the precision matrix of high-dimensional sparse Gaussian graphical models in the sequential setting. Non-parametric approaches such as graph-based methods also appeared in recent years of study. Dubey et al., 2021 consider online change point detection in dynamic networks with missing values.

1.4 Organization of this dissertation

The flow of this proposal is as follow. In this chapter, we introduce the background of change point detection and review existing research work in the literature. In chapter 2, we proose an offline autocovariance change point detection approach. We construct a moving sum statistic with element-wise truncated autocovariance estimator and apply binary segmentation to detect the number and locations of change points. The consistency is shown under mild moments, dependence and signal-to-noise ratio condition.

In chapter 3, we consider the online monitoring scenario and propose CUSUM-type statistic with both element-wise truncated autocovariance and spectrum-wise truncated estimators. In addition, chapter 4, pandemic time series data related to Covid-19 mortality is investigated using both offline and online techniques. We conduct retrospective analysis on state-level mortality data and study detected change points. Under online monitoring setting, we use change points in a sequential manner to facilitate the forecasting of time series.

1.5 Notation

Let $\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{R}$ denote the set of integers, natural numbers and real numbers, respectively. Let $|S|$ denotes the cardinality of a set S. For $n_1, n_2 \in \mathbb{N}$ and $n_1 \leq n_2$, we denote $[n_1, n_2] = \{n_1, n_1 + 1, \ldots, n_2\}$. The superscript $^\intercal$ denotes the transpose of a matrix or a vector. Given a vector $\boldsymbol{x} = (x_1, \ldots, x_d)^\intercal \in \mathbb{R}^d$, we write the vector l_q-norm as $|\boldsymbol{x}|_q = \big(\sum_{j=1}^d |x_j|^q\big)^{1/q}$ for $1 \leq q < \infty$ and the vector l_∞-norm as $|\boldsymbol{x}|_\infty = \max_{j\in[d]} |x_j|$. Let $\mathbb{S}^{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d : |\boldsymbol{x}|_2 = 1\}$ denotes the d-dimensional unit sphere. Given two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, we write the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{j=1}^d x_j y_j$. Given a matrix $\mathbf{A} = \big(A_{kl}\big)_{k\in[d_1];l\in[d_2]} \in \mathbb{R}^{d_1\times d_2}$, if $d_1 = d_2 = d$, $\mathrm{tr}(\mathbf{A})$ and $\det(\mathbf{A})$ denote the trace and the determinant of $\mathbf{A}$, respectively. If $\mathbf{A}$ is a symmetric matrix, $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote the largest and smallest eigenvalues of $\mathbf{A}$, respectively. The spectral-norm, Frobenius-norm, 1-norm, ∞-norm and max-norm of $\mathbf{A}$ are respectively $\|\mathbf{A}\|_{\mathrm{F}} = \sqrt{\mathrm{tr}(\mathbf{A}^\intercal\mathbf{A})}$, $\|\mathbf{A}\|_1 = \max_l \sum_{k=1}^d |A_{kl}|$, $\|\mathbf{A}\|_\infty = \max_k \sum_{l=1}^d |A_{kl}|$ and $\|\mathbf{A}\|_{\max} = \max_{k,l} |A_{kl}|$. For a sequence of matrices $\{\mathbf{A}_i\}_{i\in S}$ with $S \subseteq \mathbb{Z}$, we write $A_{i,(kl)}$ as the (k, l)-th entry of matrix $\mathbf{A}_i$. Let $\mathbf{I}_d$ denote the d-dimensional identity matrix. For an $\mathbb{R}$-valued random variable X with mean μ and variance σ^2, let $\mathrm{kurt}(X) = \mathbb{E}[(X - \mu)^4]/\sigma^4$ be the kurtosis. For $q > 0$,

we write the L_q-norm of X as $\|X\|_q = \left(\mathbb{E}[|X|^q]\right)^{1/q}$. For $a \in \mathbb{R}$, let $\lfloor a \rfloor = \max\{z \in \mathbb{Z}, z \leq a\}$ and $\lceil a \rceil = \min\{z \in \mathbb{Z}, z \geq a\}$. For $a, b \in \mathbb{R}$, let $\mathrm{sign}(a)$ be the sign of a, and denote $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. For two positive values a and b, we write $a \asymp b$ (resp. $a \lesssim b$) if there exists a positive constant C such that $C^{-1} \leq a/b \leq C$ (resp. $a/b \leq C$). Let $C, C_1, C_2, \ldots$ be positive absolute constants which may be different in each place.

Chapter 2

Autocovariance change point analysis for high-dimensional time series

We establish a framework to study multiple autocovariance change-points problems in high-dimensional, piece-wise stationary, and heavy-tailed time series. First, we propose an element-wise truncated autocovariance estimator for high dimensional and nonstationary time series. We prove the estimator enjoys nice nonasymptptic and asymptotic properties when the time series data exhibits nonlinear temporal dependency and heavy-tailedness. Next, we introduce a moving sum statistic and a binary segmentation algorithm to consistently detect the number and locations of autocovariance change-points in high-dimensional time series. The detection threshold in the algorithm is selected by a block-wise Gaussian multiplier bootstrap method. Further, we study the inference for the existence of a change-point around a pre-specified location and false discovery rate control for multiple autocovariance change-points detection.

In Section 2.1, we illustrate the setup for multiple autocovariance change-points problems. In Section 2.2, we introduce a tail-robust autocovariance matrix estimation method for nonstationary time series and study its properties. In Section 2.3, we propose a multiple autocovariance change-points detection algorithm and a data-driven threshold selection strategy. In Section 2.4, we discuss several inference problems with asymp- totic analysis. In Sections 2.5 and 2.6, we use simulation experiments to assess the empirical performance of the proposed methods.

2.1 Problem Setup

We define $\{\boldsymbol{Y}_t\}_{t=1}^T$ as an $\mathbb{R}^d$-valued time series of the following form

$$\boldsymbol{Y}_t = G_t(\mathcal{F}_t), \tag{2.1.1}$$

where $G_t(\cdot) = \big(g_{t1}(\cdot), g_{t2}(\cdot), \ldots, g_{td}(\cdot)\big)^\intercal$ is an $\mathbb{R}^d$-valued measurable function, and $\mathcal{F}_t = \sigma(\ldots, \epsilon_{t-1}, \epsilon_t)$ is a filtration with $\{\epsilon_t\}_{t\in\mathbb{Z}}$ being a sequence of i.i.d. random variables. In this paper, we allow $\{\boldsymbol{Y}_t\}_{t=1}^T$ to follow various heavy-tailed distributions. Besides, we would like to emphasize that the function G_t is time-dependent, thus the representation (3.1.1) covers a large amount of nonstationary time series models, including the second-order piece-wise stationary model that we will study in this paper. The representation (3.1.1) can also be regarded as a generalization of the stationary process $\widetilde{\boldsymbol{Y}}_t = G(\mathcal{F}_t)$ with the measurable function $G(\cdot)$ being independent of time.

Let $\{\epsilon_{t-k+1}, \ldots, \epsilon_t\} = \emptyset$ when $k \leq 0$. We define $\boldsymbol{Y}_{t,\{t-k\}} = G_t(\mathcal{F}_{t,\{t-k\}})$ as a coupled version of $\boldsymbol{Y}_t$, where $\mathcal{F}_{t,\{t-k\}} = \sigma(\ldots, \epsilon_{t-k-1}, \epsilon_{t-k}^{'}, \epsilon_{t-k+1}, \ldots, \epsilon_t)$ and $\epsilon_{t-k}^{'}$ is an i.i.d. copy of ϵ_{t-k}. To measure the dependence of $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$, we define the functional dependence measure in Wu and Zhou, 2011 for a

nonstationary process as

$$\delta_{k,q,j} = \sup_t \|Y_{t,j} - Y_{t,\{t-k\},j}\|_q, \text{ for } k \geq 0,\ q \geq 1, \text{ and } j \in [1, d],$$

where $Y_{t,j}$ is the j-th component of $\boldsymbol{Y}_t$. As a generalization of the classical functional dependence measure **Wu2** for stationary time series, $\delta_{k,q,j}$ uniformly quantifies the lag-k dependence by the moment of order q. As the lag k increases, $\delta_{k,q,j}$ would decrease in general, and we will further impose dependence conditions by restricting the decay rates of $\delta_{k,q,j}$ with respective to k.

Next, we introduce a general second-order piece-wise stationary model for $\{\boldsymbol{Y}_t\}_{t=1}^T$. For the simplicity of presentation, we assume $\mathbb{E}[\boldsymbol{Y}_t] = \mathbf{0}$ for $1 \leq t \leq T$. Let $K \in \mathbb{N}$ be the number of underlying change-points, and the change-points $\{t_k\}_{k=1}^K$ satisfy

$$1 = t_0 < t_1 < t_2 < \cdots < t_K < t_{K+1} = T.$$

Here, t_0 and t_{K+1} are defined only for notational convenience. For a nonnegative integer ℓ, we define the lag-ℓ autocovariance matrices as $\boldsymbol{\Sigma}_\ell(t) \equiv \mathbb{E}[\boldsymbol{Y}_{t-\ell}\boldsymbol{Y}_t^\intercal] \in \mathbb{R}^{d\times d}$. Note that for any $\ell \geq 0$, $k \in [1, K+1]$ and $t \in [t_{k-1} + \ell, t_k]$, we have $\boldsymbol{\Sigma}_\ell(t) = \boldsymbol{\Sigma}_{-\ell}^\intercal(t-\ell)$. Without loss of generality, we only consider $\boldsymbol{\Sigma}_\ell(t)$ with $\ell \geq 0$ throughout this paper. We say $\{\boldsymbol{Y}_t\}_{t=1}^T$ is second-order piece-wise stationary, if the following three conditions are satisfied:

(a) $\{\boldsymbol{Y}_t\}_{t=t_{k-1}+1}^{t_k}$ is a stationary process, for any $k \in [1, K+1]$;

(b) $\boldsymbol{\Sigma}_\ell(t)$ depends on t such that

$$\boldsymbol{\Sigma}_\ell(t) = \begin{cases} \boldsymbol{\Sigma}_\ell^1, & t \in [1+\ell, t_1], \\ \boldsymbol{\Sigma}_\ell^2, & t \in [t_1+1+\ell, t_2], \\ \ldots, & \\ \boldsymbol{\Sigma}_\ell^{K+1}, & t \in [t_K+1+\ell, T], \end{cases} \tag{2.1.2}$$

where $\boldsymbol{\Sigma}_\ell^{k-1} \neq \boldsymbol{\Sigma}_\ell^k$ for any $k \in [2, K+1]$.

2.2 Element-wise truncated estimator

Let $\{\boldsymbol{Y}_t\}_{t=1}^T$ be a centered $\mathbb{R}^d$-valued time series following the representation (3.1.1). For $\ell \geq 0$, denote $\mathcal{I}_\ell = [\ell + s, e] \subseteq [1, T]$ with $s, e \in [1, T]$ and $e - s \geq \ell$. For any $t \in \mathcal{I}_\ell$, the lag-ℓ outer product of $\boldsymbol{Y}_t$ is defined as

$$\mathbf{H}_{t,\ell} = \boldsymbol{Y}_{t-\ell}\boldsymbol{Y}_t^\intercal = \left(H_{t,\ell,(jk)}\right)_{j,k=1}^d. \tag{2.2.1}$$

Define the truncation function $\psi_\tau : \mathbb{R} \mapsto \mathbb{R}$ as

$$\psi_\tau(u) = \operatorname{sign}(u)(|u| \wedge \tau), \tag{2.2.2}$$

where $\tau > 0$ is a robustification parameter. It is easy to see that $\psi_\tau(u)$ is Lipschitz continuous and the first order derivative of the Huber loss function defined as follows

$$\ell_\tau(u) \equiv \begin{cases} u^2/2, & \text{if } |u| \le \tau \\ \tau|u| - \tau^2/2 & \text{if } |u| > \tau \end{cases}, \ u \in \mathbb{R}.$$

Using the truncation operator $\psi_\tau(\cdot)$, the truncated estimator of $\mathbb{E}[H_{t,\ell,(jk)}]$ is defined as

$$\widehat{\gamma}_{\mathcal{I}_\ell,(jk)} = \frac{1}{|\mathcal{I}_\ell|} \sum_{t \in \mathcal{I}_\ell} \psi_{\tau_\ell}(H_{t,\ell,(jk)}), \quad \text{for } 1 \le j,k \le d.$$

By collecting the element-wise truncated estimators, the tail-robust lag-ℓ autocovariance matrix estimator is formed as

$$\widehat{\mathbf{\Sigma}}_{\mathcal{I}_\ell} = \big(\widehat{\gamma}_{\mathcal{I}_\ell,(jk)}\big)_{j,k=1}^{d}. \tag{2.2.3}$$

Assumption 1. *Suppose that the marginal fourth moment of* $\{\boldsymbol{Y}_t\}_{t=1}^T$ *satisfies*

$$\omega_4 \equiv \max_{1 \le j \le d} \sup_{t \in \mathbb{Z}} \|Y_{t,j}\|_4 < \infty.$$

Assumption 2. *There exists some* $\rho \in (0,1)$*, such that the coordinate-wise dependence adjusted fourth moment of* $\{\boldsymbol{Y}_t\}_{t=1}^T$ *satisfies*

$$\|Y_\cdot\|_4 \equiv \max_{1 \le j \le d} \sup_{m \ge 0} \rho^{-m} \sum_{k=m}^{\infty} \delta_{k,4,j} < \infty.$$

Theorem 2.2.1 (Nonasymptotic error bound). *Let $\{\boldsymbol{Y}_t\}_{t\in\mathbb{Z}}$ be a centered $\mathbb{R}^d$-valued time series following* (3.1.1). *Denote the long-run variance of the lag-ℓ cross term as*

$$\nu_H^2 \equiv \max_{1\le j,k\le d} \sum_{l=-\infty}^{\infty} \sup_{t\in\mathbb{Z}} \big| \operatorname{cov}\left(\psi_{\tau_\ell}(H_{t,\ell,(jk)}), \psi_{\tau_\ell}(H_{t+\ell,\ell,(jk)})\right)\big|.$$

For any integer interval $\mathcal{I}_\ell = [\ell + s, e]$ with $|\mathcal{I}_\ell| \ge 2$ and for any $m > 0$, we choose the robustification parameter as

$$\tau_\ell = C_\tau \frac{|\mathcal{I}_\ell|^{1/2} \max\{\omega_4^2, \nu_H\}}{\log(|\mathcal{I}_\ell|)(m + \log d)^{1/2}}, \tag{2.2.4}$$

where $C_\tau > 0$ is an absolute constant. Then, under Assumptions 7 and 8, we have $\nu_H < \infty$ and the following result holds with a probability at least $1 - 2e^{-m}$,

$$\|\widehat{\boldsymbol{\Sigma}}^{\diamond}_{\mathcal{I}_\ell}\|_{\max} \lesssim \frac{\max\{\omega_4^2, \nu_H\} \log(|\mathcal{I}_\ell|)(m + \log d)^{1/2}}{|\mathcal{I}_\ell|^{1/2}},$$

where $\widehat{\boldsymbol{\Sigma}}^{\diamond}_{\mathcal{I}_\ell} \equiv \widehat{\boldsymbol{\Sigma}}_{\mathcal{I}_\ell} - |\mathcal{I}_\ell|^{-1} \sum_{t\in\mathcal{I}_\ell} \boldsymbol{\Sigma}_\ell(t)$.

2.2.1 Gaussian approximation

Theorem 2.2.1 has shown that, with a proper chosen robustification parameter τ_ℓ, the element-wise truncated autocovariance estimator possesses an exponential type error bound even for nonstationary and heavy-tailed time series. In this subsection, we study the Gaussian approximation for our element-wise truncated autocovariance estimator. To be specific, we aim to show that the limiting distribution of $|\mathcal{I}_\ell|^{1/2}\big\|\widehat{\boldsymbol{\Sigma}}^{\diamond}_{\mathcal{I}_\ell}\big\|_{\max}$ can be approximated well by the l_∞ norm of an $\mathbb{R}^{d^2}$-valued Gaussian vector

$\boldsymbol{Z}_{\mathcal{I}_\ell} \sim N(\mathbf{0}, \boldsymbol{\Gamma})$. Moreover, $\boldsymbol{\Gamma}$ can be defined as a long-run covariance matrix, i.e.

$$\boldsymbol{\Gamma} \equiv \lim_{|\mathcal{I}_\ell| \to \infty} \operatorname{var}\left(\boldsymbol{U}_{\mathcal{I}_\ell}\right), \tag{2.2.5}$$

where $\boldsymbol{U}_{\mathcal{I}_\ell}$ is the vectorization (i.e. staking the columns into a vector) of $|\mathcal{I}_\ell|^{1/2}\widehat{\boldsymbol{\Sigma}}_{\mathcal{I}_\ell}$. We give the Gaussian approximation result in Corollary 2.2.1. The proof of Corollary 2.2.1 is a direct application of Theorem **??**. Next, we introduce and discuss several assumptions before stating the corollary. Note that instead of pursuing the minimum moment condition ($\theta \in (0, 1]$) required by Theorem **??**, the following assumptions are based on $\theta = 1$ for the simplicity of presentation.

Assumption 3. *Suppose that the marginal sixth moment* $\{\boldsymbol{Y}_t\}_{t=1}^T$ *satisfies*

$$\omega_6 \equiv \max_{1 \le j \le d} \sup_{t \in \mathbb{Z}} \|Y_{t,j}\|_6 < \infty.$$

Assumption 4. *There exists some* $\rho \in (0, 1)$*, such that the coordinate-wise dependence adjusted sixth moment of* $\{\boldsymbol{Y}_t\}_{t=1}^T$ *satisfies*

$$\|Y_\cdot\|_6 \equiv \max_{1 \le j \le d} \sup_{m \ge 0} \rho^{-m} \sum_{k=m}^{\infty} \delta_{k,6,j} < \infty.$$

Assumption 5. *There exists an absolute constant* $b > 0$*, such that*

$$\min_{1 \le j,k \le d} \inf_{\mathcal{S} \subseteq [n]} \frac{1}{|\mathcal{S}|} \operatorname{var}\Big(\sum_{t \in \mathcal{S}} Y_{t-\ell,j} Y_{t,k} \Big) > b.$$

Remark 2.2.1. *Assumption 3 requires finite coordinate-wise moments up to sixth order for all dimensions of* $\{\boldsymbol{Y}_t\}_{t=1}^T$. *Assumption 4 requires an exponential decay of dependence measure for all dimensions of* $\{\boldsymbol{Y}_t\}_{t=1}^T$. *Assumptions 3 and 4 are imposed on higher order moments of* $\{\boldsymbol{Y}_t\}_{t=1}^T$ *and hence can imply Assumptions 7 and 8. Assumption 5 ensures the nondegeneracy of the partial sums of lag-*ℓ *cross products, which is a very mild condition.*

Corollary 2.2.1 (Gaussian approximation)**.** *Suppose Assumptions 3, 4 and 5 hold. Let* $C, C_\tau > 0$ *be some absolute constants. Assume that* $\log d = C|\mathcal{I}_\ell|^\beta$ *for some* $\beta < 1/19$. *Choose the robustafication parameter* $\tau = C_\tau\big[|\mathcal{I}_\ell|^{16/19}(\log d)^{-1}\big]^{1/3}$. *Then, as* $|\mathcal{I}_\ell| \to \infty$, *we have that*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\Big(|\mathcal{I}_\ell|^{1/2}\big\|\widehat{\boldsymbol{\Sigma}}^{\diamond}_{\mathcal{I}_\ell}(c)\big\|_{\max} \le t\Big) - \mathbb{P}\Big(\big|\boldsymbol{Z}_{\mathcal{I}_\ell}\big|_\infty \le t\Big)\right| \lesssim |\mathcal{I}_\ell|^{-2(1/19-\beta)/3} \to 0,$$

where $\boldsymbol{Z}_{\mathcal{I}_\ell} \sim N(\boldsymbol{0}, \boldsymbol{\Gamma})$.

The Gaussian approximation result in Corollary 2.2.1 paves the way for the inference problems to be discussed in Section **??**. Note that the asymptotic covariance matrix $\boldsymbol{\Gamma}$ is usually unknown to us, and hence need to be estimated. We propose a block-wise Gaussian multiplier bootstrap method in Section 2.5.1 to address this issue.

2.3 Tail-robust moving sum statistic

Suppose we observe $\{\boldsymbol{Y}_t\}_{t=1}^T$ as a centered $\mathbb{R}^d$-valued time series following the second-order piece-wise stationary model defined in Section **??**. Let c be a checkpoint of interest and W be a pre-specified window size, such that $1 < c - W + 1 < c < c + W < T$. Further, for a lag $\ell \ge 0$, we denote the integer

intervals $\mathcal{I}_\ell^{\mathrm{B}}(c) \equiv [\ell + c - W + 1, c]$ and $\mathcal{I}_\ell^{\mathrm{A}}(c) \equiv [\ell + c + 1, c + W]$ as two windows before and after the checkpoint c, respectively. Follow the definition in (3.2.3), we can construct element-wise truncated lag-ℓ autocovariance matrix estimators based on $\mathcal{I}_\ell^{\mathrm{B}}(c)$ and $\mathcal{I}_\ell^{\mathrm{A}}(c)$ as

$$\widehat{\mathbf{\Sigma}}_\ell^{\mathrm{B}}(c) \equiv \widehat{\mathbf{\Sigma}}_{\mathcal{I}_\ell^{\mathrm{B}}(c)} \text{ and } \widehat{\mathbf{\Sigma}}_\ell^{\mathrm{A}}(c) \equiv \widehat{\mathbf{\Sigma}}_{\mathcal{I}_\ell^{\mathrm{A}}(c)}.$$

Note that $|\mathcal{I}_\ell^{\mathrm{B}}(c)| = |\mathcal{I}_\ell^{\mathrm{A}}(c)| = W - \ell$. Next, we define a lag-$\ell$ moving sum difference matrix at checkpoint c as

$$\begin{aligned}
\mathbf{S}_\ell(c) =& [(W-\ell)/2]^{1/2}\big[\widehat{\mathbf{\Sigma}}_\ell^{\mathrm{B}}(c) - \widehat{\mathbf{\Sigma}}_\ell^{\mathrm{A}}(c)\big] \\
=& [(W-\ell)/2]^{1/2} \left\{ \Big[\widehat{\mathbf{\Sigma}}_\ell^{\mathrm{B}}(c) - \frac{1}{|\mathcal{I}_\ell^{\mathrm{B}}(c)|} \sum_{t \in \mathcal{I}_\ell^{\mathrm{B}}(c)} \mathbf{\Sigma}_\ell(t)\Big] - \Big[\widehat{\mathbf{\Sigma}}_\ell^{\mathrm{A}}(c) - \frac{1}{|\mathcal{I}_\ell^{\mathrm{A}}(c)|} \sum_{t \in \mathcal{I}_\ell^{\mathrm{A}}(c)} \mathbf{\Sigma}_\ell(t)\Big] \right\} \\
&+ [2(W-\ell)]^{-1/2}\Big[\sum_{t \in \mathcal{I}_\ell^{\mathrm{B}}(c)} \mathbf{\Sigma}_\ell(t) - \sum_{t \in \mathcal{I}_\ell^{\mathrm{A}}(c)} \mathbf{\Sigma}_\ell(t)\Big] \\
\equiv& \mathbf{S}_\ell^\diamond(c) + \mathbf{S}_\ell^*(c),
\end{aligned} \tag{2.3.1}$$

where $\mathbf{S}_\ell^*(c) = [2(W-\ell)]^{-1/2}\Big[\sum_{t=c-W+1+\ell}^{c} \mathbf{\Sigma}_\ell(t) - \sum_{t=c+1+\ell}^{c+W} \mathbf{\Sigma}_\ell(t)\Big]$ is the population counterpart of $\mathbf{S}_\ell(c)$ and $\mathbf{S}_\ell^\diamond(c) = \mathbf{S}_\ell(c) - \mathbf{S}_\ell^*(c)$. Note that $\mathbf{S}_\ell^\diamond(c)$ is a re-scaled difference of the estimation errors before and after c. The expectation of $\mathbf{S}_\ell^\diamond(c)$ is not zero in general as the element-wise truncated autocovariance estimators are biased. However, the biases can be well controlled with a properly chosen robustifcation parameter. The next corollary gives a nonasymptotic error bound for the max-norm of $\mathbf{S}_\ell^\diamond(c)$ under general heavy-tailed settings.

Corollary 2.3.1 (Max-norm of re-scaled estimation error difference matrix)**.** *For any window size W such that $W - \ell \geq 2$ and for any $m > 0$, we choose the robustification parameter as*

$$\tau_\ell = C_\tau \frac{(W-\ell)^{1/2} \max\{\omega_4^2, \nu_H\}}{\log(W-\ell)(m + \log d)^{1/2}}, \tag{2.3.2}$$

where $C_\tau > 0$ is an absolute constant. Then, under Assumptions 7 and 8, we have $\nu_H < \infty$ and the following result holds with a probability at least $1 - 2e^{-m}$,

$$\left\|\mathbf{S}_\ell^\diamond(c)\right\|_{\max} \lesssim \max\{\omega_4^2, \nu_H\} \log(W-\ell)(m + \log d)^{1/2}.$$

Provided the max-norm of $\mathbf{S}_\ell^\diamond(c)$ is well controlled, we use the max-norm of $\mathbf{S}_\ell(c)$ to construct a tail-robust lag-ℓ moving sum statistic at the checkpoint c, i.e.

$$T_\ell(c) \equiv \left\|\mathbf{S}_\ell(c)\right\|_{\max}. \tag{2.3.3}$$

Note that $T_\ell(c)$ is a non-negative quantity. When there is no change-point in $\mathcal{I}_\ell^{\mathrm{B}}(c) \cup \mathcal{I}_\ell^{\mathrm{A}}(c)$, we have $T_\ell(c) = \left\|\mathbf{S}_\ell^\diamond(c)\right\|_{\max}$. On the other hand, when there is a significant enough autocovariance change in $\mathcal{I}_\ell^{\mathrm{B}}(c) \cup \mathcal{I}_\ell^{\mathrm{A}}(c)$, we have that $T_\ell(c)$ is dominated by $\left\|\mathbf{S}_\ell^*(c)\right\|_{\max}$ which should be distinct from 0. Therefore, a large value of $T_\ell(c)$ provides some evidence that an autocovariance change occurs near c and vice versa.

2.4 Detection algorithm

In this subsection, we introduce a generic algorithm to detect all autocovariance chagnepoints that lie in the observed time series. To begin with, we consider a simple scenario by assuming there exists only one change-point for the lag-ℓ autocovariance of $\{\boldsymbol{Y}_t\}_{t=1}^T$. In other words, for a reasonable W and some t_1 satisfy $W \leq t_1 \leq T - W$, we assume that

$$\boldsymbol{\Sigma}_\ell(t) = \begin{cases} \boldsymbol{\Sigma}_\ell^1, & t \in [1+\ell, t_1], \\ \boldsymbol{\Sigma}_\ell^2, & t \in [t_1 + 1 + \ell, T]. \end{cases} \tag{2.4.1}$$

Note that (2.4.1) is a simplified version of (2.1.2) with only one change-point. It is easy to check that, $\big\|\mathbf{S}_\ell^*(c)\big\|_{\max}$ is maximized at $c = t_1$. Empirically, we slide the checkpoint c over the interval $[W, T-W]$. For each c, we construct two windows $\mathcal{I}_\ell^{\mathrm{B}}(c)$ and $\mathcal{I}_\ell^{\mathrm{A}}(c)$, and compute the moving sum statistic $T_\ell(c) = \big\|\mathbf{S}_\ell(c)\big\|_{\max}$. Let $\widehat{t}_1 = \operatorname{argmax}_c T_\ell(c)$ and $\chi > 0$ be a suitable threshold to represent the minimal signal strength. If $T_\ell\big(\widehat{t}_1\big) > \chi$, we estimate t_1 by $\widehat{t}_1$.

When there are multiple change-points, say $K > 1$ in (2.1.2), we propose a recursive segmentation procedure to detect all change-points, which is summarized in Algorithm 1 below. The algorithm starts with finding one change-point in the interval $[W, T-W]$. Let $\mathcal{A}$ be an empty set and $\widehat{t}_1 = \operatorname{argmax}_c T_\ell(c)$. If $T_\ell\big(\widehat{t}_1\big) > \chi$, we add $\widehat{t}_1$ to $\mathcal{A}$ as the first detected change-point. Then, the algorithm divides the interval $[W, T-W]$ into two sub-intervals $[W, \widehat{t}_1]$ and $[\widehat{t}_1 + 1, T - W]$. For each sub-interval, the algorithm recursively detects if there exist at least one change-point. If so, the algorithm adds the newly detected change-point to $\mathcal{A}$ and further divides the current sub-interval into two smaller ones. The algorithm

stops dividing an sub-interval if there is no change-point to be detected in it or its size is no larger than $2W$. The algorithm stops if no new change-point can be added to $\mathcal{A}$. The algorithm outputs the set $\mathcal{A}$ which contains the locations of all detected change-points.

Algorithm 1 Multiple Autocovariance Change-points Detection. MACD(s, e, W, ℓ, χ)

Input: $\{\boldsymbol{Y}_t\}_{t=s}^{e}$, a window size W, a lag $\ell \geq 0$, and a detection threshold χ.
Initialize FLAG $\leftarrow 0$, $\mathcal{A} \leftarrow \emptyset$, $s = 1$, and $e = T$.
while $e - s > 2W$ and FLAG $= 0$ **do**
    Compute $T_\ell(c) = \big\|\mathbf{S}_\ell(c)\big\|_{\max}$ for $c \in [s + W - 1, e - W]$, and find $\hat{t} \leftarrow \operatorname{argmax}_c T_\ell$.
    if $T_\ell(\hat{t}) \leq \chi$ **then**
        FLAG $\leftarrow 1$.
    else
        Add $\hat{t}$ to $\mathcal{A}$; Do MACD$(s, \hat{t} - 1, W, \ell, \chi)$ and MACD$(\hat{t}, e, W, \ell, \chi)$.
Output: $\mathcal{A}$.

Next, we analyze the properties of Algorithm 1. We define an event as follows

$$\mathcal{E}_\ell^{[1,T]}(\lambda) \equiv \Big\{ \max_{W \leq c \leq T-W} \big\|\mathbf{S}_\ell^{\diamond}(c)\big\|_{\max} \leq \lambda \Big\}. \tag{2.4.2}$$

If the event $\mathcal{E}_\ell^{[1,T]}(\lambda)$ holds, the estimation errors of autocovariance matrices are uniformly bounded by λ over all checkpoints in the interval $[W, T - W]$. The probability of the event $\mathcal{E}_\ell^{[1,T]}(\lambda)$ to hold can be discussed in low and high dimensional regimes, respectively.

I. (Fixed/low-dimensional regime: $d \lesssim T$) By Corollary 2.3.1 and let $m = \xi \log T$ for some absolute constant $\xi > 1$, we have

$$\mathbb{P}\big(\mathcal{E}_\ell^{[1,T]}(\lambda_0)\big) \leq 1 - 2T^{1-\xi},$$

where $\lambda_0 = C \max\{\omega_4^2, \nu_H\} \log(W-\ell)(\log T)^{1/2}$, for some sufficiently large absolute constant $C > 0$.

II. (High-dimensional regime: $d \gg T$) By Corollary 2.3.1 and let $m = \log d$, we have

$$\mathbb{P}\big(\mathcal{E}_\ell^{[1,T]}(\lambda_0)\big) \leq 1 - 2Td^{-1},$$

where $\lambda_0 = C \max\{\omega_4^2, \nu_H\} \log(W-\ell)(\log d)^{1/2}$, for some sufficiently large absolute constant $C > 0$.

In either regime, we can uniformly upper bound $\big\|\mathbf{S}_\ell^\diamond(c)\big\|_{\max}$ with a high probability. Next, we prove that Algorithm 1 can consistently detect the number and locations of all change-points under mild conditions on δ and $\underline{\kappa}$, as well as a properly chosen threshold χ. The detailed results are stated in the theorem below.

Theorem 2.4.1 (Detection consistency)**.** *Assume that the event $\mathcal{E}_\ell^{[1,T]}(\lambda_0)$ holds with probability at least $a_T \to 1$. Let b_T be any slowly divergent sequence. Assume that the minimal jump size $\underline{\kappa}$ and the minimal spacing δ satisfy that*

$$\underline{\kappa}^{-1} = o(b_T) \quad \textit{and} \quad \delta\underline{\kappa} \geq C_{snr} b_T \lambda_0 \sqrt{W-\ell}, \tag{2.4.3}$$

for some sufficiently large absolute constant $C_{snr} > 0$. Choose the threshold

$$\chi = C_\chi \underline{\kappa}^{-1} \lambda_0, \tag{2.4.4}$$

for some sufficiently large absolute constant $C_\chi > 3$. Let $\mathcal{A} = \{\widehat{t}_k\}_{k=1}^{\widehat{K}}$ be the locations of change-points detected by Algorithm 1. We have that

$$\mathbb{P}\Big(\widehat{K} = K \text{ and } \max_{1\le k\le K} |t_k - \widehat{t}_k| \le 2\sqrt{2}\underline{\kappa}^{-1}\lambda_0\sqrt{W-\ell}\Big) \ge a_T.$$

Remark 2.4.1. *In Theorem 2.4.1, we restrict the ratio between the maximal and minimal jump sizes by the diverging sequence b_T. Since b_T appears also in the signal-to-noise condition in* (2.4.3)*, the stronger the signal-to-noise condition one imposes, the weaker the restriction will be. Note that we allow the scenarios that some (or all) of κ_k vanishes.*

Further, the discussion of $\mathcal{E}_\ell^{[1,T]}(\lambda)$ above suggests that we can choose m and λ_0 such that $a_T \to 1$ in both low and high dimensionality regimes. Therefore, we have the following corollary.

Corollary 2.4.1. *Let $C_{snr}, C > 0$ and $C_\chi > 3$ be sufficiently large absolute constants, and b_T be any slowly divergent sequence. Under the same conditions as in Corollary 2.3.1 and Theorem 2.4.1. Assume that*

$$\delta\underline{\kappa} \ge C_{snr} b_T \log(W-\ell)\sqrt{(W-\ell)\log(\max\{T,d\})}.$$

For any window size W such that $W - \ell \ge 2$, we choose the robustification parameter and the threshold respectively as

$$\tau_\ell = C_\tau \frac{\sqrt{W-\ell}}{\log(W-\ell)\sqrt{\log(\max\{T,d\})}} \quad \text{and} \quad \chi = C_\chi \frac{\log(W-\ell)\sqrt{\log(\max\{T,d\})}}{\underline{\kappa}}.$$

Let $\mathcal{A} = \{\widehat{t}_k\}_{k=1}^{\widehat{K}}$ *be the locations of change-points detected by Algorithm 1. It holds with probability at least* $1 - (\max\{T, d\})^{-2}$ *that*

$$\widehat{K} = K \ \ \textit{and} \ \ \max_{1 \le k \le K} |t_k - \widehat{t}_k| \le C \frac{\log(W-\ell)\sqrt{(W-\ell)\log(\max\{T,d\})}}{\underline{\kappa}}.$$

Corollary 2.4.1 follows directly from Corollary 2.3.1 and Theorem 2.4.1 by letting $m = 4\log(\max\{T, d\})$. Then, it follows that $\lambda_0 = C\log(W-\ell)(\log(\max\{T,d\}))^{1/2}$ and

$$\mathbb{P}\big(\mathcal{E}_\ell^{[1,T]}(\lambda_0)\big) \ge 1 - 2T(\max\{T,d\})^{-4}.$$

2.4.1 Threshold selection by block-wise Gaussian multiplier bootstrap

The threshold χ plays an important role in Algorithm 1. If χ is chosen too large, the algorithm may miss some underlying change-points. On the other hand, if χ is chosen too small, the algorithm can commit several false discoveries. The theoretical order of the optimal choice of χ, provided in Theorem 2.4.1, involves some unknown quantities and hence may not readily guide the practice. In this subsection, we propose a data-driven threshold selection method based on a novel block-wise Gaussian multiplier bootstrap idea.

Recall that, $\mathbf{H}_{t,\ell}$ is the lag-ℓ outer products at time t defined as in (3.2.1). For a given checkpoint c, we construct two sets of lag-ℓ outer products $\{\mathbf{H}_{t,\ell}\}_{t\in\mathcal{I}_\ell^{\mathrm{B}}(c)}$ and $\{\mathbf{H}_{t,\ell}\}_{t\in\mathcal{I}_\ell^{\mathrm{A}}(c)}$ based on the intervals $\mathcal{I}_\ell^{\mathrm{B}}(c)$ and $\mathcal{I}_\ell^{\mathrm{A}}(c)$ defined in Section 2.3. Let $R \ll (W-\ell)/2$ be a positive integer. We divide each window into $2R$ blocks. For simplicity, we assume the window length is divisible and let the block size

$S = (W - \ell)/(2R)$. The total $4R$ blocks, denoted by $\mathcal{S}_1, \ldots, \mathcal{S}_{4R}$, can be expressed as

$$\mathcal{S}_r = \begin{cases} [c - W + \ell + (r-1)S, c - W + \ell + rS], & r = 1, \ldots, 2R, \\ [c - W + 2\ell + 1 + (r-1)S, c - W + 2\ell + 1 + rS], & r = 2R+1, \ldots, 4R. \end{cases}$$

For $r \in \{1, \ldots, 2R\}$, the local tail-robust autocovariance estimators in odd and even blocks can be written as follows

$$\widehat{\mathbf{\Sigma}}_{\mathcal{S}_{2r-1}}(c) = \frac{1}{S} \sum_{t \in \mathcal{S}_{2r-1}} \psi_{\tau_\ell}(\mathbf{H}_{t,\ell}) \quad \text{and} \quad \widehat{\mathbf{\Sigma}}_{\mathcal{S}_{2r}}(c) = \frac{1}{S} \sum_{t \in \mathcal{S}_{2r}} \psi_{\tau_\ell}(\mathbf{H}_{t,\ell}).$$

With block-wise quantities, we can rewrite the lag-ℓ moving sum difference matrix $\mathbf{S}_\ell(c)$ as

$$\mathbf{S}_\ell(c) = [2(W-\ell)]^{-1/2} S \Big\{ \sum_{r=1}^{R} \big[\widehat{\mathbf{\Sigma}}_{\mathcal{S}_{2r-1}}(c) + \widehat{\mathbf{\Sigma}}_{\mathcal{S}_{2r}}(c)\big] - \sum_{r=R+1}^{2R} \big[\widehat{\mathbf{\Sigma}}_{\mathcal{S}_{2r-1}}(c) + \widehat{\mathbf{\Sigma}}_{\mathcal{S}_{2r}}(c)\big] \Big\}.$$

Next, we describe the block-wise Gaussian multiplier bootstrap. Let M be the number of bootstrap samples. Let $\{e_r\}_{r=1}^{2R}$ be a sequence of i.i.d. standard normal random variables, and $\{e_r^{(m)}\}_{r=1}^{2R}$ be an i.i.d. copy of $\{e_r\}_{r=1}^{2R}$ used in the m-th bootstrap sample, for $m = 1, \ldots, M$. The m-th Gaussian multiplier lag-ℓ moving sum difference matrix can be constructed by

$$\mathbf{S}_\ell^{(m)}(c) = [2(W-\ell)]^{-1/2} S \sum_{r=1}^{2R} e_r^{(m)} [\widehat{\mathbf{\Sigma}}_{\mathcal{S}_{2r-1}}(c) - \widehat{\mathbf{\Sigma}}_{\mathcal{S}_{2r}}(c)], \quad \text{for} \quad m = 1, \ldots, M.$$

Then, we define the m-th Gaussian multiplier bootstrapped sample of $T_\ell(c)$ as

$$T_\ell^{(m)}(c) \equiv \left\|\mathbf{S}_\ell^{(m)}(c)\right\|_{\max} = \left|\boldsymbol{U}_\ell^{(m)}(c)\right|_\infty, \quad \text{for} \quad m = 1, \ldots, M, \tag{2.4.5}$$

where $\boldsymbol{U}_\ell^{(m)}(c)$ is the vectorization of $\mathbf{S}_\ell^{(m)}(c)$. The block-wise differences $[\widehat{\boldsymbol{\Sigma}}_{\mathcal{S}_{2r-1}}(c) - \widehat{\boldsymbol{\Sigma}}_{\mathcal{S}_{2r}}(c)]$ are tailored for piece-wise stationary time series. When there is a change point in $[c-W+l+1, c+W]$, it will cause an autocovariance structure change in at most one pair of consecutive odd and even blocks, while the rest $2R-1$ pairs remain unaffected. Thus, the bootstrap sample $\{T_\ell^{(m)}(c)\}_{m=1}^M$ well approximates the empirical distribution of $T_\ell(c)$ under the "null case" when R is reasonable and M is large.

Here we use a toy example to illustrate the intuition of block-wise Gaussian multiplier bootstrap. We generate a two-dimensional time series by concatenating three stationary VAR(1) segments with different transition matrices. Each VAR(1) segment is of length 300. Hence, the autocovariance changes occur at time 301 and 601. The generated time series are shown in Figure 2.1(a). Figure 2.1(b) plots the lag-1 moving sum statistic of each checkpoint and highlights the values at two true change-points. In Figure 2.1(c), we plot the ingredients of computing the moving sum statistics, i.e each entry of the 2 by 2 lag-1 outer products at each checkpoint, and highlights the checkpoint (and the windows around it) which outputs the largest moving sum statistic. Figure 2.1(d) enlarges the segment highlighted in Figure 2.1(c) and illustrates the $4R$ blocks to be divided. For each divided block, we multiply it by an independent standard norm random. Two Gaussian multiplied copies of Figure 2.1(d) are presented in Figure 2.1(e) and (f). Clearly, the moving sum statistics based on the block-wise Gaussian multiplier bootstrap sample are significant smaller than the actual moving sum, and thus can be utilized to identify the existence of a change-point.

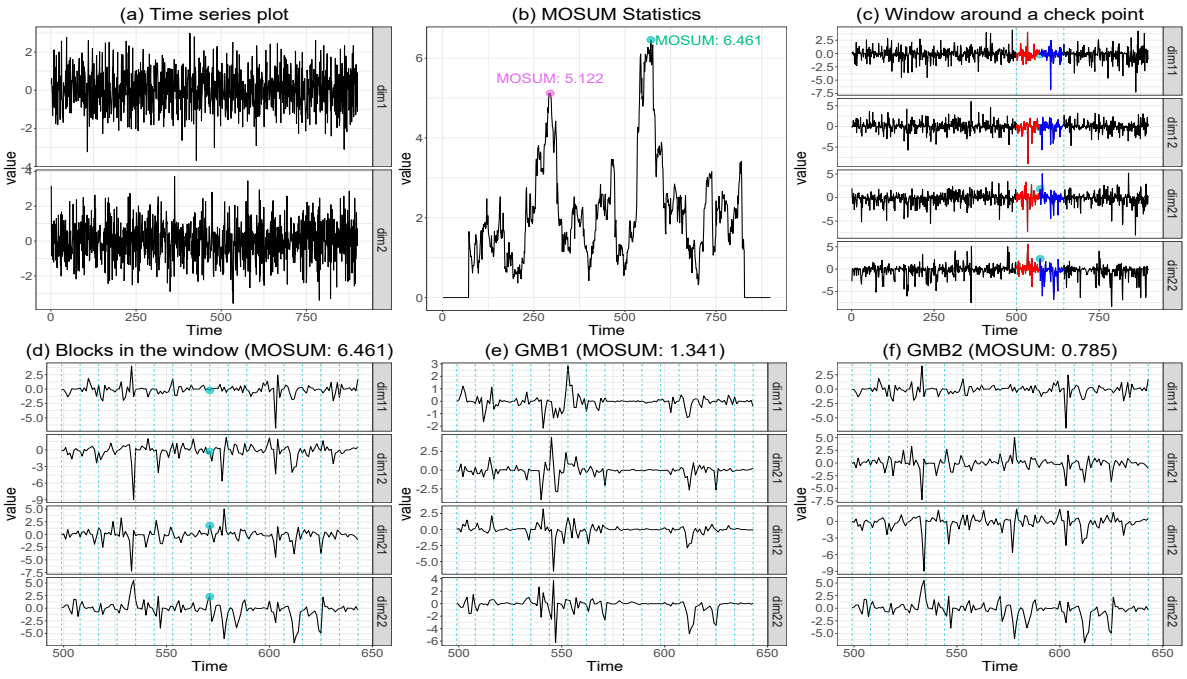

Figure 2.1: An illustrative example ($d = 2$) for block-wise Gaussian multiplier bootstrap. The bi-variate time series contains two autocovariance change-points which are correctly detected.

Since $\{e_r^{(m)}\}_{r=1}^{2R}$ are i.i.d. standard normal, we have the conditional distribution of $\boldsymbol{U}_\ell^{(m)}(c)$ given the sample $\mathcal{Y}(c, W) \equiv \{\boldsymbol{Y}_t\}_{t\in\mathcal{I}_\ell^{\mathrm{B}}(c)\cup\mathcal{I}_\ell^{\mathrm{A}}(c)}$ satisfies

$$\boldsymbol{U}_\ell^{(m)}(c)|\mathcal{Y}(c, W) \sim \mathcal{N}\Big(\boldsymbol{0}, \frac{S}{2(W-\ell)}\sum_{r=1}^{2R}\boldsymbol{D}_r(c)\boldsymbol{D}_r(c)^\intercal\Big), \tag{2.4.6}$$

where $\boldsymbol{D}_r(c) = (D_{r,j})_{j=1}^{d^2}$ is the vectorization of $\sqrt{S}[\widehat{\boldsymbol{\Sigma}}_{\mathcal{S}_{2r-1}}(c) - \widehat{\boldsymbol{\Sigma}}_{\mathcal{S}_{2r}}(c)]$. Then, for a pre-specified $\alpha \in (0, 1)$, we propose to choose the data-driven threshold at the checkpoint c, denoted by $\chi_\alpha(c)$, as the $(1-\alpha)$-th empirical quantile of $\{T_\ell^{(m)}(c)\}_{m=1}^M$. In practice, we can choose $\chi_\alpha(c)$ as the $\lfloor \alpha M \rfloor$-th largest value in $\{T_\ell^{(m)}(c)\}_{m=1}^M$ when M is large enough. On the other hand, if we sort $\{T_\ell(c), T_\ell^{(1)}(c), \ldots, T_\ell^{(M)}(c)\}$ in

descending order, we can compute an empirical p-value for $T_\ell(c)$ as $\hat{p}(c) = \mathrm{Rank}(c)/M$, where $\mathrm{Rank}(c)$ is the rank of $T_\ell(c)$ in the sorted sequence.

2.5 Inference

2.5.1 Test the existence of a lag-ℓ change-point

A fundamental inference problem in multiple change-points detection is to test if there exist a change-point in a neighborhood of a checkpoint. Following our model setup, we formalize this testing problem as follows. Let $c \in [W, T - W]$ be a checkpoint of interest and $t^* \equiv t^*(c; \ell) \in \{t_k\}_{k=1}^K$ be the lag-ℓ autocovariance change-point that is the closest to c. The null and alternative hypotheses are formulated as

$$H_0 : |c - t^*| > W \quad \text{vesus} \quad H_1 : |c - t^*| \leq (1 - \eta_W)W, \tag{2.5.1}$$

where $\eta_W \in (0, 1)$ is some slowly vanishing sequence to be discussed later.

We use the tail-robust moving sum statistic $T_\ell(c)$ defined in (2.3.3) as our test statistic. The following corollary provides the limiting distribution of $T_\ell(c)$ when the null hypothesis is true.

Corollary 2.5.1 (Limiting distribution under the null)**.** *Suppose Assumptions 3, 4 and 5 hold. Let $C, C_\tau > 0$ be some absolute constants. Assume that $\log d = CW^\beta$ for some $\beta < 1/19$. Choose the robustafication parameter as*

$$\tau_\ell = C_\tau \Big(\frac{(W - \ell)^{16/19}}{\log d} \Big)^{1/3}. \tag{2.5.2}$$

For any checkpoint $c \in [W, T-W]$, when the null hypothesis H_0 in (2.5.1) *is true, we have that*

$$\sup_{t\in\mathbb{R}} \Big| \mathbb{P}\big(T_\ell(c) \le t\big) - \mathbb{P}\big(|\boldsymbol{Z}|_\infty \le t\big) \Big| \lesssim (W-\ell)^{-2(1/19-\beta)/3} \to 0, \quad \textit{as} \quad W \to \infty,$$

where $\boldsymbol{Z} \sim N(\mathbf{0}, \boldsymbol{\Gamma})$ and $\boldsymbol{\Gamma}$ is defined as in (2.2.5).

Remark 2.5.1. *Corollary 2.5.1 follows directly from Corollary 2.2.1. Please notice that the choice of the robust parameter τ_ℓ in* (2.5.2) *is slightly different from the choice in* (2.3.2). *Both choices try to balance the bias and robustness trade-off but under different measurement. The bias and robustness are measured by the Kolmogorov distance in Corollary 2.5.1, while they are measured by the matrix max norm in Corollary 2.3.1. However, these two choices of τ_ℓ can of the same order if we set m in* (2.3.2) *as*

$$m = C_1 \frac{W^{25/57}(\log d)^{2/3}}{[\log(W-\ell)]^2}.$$

The above choice of m leads to $m \to \infty$ as $W \to \infty$. Therefore, Corollary 2.3.1 together with $\lambda_0 = C_2 W^{25/114}(\log d)^{1/3}$ can still guarantee

$$\mathbb{P}\big(\mathcal{E}_\ell^{[1,T]}(\lambda_0)\big) \ge 1 - 2\exp(-m) \to 1 \quad \textit{as} \quad W \to \infty.$$

Moreover, with this choice of m and λ_0, we can update the signal-to-noise condition and the choice of the threshold in Theorem 2.4.1 accordingly, and show that Algorithm 1 achieves consistency of detecting multiple change-points.

For a pre-specified significance level $\alpha \in (0, 1)$, Corollary 2.5.1 suggests us to choose the critical value for the test as the $(1 - \alpha)$-th quantile of $|\boldsymbol{Z}|_\infty$, which can asymptotically control the probability of committing a type-I error at α.

2.6 Simulation

In this section, we use simulated experiments to assess the empirical performance of the proposed multiple autocovariance change-points detection method. Throughout this section, we use MACD to represent the autocovarience change-points detection method proposed in Algorithm 1 . The threshold parameter χ_α is selected by the block-wise Gaussian multiplier bootstrap procedure introduced in Section 2.4.1. The robustificaion parameters in tail-robust autocovariance matrix estimations are chosen by the gap-block cross-validation (Shu & Nan, 2019). In this section, we set the window size $W = 300$, the number of blocks as $4R = 24$, the number of bootstrap samples $M = 100$, and $\alpha = 0.01$ unless otherwise stated. We also compare MACD with three state-of-the-art second-order change-points detection methods: the sparsified binary segmentation algorithm (SBS, Cho & Fryzlewicz, 2015) and the double CUSUM algorithm (D-CUSUM, Cho, 2016b). The implementation of the competing methods follows the instructions in their respective literature. Due to the limitation of space, we present selective simulation results in this section and defer more results to an appendix in the supplementary material.

2.6.1 Experiment 1: covariance change-points detection

Consider an $\mathbb{R}^d$-valued VAR(1) model

$$\boldsymbol{Y}_t = \rho \boldsymbol{Y}_{t-1} + \boldsymbol{Z}_t, \quad t = 1, \ldots, T, \tag{2.6.1}$$

where $\rho = 0.5$, $\boldsymbol{Z}_t = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$, and $\boldsymbol{\Sigma}^{1/2} = \big(\sigma_{ij}\big)_{i\in[1,d];j\in[1,d]} \in \mathbb{R}^{d\times d}$ is a symmetric and deterministic matrix. Further, $\{\boldsymbol{\epsilon}_t\}_{t=1}^T$ is an i.i.d. random sequence with $\mathbb{E}(\boldsymbol{\epsilon}_t) = \boldsymbol{0}$ and $\text{var}(\boldsymbol{\epsilon}_t) = \boldsymbol{I}_d$. In this experiment, we set $d = 100$ and 300, and we set the total length of time series as $T = 3,200$ with three covariance change-points located at $t_1 = 800$, $t_2 = 1600$ and $t_3 = 2400$, respectively. For $1 \le t < t_1$ and $t_2 \le t < t_3$, we set $\sigma_{ij} = 0.5^{|i-j|}$. For $t_1 \le t < t_2$, we set $\sigma_{ij} = 2$ if $1 \le i = j \le d/10$, and $\sigma_{ij} = 0.5^{|i-j|}$ otherwise. For $t_3 \le t \le T$, we set $\boldsymbol{\Sigma}^{1/2} = 2\boldsymbol{I}_d$

Denote $\epsilon_{t,j}$ the j-th entry of $\boldsymbol{\epsilon}_t$ for $t \in [1, T]$ and $j \in [1, d]$. We generate $\epsilon_{t,j}$ from one of the following two heavy-tailed distributions.

(1) (Student's t). $\epsilon_{t,j}$ follows a standardized Student's t_4 distribution, i.e. $\epsilon_{t,j} = 2^{-1/2}X_{t,j}$ where $X_{t,j}$'s are i.i.d. from a t_4 distribution.

(2) (Log-Normal). $\epsilon_{t,j}$ follows a standardized Log-normal distribution, i.e. $\epsilon_{t,j} = (e^2-e)^{-1/2}[\exp(X_{t,j}) - \exp(1/2)]$ where $X_{t,j}$'s are i.i.d. from a standard Normal distribution.

In each scenario, we simulate 200 replications. For each competing method, we report the number of detected change-points in Table 2.1. According to Table 2.1, the histograms of MACD correctly detects 3 change-points in more than 95% of replications in all scenarios. In contrast, all three competing methods tend to under/over-estimate the number of change-points in a good proportion of replications in most

scenarios. Further, to evaluation the estimation accuracy of change-point locations, we report the sample mean and sample standard deviation of adjusted Rand index hubert1985comparing over 200 replications in Table 2.2. Adjusted Rand index uses the permutation model to measure the similarity between estimated set of change-point locations and the ground truth. The results in Table 2.2 show that MACD has a high sample mean and a low sample standard deviation of adjusted Rand index in every scenario. All three competing methods fail to achieve high adjust Rand index when the dimension is high and the data is heavy-tailed. The experiment results suggest MACD can accurately detect the number and locations of covariance changes in high-dimensional and heavy-tailed time series, while all three state-of-the-art second-order change-points detection methods perform poorly.

Table 2.1: Experiment 1: number of detected change-points over 200 replications.

Noises	2*Dimension	Methods	Number of Change Points				
			≤ 1	2	3	4	≥ 5
t_4	100	MACD	0	0	97%	3%	0
		SBS	0	0	27%	32%	41%
		D-CUSUM	44.5%	0	53.5%	1%	1%
	300	MACD	0	0	97%	3%	0
		SBS	0	0	0%	0%	100%
		D-CUSUM	95.5%	0	4%	0	0.5%
Log-Normal	100	MACD	0	1.5%	95%	3.5%	0
		SBS	39%	11%	40%	8%	2%
		D-CUSUM	93.5%	2.5%	1.5%	0.5%	2%
	300	MACD	0	0	99%	1%	0
		SBS	1.5%	3%	15%	27.5%	53%
		D-CUSUM	98%	1.5%	0	0	0.5%

2.6.2 Experiment 2: autocovariance change-points detection

In this experiment, we generate an $\mathbb{R}^d$-valued time series $\{\boldsymbol{Y}_t\}_{t=1}^{3200}$ with three autocovariance change-points located at $t_1 = 800$, $t_2 = 1600$ and $t_3 = 2400$. We design the data generating process such that

Table 2.2: Experiment 1: sample mean and sample standard deviation (in parentheses) of adjusted Rand index over 200 replications.

Noises	Dimension	MACD	SBS	D-CUSUM
t_4	100	0.96 (0.03)	0.93 (0.05)	0.70 (0.33)
	300	0.96 (0.03)	0.83 (0.04)	0.36 (0.14)
Log-Normal	100	0.96 (0.04)	0.63 (0.33)	0.33 (0.02)
	300	0.97 (0.03)	0.84 (0.22)	0.33 (0.01)

the covariance matrix of $\{\boldsymbol{Y}_t\}_{t=1}^{3200}$ is unchanged over the whole period, while the autocovariance matrices with lage $\ell \geq 1$ changes at $t_1 = 800$, $t_2 = 1600$ and $t_3 = 2400$.

To be specific, we generate the four segments $\{\boldsymbol{Y}_t\}_{t=1}^{800}$, $\{\boldsymbol{Y}_t\}_{t=801}^{1600}$, $\{\boldsymbol{Y}_t\}_{t=1601}^{2400}$ and $\{\boldsymbol{Y}_t\}_{t=2401}^{3200}$ sequentially as follows. First, we generate $\{\boldsymbol{Y}_t\}_{t=1}^{800}$ from the VAR(1) model (2.6.1) with $\rho = 0.5$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_d$. Also, we generate $\epsilon_{t,j}$ from either the standardized Student's t_4 or standardized Log-normal distributions as described in Section 2.6.1. For the second segment, we first generate 4000 observations $\{\boldsymbol{Z}_s\}_{s=1}^{4000}$ using the same setting as we generate the first segment. Denote $Y_{t,j}$ and $Z_{t,j}$ the j-th entries of $\boldsymbol{Y}_t$ and $\boldsymbol{Z}_t$, respectively. We obtain $\{\boldsymbol{Y}_t\}_{t=801}^{1600}$ as

$$\{Y_{t,j}\}_{t=801}^{1600} = \begin{cases} \bigcup\limits_{s \bmod 5 = 0} Z_{s,j}, & \text{if } 1 \leq j \leq d/10, \\ \{Z_{s,j}\}_{s=801}^{1600}, & \text{if } d/10 < j \leq d. \end{cases}$$

In other words, the first $10/d$ entries of $\{\boldsymbol{Y}_t\}_{t=801}^{1600}$ are obtained by removing the first four observations per five from the the first $d/10$ entries of $\{\boldsymbol{Z}_s\}_{s=1}^{4000}$, while the rest entries of $\{\boldsymbol{Y}_t\}_{t=801}^{1600}$ are the first 800 observations of corresponding entries in $\{\boldsymbol{Z}_s\}_{s=1}^{4000}$. We generate the third segment $\{\boldsymbol{Y}_t\}_{t=1601}^{2400}$ using the same setting as we generate the first segment. Finally, for the fourth segment, we generate $\{\tilde{\boldsymbol{Z}}_t\}_{t=1}^{4000}$ use the

same setting as we generate the first segment. We obtain $\{\boldsymbol{Y}_t\}_{t=2401}^{3200}$ by removing the first four observations per five in $\{\tilde{\boldsymbol{Z}}_t\}_{t=1}^{4000}$.

We use MACD with lags $\ell = 0, 1$ as well as the two competing methods to detect change-points in $\{\boldsymbol{Y}_t\}_{t=1}^{3200}$. We set $d = 300$ and 500. For each scenario, we simulate 200 replications. In Table 2.3, we report the number of detected change-points for all competing methods. For all scenarios, MACD correctly detects all three change-points at lag $\ell = 1$ and identifies there is no change point at lag $\ell = 0$ in most replications. In contrast, the competing methods completely fail when the second-order change-points do not appear in the covariance matrix. In Table 2.4, we report the sample mean and sample standard deviation of adjusted Rand index over 200 replications for MACD with lag $\ell = 1$ and the two competing methods. Again, MACD achieves the highest adjusted Rand index in all scenarios, while the competing methods struggle.

Table 2.3: Experiment 2: number of detected change-points over 200 replications.

Noises	Dimension	Methods	Number of Change Points					
			0	1	2	3	4	≥ 5
t_4	300	MACD($\ell = 1$)	0	2.5%	10%	82%	5%	0.5%
		MACD($\ell = 0$)	84.7%	14.7%	0	0.6%	0	0
		SBS	0	0	0	0	0	100%
		D-CUSUM	0	97.5%	0	1.5%	0	1%
	500	MACD($\ell = 1$)	0	2%	9.5%	84.5%	4%	0
		MACD($\ell = 0$)	85.6%	13.6%	0.8%	0	0	0
		SBS	0	0	0	0	0	100%
		D-CUSUM	0	62%	0.5%	34.5%	1.5%	1.5%
Log-Normal	300	MACD($\ell = 1$)	0	3%	6%	89%	2%	0
		MACD($\ell = 0$)	88.6%	10.9%	0.5%	0%	0%	0%
		SBS	0	17%	6.5%	35%	26.5%	15%
		D-CUSUM	0	100%	0	0	0	0
	500	MACD($\ell = 1$)	0	2.5%	7.5%	85.5%	4.5% 0	0
		MACD($\ell = 0$)	88.7%	12%	1.3%	0	0	0
		SBS	0	1.5%	0.5%	18.5%	27.5%	52%
		D-CUSUM	0	100%	0	0	0	0

Table 2.4: Experiment 2: sample mean and sample standard deviation (in parentheses) of adjusted Rand index over 200 replications.

Noises	Dimension	MACD($\ell = 1$)	SBS	D-CUSUM
t_4	300	0.91 (0.13)	0.84 (0.04)	0.35 (0.10)
	500	0.92 (0.12)	0.81 (0.05)	0.58 (0.32)
Log-Normal	300	0.92 (0.12)	0.77 (0.28)	0.33 (2e-05)
	500	0.92 (0.12)	0.87 (0.15)	0.33 (1e-05)

2.7 Proofs for chaper 2

2.7.1 Proof of Theorem 2.2.1

Proof of Theorem 2.2.1. Under Assumption 8, for any $s \geq 0$, $1 \leq j, k \leq d$ and $\tau_l > 0$, the functional dependence measure of $\{\psi_{\tau_l}(H_{t,\ell,(jk)})\}_{i\in\mathbb{Z}}$, satisfies that

$$\begin{aligned}&\sup_t \|\psi_{\tau_l}(H_{t,\ell,(jk)}) - \psi_{\tau_l}(H_{t,\{t-s\},\ell,(jk)})\|_2\\ \leq &\sup_t \|Y_{t-\ell,j}Y_{t,k} - Y_{t-\ell,\{t-s\},j}Y_{t,\{t-s\},k}\|_2 \leq \omega_4(\delta_{s-\ell,4} + \delta_{s,4}),\end{aligned}$$

where the first inequality follows from the Lipschitz continuity of $\psi_\tau(\cdot)$, the second inequality follows from Hölder's and the triangle inequalities. So,

$$\sup_{m\geq 0} \rho^{-m} \sum_{s=m}^{\infty} \sup_t \|\psi_{\tau_\ell}(H_{t,\ell,(jk)}) - \psi_{\tau_\ell}(H_{t,\{t-s\},\ell,(jk)})\|_2 \leq (1+\rho^{-\ell})\omega_4\|Y_\cdot\|_4.$$

Since $|\psi_{\tau_\ell}(H_{t,\ell,(jk)})| \leq \tau_\ell$, the conditions of Theorem **??** are fulfilled. Moreover, by Lemma **??**, we have that the long-run variance of $\{\psi_{\tau_l}(H_{t,\ell,(jk)})\}_{i\in\mathbb{Z}}$ satisfies that

$$\begin{aligned}
&\lim_{|\mathcal{I}_\ell|\to\infty} \operatorname{var}\Big(\frac{1}{\sqrt{|\mathcal{I}_\ell|}}\sum_{t\in\mathcal{I}_\ell}\psi_{\tau_\ell}(H_{t,\ell,(jk)})\Big) \\
\leq &\max_{1\leq j,k\leq d}\sum_{l=-\infty}^{\infty}\sup_{t\in\mathbb{Z}}\big|\operatorname{cov}\big(\psi_{\tau_\ell}(H_{t,\ell,(jk)}),\psi_{\tau_\ell}(H_{t+\ell,\ell,(jk)})\big)\big| = \nu_H^2 < \infty.
\end{aligned}$$

For any $u > 0$, set $x = \max\big\{\sqrt{u(\nu_H^2|\mathcal{I}_\ell| + \tau_\ell^2)}, u\tau_\ell(\log n)^2\big\}$. Then, for any $|\mathcal{I}_\ell| \geq 2$ and any $u > 0$, we have with probability at least $1 - 2e^{-u/C}$ that

$$\begin{aligned}
|\mathcal{I}_\ell|\big|\widehat{\gamma}_{\mathcal{I}_\ell,(jk)} - \mathbb{E}[\widehat{\gamma}_{\mathcal{I}_\ell,(jk)}]\big| \leq &\max\Big\{\sqrt{u(\nu_H^2|\mathcal{I}_\ell| + \tau_\ell^2)}, u\tau_\ell(\log|\mathcal{I}_\ell|)^2\Big\} \\
\leq &\sqrt{u|\mathcal{I}_\ell|}\nu_H + \sqrt{u}\tau_\ell + u\tau_\ell(\log|\mathcal{I}_\ell|)^2, \qquad (2.7.1)
\end{aligned}$$

where the second inequality follows from Lemma **??**. Moreover, we have for any $1 \leq j, k \leq d$ that

$$\begin{aligned}
&\big|\mathbb{E}[H_{t,\ell,(jk)}] - \mathbb{E}[\psi_{\tau_\ell}(H_{t,\ell,(jk)})\big| \\
=&\big|\mathbb{E}\big[(Y_{t-\ell,j}Y_{t,k} - \tau_\ell)\mathbb{1}\{Y_{t-\ell,j}Y_{t,k} > \tau_\ell\}\big] + \mathbb{E}\big[(Y_{t-\ell,j}Y_{t,k} + \tau_\ell)\mathbb{1}\{Y_{t-\ell,j}Y_{t,k} < -\tau_\ell\}\big]\big| \\
\leq&\mathbb{E}\big[Y_{t-\ell,j}Y_{t,k}\mathbb{1}\{Y_{t-\ell,j}Y_{t,k} > \tau_\ell\}\big] + \mathbb{E}\big[-Y_{t-\ell,j}Y_{t,k}\mathbb{1}\{Y_{t-\ell,j}Y_{t,k} < -\tau_\ell\}\big] \\
=&\mathbb{E}\big[|Y_{t-\ell,j}Y_{t,k}|\mathbb{1}\{|Y_{t-\ell,j}Y_{t,k}| > \tau_\ell\}\big] \leq \mathbb{E}[(Y_{t-\ell,j}Y_{t,k})^2]/\tau_\ell \leq \omega_4^4/\tau_\ell,
\end{aligned}$$

where the second inequality follows from Markov's and Hölder's inequalities. Thus, it follows that

$$\Big|\sum_{t\in\mathcal{I}_\ell}\mathbb{E}[H_{t,\ell,(jk)}] - |\mathcal{I}_\ell|\mathbb{E}[\widehat{\gamma}_{\mathcal{I}_\ell,(jk)}]\Big| \leq \omega_4^4|\mathcal{I}_\ell|/\tau_\ell, \qquad (2.7.2)$$

By the triangle inequality, it follows for any $j, k \in [1, d]$ that

$$\Big||\mathcal{I}_\ell|\widehat{\gamma}_{\mathcal{I}_\ell,(jk)} - \sum_{t\in\mathcal{I}_\ell}\mathbb{E}[H_{t,\ell,(jk)}]\Big| \leq |\mathcal{I}_\ell|\big|\widehat{\gamma}_{\mathcal{I}_\ell,(jk)} - \mathbb{E}[\widehat{\gamma}_{\mathcal{I}_\ell,(jk)}]\big| + \Big|\sum_{t\in\mathcal{I}_\ell}\mathbb{E}[H_{t,\ell,(jk)}] - |\mathcal{I}_\ell|\mathbb{E}[\widehat{\gamma}_{\mathcal{I}_\ell,(jk)}]\Big|.$$

To balance the upper bounds given in (2.7.1) and (2.7.2), we set

$$\tau_\ell = \frac{|\mathcal{I}_\ell|^{1/2}\max\{\omega_4^2, \nu_H\}}{2\big[\big(\log(\mathcal{I}_\ell)\big)^2 u + \sqrt{u}\big]^{1/2}} = C_{\tau_\ell}\frac{|\mathcal{I}_\ell|^{1/2}\max\{\omega_4^2, \nu_H\}}{\log(\mathcal{I}_\ell)u^{1/2}},$$

we have for $u > 0$, with probability at least $1 - 2e^{-u}$

$$\Big|\widehat{\gamma}_{\mathcal{I}_\ell,(jk)} - |\mathcal{I}_\ell|^{-1}\sum_{t\in\mathcal{I}_\ell}\mathbb{E}[H_{t,\ell,(jk)}]\Big| \lesssim \frac{\max\{\omega_4^2, \nu_H\}\log(|\mathcal{I}_\ell|)u^{1/2}}{|\mathcal{I}_\ell|^{1/2}}. \tag{2.7.3}$$

Finally, setting $u = m + \log d^2$ and applying the union bound concludes the proof. □

2.7.2 Proof of Theorem 2.4.1

Proof of Theorem 2.4.1. Throughout this proof, we assume the event $\mathcal{E}_\ell^{[1,T]}(\lambda_0)$ holds. Define

$$\eta = 2\sqrt{2}\underline{\kappa}^{-1}(W - \ell)^{-1/2}\lambda_0.$$

The proof is by induction.

Step 1: We discuss all the possible cases which would happen for the current segment $[s, e]$, and show that with the threshold χ, we can distinguish the undetected change-points and the detected change-points.

Case 1: There exists at least one undetected true change-point in $[s, e]$. Let $t_k \in [s, e]$ be (any of) the undetected change-point(s). Due to the fact that t_k is undetected from the previous steps, we have $\min\{t_k - s, e - t_k\} > \delta - \eta(W - \ell)$. Then we have that

$$\begin{aligned}\max_{s+W-1\leq c\leq e-W} \left\|\mathbf{S}_\ell(c)\right\|_{\max} \geq & \max_{s+W-1\leq c\leq e-W} \left\|\mathbf{S}^*_\ell(c)\right\|_{\max} - \max_{s+W-1\leq c\leq e-W} \left\|\mathbf{S}^\diamond_\ell(c)\right\|_{\max} \\ \geq & [\delta - \eta(W-\ell)][2(W-\ell)]^{-1/2}\underline{\kappa} - \lambda_0 \\ \geq & 2^{-1/2} C_{snr} b_T \lambda_0 - 3\lambda_0 > \chi,\end{aligned}$$

where the first inequality follows from (2.3.1) and the triangle inequality, the second inequality follows from the fact that $\|\mathbf{S}^*_\ell(c)\|_{\max}$ is maximized at (one of) the undetected change-point(s) and the event $\mathcal{E}^{[1,T]}_\ell(\lambda_0)$, and the third inequality follows from the assumption that $\delta\underline{\kappa} \geq C_{snr} b_T \lambda_0 \sqrt{W - \ell}$.

Case 2: There exists no true change-point in $[s, e]$. Then, we have that

$$\begin{aligned}\max_{s+W-1\leq c\leq e-W} \left\|\mathbf{S}_\ell(c)\right\|_{\max} \leq & \max_{s+W-1\leq c\leq e-W} \left\|\mathbf{S}^*_\ell(c)\right\|_{\max} + \max_{s+W-1\leq c\leq e-W} \left\|\mathbf{S}^\diamond_\ell(c)\right\|_{\max} \\ \leq & \lambda_0 < \chi,\end{aligned}$$

where the first inequality follows from (2.3.1) and the triangle inequality, the second inequality follows from the fact that $\|\mathbf{S}^*_\ell(c)\|_{\max} = 0$ for any $c \in [s + W - 1, e - W]$ and the definition of $\mathcal{E}^{[1,T]}_\ell(\lambda_0)$.

Case 3: There exists one detected true change-point from previous steps in $[s, e]$. Let $t_k \in [s, e]$ be such change-point. Since t_k is detected from the previous steps, we have that $\max\{t_k - s, e - t_k\} \leq \eta(W - \ell)$.

Then, it follows that

$$\begin{aligned}\max_{s+W-1\le c\le e-W}\left\|\mathbf{S}_\ell(c)\right\|_{\max} \le& \max_{s+W-1\le c\le e-W}\left\|\mathbf{S}^*_\ell(c)\right\|_{\max} + \max_{s+W-1\le c\le e-W}\left\|\mathbf{S}^\diamond_\ell(c)\right\|_{\max}\\ \le&[(W-\ell)/2]^{1/2}\eta\overline{\kappa}+\lambda_0 = \big(1+2\underline{\kappa}^{-1}\overline{\kappa}\big)\lambda_0 \le 3\underline{\kappa}^{-1}\overline{\kappa}\lambda_0 < \chi.\end{aligned}$$

Case 4: There exists two detected true change-points from previous steps. For instance, we consider $s \le t_k < t_{k+1} \le e$. Since t_k and t_{k+1} are detected from the previous steps, we have $\max\{t_k - s, e - t_k\} \le \eta(W-\ell)$ and $\max\{t_{k+1}-s, e-t_{k+1}\} \le \eta(W-\ell)$. Then

$$\begin{aligned}\max_{s+W-1\le c\le e-W}\left\|\mathbf{S}_\ell(c)\right\|_{\max} \le& \max_{s+W-1\le c\le e-W}\left\|\mathbf{S}^*_\ell(c)\right\|_{\max} + \max_{s+W-1\le c\le e-W}\left\|\mathbf{S}^\diamond_\ell(c)\right\|_{\max}\\ \le&[(W-\ell)/2]^{1/2}\eta\overline{\kappa}+\lambda_0 \le 3\underline{\kappa}^{-1}\overline{\kappa}\lambda_0 < \chi.\end{aligned}$$

Step 2: Consider an undetected change-point $t_k \in [s,e]$ which satisfies that $\min\{t_k - s, e - t_k\} > \delta - \eta(W-\ell)$ and that $t_k = \operatorname{argmax}_{s+W-1\le c\le e-W}\|\mathbf{S}^*_\ell(c)\|_{\max}$. In order to complete the loop of induction, we need to show that if a checkpoint b such that $b \in \operatorname{argmax}_{s+W-1\le c\le e-W}\left\|\mathbf{S}_\ell(c)\right\|_{\max}$, then it must follows that $|b - t_k| \le \eta(W-\ell)$. For t_k, we have

$$\begin{aligned}\|\mathbf{S}^*_\ell(t_k)\|_{\max} =&[2(W-\ell)]^{-1/2}\max_{s+W-1\le c\le e-W}\Big\|\sum_{t\in\mathcal{I}^B_\ell(c)}\mathbf{\Sigma}(t) - \sum_{t\in\mathcal{I}^A_\ell(c)}\mathbf{\Sigma}(t)\Big\|_{\max}\\ \le& \max_{s+W-1\le c\le e-W}\left\|\mathbf{S}_\ell(c)\right\|_{\max} + \lambda_0 = \left\|\mathbf{S}_\ell(b)\right\|_{\max} + \lambda_0\end{aligned}$$

$$\leq [2(W-\ell)]^{-1/2} \Big\| \sum_{t=b-W+1+\ell}^{b} \mathbf{\Sigma}(t) - \sum_{t=b+1+\ell}^{b+W} \mathbf{\Sigma}(t) \Big\|_{\max} + 2\lambda_0, \tag{2.7.4}$$

and by triangle inequality, we have that

$$\begin{aligned} &\Big\| \sum_{t=t_k-W+1+\ell}^{t_k} \mathbf{\Sigma}(t) - \sum_{t=t_k+1+\ell}^{t_k+W} \mathbf{\Sigma}(t) \Big\|_{\max} - \Big\| \sum_{t=b-W+1+\ell}^{b} \mathbf{\Sigma}(t) - \sum_{t=b+1+\ell}^{b+W} \mathbf{\Sigma}(t) \Big\|_{\max} \\ =&(W-\ell)\kappa_k - \big(W-\ell-|t_k-b|\big)\kappa_k = |t_k-b|\kappa_k. \end{aligned} \tag{2.7.5}$$

Combining (2.7.4) and (2.7.5), we have

$$|t_k - b| \leq 2\sqrt{2}\lambda_0 \kappa_k^{-1} \sqrt{W-\ell} \leq 2\sqrt{2}\lambda_0 \underline{\kappa}^{-1} \sqrt{W-\ell} \leq \eta(W-\ell).$$

Combining **Step 1** and **Step 2** concludes the proof. □

Chapter 3

Autocovariance change point detection for streaming time series

In this chapter, we study the second-order autocovariance change point detection in an online manner, i.e. with each data point being monitored and processed, we determine if the data collected so far or in the near past present enough evidence for ocurance of change point. We construct a CUSUM-type statistics which is robust against heavy-tailed distribution. The quality of a online monitoring scheme is measured by false alarm rate or average run length and the detection delay. A good online procedure is considered to have less chance of making type I error or incorrect identifications if no shift happened as well as shorter delay after the true change point occurs. We show that with proper choice of threshold, false alarm rates of our procedure can be controlled and also the high-probability upper bound for the detection delay. In the end, we present results for simulation study.

3.1 Problem Setup

In this section, we introduce the sequential change point problem associated with second-order structure change in autocovariance matrix $\mathbf{\Sigma}_\ell$, which is characterized by its lag-ℓ, $l = 0, 1, 2 \ldots$. We begin this section with conditions of temporal dependence on the nonstationary process $\{\boldsymbol{Y}_t\}_{t\in\mathbb{N}^+}$.

Consider an $\mathbb{R}^d$-valued process

$$\boldsymbol{Y}_t = G_t(\mathcal{F}_t), \tag{3.1.1}$$

where $G_t(\cdot) = \big(g_{t1}(\cdot), g_{t2}(\cdot), \ldots, g_{td}(\cdot)\big)^\intercal$ is an $\mathbb{R}^d$-valued measurable function, and $\mathcal{F}_t = \sigma(\epsilon_1, \epsilon_2, \ldots, \epsilon_t)$ is a filtration with $\{\epsilon_t\}_{t\in\mathbb{N}^+}$ being a sequence of i.i.d. random variables. Again, in this paper, $\{\boldsymbol{Y}_t\}_{t\in\mathbb{N}^+}$ is allowed to follow various heavy-tailed distributions. Besides, we would like to emphasize that the function G_t is time-dependent, thus the representation (3.1.1) covers a large amount of nonstationary time series models, including the piece-wise stationary model in 6. The representation (3.1.1) can also be regarded as a generalization of the stationary process $\widetilde{\boldsymbol{Y}}_t = G(\mathcal{F}_t)$ with the measurable function $G(\cdot)$ being independent of time.

Let $\{\epsilon_{t-k+1}, \ldots, \epsilon_t\} = \emptyset$ when $k \leq 0$. We define $\boldsymbol{Y}_{t,\{t-k\}} = G_t(\mathcal{F}_{t,\{t-k\}})$ as a coupled version of $\boldsymbol{Y}_t$, where $\mathcal{F}_{t,\{t-k\}} = \sigma(\epsilon_1, \ldots, \epsilon_{t-k-1}, \epsilon_{t-k}^{'}, \epsilon_{t-k+1}, \ldots, \epsilon_t)$ and $\epsilon_{t-k}^{'}$ is an i.i.d. copy of ϵ_{t-k}. To measure the dependence, we define the functional dependence measure Wu2 for a nonstationary process as

$$\delta_{k,q,j} = \sup_t \|Y_{t,j} - Y_{t,\{t-k\},j}\|_q, \text{ for } k \geq 0,\ q \geq 1, \text{ and } j \in [1, d],$$

where $Y_{t,j}$ is the j-th component of $\boldsymbol{Y}_t$. As a generalization of the classical functional dependence measure (Wu, 2005) for stationary time series, $\delta_{k,q,j}$ uniformly quantifies the lag-k dependence by the moment of

order q. As the lag k increases, $\delta_{k,q,j}$ would decrease in general, and we will further impose dependence conditions by restricting the decay rates of $\delta_{k,q,j}$ with respective to k.

For simplicity, we assume $\mathbb{E}\boldsymbol{Y}_t = \mathbf{0}$ throughout rest of the chapter. For $\ell \in \mathbb{N}$, we define the lag-ℓ autocovariance matrices $\boldsymbol{\Sigma}_\ell(t) \coloneqq \mathbb{E}(\boldsymbol{Y}_{t-\ell}\boldsymbol{Y}_t^\intercal) \in \mathbb{R}^{d\times d}$. We unfold this problem by a general assumption as follows.

Assumption 6. *Assume that there exists a positive integer Δ such that*

$$\boldsymbol{\Sigma}_\ell^1 = \boldsymbol{\Sigma}_\ell^2 = ... = \boldsymbol{\Sigma}_\ell^\Delta \neq \boldsymbol{\Sigma}_\ell^{\Delta+1} = \boldsymbol{\Sigma}_\ell^{\Delta+2} = ... \tag{3.1.2}$$

In addition, define the jump sizes with respect to matrix max and spectral norms respectively as

$$\kappa_{max} = \|\boldsymbol{\Sigma}_\ell^{\Delta+1} - \boldsymbol{\Sigma}_\ell^\Delta\|_{max}$$

and

$$\kappa_{spec} = \|\boldsymbol{\Sigma}_\ell^{\Delta+1} - \boldsymbol{\Sigma}_\ell^\Delta\|_{spec}.$$

Note that $\Delta = \infty$ indicates that change point does not exist, and without loss of generality, we consider ℓ to be fixed.

In this paper, we are concerned with change of second-order structure in autocovariance matrix. Temporal dependence undoutedly brings difficulties to sequential detection. Besides that, we allow $\{\boldsymbol{Y}_t\}_{t\in\mathbb{N}^+}$ to follow various heavy-tailed distributions. This will considerably affect the detection as well. Before presenting our detection procedure, we introduce two tail-robust autocovariance estimators studied in Xu et al., 2021.

3.2 Tail-robust Autocovariance Estimators

3.2.1 Element-wise truncated estimator

Let $\{\boldsymbol{Y}_t\}_{t\in\mathbb{N}^+}$ be a centered $\mathbb{R}^d$-valued time series following the representation (3.1.1). For a segment of observations $\{\boldsymbol{Y}_t\}_{t=s}^e$ with starting and ending indices s, e satisfying $e-s \geq \ell \geq 0$, denote $\mathcal{I}_\ell = [\ell+s, e]$. The lag-ℓ outer product of $\boldsymbol{Y}_t$ is defined as

$$\mathbf{H}_{t,\ell} = \boldsymbol{Y}_{t-\ell}\boldsymbol{Y}_t^\intercal = \left(H_{t,\ell,(jk)}\right)_{j,k=1}^d, \text{ for } t = s+\ell, \ldots, e. \tag{3.2.1}$$

Define the truncation function $\psi_\tau : \mathbb{R} \mapsto \mathbb{R}$ as

$$\psi_\tau(u) = \text{sign}(u)(|u| \wedge \tau), \tag{3.2.2}$$

where $\tau > 0$ is a robustification parameter. Notice that $\psi_\tau(u)$ is the first order derivative of the Huber loss function huber1984finite defined as follows

$$L_\tau(u) \equiv \begin{cases} u^2/2, & \text{if } |u| \leq \tau \\ \tau|u| - \tau^2/2 & \text{if } |u| > \tau \end{cases}, \ u \in \mathbb{R}.$$

Using the truncation operator $\psi_\tau(\cdot)$, the truncated estimator of $\mathbb{E}[H_{t,\ell,(jk)}]$ is defined as

$$\widehat{\gamma}_{(jk)}^{[s,e]} = (e-s-\ell+1)^{-1} \sum_{t=s+\ell}^{e} \psi_{\tau_\ell}(H_{t,\ell,(jk)}), \quad \text{for } 1 \leq j,k \leq d,$$

where the robustification parameter τ_ℓ needs to be chosen properly to balance the tail-robustness and the bias due to the truncation. By collecting the element-wise estimations, the tail-robust lag-ℓ autocovariance matrix estimator is formed as

$$\widehat{\boldsymbol{\Sigma}}_{ele,\ell}^{[s,e]} = \big(\widehat{\gamma}_{(jk)}^{[s,e]}\big)_{j,k=1}^{d}. \tag{3.2.3}$$

The nonasymptotic properties of $\widehat{\boldsymbol{\Sigma}}_{ele,\ell}$ has been studied in Xu et al., 2021 when $\{\boldsymbol{Y}_t\}_{t=1}^{T}$ is stationary, i.e. $G_t(\cdot) = G(\cdot)$ in (3.1.1).

Assumption 7. *Suppose that the marginal fourth moment of $\{\boldsymbol{Y}_t\}_{t\in\mathbb{N}^+}$ satisfies*

$$\omega_4 \equiv \max_{1\le j\le d} \sup_{t\in\mathbb{N}^+} \|Y_{t,j}\|_4 < \infty.$$

Assumption 8. *There exists some $\rho \in (0,1)$, such that the coordinate-wise dependence adjusted fourth moment of $\{\boldsymbol{Y}_t\}_{t\in\mathbb{N}^+}$ satisfies*

$$\|Y_\cdot\|_4 \equiv \max_{1\le j\le d} \sup_{m\ge 0} \rho^{-m} \sum_{k=m}^{\infty} \delta_{k,4,j} < \infty.$$

Theorem 3.2.1 (Theorem 2 in Xu et al., 2021)**.** *Consider the setting described in 3.1.1. For any integers $0 < s < e$ and $e - s \ge \ell$, and for any $m > 0$, choose the robustification parameter*

$$\tau_{\ell,(jk)} \asymp \omega_4^2 (\log(e - s + 1 - \ell))^{-1} \sqrt{\frac{e - s + 1 - \ell}{m + 2\log d}}. \tag{3.2.4}$$

Then, under Assumptions 7 and 8, if $e - s + 1 - \ell \geq 4 \bigvee \log(\rho^{-1})/2$*, we have with probability at least* $1 - 4e^{-m}$*,*

$$\left\| \hat{\mathbf{\Sigma}}_{ele,\ell}^{[s,e]} - \mathbf{\Sigma}_{\ell} \right\|_{max} \leq C_e[\rho^{-l}\|X_{\cdot}\|_4 + \omega_4 \log(n-\ell)]\omega_4 \sqrt{\frac{m + 2\log d}{e - s + 1 - \ell}}. \tag{3.2.5}$$

3.2.2 Spectrum-wise truncated estimator

In contrast to the previous element-wise truncated estimator, the spectrum-wise truncated estimator is robust in the spectrum domain against extreme deviations from the mean, and it results in a bound of error with respect to the spectral norm. We begin introducing the estimator with two definitions as follows.

Definition 3.2.1. *Given a function* f *defined on* $\mathbb{R}$ *and a symmetric matrix* $A \in \mathbb{R}^{d\times d}$ *and the eigenvalue decomposition* $A = U\Lambda U^T$ *such that* $\lambda_i(A) \in \mathbb{R}$*,* $i = 1, 2, \ldots, d$*, define* $f(A) = U f(\Lambda) U^T$ *where* $f(\Lambda) = diag(f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_d))$*.*

Definition 3.2.2. *Consider any square matrix* $A \in \mathbb{R}^{d\times d}$*, define the Hermitian dilation of* A *as*

$$\bar{A} := \begin{pmatrix} 0 & A \\ A^{\intercal} & 0 \end{pmatrix}.$$

Remark 3.2.1. *Definition 3.2.1 cannot be applied to asymmetric matrices. To address this issue, we transform the asymmetric square matrix into a symmetric one by its Hermitian dilation* $\bar{A}$*. In addition,* $\bar{A}$ *is a block anti-diagonal matrix and therefore* $\|\bar{A}\|_{spec} = \|A\|_{spec}$*.*

Hence, we have the spectrum-wise truncation by definition 3.2.1,

$$\psi_\tau(\bar{A}) = \psi_\tau(\|A\|_{spec})\frac{\bar{A}}{\|A\|_{spec}},$$

and the eigenvalues of $\bar{A}$ are now bounded by τ. In general, the lag-ℓ outer product $\mathbf{H}_{t,\ell} = \boldsymbol{Y}_{t-\ell}\boldsymbol{Y}_t^\intercal$ is asymmetric when $\ell > 0$, and we then apply the Hermitian dilation on $\mathbf{H}_{t,\ell}$ for $\ell > 0$.

Let $\{\boldsymbol{Y}_t\}_{t\in\mathbb{N}^+}$ be a centered $\mathbb{R}^d$-valued time series following the representation (3.1.1). For a segment of observations $\{\boldsymbol{Y}_t\}_{t=s}^e$ with starting and ending indices s, e satisfying $e - s \geq \ell \geq 0$, we define the spectrum-wise truncated estimator by

$$\widehat{\boldsymbol{\Sigma}}_{spec,\ell}^{[s,e]} = \begin{cases} (e-s+1)^{-1}\sum_{t=s}^e \psi_\tau(\mathbf{H}_{t,0}), & \text{if } \ell = 0 \\ (e-s-\ell+1)^{-1}\sum_{t=s+\ell}^e \mathbf{T}\psi_\tau(\bar{\mathbf{H}}_{t,\ell})\mathbf{T}^\intercal, & \text{if } \ell > 0 \end{cases} \tag{3.2.6}$$

where $\mathbf{T}_{d\times 2d} = (\mathbf{0}_{d\times d}, \mathbf{I}_{d\times d})$ is to take the upper-right $d \times d$ matrix of the $2d \times 2d$ block anti-diagonal matrix $\psi_\tau(\bar{\mathbf{H}}_{t,\ell})$.

Later in this subsection, we provide the nonasymptotic properties of the spectrum-wise truncated estimator studied in Xu et al., 2021. Before that, we first state several necessary assumptions. Denote $\sigma^2 := \sigma_j^2 = (\gamma_{0,jj})_{j\in[d]}$ as the diagonal of $\boldsymbol{\Sigma}_0$.

Assumption 9. $\sigma^2 := max_{j\in[d]}\sigma_j^2 < \infty$.

Assumption 10. *The kurtosis of linear forms:*

$$K := \sup_{u\in\mathbb{R}^d}[kurt(\langle \mathbf{u}, \mathbf{Y}_0\rangle)]^{\frac{1}{4}} = \sup_{\mathbf{u}\in\mathbb{R}^d}\frac{\|\langle \mathbf{u}, \mathbf{Y}_0\rangle\|_4}{\|\langle \mathbf{u}, \mathbf{Y}_0\rangle\|_2^2} < \infty.$$

Assumption 11. *There exist constants* $C_1,\ C_2 > 0$ *and some* $\rho_1,\ \rho_2 \in (0,1)$ *such that for all* $i \geq 0$

$$\||Y_i - Y_{i,0,-\infty}|_2\|_2 \leq C_1\sqrt{tr(\Sigma_0)}\rho_1^i,$$

and

$$\sup_{v\in\mathbb{S}^{d-1}} \|\langle v, Y_i - Y_{i,0,-\infty}\rangle_2\|_2 \leq C_2\|\Sigma_0\|^{\frac{1}{2}}\rho_2^i.$$

Assumption 12. *There exists a constant* $K^* > 0$ *such that*

$$\begin{aligned} K^* &:= \sup_{i\geq 0}\sup_{\mathbf{u}\in\mathbb{R}^d}[kurt(\langle \mathbf{u}, \mathbf{Y}_i - \mathbf{Y}_{i,\{0,-\infty\}}\rangle)]^{\frac{1}{4}} \\ &= \sup_{i\geq 0}\sup_{u\in\mathbb{R}^d}\frac{\|\langle \mathbf{u}, \mathbf{Y}_i - \mathbf{Y}_{i,\{0,-\infty\}}\rangle)]\rangle\|_4}{\|\langle \mathbf{u}, \mathbf{Y}_i - \mathbf{Y}_{i,\{0,-\infty\}}\rangle)]\rangle\|_2^2} < \infty \end{aligned}$$

In comparison to the coordinate-wise conditions on moment and dependence measure in matrix norm, Assumptions 9-12 consider the projected sequences by l_2-norm and linear forms.

Theorem 3.2.2. *(cf. Theorem 6 in Xu et al., 2021) Let* $\gamma_\ell(\rho) := \frac{\log n}{\log 2}\, max(1, \frac{8log(d\eta_\ell n^6)}{log\rho^{-1}})$ *,* $\eta_\ell := C(1 + \rho^{-\ell})\sqrt{tr(\Sigma_0)}$ *and* $r(\mathbf{\Sigma}_0) := tr(\mathbf{\Sigma}_0)/\|\mathbf{\Sigma}_0\|_{spec}$ *be the effective rank of the covariance matrix* $\mathbf{\Sigma}_0$*. For* $m \geq 0$*, if we choose the robustification parameter*

$$\tau_\ell \asymp K^2\|\mathbf{\Sigma}_0\|_{spec}\gamma_\ell(\rho)^{-\frac{1}{2}}\sqrt{\frac{e - s - \ell + 1}{m + log(2d)}},$$

then under assumptions 9-12, we have with probability at least $1 - 2e^{-m}$,

$$\|\widehat{\mathbf{\Sigma}}^{[s,e]}_{spec,\ell} - \mathbf{\Sigma}_\ell\|_{spec} \leq C_e K^2 \|\mathbf{\Sigma}_0\|_{spec}(\sqrt{\gamma_\ell(\rho)} + (K^*/K)^{\frac{1}{2}}\rho^{\frac{\ell}{2}}\sqrt{r(\mathbf{\Sigma}_0)})\sqrt{\frac{m + \log(2d)}{e - s - \ell + 1}}$$

3.3 CUSUM-type Statistic

In this section, we adopt the CUSUM statistic with the element-wise truncated estimator and spectrum-wise truncated estimator to detect change points associated with autocovariance matrix. Following the representation in (3.1.1), for any pair of integers $0 < s < t$, we define the CUSUM statistics with respect to max norm and spectral norm as

$$\hat{D}^{max}_{s,t} = \|\hat{\mathbf{\Sigma}}^{[s,t]}_{ele,\ell} - \hat{\mathbf{\Sigma}}^{[1,s]}_{ele,\ell}\|_{max}, \tag{3.3.1}$$

and

$$\hat{D}^{spec}_{s,t} = \|\hat{\mathbf{\Sigma}}^{[s,t]}_{spec,\ell} - \hat{\mathbf{\Sigma}}^{[1,s]}_{spec,\ell}\|_{spec}. \tag{3.3.2}$$

Algorithm 2 run through the current data sequence with a pre-specified and time-dependent threshold. Alarm will be triggered as long as there exists an integer $s \in (1, t - \ell)$ such that the statistic $\hat{D}_{s,t} > \xi_{s,t}$. In the rest of the section, we analyze the algorithm 2 and provide theoretical guarantees for the two statistics. We begin with the signal-to-noise ratio for $\hat{D}^{max}_{s,t}$ in assumption 13.

Assumption 13 (Signal-to-noise ratio)**.**

$$\kappa\sqrt{\Delta - \ell} = C_{SNR}\omega_4\sqrt{\log(\frac{d\Delta}{\alpha})} \tag{3.3.3}$$

Algorithm 2 CUSUM

Input: $\{\boldsymbol{Y}_k\}_{k=1,2,\ldots}, \ell, \xi$
$t \leftarrow 1 + \ell$
while FLAG = 0 **do**
    $t \leftarrow t + 1$;
    $s = 1$
    while FLAG = 0 and $s < t - \ell$ **do**
        $s \leftarrow s + 1$
        FLAG = $\mathbb{1}\{\hat{D}_{s,t} > \xi_{s,t}\}$
    end while
end while

In 3.3.3, the signal is characterized by the jump size κ and the pre-change length Δ while the noise is related to the dimension d, fourth moment ω_4 and adjusted fourth moment $\|Y.\|_4$.

Theorem 3.3.1. *Consider the setting described in 3.1.1. Let $\alpha \in (0,1)$ and $\hat{\Delta}$ from 2. For any integers $0 < s < t$ and $t - s \geq \ell$, under Assumptions 7 and 8, choose the robustification parameter*

$$\tau_{\ell,(jk)} \asymp \omega_4^2 (\log(e - s + 1 - \ell))^{-1} \sqrt{\frac{e - s + 1 - \ell}{\log(\frac{dt}{\alpha})}}.$$

With the choice of threshold being

$$\xi_{s,t} = C_e \omega_4 \Big\{ \sqrt{\frac{\log(\frac{dt}{\alpha})}{t - s + 1 - \ell}} + \sqrt{\frac{\log(\frac{ds}{\alpha})}{s - \ell}} \Big\}$$

under $\Delta = \infty$,

$$\mathbb{P}_\infty(\hat{\Delta} < \infty) \leq \alpha. \tag{3.3.4}$$

Under Assumption 6,

$$\mathbb{P}_{\Delta}(\hat{\Delta} < \Delta) \leq \alpha. \tag{3.3.5}$$

If Assumptions 6 and 13 holds,

$$\mathbb{P}_{\Delta}(\Delta < \hat{\Delta} < \Delta + \epsilon) \geq 1 - \alpha, \tag{3.3.6}$$

where

$$\epsilon = \ell + C_{\epsilon}\kappa^{-2}\log(\frac{2d\Delta}{\alpha}).$$

The bound 3.3.4 guarantees that with the absence of change point, the detection procedure will continue indefinitely with probability at lease $1 - \alpha$. Under Assumtion 6, when the change point exists, 3.3.12 ensures that the false alarm probability of our procedure is no larger than α. As far as the detection delay is concerned, 3.3.6 provides a high probability bound on the delay of order $\kappa^{-2}\log(\frac{2d\Delta}{\alpha})$. It suggests that the delay shrinks with higher jump size κ and slowly increases with dimension d. Additionally, there is a trade-off between the desired false alarm rate α and the delay $(\hat{\Delta} - \Delta)_{+}$.

Similarly, we show the results for $\hat{D}_{s,t}^{spec}$ as follows.

Assumption 14 (Signal-to-noise ratio)**.**

$$\kappa_{spec}\sqrt{\Delta - \ell} = C_{SNR}\|\mathbf{\Sigma}_0\|_{spec}\sqrt{\log(\frac{d\Delta}{\alpha})}, \tag{3.3.7}$$

In 3.3.7, the signal is characterized by the jump size in spectral norm κ_{spec} and the pre-change length Δ. On the other hand, the noise is captured by dimension d, the spectral norm of covariance matrix $\|\mathbf{\Sigma}_0\|_{spec}$.

Theorem 3.3.2. *Consider the setting described in 3.1.1. Let* $\alpha \in (0,1)$ *and* $\hat{\Delta}$ *from 2. For any integers* $0 < s < t$ *and* $t - s > \ell$*, under Assumptions 9- 12, we choose the robustification parameter*

$$\tau_\ell \asymp K^2 \|\mathbf{\Sigma}_0\|_{spec} \gamma_\ell(\rho)^{-\frac{1}{2}} \sqrt{\frac{e-s-\ell+1}{m+log(2d)}}.$$

Then with the choice of threshold being

$$\xi_{s,t} = C_E K^2 \|\mathbf{\Sigma}_0\|_{spec} \Big\{ \sqrt{\gamma_l^*(\rho)} + \sqrt{\frac{K^* tr(\mathbf{\Sigma}_0)}{K\|\mathbf{\Sigma}_0\|_{spec}}} \Big\} \Big\{ \sqrt{\frac{\log(\frac{dt}{\alpha})}{t-s+1-\ell}} + \sqrt{\frac{\log(\frac{ds}{\alpha})}{s-\ell}} \Big\}$$

under $\Delta = \infty$*,*

$$\mathbb{P}_\infty(\hat{\Delta} < \infty) \leq \alpha. \tag{3.3.8}$$

and under Assumption 6,

$$\mathbb{P}_\Delta(\hat{\Delta} < \Delta) \leq \alpha. \tag{3.3.9}$$

If Assumptions 6 and 14 holds,

$$\mathbb{P}_\Delta(\Delta < \hat{\Delta} < \Delta + \ell + \epsilon) \geq 1 - \alpha, \tag{3.3.10}$$

where

$$\epsilon = C_\epsilon \kappa_{spec}^{-2} \log(\frac{2d\Delta}{\alpha}).$$

In practice, as it is often the case in the applications, the threshold $\xi_{s,t}$ needs to be calibrated from observations obtained from the pre-change distribution. One major challenge here is the presence of dependence. Regular resampling method such as bootstrap and permutation will not work. Hence, we adopt the block-wise permutation to obtain the threshold from the calibration data set since it can preserve the dependence of the data sequence.

Another practical issue is computational complexity. At each time t, one can store partial sums of truncated outer products $\big\{ \sum_{k=1+\ell}^{s} \psi_{\tau_\ell}(H_{k,\ell}) \big\}_{s=1}^{t}$ The computational cost of algorithm 1 is $O(t^2)$ as it needs one more loop for all possible s.

In order to reduce the computational burden, we propose to compute $\hat{D}_{t-h,t}$ only over a geometrically increasing sequence of values for h, stated in Algorithm 3 below. It takes time in a linear order, and more importantly, it can hold the same theoretical property as before.

Corollary 3.3.1. *Consider the setting described in 3.1.1. Let $\alpha \in (0,1)$ and $\hat{\Delta}$ from Algorithm 3. For any integers $0 < s < t$ and $t - s > \ell$, under Assumptions 7 and 8, choose the robustification parameter*

$$\tau_{\ell,(jk)} \asymp \omega_4^2 (\log(e - s + 1 - \ell))^{-1} \sqrt{\frac{e - s + 1 - \ell}{\log(\frac{dt}{\alpha})}}.$$

With the choice of threshold being

$$\xi_{s,t} = C_e \omega_4 \Big\{ \sqrt{\frac{\log(\frac{dt}{\alpha})}{t - s + 1 - \ell}} + \sqrt{\frac{\log(\frac{ds}{\alpha})}{s - \ell}} \Big\}$$

under $\Delta = \infty$*,*

$$\mathbb{P}_\infty(\hat{\Delta} < \infty) \leq \alpha. \tag{3.3.11}$$

Under Assumption 6,

$$\mathbb{P}_\Delta(\hat{\Delta} < \Delta) \leq \alpha. \tag{3.3.12}$$

If Assumptions 6 and 13 holds,

$$\mathbb{P}_\Delta(\Delta < \hat{\Delta} < \Delta + \epsilon) \geq 1 - \alpha, \tag{3.3.13}$$

where

$$\epsilon = \ell + C'_\epsilon \kappa^{-2} \omega_4 \log(\frac{2d\Delta}{\alpha}) and C'_\epsilon \geq C_\epsilon.$$

The major benefit of Algorithm 3 is ease of computation. It will reduce the computational cost to $\mathcal{O}(t \log t)$. Besides, this scheme yields the same nearly optimal guarantees of Theorem 3.3.1, as shown in the corollary 3.3.1. Similarly, it can also be extended for statistic with spectrum-wise truncated estimator and measured in matrix spectral norm.

Algorithm 3 CUSUM

```
Input: {Y_k}_{k=1,2,...}, ℓ, ξ
t ← 1 + ℓ
while FLAG = 0 do
    t ← t + 1;
    J = ⌊log(t)/ log(2)⌋;
    j = 0
    while FLAG = 0 and j < J do
        j = j + 1
        s ← t − 2^{j−1}
        FLAG = 𝟙{D̂_{s,t} > ξ_{s,t}}
    end while
end while
```

3.4 Lower bound for the detection delay

In the proceeding section, we find a upper bound on the detection delay in Theorem 3.3.1 and 3.3.2 with high probability. In this section, we derive the minimax lower bound on the expected detection delay. The proof adapts arguments used for mean change in Proposition 7 of Y. Yu et al., 2020 and Theorem 2 of Lai, 1998.

Proposition 3.4.1. *Assume that* $\{X_i\}_{i=1,2,\dots}$ *is a sequence of independent Gaussian random variables with* $\mathbb{E}(X_i) = 0$ *and* $Var(X_i) = \sigma^2$. *Denote the joint distribution of* $\{X_i\}_{i=1,2,\dots}$ *as* $P_{\kappa,\sigma,\Delta}$. *For* $\alpha \in (0,1)$, *under Assumption 6, consider the following estimators*

$$\mathcal{D}_T = \{T \mid T \text{ is a stopping time with respect to the natural filtration and satisfies } \mathbb{P}_\infty(T < \infty) \leq \alpha\}.$$

Define

$$\mathcal{S}_\alpha = \left\{0 < \alpha < 1 \mid 18\log(\frac{1}{\alpha})\alpha \leq \alpha^{\frac{1}{4}} \text{ and } \alpha + \alpha^{\frac{7}{2}} + \alpha^{\frac{1}{4}} < \frac{1}{2}\right\}$$

and

$$\mathcal{S}_{\kappa,\sigma_0} = \left\{(\kappa, \sigma_0) \mid \frac{1}{6}\sigma_0^2 < \kappa < \frac{1}{4}\sigma_0^2\right\}.$$

Then, for sufficiently small $\alpha \in \mathcal{S}_\alpha$, $(\kappa, \sigma_0^2) \in \mathcal{S}_{\kappa,\sigma_0^2}$, *and any change point time* Δ,

$$\inf_{\hat{\Delta} \in \mathbb{D}(\alpha)} \sup_{P_{\kappa,\sigma,\Delta}} \mathbb{E}_P\{(\hat{\Delta} - \Delta)_+\} \geq \frac{\sigma_0^4}{4\kappa^2} \log\left(\frac{1}{\alpha}\right).$$

3.5 Simulation Study

In this section, we use simulated experiments to evaluate the performance of the two proposed robust online change point detection methods based on CUSUM process and measured in spectral and max norm respectively, denoted as CUSUM-spec-spec and CUSUM-ele-max. For comparison, replacing the robust estimators by the sample autocovariance estimators, we get two competitors, denoted as CUSUM-sa-spec and CUSUM-sa-max.

3.5.1 Experiment 1

In this experiment, we generate the $\mathbb{R}^d$-valued time series $\{\boldsymbol{Y}_t\}_{t=1}^{2L}$ composed of pre-change and post-change segments with same length L. We consider an $\mathbb{R}^d$-valued VAR(1) model

$$\boldsymbol{Y}_t = \rho \boldsymbol{Y}_{t-1} + \boldsymbol{Z}_t, \quad t = 1, \ldots, 2L,$$

where $\rho = 0.5$, $\boldsymbol{Z}_t = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$, and $\boldsymbol{\Sigma} = \left(\sigma_{ij}\right)_{i\in[1,d];j\in[1,d]} \in \mathbb{R}^{d\times d}$ is a symmetric and deterministic matrix. Further, $\{\boldsymbol{\epsilon}_t\}_{t=1}^{2L}$ is an i.i.d. random sequence with $\mathbb{E}(\boldsymbol{\epsilon}_t) = \boldsymbol{0}$ and $\text{var}(\boldsymbol{\epsilon}_t) = \boldsymbol{I}_d$. We set the length of calibration to be $2d$ and allow the variation of pre-change and post-change segments by choosing $L = 2d + Unif(0, 50)$. For $1 \leq t \leq L$, we set $\sigma_{ij} = 0.5^{|i-j|}$. For the post-change segment, where $L < t \leq 2L$, we only change the first $\lfloor d/5 \rfloor$ diagonal entries to 2, i.e. we set $\sigma_{ij} = 2$ if $1 \leq i = j \leq d/5$, and $\sigma_{ij} = 0.5^{|i-j|}$ otherwise.

Denote $\epsilon_{t,j}$ the j-th entry of $\boldsymbol{\epsilon}_t$ for $t \in [1, 2L]$ and $j \in [1, d]$. We generate $\epsilon_{t,j}$ from one of the following three standardized distributions.

(1) (Normal). $\epsilon_{t,j}$ follows a standard Normal distribution.

(2) (Student's t). $\epsilon_{t,j}$ follows a standardized Student's t_4 distribution, i.e. $\epsilon_{t,j} = 2^{-1/2}X_{t,j}$ where $X_{t,j}$'s are i.i.d. from a t_4 distribution.

(3) (Log-Normal). $\epsilon_{t,j}$ follows a standardized Log-normal distribution, i.e. $\epsilon_{t,j} = (e^2-e)^{-1/2}[\exp(X_{t,j})-\exp(1/2)]$ where $X_{t,j}$'s are i.i.d. from a standard Normal distribution.

In the calibration process, we run blockwise permutation $K = 200$ times and compute the CUSUM statistics $\{\|\hat{D}_{s,t}^{(k)}\|\}_{k=1}^{200}$ to evaluate the unknown constant C_E in the detection threshold. It is chosen such that the proportion of $\{\|\hat{D}_{s,t}^{(k)}\|\}_{k=1}^{200}$ cross $\xi_{s,t}$ is capped at α. In addtion to the threshold, we use the calibration data to obtain the robustification parameter τ by gap-block validation, which is the same strategy of offline procedure in chapter 2.

To fairly evaluate the performance of online change point approach, we consider three metrics: proportion of false alarm, power and average detection delay. The proportion of false alarm is given by

$$N^{-1}\sum_{j=1}^{N}\mathbb{1}\{\hat{\Delta} \leq L\},$$

where $N = 200$ is the number of replicates and $\hat{\Delta}$ is the estimated change point location (true location $\Delta = L$). Again in this experiment, the pre-change and post-change length equal to L, and we say the dection scheme commits a type II error if it fails to discover a change point within the total run length $2L$, i.e., $\hat{\Delta} \geq 2L$. Thus, the power is defined by

$$N^{-1}\sum_{j=1}^{N}\mathbb{1}\{L < \hat{\Delta} < 2L\},$$

and the average detection delay is given by

$$\frac{\sum_{j=1}^{N}(\hat{\Delta}-L)\mathbb{1}\{L<\hat{\Delta}<2L\}}{\sum_{j=1}^{N}\mathbb{1}\{L<\hat{\Delta}<2L\}}.$$

Table 3.1: Proportion of false alarm over 200 replications Considered in Scenario 2 when p = 20.

	normal	t4	lognorm
CUSUM-sa-spec	0.02	0.05	0.02
CUSUM-spec-spec	0.03	0.05	0.04
CUSUM-sa-max	0.01	0.03	0.07
CUSUM-ele-max	0.00	0.03	0.04

Table 3.2: Average delay and power(in parentheses) over 200 replications when d = 20 in experiment 1.

	normal	t4	lognorm
CUSUM-sa-spec	7.99(0.97)	28.90(0.42)	27.48(0.59)
CUSUM-spec-spec	8.01(0.97)	24.26(0.69)	27.17 (0.61)
CUSUM-sa-max	1.95(1.00)	12.57(0.97)	10.91(0.92)
CUSUM-ele-max	2.64(1.00)	4.53(0.97)	4.30(0.95)

Table 3.3: Porpotion of false alarm over 200 replications when d = 50.

	normal	t4	lognorm
CUSUM-sa-spec	0.15	0.00	0.00
CUSUM-spec-spec	0.01	0.01	0.03
CUSUM-sa-max	0.05	0.00	0.02
CUSUM-ele-max	0.04	0.0	0.00

Table 3.4: Average Delay and Power(in parentheses) over 200 replications Considered in Scenario 2 When p = 50.

	normal	t4	lognorm
CUSUM-sa-spec	4.91(1.00)	38.64(0.53)	40.49(0.59)
CUSUM-spec-spec	2.92(1.00)	33.54(0.55)	34.45 (0.66)
CUSUM-sa-max	3.67(1.00)	16.24(0.97)	34.80(0.96)
CUSUM-ele-max	3.13(1.00)	10.73(0.98)	18.75(0.97)

3.5.2 Experiment 2: Autocovariance Change

In the second experiment, we want to restrict our attention to the autocovariance change only while keeping the same covariance structure before and after the change point. This motivates us to consider the following setup.

$$Y_t = \begin{cases} \sqrt{\frac{1}{1-\rho^2}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t, & \text{if } 0 < t < L \\ \rho \boldsymbol{Y}_{t-1} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t & \text{if } L < t < 2L, \end{cases}$$

where $\boldsymbol{\Sigma} = I_p$, $L = Unif(50, 100)$, and $\boldsymbol{\epsilon}_t$ still follows one of the three standardized distributions as described in the previous experiment. Under this scenario, the covariance matrix $\boldsymbol{\Sigma_0}$ remains $\sqrt{\frac{1}{1-\rho^2}}\boldsymbol{\Sigma}$ along the series, but the lag-1 autocovariance matrix $\boldsymbol{\Sigma_1}$ alters from 0 to $\sqrt{\frac{\rho}{1-\rho^2}}\boldsymbol{\Sigma}$ at change point L. Compared to the previous experiment, the signal in this case is definitely weaker.

Table 3.5: Porpotion of false alarm over 200 replications when p = 20 in experiment 2.

	normal	t4	lognorm
CUSUM-sa-spec	0.02	0.01	0.02
CUSUM-spec-spec	0.01	0.01	0.01
CUSUM-sa-max	0.04	0.00	0.03
CUSUM-ele-max	0.04	0.00	0.04

Table 3.6: Average Delay and Power(in parentheses) over 200 replications when p = 20 in experiment 2.

	normal	t4	lognorm
CUSUM-sa-spec	18.08(0.94)	47.90(0.05)	37.85(0.03)
CUSUM-spec-spec	12.75(0.96)	26.46(0.74)	20.51 (0.94)
CUSUM-sa-max	27.80(0.61)	50.67(0.05)	37.69(0.24)
CUSUM-ele-max	26.75(0.69)	31.59(0.32)	25.71(0.84)

Table 3.7: Porpotion of false alarm over 200 replications when p = 50 in experiment 2 .

	normal	t4	lognorm
CUSUM-sa-spec	0.15	0.00	0.00
CUSUM-spec-spec	0.01	0.01	0.03
CUSUM-sa-max	0.05	0.00	0.02
CUSUM-ele-max	0.04	0.00	0.00

Table 3.8: Average Delay and Power(in parentheses) over 200 replications when p = 50 in experiment 2.

	normal	t4	lognorm
CUSUM-sa-spec	4.91(1.00)	56.64(0.18)	59.61(0.19)
CUSUM-spec-spec	2.92(1.00)	9.94(1.00)	15.34 (1.00)
CUSUM-sa-max	21.49(1.00)	59.81(0.16)	51.80(0.39)
CUSUM-ele-max	21.79(1.00)	15.15(0.58)	14.75(0.76)

Table 3.9: Porpotion of false alarm over 200 replications when p = 100 in experiment 2.

	normal	t4	lognorm
CUSUM-sa-spec	0.00	0.00	0.01
CUSUM-spec-spec	0.00	0.04	0.06
CUSUM-sa-max	0.00	0.0	0.20
CUSUM-ele-max	0.04	0.00	0.00

Table 3.10: Average Delay and Power(in parentheses) over 200 replications Considered in Scenario 2 When p = 100.

	normal	t4	lognorm
CUSUM-sa-spec	5.32(1.00)	90.671(0.16)	95.48(0.48)
CUSUM-spec-spec	2.69(1.00)	6.43(1.00)	7.28 (1.00)
CUSUM-sa-max	32.39(1.00)	97.14(0.11)	63.54(0.96)
CUSUM-ele-max	23.83 (1.00)	12.22(0.50)	14.59(1.00)

3.6 Proofs for chaper 3

3.6.1 Proof of Theorem 3.3.1

Lemma 3.6.1. *Assume* $\Delta = \infty$. *Define*

$$\mathcal{A} = \big\{\forall s,t \in \mathbb{N}^+, t > 2+\ell \, and \, t-s \geq 1+\ell : \, \|\hat{\mathbf{\Sigma}}^{[s:t]}_{ele,\ell} - \mathbf{\Sigma}_\ell\|_{max} \leq a_{s,t}(\delta)\big\},$$

where

$$a_{s,t}(\delta) = C_e[\rho^{-l}\|X.\|_4 + \omega_4 \log(t-s+1-\ell)]\omega_4\sqrt{\frac{\delta + 2\log d}{t-s+1-\ell}},$$

with $0 < \alpha < 1$,

$$\delta = -2\log(\frac{\alpha}{t}).$$

Then, it holds that

$$\mathbb{P}(\mathcal{A}) \geq 1-\alpha.$$

Proof of lemma 3.6.1. Consider $\mathcal{A}^c$, the complement of $\mathcal{A}$.

$$\begin{aligned}
\mathbb{P}(\mathcal{A}^c) =& \mathbb{P}\{\exists s,t \in \mathbb{N}^+, t \geq 2+\ell \, and \, t-s \geq 1+\ell : \, \|\hat{\mathbf{\Sigma}}^{[s:t]}_{ele,\ell} - \mathbf{\Sigma}_\ell\|_{max} \geq a_{s,t}(\delta)\} \\
\leq& \sum_{j=1}^{\infty} \mathbb{P}\Big\{ \prod_{2^j \leq t < 2^{j+1}} \prod_{0<s\leq t-1-\ell} \mathbb{1}\{\|\hat{\mathbf{\Sigma}}^{[s:t]}_{ele,\ell} - \mathbf{\Sigma}_\ell\|_{max} \geq a_{s,t}(\delta)\} = 1\Big\} \\
\leq& \sum_{j=1}^{\infty} 2^j \max_{2^j \leq t < 2^{j+1}} \mathbb{P}\Big\{ \prod_{0<s\leq t-1-\ell} \mathbb{1}\{\|\hat{\mathbf{\Sigma}}^{[s:t]}_{ele,\ell} - \mathbf{\Sigma}_\ell\|_{max} \geq a_{s,t}(\delta)\} = 1\Big\} \\
\leq& \sum_{j=1}^{\infty} 2^{2j+1} \max_{2^j \leq t < 2^{j+1}} \max_{0<s\leq t-1-\ell} \mathbb{P}\Big\{ \mathbb{1}\{\|\hat{\mathbf{\Sigma}}^{[s:t]}_{ele,\ell} - \mathbf{\Sigma}_\ell\|_{max} \geq a_{s,t}(\delta)\} = 1\Big\}
\end{aligned}$$

By the estimation error provided in Theorem 3.2.1, choose

$$\tilde{\delta} = -\frac{1}{4}\log\Big\{\frac{\alpha\log^2(2)}{2[\log(t)+\log(2)]^2t^2}\Big\}.$$

Then

$$\begin{aligned}
\mathbb{P}(\mathcal{A}^c) &= \leq \sum_{j=1}^{\infty} 2^{2j+1} \max_{2^j\leq t<2^{j+1}} \max_{0<s\leq t-1} \mathbb{P}\Big\{\|\hat{\mathbf{\Sigma}}_{ele,\ell}^{[s:t]} - \mathbf{\Sigma}_\ell\|_{max} \geq a_{s,t}(\tilde{\delta})\Big\} \\
&\leq \sum_{j=1}^{\infty} 2^{2j+1} \max_{2^j\leq t<2^{j+1}} \max_{0<s\leq t-1} \frac{\alpha\log^2(2)}{2[\log(t)+\log(2)]^2t^2} \\
&\leq \sum_{j=1}^{\infty} 2^{2j+1} \frac{\alpha}{2(j+1)^2 2^{2j}} \\
&\leq \alpha\sum_{j=1}^{\infty}\frac{1}{(1+j)^2} \leq \alpha\sum_{j=1}^{\infty}\frac{1}{(1+j)j} = \alpha
\end{aligned}$$

Because

$$\begin{aligned}
-\frac{1}{4}\log\Big\{\frac{\alpha\log^2 2}{2[\log(t)+\log(2)]^2t^2}\Big\} &\leq -\frac{1}{4}\log\Big\{\frac{\alpha\log^2 2}{8\log(t)^2t^2}\Big\} \\
&\leq -2\log(\frac{\alpha}{t}),
\end{aligned}$$

for simplicity, we take

$$\delta = -2\log(\frac{\alpha}{t}).$$

□

Proof of Theorem 3.3.1. Recall

$$\mathcal{A} = \Big\{\forall s,t\in\mathbb{N}^+, t>2,\ t>s:\ \|\hat{\mathbf{\Sigma}}_\ell^{[s:t]} - \mathbf{\Sigma}_\ell\| \leq a_{s,t}(\delta)\Big\}$$

with

$$a_{s,t}(\delta) = C_e[\rho^{-l}\|X.\|_4 + \omega_4 \log(t-s+1-\ell)]\omega_4\sqrt{\frac{\delta + 2\log d}{t-s+1-\ell}},$$

and

$$\delta = -2\log(\frac{\alpha}{t}).$$

$$\begin{aligned} a_{s,t} &= C_e[\,\rho^{-l}\|X.\|_4 + \omega_4 \log(t-s+1-\ell)\,]\,\omega_4\sqrt{\frac{\delta + 2\log d}{t-s+1\ell}} \\ &\leq C_e'[\,\rho^{-l}\|X.\|_4 + \ d\omega_4\,]\,\omega_4\sqrt{\frac{\log(\frac{dt}{\alpha})}{t-s+1-\ell}}\ (\ C_e' = \sqrt{2}C_e, log(t) < d\ ) \\ &= C_E\sqrt{\frac{\log(\frac{dt}{\alpha})}{t-s+1-\ell}} \end{aligned}$$

For any $t \leq \Delta$,

$$\begin{aligned} &\|\hat{\mathbf{\Sigma}}_\ell^{[s:t]} - \hat{\mathbf{\Sigma}}_\ell^{[1:s]}\|_{max} \leq \|\hat{\mathbf{\Sigma}}_\ell^{[1:s]} - \mathbf{\Sigma}_\ell^{\Delta}\|_{max} + \|\hat{\mathbf{\Sigma}}_\ell^{[s:t]} - \mathbf{\Sigma}_\ell^{\Delta}\|_{max} \\ &= C_E\Big\{\sqrt{\frac{\log(\frac{dt}{\alpha})}{t-s+1-\ell}} + \sqrt{\frac{\log(\frac{ds}{\alpha})}{s-\ell}}\Big\} = \xi_{s,t} \end{aligned}$$

Consequently, we have $\hat{\Delta} > \Delta$.

For $t > \Delta$, we are seeking an upper bound for $\hat{\Delta}$ or $\hat{\Delta} - \Delta$, and it will immediately yield 3.3.12. Let $\tilde{\Delta} = min\big\{t > \Delta : \|\hat{\mathbf{\Sigma}}_\ell^{[\Delta:t]} - \hat{\mathbf{\Sigma}}_\ell^{[1:\Delta]}\| \geq \xi_{\Delta,t}\big\}$ and $\epsilon = \tilde{\Delta} - \Delta$. Then $\hat{\Delta} - \Delta \leq \epsilon$. Hence, it suffices to find an upper bound for ϵ.

By the definition of $\tilde{\Delta}$, we have

$$\begin{aligned}\|\hat{\mathbf{\Sigma}}_\ell^{[\Delta:\tilde{\Delta}]} - \hat{\mathbf{\Sigma}}_\ell^{[1:\Delta]}\|_{max} &= \|\hat{\mathbf{\Sigma}}_\ell^{[\Delta:\tilde{\Delta}]} - \mathbf{\Sigma}_\ell^{\Delta+1} + \mathbf{\Sigma}_\ell^{\Delta+1} - \mathbf{\Sigma}_\ell^{\Delta} + \mathbf{\Sigma}_\ell^{\Delta} - \hat{\mathbf{\Sigma}}_\ell^{[1:\Delta]}\|_{max} \\ &\geq \|\mathbf{\Sigma}_\ell^{\Delta+1} - \mathbf{\Sigma}_\ell^{\Delta}\|_{max} - \big(\|\hat{\mathbf{\Sigma}}_\ell^{[\Delta:\tilde{\Delta}]} - \mathbf{\Sigma}_\ell^{\Delta+1}\|_{max} + \|\mathbf{\Sigma}_\ell^{\Delta} - \hat{\mathbf{\Sigma}}_\ell^{[1:\Delta]}\|_{max}\big) \\ &\geq \kappa - \xi_{\Delta,\tilde{\Delta}}\end{aligned}$$

Thus, a proper upper bound of ϵ can be obtained if $\kappa - \xi_{\Delta,\tilde{\Delta}} \geq \xi_{\Delta,\tilde{\Delta}}$.

It now suffices to show that with the choice of

$$\sqrt{\epsilon - \ell} = \kappa^{-1}\, C_\epsilon \sqrt{\log(\frac{2d\Delta}{\alpha})},$$

it holds that $\kappa - \xi_{\Delta,\Delta+\epsilon} \geq \xi_{\Delta,\Delta+\epsilon}$, $i.e.\ \kappa \geq 2\xi_{\Delta,\Delta+\epsilon}$.

Due to the signal ratio in 13,

$$\kappa\sqrt{\Delta - \ell} = C_{SNR}\,(\rho^{-l}\|X_\cdot\|_4\omega_4 + \ d\omega_4^2)\,\sqrt{\log(\frac{d\Delta}{\alpha})},$$

then

$$
\begin{aligned}
2\xi_{\Delta,\Delta+\epsilon} &= 2C_E\Big\{\sqrt{\frac{\log(\frac{d(\Delta+\epsilon)}{\alpha})}{\epsilon+1-\ell}}+\sqrt{\frac{\log(\frac{d\Delta}{\alpha})}{\Delta-\ell}}\Big\} \\
&\leq 2C_E\Big\{\sqrt{\frac{\log(\frac{2d\Delta}{\alpha})}{\epsilon+1-\ell}}+\sqrt{\frac{\log(\frac{d\Delta}{\alpha})}{\Delta-\ell}}\Big\}\, red(\epsilon+1\leq\Delta) \\
&\leq 2C_E\sqrt{\log(\frac{2d\Delta}{\alpha})}\sqrt{\frac{1}{\epsilon+1-\ell}}+\kappa\frac{2C_e^{'}}{C_{SNR}} \\
&\leq \kappa\frac{2C_E}{C_\epsilon}+\kappa\frac{2C_e^{'}}{C_{SNR}} \\
&\leq \kappa,
\end{aligned}
$$

where $max(\frac{C_e^{'}}{C_{SNR}},\frac{C_E}{C_\epsilon})\leq\frac{1}{4}$. □

3.6.2 Proof of Theorem 3.3.2

Lemma 3.6.2. *Assume $\Delta=\infty$. Define*

$$
\mathcal{A}=\big\{\forall s,t\in\mathbb{N}^+, t\geq 2+\ell \,and\, t-s\geq 1+\ell:\ \|\hat{\mathbf{\Sigma}}^{[s:t]}_{spec,\ell}-\mathbf{\Sigma}_\ell\|\leq a_{s,t}(\delta)\big\},
$$

where

$$
a_{s,t}(\delta)=C_eK^2\|\mathbf{\Sigma}_0\|\big\{\sqrt{\gamma_\ell(\rho)}+\sqrt{\frac{tr(\mathbf{\Sigma}_0)K^*}{\|\mathbf{\Sigma}_0\|K}}\big\}\sqrt{\frac{\delta+\log(2d)}{e-s+1}}\big\},
$$

with $0<\alpha<1$,

$$
\delta=-4\log(\frac{\alpha}{t}).
$$

Then, it holds that

$$\mathbb{P}(\mathcal{A}) \geq 1 - \alpha.$$

Proof of lemma 3.6.2. Consider $\mathcal{A}^c$, the complement of $\mathcal{A}$.

$$\begin{aligned}
\mathbb{P}(\mathcal{A}^c) =& \mathbb{P}\{\exists s, t \in \mathbb{N}^+, t \geq 2 + \ell \, and \, t - s \geq 1 + \ell : \, \|\hat{\mathbf{\Sigma}}_{\ell}^{[s:t]} - \mathbf{\Sigma}_{\ell}\| \geq a_{s,t}(\delta)\} \\
\leq & \sum_{j=1}^{\infty} \mathbb{P}\Big\{ \prod_{2^j \leq t < 2^{j+1}} \prod_{0 < s \leq t-1} \mathbb{1}\{\|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[s:t]} - \mathbf{\Sigma}_{\ell}\| \geq a_{s,t}(\delta)\} = 1 \Big\} \\
\leq & \sum_{j=1}^{\infty} 2^j \max_{2^j \leq t < 2^{j+1}} \mathbb{P}\Big\{ \prod_{0 < s \leq t-1} \mathbb{1}\{\|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[s:t]} - \mathbf{\Sigma}_{\ell}\| \geq a_{s,t}(\delta)\} = 1 \Big\} \\
\leq & \sum_{j=1}^{\infty} 2^{2j+1} \max_{2^j \leq t < 2^{j+1}} \max_{0 < s \leq t-1} \mathbb{P}\Big\{ \mathbb{1}\{\|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[s:t]} - \mathbf{\Sigma}_{\ell}\| \geq a_{s,t}(\delta)\} = 1 \Big\}
\end{aligned}$$

By the estimation error provided in Theorem 3.2.2, choose

$$\tilde{\delta} = -\frac{1}{2} \log \Big\{ \frac{\alpha \log^2(2)}{2[\log(t) + \log(2)]^2 t^2} \Big\}.$$

Then

$$\begin{aligned}
\mathbb{P}(\mathcal{A}^c) = \leq & \sum_{j=1}^{\infty} 2^{2j+1} \max_{2^j \leq t < 2^{j+1}} \max_{0 < s \leq t-1} \mathbb{P}\Big\{ \|\hat{\mathbf{\Sigma}}_{\ell}^{[s:t]} - \mathbf{\Sigma}_{\ell}\| \geq a_{s,t}(\tilde{\delta}) \Big\} \\
\leq & \sum_{j=1}^{\infty} 2^{2j+1} \max_{2^j \leq t < 2^{j+1}} \max_{0 < s \leq t-1} \frac{\alpha \log^2(2)}{2[\log(t) + \log(2)]^2 t^2} \\
\leq & \sum_{j=1}^{\infty} 2^{2j+1} \frac{\alpha}{2(j+1)^2 2^{2j}} \\
\leq & \alpha \sum_{j=1}^{\infty} \frac{1}{(1+j)^2} \leq \alpha \sum_{j=1}^{\infty} \frac{1}{(1+j)j} = \alpha
\end{aligned}$$

Because

$$-\frac{1}{2}\log\{\frac{\alpha\log^2 2}{2[\log(t)+\log(2)]^2t^2}\} \leq -\frac{1}{2}\log\{\frac{\alpha\log^2 2}{8\log(t)^2t^2}\}$$
$$\leq -4\log(\frac{\alpha}{t})$$

for simplicity, we take

$$\delta = -4\log(\frac{\alpha}{t})$$

□

Proof of Theorem 3.3.2. Recall

$$\mathcal{A} = \{\forall s,t \in \mathbb{N}^+, t \geq 2+\ell \, and \, t-s \geq 1+\ell : \, \|\hat{\mathbf{\Sigma}}^{[s:t]}_{spec,\ell} - \mathbf{\Sigma}_\ell\| \leq a_{s,t}\}$$

with

$$a_{s,t} = C_e K^2 \|\mathbf{\Sigma}_0\| \{\sqrt{\gamma_l(\rho)} + \sqrt{\frac{tr(\mathbf{\Sigma}_0)K^*}{\|\mathbf{\Sigma}_0\|K}}\}\sqrt{\frac{-4\log(\frac{\alpha}{t})+\log(2d)}{t-s-l+1}}.$$

Denote $A = K^2\|\mathbf{\Sigma}_0\|$, and $B = (\frac{K^*}{K})^{1/2}\frac{\sqrt{tr(\mathbf{\Sigma}_0)}}{\|\mathbf{\Sigma}_0\|^{1/2}}$.

$$
\begin{aligned}
a_{s,t} &= C_e A\Big\{\sqrt{\gamma_l(\rho)} + B\Big\}\sqrt{\frac{\log(2d) - 4\log(\frac{\alpha}{t})}{t-s-l+1}} \\
&\leq C_E A\Big\{\sqrt{\gamma_l(\rho)} + B\Big\}\sqrt{\frac{\log(\frac{dt}{\alpha})}{t-s-l+1}}\, red(C_E = 2\sqrt{2}C_e) \\
&\leq C_E A\Big\{\sqrt{\gamma_l^*(\rho)} + B\Big\}\sqrt{\frac{\log(\frac{dt}{\alpha})}{t-s-l+1}}\, red(log(t) < d) \\
&\leq C_E'\sqrt{\frac{\log(\frac{dt}{\alpha})}{t-s-l+1}},
\end{aligned}
$$

where $\gamma_l^*(\rho) = \frac{d}{\log 2}\, max(1, \frac{8log(d\eta_\ell)+48d}{log\rho^{-1}})$ and $C_E = C_e' A\Big\{\sqrt{\gamma_l^*(\rho)} + B\Big\}$.

For any $t \leq \Delta$,

$$
\begin{aligned}
\|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[s:t]} - \hat{\mathbf{\Sigma}}_{spec,\ell}^{[1:s]}\| &\leq \|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[1:s]} - \mathbf{\Sigma}_\ell^{\Delta}\| + \|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[s:t]} - \mathbf{\Sigma}_\ell^{\Delta}\| \\
&= C_E\Big\{\sqrt{\frac{\log(\frac{dt}{\alpha})}{t-s-l+1}} + \sqrt{\frac{\log(\frac{ds}{\alpha})}{s-\ell}}\Big\} = \xi_{s,t}.
\end{aligned}
$$

Consequently, we have $\hat{\Delta} > \Delta$.

For $t > \Delta$, we are seeking an upper bound for $\hat{\Delta}$ or $\hat{\Delta} - \Delta$, and it will immediately yield **??**. Let $\tilde{\Delta} = min\big\{t > \Delta : \|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[\Delta:t]} - \hat{\mathbf{\Sigma}}_{spec,\ell}^{[1:\Delta]}\| \geq \xi_{\Delta,t}\big\}$ and $\epsilon = \tilde{\Delta} - \Delta$. Then $\hat{\Delta} - \Delta \leq \epsilon$. Hence, it suffices to find an upper bound for ϵ.

By the definition of $\tilde{\Delta}$, we have

$$
\begin{aligned}
\|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[\Delta:\tilde{\Delta}]} - \hat{\mathbf{\Sigma}}_{spec,\ell}^{[1:\Delta]}\| &= \|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[\Delta:\tilde{\Delta}]} - \mathbf{\Sigma}_\ell^{\Delta+1} + \mathbf{\Sigma}_\ell^{\Delta+1} - \mathbf{\Sigma}_\ell^{\Delta} + \mathbf{\Sigma}_\ell^{\Delta} - \hat{\mathbf{\Sigma}}_{spec,\ell}^{[1:\Delta]}\| \\
&\geq \|\mathbf{\Sigma}_\ell^{\Delta+1} - \mathbf{\Sigma}_\ell^{\Delta}\| - \big(\|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[\Delta:\tilde{\Delta}]} - \mathbf{\Sigma}_\ell^{\Delta+1}\| + \|\mathbf{\Sigma}_\ell^{\Delta} - \hat{\mathbf{\Sigma}}_{spec,\ell}^{[1:\Delta]}\|\big) \\
&\geq \kappa - \xi_{\Delta,\tilde{\Delta}}
\end{aligned}
$$

Thus, a proper upper bound of ϵ can be obtained if $\kappa - \xi_{\Delta,\tilde{\Delta}} \geq \xi_{\Delta,\tilde{\Delta}}$.

It now suffices to show that with the choice of

$$\sqrt{\epsilon - \ell} = \kappa^{-1} C_\epsilon \sqrt{\log(\frac{2d\Delta}{\alpha})},$$

it holds that $\kappa - \xi_{\Delta,\Delta+\epsilon} \geq \xi_{\Delta,\Delta+\epsilon}$, $i.e.$ $\kappa \geq 2\xi_{\Delta,\Delta+\epsilon}$.

Due to the signal ratio in 14,

$$\begin{aligned}
2\xi_{\Delta,\Delta+\epsilon} &= 2C_E\Big\{\sqrt{\frac{\log(\frac{d(\Delta+\epsilon)}{\alpha})}{\epsilon+1-\ell}} + \sqrt{\frac{\log(\frac{d\Delta}{\alpha})}{\Delta-\ell}}\Big\} \\
&\leq 2C_E\Big\{\sqrt{\frac{\log(\frac{d(2\Delta)}{\alpha})}{\epsilon+1-\ell}} + \sqrt{\frac{\log(\frac{d\Delta}{\alpha})}{\Delta-\ell}}\Big\} \\
&\leq 2C_E\sqrt{\log(\frac{2d\Delta}{\alpha})}\sqrt{\frac{1}{\epsilon+1-\ell}} + \kappa\frac{2C_E}{C_{SNR}} \\
&\leq \kappa\frac{2C_E}{C_\epsilon} + \kappa\frac{2C_E}{C_{SNR}} \\
&\leq \kappa,
\end{aligned}$$

where $max(\frac{C_E}{C_{SNR}}, \frac{C_E}{C_\epsilon}) < \frac{1}{4}$.

□

3.6.3 Proof of Corollary 3.3.1

Step 1. For any $t > 2 + \ell$, we define

$$\mathcal{S}(t) = \{t - 2^0 - \ell, t - 2^1 - \ell, \ldots, t - 2^{\lfloor \log(t)/\log 2 \rfloor} - \ell\}.$$

Assume $\Delta = \infty$. Define

$$\mathcal{B} = \big\{\forall s, t \in \mathbb{N}^+, t > 2 + \ell \; and \; s \in \mathcal{S}(t) : \; \|\hat{\mathbf{\Sigma}}_{ele,\ell}^{[s:t]} - \mathbf{\Sigma}_\ell\|_{max} \leq \xi_{s,t}\big\},$$

where $\xi_{s,t} = C_E\Big\{\sqrt{\frac{\log(\frac{dt}{\alpha})}{t-s+1-\ell}} + \sqrt{\frac{\log(\frac{ds}{\alpha})}{s-\ell}}\Big\}$. From lemma 3.6.1 and theorem 3.3.1, it follows that $\mathbb{P}(\mathcal{B}) \geq 1 - \alpha$.

Step 2. We define $\tilde{\Delta}^* = min\big\{t > \Delta : \|\hat{\mathbf{\Sigma}}_{spec,\ell}^{[\Delta:t]} - \hat{\mathbf{\Sigma}}_{spec,\ell}^{[1:\Delta]}\| \geq \xi_{\Delta,t}\big\}$. It is not hard to see that $\tilde{\Delta}^* \geq \tilde{\Delta}$ from the previous cases. With the choice of

$$\sqrt{\epsilon - \ell} = \kappa^{-1} \, C_\epsilon^{'} \sqrt{\log(\frac{2d\Delta}{\alpha})},$$

where $C_{\epsilon}^{'}\epsilon$, it holds that $\kappa - \xi_{\Delta,\Delta+\epsilon} \geq \xi_{\Delta,\Delta+\epsilon}$, $i.e.$ $\kappa \geq 2\xi_{\Delta,\Delta+\epsilon}$.

3.6.4 Proof of Proposition 3.4.1

Step 1. Let $\mathcal{F}_n$ be the σ-field generated by the observations $Y_1, Y_2, \ldots, Y_n$, and let P_n be the restriction of a distribution P^n to $\mathcal{F}_n$. For any $\nu \geq 1$ and $n \geq \nu$,

$$\frac{dP_{\kappa,\sigma_0,\nu}^n}{dP_{\kappa,\sigma_0,\infty}^n} = \exp\left(\sum_{i=\nu+1}^{n} Z_i\right), \tag{3.6.1}$$

where $P_{\kappa,\sigma_0,\infty}^n$ is the joint distribution under which there is no change point and

$$Z_i = \frac{1}{2}\log\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2}\left(\frac{Y_i^2}{\sigma_1^2}\right).$$

We assume that $\kappa = \sigma_1^2 - \sigma_0^2 > 0$, and for any $\nu > 1$, define the event

$$\mathcal{C}_\nu = \left\{ \nu < T < \nu + \frac{\sigma_0^4}{2\kappa^2} \log\left(\frac{1}{\alpha}\right), \sum_{i=\nu+1}^{T} Z_i < \frac{5}{2} \log\left(\frac{1}{\alpha}\right) \right\}.$$

By 3.6.1, we have

$$\begin{aligned}
\mathbb{P}_{\kappa,\sigma_0,\nu}(\mathcal{C}_\nu) = \int_{\mathcal{C}_\nu} dP^n_{\kappa,\sigma_0,\nu} = \int_{\mathcal{C}_\nu} \exp\left(\sum_{i=\nu+1}^{n} Z_i \right) dP^n_{\kappa,\sigma_0,\infty} \leq \exp\{(5/2)\log(1/\alpha))\} \mathbb{P}_{\kappa,\sigma_0,\infty}(\mathcal{C}_\nu) \\
\leq \exp\{(5/2)\log(1/\alpha))\}\, \mathbb{P}_{\kappa,\sigma_0,\infty}\left\{ \nu < T < \nu + \frac{\sigma_0^4}{\kappa^2} \log\left(\frac{1}{\alpha}\right) \right\} \leq \alpha^{5/2}\alpha = \alpha^{7/2},
\end{aligned} \tag{3.6.2}$$

where the last inequality follows from the definition of $\mathcal{D}(\alpha)$, $\mathbb{P}_{\kappa,\sigma_0,\infty}\left\{ \nu < T < \nu + \frac{\sigma_0^4}{\kappa^2} \log\left(\frac{1}{\alpha}\right) \right\} \leq \mathbb{P}_{\kappa,\sigma_0,\infty}\Big\{ T < \infty \Big\} \leq \alpha$.

Step 2. We assume for simplicity that $\frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})$ is an integer. Since $\{T \geq \nu\} \in \mathcal{F}_{\nu-1}$,

$$\begin{aligned}
&\mathbb{P}_{\kappa,\sigma_0,\nu}\left\{ \nu < T < \nu + \frac{\sigma_0^4}{2\kappa^2} \log\left(\frac{1}{\alpha}\right), \sum_{i=\nu+1}^{T} Z_i \geq (5/2)\log(1/\alpha) \,\Big|\, T > \nu \right\} \\
&\leq \text{ess sup}\, \mathbb{P}_{\kappa,\sigma_0,\nu}\left\{ \max_{1 \leq t \leq \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})-1} \sum_{i=\nu+1}^{t} Z_i \geq (5/2)\log(1/\alpha) \,\Big|\, Y_1, Y_2, \ldots, Y_\nu \right\} \\
&\leq \text{ess sup}\, \mathbb{P}_{\kappa,\sigma_0,\nu}\left\{ \max_{1 \leq t \leq \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})-1} \frac{\kappa}{2\sigma_0^2} \sum_{i=\nu+1}^{t} \left(\frac{Y_i^2}{\sigma_1^2}\right) \geq (5/2)\log(1/\alpha) \,\Big|\, Y_1, Y_2, \ldots, Y_\nu \right\} \\
&\leq \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})\, \text{ess sup}\, \mathbb{P}_{\kappa,\sigma_0,\nu}\left\{ \frac{\kappa}{2\sigma_0^2} \sum_{i=\nu+1}^{\frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})} \left(\frac{Y_i^2}{\sigma_1^2} - 1\right) \geq \log(1/\alpha) \left[(5/2) - \frac{\sigma_0^2}{4\kappa} \right] \,\Big|\, Y_1, Y_2, \ldots, Y_\nu \right\} \\
&\leq \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})\, \text{ess sup}\, \mathbb{P}_{\kappa,\sigma_0,\nu}\left\{ \frac{\kappa}{2\sigma_0^2} \sum_{i=\nu+1}^{\frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})} \left(\frac{Y_i^2}{\sigma_1^2} - 1\right) \geq \log(1/\alpha) \,\Big|\, Y_1, Y_2, \ldots, Y_\nu \right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})\ \exp\left\{-\min\Big[\frac{\log^2(1/\alpha)}{8\,\frac{\sigma_0^4}{2\kappa^2}\log(1/\alpha)\frac{\kappa^2}{4\sigma_0^4}},\ \frac{\log(1/\alpha)}{8\,\frac{\kappa}{2\sigma_0^2}}\Big]\right\}\\
&\leq \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})\ \exp\left\{-\min\Big[\log(1/\alpha),\ \frac{\log(1/\alpha)}{\frac{4\kappa}{\sigma_0^2}}\Big]\right\}\\
&\leq \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})\alpha\\
&\leq 18\log(\frac{1}{\alpha})\alpha\\
&\leq \alpha^{\frac{1}{4}},
\end{aligned}
$$

where the second inequality holds due to $\log(\frac{\sigma_0^2}{\sigma_1^2}) < 0$, the third one is given by the union bound argument, the fourth inequality holds for $\kappa > \frac{1}{6}\sigma_0^2$, the fifth one is by the Berinstain's inequality for sub-exponential variables (Theorem 2.8.1 in Vershynin, 2018), the seventh inequality holds provided that $\kappa < \frac{1}{4}\sigma_0^2$, and the last inequality is met for sufficiently small α.

Because the upper bound is independent of ν, then we have

$$
\sup_{\nu\geq 1}\mathbb{P}_{\kappa,\sigma_0,\nu}\left\{\nu < T < \nu + \frac{\sigma_0^4}{2\kappa^2}\log\left(\frac{1}{\alpha}\right),\ \sum_{i=\nu+1}^{T} Z_i \geq (5/2)\log(1/\alpha)\right\} \leq \alpha^{\frac{1}{4}}. \tag{3.6.3}
$$

Then, combining 3.6.2 and 3.6.3, we have

$$
\mathbb{P}_{\kappa,\sigma_0,\nu}\left\{\nu < T < \nu + \frac{\sigma_0^4}{2\kappa^2}\log\left(\frac{1}{\alpha}\right)\right\} \leq \alpha^{\frac{7}{2}} + \alpha^{\frac{1}{4}}. \tag{3.6.4}
$$

Step 3. For any change point Δ,

$$\begin{aligned}
&\mathbb{E}_{\kappa,\sigma_0,\Delta}\Big\{(T-\Delta)_+\Big\} \geq \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})\,\mathbb{P}_{\kappa,\sigma_0,\Delta}\Big\{T-\Delta \geq \frac{\sigma_0^4}{2\kappa^2}\Big\} \\
&= \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})\left\{\mathbb{P}_{\kappa,\sigma_0,\Delta}\Big\{T \geq \Delta\Big\} - \mathbb{P}_{\kappa,\sigma_0,\Delta}\Big\{\Delta \leq T \leq \Delta + \frac{\sigma_0^4}{2\kappa^2}\Big\}\right\} \\
&= \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})\left\{1 - \mathbb{P}_{\kappa,\sigma_0,\infty}\Big\{T < \Delta\Big\} - \mathbb{P}_{\kappa,\sigma_0,\Delta}\Big\{\Delta \leq T \leq \Delta + \frac{\sigma_0^4}{2\kappa^2}\Big\}\right\} \\
&\geq \frac{\sigma_0^4}{2\kappa^2}\log(\frac{1}{\alpha})\,(1-\alpha-\alpha^{\frac{7}{2}}-\alpha^{\frac{1}{2}}) \geq \frac{\sigma_0^4}{4\kappa^2}\log(\frac{1}{\alpha}),
\end{aligned} \tag{3.6.5}$$

where $\alpha + \alpha^{\frac{7}{2}} + \alpha^{\frac{1}{4}} < \frac{1}{2}$ holds for sufficiently small α, and $\mathbb{P}_{\kappa,\sigma_0,\Delta}\Big\{T < \Delta\Big\} = \mathbb{P}_{\kappa,\sigma_0,\infty}\Big\{T < \Delta\Big\}$.

Define $\mathcal{S}_\alpha = \big\{\alpha \in (0,1) \mid 18\log(\frac{1}{\alpha})\alpha \leq \alpha^{\frac{1}{4}} \text{ and } \alpha + \alpha^{\frac{7}{2}} + \alpha^{\frac{1}{4}} < \frac{1}{2}\big\}$. It is easy to verify that $\mathcal{S}_\alpha \neq \emptyset$. For example, $\alpha = 1/1000 \in \mathcal{S}_\alpha$ or $\{\alpha \mid \alpha \leq 1/1000\} \subseteq \mathcal{S}_\alpha$.

$$\begin{aligned}
&\mathbb{P}_{\kappa,\sigma_0,\nu}\left\{\nu < T < \nu + \frac{\sigma_0^2}{2\kappa}\log\left(\frac{1}{\alpha}\right),\ \sum_{i=\nu+1}^{T} Z_i \geq (3/4)\log(1/\alpha)\Big|\, T > \nu\right\} \\
&\leq \text{ess sup}\,\mathbb{P}_{\kappa,\sigma_0,\nu}\left\{\max_{1\leq t\leq \frac{\sigma_0^2}{2\kappa}\log(\frac{1}{\alpha})-1}\sum_{i=\nu+1}^{t} Z_i \geq (3/4)\log(1/\alpha)\,\Big|\, Y_1, Y_2, \ldots, Y_\nu\right\} \\
&\leq \text{ess sup}\,\mathbb{P}_{\kappa,\sigma_0,\nu}\left\{\max_{1\leq t\leq \frac{\sigma_0^2}{2\kappa}\log(\frac{1}{\alpha})-1}\frac{\kappa}{2\sigma_0^2}\sum_{i=\nu+1}^{t}\left(\frac{Y_i^2}{\sigma_1^2}\right) \geq (3/4)\log(1/\alpha)\Big|\, Y_1, Y_2, \ldots, Y_\nu\right\} \\
&\leq \frac{\sigma_0^2}{2\kappa}\log(\frac{1}{\alpha})\,\text{ess sup}\,\mathbb{P}_{\kappa,\sigma_0,\nu}\left\{\frac{\kappa}{2\sigma_0^2}\sum_{i=\nu+1}^{\frac{\sigma_0^2}{2\kappa}\log(\frac{1}{\alpha})}\left(\frac{Y_i^2}{\sigma_1^2}-1\right) \geq (1/2)\log(1/\alpha)\Big|\, Y_1, Y_2, \ldots, Y_\nu\right\} \\
&\leq \frac{\sigma_0^2}{2\kappa}\log(\frac{1}{\alpha})\ \exp\left\{-\min\Big[\frac{(1/2)^2\log^2(1/\alpha)}{8\,\frac{\sigma_0^2}{2\kappa}\log(1/\alpha)\frac{\kappa^2}{4\sigma_0^4}}, \frac{(1/2)\log(1/\alpha)}{8\,\frac{\kappa}{2\sigma_0^2}}\Big]\right\} \\
&\leq \frac{\sigma_0^2}{2\kappa}\log(\frac{1}{\alpha})\exp\left\{-(1/8)\log(1/\alpha)\frac{\sigma_0^2}{\kappa}\right\} \\
&\leq \frac{\sigma_0^2}{2\kappa}\log(\frac{1}{\alpha})\exp\{-\alpha\}
\end{aligned}$$

$$\leq \frac{\sigma_0^2}{2\kappa} \log(\frac{1}{\alpha})(\alpha + 1)$$

$$\leq \alpha^{\frac{1}{4}}$$

Chapter 4

Applications to pandemic time series data

Since early 2020, the Covid-19 pandemic has spread rapidly across the globe and brought significant impact on human's lives, leading to widespread illness, death, and economic disruption. The mortality data over the two years contain valuable information for us to understand the evolution of varus, progression of the pandemic and the influence of key public health policies. In the first part, we conduct a retrospective change point analysis on the U.S. state-level mortality data by implementing MACD, the offline procedure in chapter 2. In the second part, we assume the online setting and attempt to put online detection in a broader picture. We incorporate the change points into time series modeling to facilitate the forecasting. Results show that with the assistance of change points, the accuracy and efficiency of prediction can be greatly improved.

4.1 Retrospective change point analysis

The data set is collected from the New York Times Github repository, containing the state-level daily death data from 3/18/2020 to 8/16/2022 in the 48 adjoining U.S. states and the District of Columbia. Instead of using the original scale of the data, we use 7-day log return to adjust for the weekly seasonality. To be specific, denote $\boldsymbol{X}_t$ as the daily death counts observed at time t. We define $\boldsymbol{Y}_t = \log(\boldsymbol{X}_t) - \log(\boldsymbol{X}_{t-7})$ as the 7-day log return of $\boldsymbol{X}_t$. In addition, we choose window size W = 180 and consider three different lags 0, 1, and 7 to analyze this data set. Note that, if W =180, we assume the minimum distance between two consecutive change points is at least 180 days.

Before implementing the method, it is necessary to check if the data present the heavy-tailedness by their kurtosis. Figure 4.1 displays the presence of heavy-tailedness in most of the states and minimum kurtosis value across states is around 6.8. Hence, it is appropriate to apply the robust autocovariance change point algorithm on this data set.

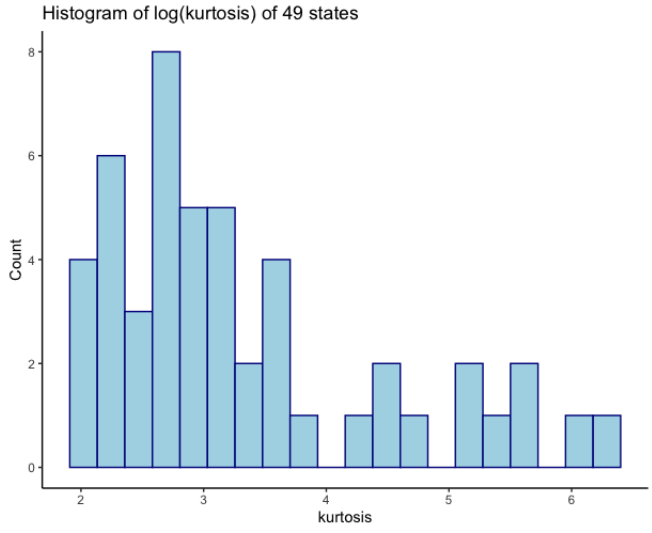

Figure 4.1: Histogram of Kurtosis of 49 states in logarithm scale

Table 4.1: Offline second-order change points in U.S. state level COVID-19 mortality data

Lag	Date (YYYY-MM-DD)
$\ell = 0$	2020-09-20, 2021-05-17, 2021-12-13
$\ell = 1$	2021-01-29
$\ell = 7$	2021-01-07

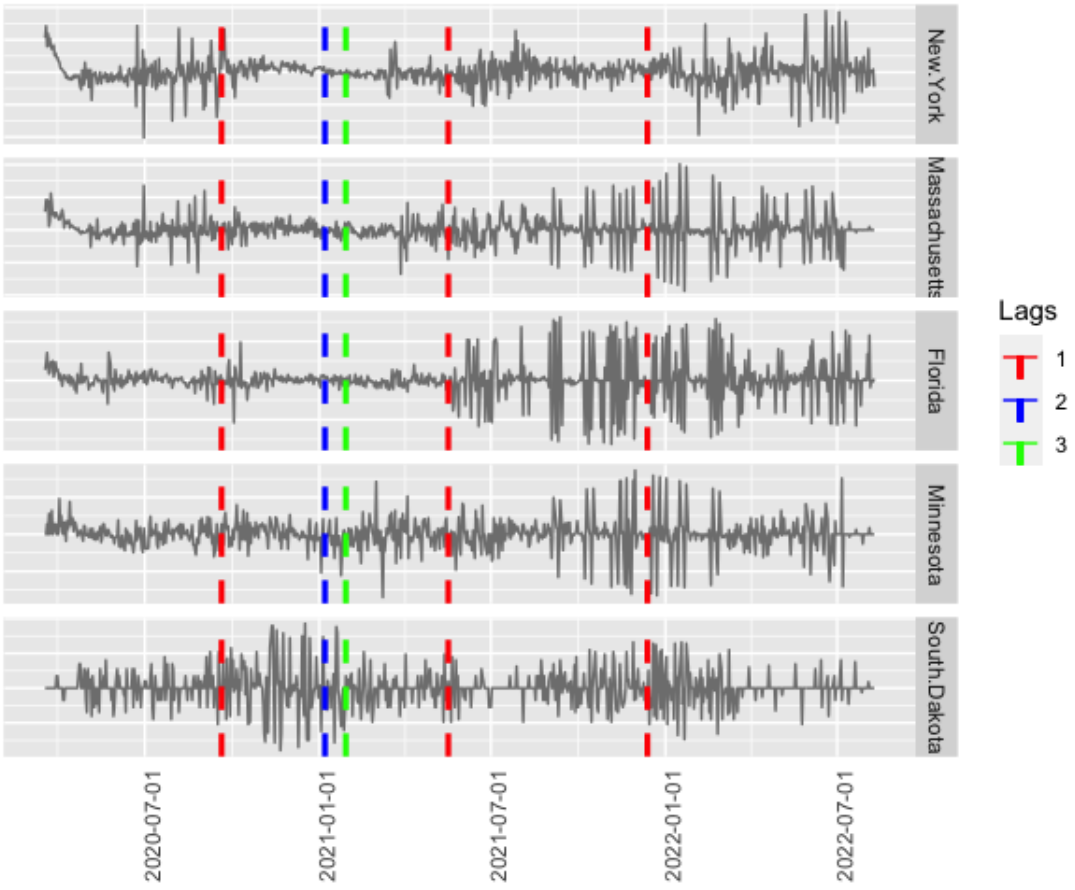

Figure 4.2: Seven-days log returns of five states with change points indicated

Fig. 4.2 displays the 7-days log returns of five representative states and the segmentation by detected change points. It contains information about the diagonal of autocovariance or covariance matrix while

the association between states cannot be visualized by this graph. As we can see, those change points discovered by MACD can capture the the variation of second-order structure well.

The first change-point on 2020-9-20 under $\ell = 0$ is likely to result from the resurgence of COVID-19 cases and deaths due to increased travel and large gathering for events related to the U.S. presidential election. FDA issues emergency use authorizations for Pfizer-BioNTech and Moderna Covid-19 vaccines in December and vaccine doses begin to be administered. Nevertheless, in the early 2021, the Alpha variant starts to become the predominant variant in US, which is more contagious and deadlier than the original virus. Consequently, it is reasonable to see a change point around January, as suggested by the results for $\ell = 1$ and $\ell = 7$. The change point 2021-05-17 discovered under $\ell = 0$ could be explained by the widespread vaccination by which more than 200 million COVID-19 vaccine doses have been administered in the U.S, leading to steady decline of mortality rate. On the other hand, the arrival of Delta variant rapidly reverse that trend which is more transmissible than Alpha and can cause more severe disease to unvaccinated people. In the midst of a Delta variant spike, CDC recommends taking the booster shot and wearing mask indoors, which mitigate the situation. However, as Omicron arrives in October, the case number in U.S. skyrocket again. This can explain why we find the last change point on 2021-12-13. The retrospective analysis above provides some hypotheses that policy changes and significant events in the COVID-19 timeline may lead to fundamental second-order structure changes in state-level mortality data. The findings may motivate a line of public health research to validate the causal relationship between policies and temporal patterns of mortality data during the pandemic.

The main purpose of offline CPD is to facilitate the retrospective study. While in the online streaming setting, new observation arrive steadily and people can take actions once a change point is identified in the system. In the next section, we attempt to put online change point detection in a broader picture

by combining it with time series modeling. Roughly speaking, we want to update the forecasting model once a change point is found instead of ignoring it.

4.2 Change point assisted online forecasting

The importance of structural stability cannot be overlooked in statistical modeling of time series. In particular, if the underlying process undergoes unexpected changes, it can cause a loss of accuracy in the forecasts. In this study, we adopt the online approach in chapter 3 to sequentially detect change points in the second order. Those change points serve as break times when the forecasting model will restart to train with the new observations. This is an ongoing project while preliminary results have shown that with the assistance of change points, the accuracy of forecasting is improved in terms of sMAPE and efficiency of training and prediction process is greatly boosted.

The data set is collected from 'Our World In Data' (Ritchie et al., n.d.), aggregated from multiple sources. The data ranges from 2020-07-24 to 2022-08-14 and contains the national level daily death data and other important information about cases, hospitalizations, ICU patients, etc. Here we focus on the daily death count for monitoring change points, some of the other variables may be of interest in building prediction models. Again, the kurtosis of the data series is around 4.1, indicating that heavy-tailedness is present.

We adopt the online second-order change point detection approach in chapter 3. We choose $\alpha = 0.01$ and take 100 data points as calibration set to obtain the threshold. To dealing with multiple change points, every time we encounter a change point, the threshold is re-calibrated. Table 4.2 displays the sequentially detected change points. Note that the results here differs from the previous offline results. Moreover,

2021-02-17, 2021-06-08 and 2021-11-26 found in this table are close to 2021-01-29, 2021-05-17 and 2021-12-13 respectively in table 4.1 whereas other dates differs a lot. It could be caused by multiple reasons. First, in the offline detection we set window size equal to 180, which, indicated by our binary segmentation algorithm that the minimum space between two consecutive change points is assumed to be at least 180. But this is not the case here for the sequential detection. The second reason is that, online procedure would be more sensitive to the abrupt changes because it make decisions based on the historical values instead of the whole sequence that is available in the offline setting. As a result, it is not surprising to see more and different change points discovered here in table 4.2.

Table 4.2: Online second-order change points in U.S. national level COVID-19 mortality data

Lag	Date (YYYY-MM-DD)
$\ell = 0$	2021-06-08, 2021-09-13, 2021-11-26, 2022-04-24, 2022-06-25
$\ell = 1$	2021-02-17, 2021-07-13, 2021-11-26, 2022-05-04, 2022-07-03
$\ell = 7$	2021-06-10, 2022-04-17, 2022-06-22

Unstable underlying structure of the time series poses challenges on forecasting. We consider a hybrid modeling process that can adapt to change points. The idea is simple, every time when we encounter a shift, we refit the model with new observations after that. In table 4.3, we compare in terms of sMAPE the CP model using CPs at $\ell = 0$ with baseline model which ignore the existence of CPs. The choices of forecasting approaches has been rapidly grown in recent years. Here we choose eight predictors coming of three major types: traditional statistical model (AR and AR(L1)), tree-based model (RF,XGB) and neural network model (TCN, LSTM, GRU and TRF). Results show that after introducing CPs, the accuracy of forecasting is improved. Additionally, the time spent for each method is dramatically reduced.

Table 4.3: Comparison of baseline model and change point assisted model on Covid-19 mortality data by sMAPE and time spent

sMAPE	AR	AR(L1)	RF	XGB	TCN	LSTM	GRU	TRF
Baseline Model	29.71	29.30	29.50	31.98	27.26	27.27	27.36	27.32
CP Model	26.70	26.96	27.69	30.63	26.91	26.72	26.35	26.07
WALL-CLOCK TIME(s)	AR	AR(L1)	RF	XGB	TCN	LSTM	GRU	TRF
Baseline Model	304.33	667.93	40228.61	46165.92	5068.33	40314.98	91233.22	27741.20
CP Model	2.46	2.59	414.21	234.18	21.62	178.33	177.16	113.98
# of Parameter Tuned	AR	AR(L1)	RF	XGB	TCN	LSTM	GRU	TRF
	2	3	4	6	4	3	3	7

Bibliography

Adams, R. P., & MacKay, D. J. C. (2007). Bayesian online changepoint detection [https://arxiv.org/abs/0710.3742].

Aue, A., Horva, L., Kühn, M., & Steinebach, J. (2012). On the reaction time of moving sum detectors. *Journal of Statistical Planning and Inference*, *142*, 2271–2288.

Avanesov, V., & Buzun, N. (2021). Change-point detection in high-dimensional covariance structure. *Bernoulli*, *27*(1), 554–575.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory*, *13*(3), 315–352.

Bellman, E., & Dreyfus, S. E. (1962). *Applied dynamic programming*. Princeton.

Chen, H. (2019). Sequential change-point detection based on nearest neighbors. *The Annals of Statistics*, *47*(3), 1381–1407.

Chen, H., & Zhang, N. (2015). Graph-based change point detection. *The Annals of Statistics*, *43*(1), 139–176.

Chen, J., & Gupta, A. K. (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association*, *92*(438), 739–747.

Chen, Y., Wang, T., & Samworth, R. J. (2022). High-dimensional, multiscale online changepoint detection. *Journal of the Royal Statistical Society (Series B)*, *84*, 234–266.

Chen, Z., & Tian, Z. (2010). Modified procedures for change point monitoring in linear models. *Mathematics and Computers in Simulation*, *81*(1), 62–75.

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

Cho, H. (2016b). Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics*, *10*(2), 2000–2038. https://doi.org/10.1214/16-EJS1155

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.

Choi, H., Ombao, H., & Ray, B. (2008). Sequential change-point detection methods for nonstationary time series. *Technometrics*, *50*(1), 40–52.

Chowdhury, M., Selouani, S., & O'Shaughnessy. (2012). Bayesian on-line spectral change point detection: A soft computing approach for on-line asr. *International Journal of Speech Technology*, *15*.

Chu, C.-S. J., Stinchcombe, M., & White, H. (1996). Monitoring structural change. *Econometrica*, *64*(5), 1045–1065.

Cook, D. J., & Krishnan, N. C. (2015). Activity learning: Discovering, recognizing, and predicting human behavior from sensor data. *Wiley*.

Detecting changes in the covariance structure of functional time series with application to fmri data. (2021). *Econometrics and Statistics*, *18*, 44–62.

Dubey, P., Xu, H., & Yu, Y. (2021). Online network change point detection with missing values. https://doi.org/10.48550/ARXIV.2110.06450

Ducré-Robitaille, J.-F., Vincent, L. A., & Boulet, G. (2003). Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology*, *23*(9), 1087–1101. https://doi.org/https://doi.org/10.1002/joc.924

Eiauer, P., & Hackl, P. (1978). The use of mosums for quality control. *Technometrics*, *20*(4), 431–436.

Fearnhead, P., & Rigaill, G. (2019). Changepoint detection in the presence of outliers. *Journal of the American Statistical Association*, *114*(525), 169–183.

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

Gao, Z., Lu, G., Yan, P., Lyu, C., Li, X., Shang, W., Xie, Z., & Zhang, W. (2018). Automatic change detection for real-time monitoring of eeg signal. *Frontiers in physiology*.

Gösmann, J., Stoehr, C., Heiny, J., & Dette, H. (2022). Sequential change point detection in high dimensional time series. *Electronic Journal of Statistics*, *16*, 3608–3671.

Harchaoui, Z., Vallet, F., Lung-Yut-Fong, A., & Cappe, O. (2009). A regularized kernel-based approach to unsupervised audio segmentation. *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1665–1668. https://doi.org/10.1109/ICASSP.2009.4959921

Horvath, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *The Annals of Statistics*, *21*(2), 671–680.

Horváth, L., Kühn, M., & Steinebach, J. (2008). On the performance of the fluctuation test for structural change. *Sequential Analysis*, *27*(2), 126–160.

Inclan, C., & Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, *89*(427), 913–923.

Itoh, N., & Kurths, J. (2010). Change-point detection of climate time series by nonparametric method. *Proceedings of the world congress on engineering and computer science*, *1*, 445–448.

Jackson, B., Scargle, J., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., & Tsai, T. (2005). An algorithm for optimal partitioning of data on an interval. *Signal Processing Letters, IEEE*, *12*, 105–108.

Jandhyala, V., Fotopoulos, S. B., Macneill, I., & Liu, P. (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, *34*.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics*, *43*(6), 2451–2483.

Kawahara, Y., & Sugiyama, M. (2009). Sequential change-point detection based on direct density-ratio estimation. *SIAM International Conference on Data Mining*, 389–400.

KD, F., Cook DJ, R. C., K, R., & M, S.-E. (2015). Automated detection of activity transitions for prompting. *IEEE Trans Hum Mach Syst*, (47), 575–585.

Keogh, E., Chu, S., Hart, D., & Pazzani, M. (2001). An online algorithm for segmenting time series. *Proceedings 2001 IEEE International Conference on Data Mining*, 289–296. https://doi.org/10.1109/ICDM.2001.989531

Keshavarz, H., Michailidis, G., & Atchad'e, Y. (2020). Sequential change-point detection in high-dimensional gaussian graphical models. *Journal of Machine Learning Research*, *21*(8), 1–57. https://doi.org/10.1109/LSP.2014.2381553

Keshavarz, H., Scott, C., & Nguyen, X. (2018). Optimal change point detection in gaussian processes. *Journal of Statistical Planning and Inference*, *193*, 151–178.

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*(500), 1590–1598.

Kirch, C., & Weber, S. (2018). Modified sequential change point procedures based on estimating functions. *Electronic Journal of Statistics*, *12*(1), 1579–1613.

Lai, T. L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory*, (44), 2917–2929.

Lavielle, M., & Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, *21*(1), 33–59.

Li, L., & Li, J. (2019). Online change-point detection in high-dimensional covariance structure with application to dynamic networks.

Li, S., Xie, Y., Dai, H., & Song, L. (2015a). M-statistic for kernel change-point detection. *Advances in Neural Information Processing Systems*, *28*.

Li, S., Xie, Y., Dai, H., & Song, L. (2015b). Scan b-statistic for kernel change-point detection. *Advances in Neural Information Processing Systems*.

Lindquist, M. A., Waugh, C., & Wager, T. D. (2007). Modeling state-related fmri activity using change-point theory. *NeuroImage*, *35*(3), 1125–1141.

Liu, X., Wu, X., Wang, H., Zhang, R., Bailey, J., & Ramamohanarao, K. (2010). Mining distribution change in stock order streams. *2010 IEEE 26th International Conference on Data Engineering*.

Lung-Yut-Fong, A., Lévy-Leduc, C., & Cappé, O. (2015). Homogeneity and change-point detection tests for multivariate data using rank statistics. *Journal de la société française de statistique*, *156*(4), 133–162.

Lynch, C., & Mestel, B. (2019). Change-point analysis of asset price bubbles with power-law hazard function. *International Journal of Theoretical and Applied Finance*, *22*(07).

Maidstone, R., Hocking, T., Rigaill, G., & Fearnhead, P. (2017). On optimal multiple changepoint algorithms for large data. *Statistics and Computing*.

Marangoni-Simonsen, D., & Xie, Y. (2015). Sequential changepoint approach for online community detection. *IEEE Signal Processing Letters*, *22*(8), 1035–1039. https://doi.org/10.1109/LSP.2014.2381553

Mei, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika*, *97*(2), 419–433.

Niu, Y. S., & Zhang, H. (2012). The screening and ranking algorithm to detect dna copy number variations. *The Annals of Applied Statistics*, *6*(3), 1306–1326.

Olshen, A. B., Venkatraman, E. S., Lucito, R., & Wigler, M. (2004). Circular binary segmentation for the analysis of array-based dna copy number data. *Biostatistics*, *5*(4), 557–572.

Page, E. (1954). Continuous inspection schemes. *Biometrika*, *41*.

Pein, F., Sieling, H., & Munk, A. (2017). Heterogeneous change point inference. *Journal of the Royal Statistical Society*, *79*(4), 1207–1227.

Ping, Y., G, D., & JM, A. (2006). Adaptive change detection in heart rate trend monitoring in anesthetized children. *IEEE transactions on bio-medical engineering*, *53*(11), 2211–2219.

Reeves, J., Chen, J., Wang, X. L., Lund, R., & Lu, Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, *46*. http://www.jstor.org/stable/26171947

Ritchie, H., Mathieu, E., Rodes-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, B. J., Beltekian, D., & Roser, M. (n.d.). *Coronavirus pandemic (covid-19)*. Our World in Data. https://ourworldindata.org/coronavirus

Romano, G., Eckley, I., Fearnhead, P., & Rigaill, G. (2021). Fast online changepoint detection via functional pruning cusum statistics. https://doi.org/10.48550/ARXIV.2110.08205

Rybach, D., Gollan, C., Schluter, R., & Ney, H. (2009). Audio segmentation for speech recognition using segment features. *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 4197–4200. https://doi.org/10.1109/ICASSP.2009.4960554

Saatchi, Y., Turner, R., & Rasmussen, C. (2010). Gaussian process change point models, 927–934.

Scott, A. J., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, *30*(3), 507–512.

Sen, A., & Srivastava, M. S. (1975). On tests for detecting change in mean. *The Annals of Statistics*, *3*(1), 98–108.

Shewart, W. A. (1931). *Economic control of quality of manufactured product*. New York: D. Van Norstrand.

Shi, X., Wu, Y., & Rao, C. R. (2017). Consistent and powerful graph-based change-point test for high-dimensional data. *Proceedings of the National Academy of Sciences*, *114*(15), 3873–3878.

Shu, H., & Nan, B. (2019). Estimation of large covariance and precision matrices from temporally dependent observations. *The Annals of Statistics*.

Takayasu, H. (2015). Basic methods of change-point detection of financial fluctuations. *2015 International Conference on Noise and Fluctuations (ICNF)*, 1–3. https://doi.org/10.1109/ICNF.2015.7288606

Truong, C., Oudre, L., & Vayatis, N. (2018). Selective review of change point detection methods. *CoRR*, *abs/1801.00718*. http://arxiv.org/abs/1801.00718

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*. Cambridge University Press. https://doi.org/10.1017/9781108231596

Wang, D., Yu, Y., & Rinaldo, A. (2020). Univariate mean change point detection: Penalization, cusum and optimality. *Electric Journal of Statistics*, *14*, 1917–1961.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, *102*(40), 14150–14154.

Wu, W. B., & Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statistica Sinica*, 1397–1413.

Xu, H., Ke, Y., Guerrier, S. e., & Li, R. (2021). Nonasymptotic theories for tail-robust autocovariance matrix estimation methods.

Yao, Y.-C., & Au., S. T. (1989). Least-squares estimation of a step function. *Sankhyā: The Indian Journal of Statistics*.

Yu, M., & Chen, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *Journal of the Royal Statistical Society*, *83*(2), 247–270.

Yu, Y., Padilla, O. H. M., Wang, D., & Rinaldo, A. (2020). A note on online change point detection [https://arxiv.org/abs/2006.03283v3].

Zhang, T., & Lavitas, L. (2018). Unsupervised self-normalized change-point testing for time series. *Journal of the American Statistical Association*, *113*, 637–648.

Zou, C., Yin, G., Feng, L., & Wang, Z. (2014)). Nonparametric maximum likelihood approach to multiple change-point problems. *The Annals of Statistics*, *42*(3), 970–1002.