THE UNIVERSITY OF CHICAGO

NON-PARAMETRIC CHANGE POINT ANALYSIS: TACKLING IRREGULAR CONDITIONS IN UNI-VARIATE AND MULTIVARIATE DATA

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS

BY
YUHAN PHILIP LIU

CHICAGO, ILLINOIS
JUNE 2024

To the innocent souls who suffered during the plague and wars of the past five years.

Remember my affliction and my wanderings,

the wormwood and the gall!

My soul continually remembers it

and is bowed down within me.

— Lamentations 3:19-20

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to express my profound gratitude to my esteemed advisor, Prof. Wei-Biao Wu, for his meticulous supervision and steadfast support throughout the course of my five-year academic journey. Our discussions have consistently served as a source of inspiration, catalyzing novel ideas and fostering critical thinking.

In the words of Edwin Arlington Robinson's "The Flying Dutchman":

"Alone, by the one light of his one thought,

He steers to find the shore from which he came,

Fearless of in what coil he may be caught

On seas that have no name."

Prof. Wu's encouragement to transcend my limitations and pursue academic excellence has been pivotal in my professional growth, enabling me to expand my intellectual horizons and attain new heights in my field of study. His resolute dedication to my scholarly development has been instrumental in shaping my academic trajectory, and I stand in humble acknowledgment that my current position is a testament to his mentorship.

I would like to convey my sincere appreciation to my committee members, Prof. Rebecca Willett, Prof. Mei Wang and Prof. Ekaterina Smetanina, for their profound insights and constructive critique of my work.

I am grateful to Prof. Claire Donnat, who advised my reading class during my second year, amidst the darkest days of the pandemic lockdown. I also thank Prof. Peter McCullagh and Prof. Mei Wang who, along with Prof. Claire Donnat, taught me a great deal through the Consulting Program and provided me with the excellent opportunity to lead a team in tackling real data problems across various domains.

I thank Prof. Rina Foygel Barber, Prof. Dan Nicolae, Prof. Matthew Stephens, Prof. Yali Amit, Prof. Chao Gao, Prof. Fei Liu, Prof. Mary Sara McPeek, Prof. Lek-Heng Lim and Prof. Yuxin Chen, for enriching my experience here and helping me better prepare for

the next stage of my career in industry. I also thank the faculty, staff, and my fellow students whom I have met at UChicago, especially Keisha Prowoznik, Mitzi Nakatsuka, and Kirsten Wellman.

Above all, I am forever indebted to my parents for their unconditional love. They are my guiding lights, epitomizing an ardent passion for family, work, and life.

ABSTRACT

Change point analysis has emerged as a critical area of research in Statistics over the past seven decades, with applications spanning various domains, including Physics, Epidemiology, Finance, and more. This thesis aims to develop innovative, non-parametric methodologies for change point analysis that can effectively operate without relying on stringent regularity conditions for the data, which are often violated in real-world scenarios. We introduce a novel framework for change point analysis that is liberated from strong regularity conditions for the signal of the data following the potential change point and propose a pioneering framework for statistical inference of change points in the presence of temporally dependent and non-stationary noise processes. Furthermore, we extend these frameworks to a multivariate setting, addressing the growing interest in detecting change points in multivariate or high-dimensional data.

We derive theoretical results to validate our proposed methodologies and demonstrate their efficacy through both synthetic data analysis and real data applications. While our proposed methods exhibit robust performance in low-dimensional settings, they are susceptible to the curse of dimensionality in high-dimensional scenarios. We encourage future research to build upon these contributions to overcome the challenges associated with high-dimensional change point analysis under these irregular conditions, paving the way for more reliable and accurate detection of change points in complex, real-world data.

CHAPTER 1

INTRODUCTION

In 1955, Ewan Stafford Page, a British academic, published two seminal papers in Biometrika that laid the foundation for the field of change point analysis in modern Statistics. The first paper, Page [1955a], introduces a statistical test to detect a shift in the distribution parameter of a sequence of independent observations at an unknown point. The second paper, Page [1955b], presents the innovative concept of using warning lines in addition to action lines on control charts, enabling the detection of process shifts based on a sequence of observations rather than a single point. Although Page's work primarily focuses on industrial quality control applications, the concepts and methods introduced have had a profound impact on the development of change point analysis as a broader statistical field.

Over the past seven decades, change point analysis has become an active and popular research area in Statistics due to its important applications and close connections across various research domains. Recent examples showcase the wide-ranging applications of change point analysis. In Environmental Sciences, Dar and Asif [2017] apply change point analysis techniques to investigate the long-run impact of financial development, energy consumption, and economic growth on greenhouse gas emissions in India, revealing two regime shifts and highlighting the importance of incorporating these methods in environmental economic research. In the area of Experimental Psychology, Williams et al. [2022] introduce a simple signal detection model that accurately captures how we detect changes in visual memory tasks. In Finance and Economics Research, Abosedra et al. [2021] explore the dynamics of financial deepening and business cycle volatility in the UAE, using cointegration tests with change points to reveal asymmetrical responses in the business cycle to changes in financial deepening. In the field of Computational Neuroscience, Liu and Li [2022] propose a multi-sensory pathway network (MSPN) framework, inspired by the parallel processing mechanism of human visual information, for change detection in multi-temporal images.

A critical challenge in change point analysis is identifying shifts in signals. Extensive studies have focused on the scenario where only a single change point is assumed to exist. The initial step involves testing for the presence of any change point. Upon rejecting the null hypothesis, which asserts the absence of a change point, the analysis proceeds to estimate the precise location of the change point, as pioneered by Hawkins [1977]. This subsequent task is non-trivial, even within the framework of the normal and homoskedastic model or the simpler one-parameter exponential family, as discussed in seminal works by Sen and Srivastava [1975], Hinkley [1970], Worsley [1986], Siegmund [1988].

Despite the impressive performance of these methodologies in their synthetic data analysis, their effectiveness cannot be assured when applied to real-world data, particularly in domains such as financial market analysis or pandemic research. Data from these sources are often characterized by highly volatile signals and noise that is not only strong but also non-stationary and temporally dependent. The classical methods discussed above typically rely on stringent regularity conditions for both the signal and the noise components, which are often violated in these real-world scenarios. Consequently, the failure to meet these assumptions leads to suboptimal performance and unreliable results when these methods are employed in practice. The inherent complexity and unpredictability of real-world data pose significant challenges for traditional change point analysis techniques, necessitating the development of more robust and adaptive approaches that can effectively handle the intricacies of such data without being constrained by rigid assumptions.

The primary objective of this thesis is to develop innovative frameworks for change point analysis that can operate effectively without relying on stringent regularity conditions for the data. The methodologies presented in this work are non-parametric, ensuring that their strong theoretical foundations and empirical performance are not contingent upon specific parametric model assumptions for the noise sequence. This approach stands in stark contrast to classical change point analysis methods, such as CUSUM, which require the noise process

to adhere to an i.i.d. Gaussian distribution. In practice, particularly in the domains of our interest, such rigid parametric assumptions are often untenable, underlining the superiority of our novel methodologies in addressing real-world challenges.

Chapter 2 introduces a novel framework for change point analysis that is liberated from the confines of strong regularity conditions for the signal of the data following the potential change point. This allows for the data to exhibit highly erratic behavior after a potential change point, a scenario frequently encountered in real-world applications. We present a sophisticated locating algorithm designed to pinpoint the initial change point with precision, capturing the critical moment when the data transitions from a state of equilibrium to one of flux. Traditional methods, heavily reliant on the regularity condition of the signals, are shown to produce significantly biased estimates when applied to this task, highlighting the superiority of our approach.

In Chapter 4, we propose a pioneering framework for statistical inference of change points in the presence of temporally dependent and non-stationary noise processes. Our non-parametric approach sets us apart as one of the first to tackle this formidable challenge head-on, recognizing that the non-stationary nature of the noise process can render signal inference an exceptionally arduous task without resorting to parametric models. Furthermore, our framework accommodates temporally varying signals, a notable advancement over classical methods that often assume the signal to be multi-level constant.

Recognizing the burgeoning interest in detecting change points in multivariate or high-dimensional data, such as graph data or network data, across both academic and industrial domains, Chapter 3 and Chapter 5 endeavor to extend the frameworks established in Chapter 2 and Chapter 4 to a multivariate setting. This aspect of the work presents formidable challenges, particularly in developing a valid testing procedure for data characterized by high dimensionality, temporal and spatial dependence, and non-stationarity in the noise component. The research presented in Chapter 5 represents a pioneering attempt to address this

complex problem. While our proposed method demonstrates robust performance in low-dimensional settings (large sample size, small dimensionality), it is susceptible to the curse of dimensionality when confronted with high-dimensional data (small sample size, large dimensionality). We remain optimistic that future generations of brilliant researchers in this field will build upon our contributions and ultimately overcome this challenge.

CHAPTER 2

UNI-VARIATE CHANGE POINT ANALYSIS UNDER IRREGULAR SIGNALS

2.1 Introduction

Recently, the detection of multiple change points has garnered significant attention, as evidenced by studies such as Frick et al. [2014], Fryzlewicz [2014, 2018], Baranowski et al. [2019]. Notably, the R package **changepoint** offers functions that maintain linear computational complexity as the number of observations and change points increases [Killick et al., 2012, Killick and Eckley, 2014a], assuming a piecewise constant mean function under the alternative hypothesis. However, for many applications, it is more realistic to assume that the functions between change points vary smoothly and/or are subject to dependent errors. This perspective is supported by various statistical methods and theories [Muller, 1992, Horváth and Kokoszka, 2002, Mallik et al., 2011, 2013, Vogt and Dette, 2015, Dette et al., 2020, Bücher et al., 2021].

In contrast to the existing literature, our research is motivated by scenarios where signals become highly irregular following a change point, diverging significantly from the assumptions of constant mean or smooth variations posited under the alternative hypothesis. These signals may exhibit abrupt and erratic variations. An exemplary case involves the analysis of irregular signals found in data tracking Baidu searches for COVID-19 related symptoms, such as fever, in Hubei Province from 1 October 2019 to 31 January 2020 (see Figure 2.1). Determining the pandemic's start date is a critical epidemiological challenge, compounded by restricted data access, leading to diverse findings among researchers Worobey [2021], Huang et al. [2020], and Centre for Disease Control and Prevention [2022]. This study aims to address this crucial issue by analyzing indirect Baidu search data, specifically focusing on identifying the pandemic's emergence through changes in the Baidu search index, employing

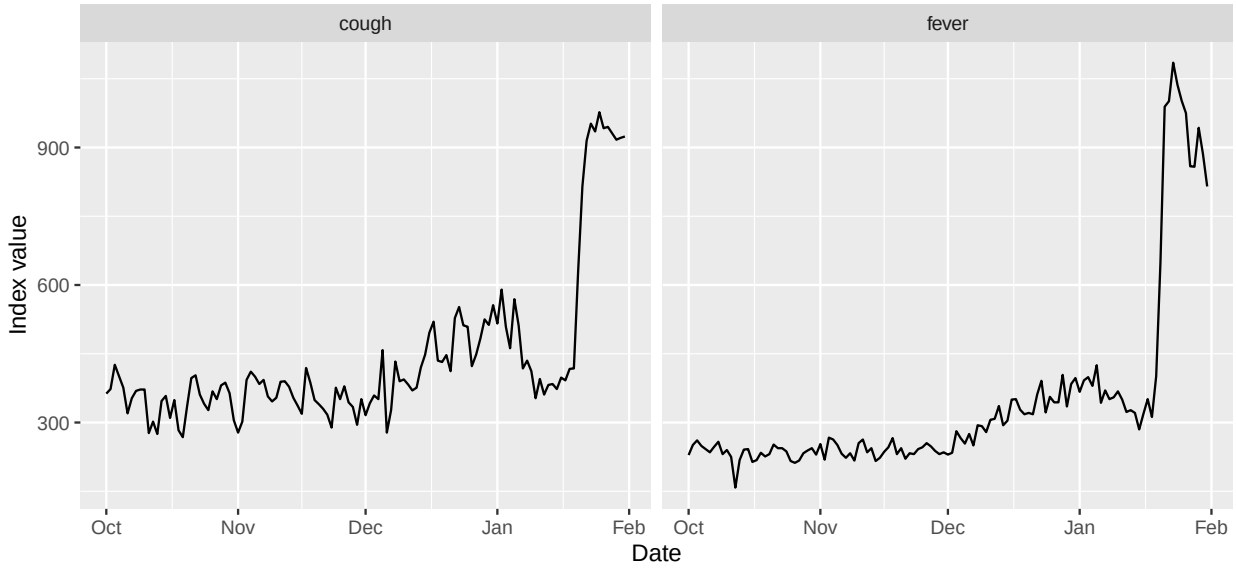

Figure 2.1: Daily Baidu search index values for COVID-related symptoms keywords from 1 October 2019 to 31 January 2020. Keywords "cough" and "fever" are depicted in the left and right panels, respectively.

the change point paradigm, as presented in (2.2.3) and (2.2.4). Given the nature of the data, it is reasonable to assume a constant mean before the change point, while post-change data, characterized by significant variability and irregularity, defies the assumptions of constancy or smooth trends. To the best of our knowledge, the existing change point literature does not adequately accommodate such a scenario.

In this chapter, we focus on the testing and the estimation of the first change point, where the null hypothesis assumes a constant mean and the signals under the alternative hypothesis can be quite general. Our method is offline, meaning that we assume the data are fully observed, which differs from online change point detection, where information accrues over time and only data available before the current time are considered. Differently from Cao and Wu [2015], where the focus is on multiple testing with clustered signals, we propose novel test statistics and a two-step method for detecting the first irregular signals after the change point. We employ a CUSUM-type statistic to test the global null hypothesis that there is no change point. If this global null is rejected, we develop a two-step method to

locate the change point. In the first step, we use the minimum of the batched means as a rough estimation of the change point location. Intuitively, since data before the change point have a constant mean, this minimum falls in the middle between the time origin and the true change point. The batched mean effectively smooths the data and increases the signal-to-noise ratio. Equipped with the preliminary estimation from the first step, we estimate the constant mean before the change point and the minimum distance between signals and the constant mean. This allows us to construct a new test statistic to obtain a refined estimation of the change point in the second step. Under suitable conditions, we achieve an $\mathcal{O}_{\mathbb{P}}(1)$ rate of convergence of the estimated change point to the true change point, which fundamentally improves the results in Cao and Wu [2022], where multiple sequences are needed to estimate the variance due to heteroscedasticity.

The chapter is structured as follows: Section 2.2 presents our primary methodology for the global null hypothesis test and the two-step change point localization. Theoretical results are developed in Section 2.3, where we demonstrate the type-one error control and power analysis of the testing procedure and establish the desirable $\mathcal{O}_{\mathbb{P}}(1)$ rate of convergence of the estimated change point to the true change point. The empirical efficacy of our method is evaluated through simulations in Section 2.4 and applied to a real-world dataset, specifically the Baidu search indices for COVID-19 symptoms in 2019-2020, in Section 2.5. Comprehensive proofs are provided in Section 2.6.

2.2 Methodology

2.2.1 Model for the Data and the Problems Considered

Let us consider a data model characterized by the presence of noise, formally defined as

$$X_t = \mu_t + Z_t, \quad t = 1, \ldots, n, \tag{2.2.1}$$

where μ_t represents the mean or signal at time t, and $(Z_t)_{t\in\mathbb{Z}}$ denotes a stationary stochastic process with zero mean, an auto-covariance function $\gamma(k) = \mathrm{Cov}(Z_{t+k}, Z_t)$, and a finite long-run variance specified by

$$0 < \sigma_\infty^2 := \sum_{k=-\infty}^{\infty} \gamma(k) < \infty. \tag{2.2.2}$$

We consider the following null hypothesis

$$H_0 : \mu_1 = \ldots = \mu_n, \tag{2.2.3}$$

indicating a constant signal across observations, and an alternative hypothesis

$$H_1 : \exists\, \tau \in \{2, \ldots, n\}, d > 0 : \mu_1 = \ldots = \mu_{\tau-1}, \quad \mu_\tau, \ldots, \mu_n \geq \mu_1 + d, \tag{2.2.4}$$

which represents an upward shift in the signal from a constant level, detectable after a certain observation τ and exceeding a minimum difference d. This scenario, capturing shifts to a higher and non-constant level is often encountered empirically; cf. Figure 2.1. Despite its practical relevance, there appears to be a lacuna in the literature regarding methodologies specifically designed for this scenario.

Moreover, this framework encompasses instances where the signal from the τth observation onwards may either remain constant or exhibit multiple changes, provided that it remains at least d units above the initial level. Here, τ signifies the earliest point of change,

illustrating the comprehensive nature of hypothesis (2.2.4) that extends beyond conventional cases by not imposing any further structure on the signal post-change.

Given the observations $X_1, \ldots, X_n$, we aim to develop

- a hypothesis test to determine whether H_0 or H_1 holds, and
- a procedure that, under H_1, will estimate the change point τ.

It is noteworthy that existing literature, such as Dette and Wu [2019], Heinrichs and Dette [2021], Vogt and Dette [2015], Bücher et al. [2021], has explored the detection of changes in sequences of means, typically relying on smoothness conditions for the applicability of their methods. In contrast, our proposed methodology distinctively does not necessitate such smoothness conditions.

2.2.2 Testing Procedure

We propose a test to distinguish between H_0 and H_1, defined in (2.2.3) and (2.2.4), respectively, given observations $X_1, \ldots, X_n$ that follow (2.2.1). Consider the test statistic

$$\hat{T} := \min_{j=1,2,\ldots,n} \frac{1}{\sqrt{n}\hat{\sigma}_\infty} \sum_{i=1}^{j} (X_i - \bar{X}_n), \quad \text{where } \bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{2.2.5}$$

and $\hat{\sigma}^2_\infty$ is a consistent estimator of the long-run variance σ^2_∞, defined in (2.2.2). Section 2.2.3 proposes an estimator for σ^2_∞ that is consistent under both H_0 and H_1. Section 2.3.2 shows that, under H_0, the distribution of $\hat{T}$ is asymptotically close to the distribution of the minimum of a standard Brownian bridge, whose quantiles can be obtained via simulation or asymptotic approximation. Furthermore, under H_1, $\hat{T} \xrightarrow{\mathbb{P}} -\infty$ as $n \to \infty$. Therefore, we reject H_0 if $\hat{T} < c$, where c is the α-quantile of the minimum of a standard Brownian bridge; cf. Section 2.3.2.

2.2.3 A Two-Step Locating Algorithm

Data Blocking

To reduce noise and focus on the signal, we split the data into $m := \lfloor n/k \rfloor$ blocks of size k, where $k \to \infty$ and $k/n \to 0$. The block sample means are

$$R_j := \frac{1}{k} \sum_{i=(j-1)k+1}^{jk} X_i, \quad j = 1, 2, \ldots, m. \tag{2.2.6}$$

Section 2.3.3 provides guidance on choosing k. Define

$$\hat{L} := \underset{i=1,\ldots,m}{\arg\min}\, R_i, \quad \hat{\ell} := k\hat{L}, \tag{2.2.7}$$

where $\hat{L}$ indicates a block likely to have all observations prior to the change point and $\hat{\ell}$ is the last observation in the $\hat{L}$th block.

Estimating the Long-Run Variance

After blocking the data, the index $\hat{\ell}$ satisfies $\hat{\ell} < \tau$ with high probability. We estimate the long-run variance by

$$\hat{\sigma}^2 := \frac{k}{\hat{\ell}} \sum_{s=k}^{\hat{\ell}} \Big(R_{s/k} - \hat{\mu}_0 \Big)^2, \quad R_{s/k} := \frac{1}{k} \sum_{i=s-k+1}^{s} X_i, \quad \hat{\mu}_0 := \frac{1}{\hat{\ell}} \sum_{i=1}^{\hat{\ell}} X_i, \tag{2.2.8}$$

where $R_{s/k}$ extends (2.2.6) to overlapping blocks and $\hat{\mu}_0$ is a preliminary estimate of μ_1. Overlapping blocks reduce asymptotic mean squared error Lahiri [1999]. The estimator (2.2.8) is motivated by $\mathbb{E}[(\sqrt{k}(R_{s/k} - \mathbb{E}R_{s/k}))^2] \to \sigma_\infty^2$ for $s = 1, \ldots, \tau - 1$, together with $\tau > \hat{\ell} \to \infty$ with high probability (Lemma 2.6.1) and the consistency of $\hat{\mu}_0$ for μ_1 (Lemma 2.6.3). Similar estimates were considered by Mies and Steland [2023], Zhou [2013], and Peligrad and

Shao [1995]. The key novelty is the data-dependent segment selection via $\hat{\ell}$, ensuring consistency under H_0 and H_1 but presenting technical challenges. Proving consistency requires a maximal inequality for quadratic forms.

Locating Algorithm: Step 1

Step 1 aims to improve the estimate of μ_1 and provide an estimate of d. With the block averages R_j, index $\hat{L}$ indicating a pre-change block, index $\hat{\ell}$ pointing to the last observation in the $\hat{L}$th block, and the preliminary estimate $\hat{\mu}_0$ for μ_1 from (2.2.8), compute the 'test statistics' $\hat{D}_j$ and 'test decisions' $\hat{I}_j$:

$$\hat{D}_j := \frac{\sqrt{k}(R_j - \hat{\mu}_0)}{\hat{\sigma}_\infty}, \quad \hat{I}_j = \begin{cases} 1 & \text{if } \hat{D}_j \geq z_{1-1/m}, \\ 0 & \text{otherwise,} \end{cases} \tag{2.2.9}$$

where $\hat{\sigma}^2_\infty$ is a consistent long-run variance estimate (e.g., $\hat{\sigma}^2$ from (2.2.8)), z_α is the α-quantile of the standard normal distribution, and $m := \lfloor n/k \rfloor$. Then compute

$$\hat{\eta} := \underset{t=1,\ldots,m-1}{\arg\min} \sum_{j=1}^{m} \left(\hat{I}_j - 1_{[t+1,m]}(j) \right)^2 = \underset{t=1,\ldots,m-1}{\arg\min} \left[\sum_{j=1}^{t} \hat{I}_j + \sum_{j=t+1}^{m} \left(1 - \hat{I}_j \right) \right]. \tag{2.2.10}$$

Finally, obtain preliminary estimates for μ_1 and d:

$$\hat{\mu}_1 := \frac{1}{k\hat{\eta}} \sum_{i=1}^{k\hat{\eta}} X_i, \quad \hat{d} := \underset{\substack{i=k(\hat{\eta}+1)+1, \\ \ldots, n-k+1}}{\min} \frac{1}{k} \sum_{j=i}^{i+k-1} (X_j - \hat{\mu}_1). \tag{2.2.11}$$

Let $\eta := \lfloor \tau/k \rfloor$ be the index of the last block before the change, so $\eta k + 1 \leq \tau \leq (\eta+1)k$. Then $\mathbb{E}R_j = \mu_1$ for $j = 1, \ldots, \eta$ and $\mathbb{E}R_j > \mu_1$ for $j = \eta+1, \ldots, m$. Since $R_1, \ldots, R_\eta$ fluctuate around μ_1 but $R_{\eta+1}, \ldots, R_m$ have means strictly larger than μ_1, $\hat{L}$ is approximately uniform on $\{1, \ldots, \eta\}$. Thus, $\hat{\mu}_0$ averages about $k\eta/2$ pre-change observations, yielding a

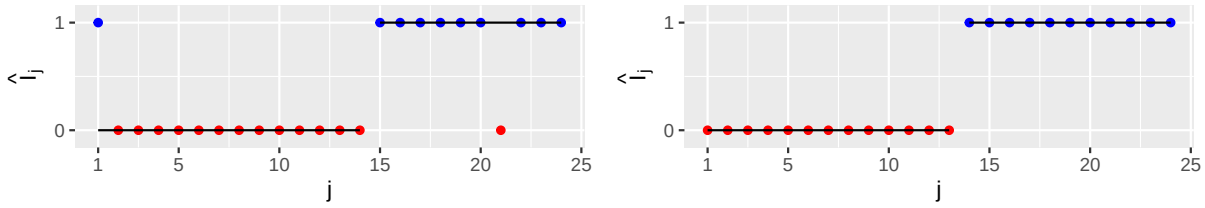

Figure 2.2: Test decisions $\hat{I}_j$ for the cough (left) and fever data (right) analysis described in Section 2.5.1. Fitted step function $j \mapsto 1_{[\hat{\eta}+1,m]}(j)$ is indicated by solid black line.

$\sqrt{n}$-consistent estimate of μ_1 if τ diverges at the same rate as n.

The estimate $\hat{\mu}_1$ improves upon $\hat{\mu}_0$ by using $\hat{\eta}$ instead of $\hat{L}$, as $\hat{\eta}$ is expected to be closer to η than $\hat{L}$. The test decisions $\hat{I}_j$ indicate whether a block is from before ($\mathbb{E}R_j = \mu_1$) or after ($\mathbb{E}R_j > \mu_1$) the change. The sequence $\hat{I}_1, \ldots, \hat{I}_m$ empirically approximates the step function $I_1, \ldots, I_m$ with $I_j := 0$ for $j \leq \eta$ and $I_j := 1$ for $j > \eta$. The estimate $\hat{\eta}$ fits a step function to $\hat{I}_1, \ldots, \hat{I}_m$ that jumps from 0 to 1 at the block including the change, providing a smoothed estimate. Figure 2.2 illustrates this smoothing for the example in Section 2.5.1. The threshold z_α in (2.2.9) uses $\alpha = 1 - 1/m \to 1$ to avoid false rejections in blocks prior to the change.

Locating Algorithm: Step 2

We propose a novel estimate for the change point τ in (2.2.4):

$$\hat{\tau} := \arg \min_{j=2,\ldots,n} \sum_{t=1}^{j-1} (X_t - \hat{\mu}_1 - \rho\hat{d}), \tag{2.2.12}$$

where $\rho \in (0, 1)$ is a tuning parameter. Section 2.3.3 provides theory for $\hat{\tau}$ and guidance on choosing ρ. Under weak conditions, $\rho := 1/2$ yields a consistent estimate, similar to Chen et al. [2022].

In Section 2.5.1, we employ $\hat{\tau}$ to infer the beginning of the COVID-19 pandemic from Baidu search indices, revealing a plausible date where traditional methods fail.

2.3 Theory

2.3.1 Assumptions on the Noise Process

To derive meaningful results regarding the statistical properties of our proposed methods, we impose the following assumptions on the noise process $(Z_t)_{t\in\mathbb{Z}}$ in model (2.2.1).

We employ the framework of functional dependence measures introduced by Wu [2005]. In this framework, we view the causal stationary process $(Z_t)_{t\in\mathbb{Z}}$ as outputs from a physical system:

$$Z_t = G(\ldots, \varepsilon_{t-1}, \varepsilon_t), \tag{2.3.13}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ are i.i.d. with mean zero and variance one, representing the input information, and G is an $\mathbb{R}$-valued measurable function that can be thought of as a filter or mechanism of the system. This representation encompasses a wide range of linear and nonlinear time series models, including ARCH, threshold autoregressive, random coefficient autoregressive, and bilinear autoregressive processes; see, for example, Tong [1990], Priestley [1988], Wu [2011]. The functional dependence measure quantifies the dependence of the output Z_t on the input ε_0 by assessing the change in the output when ε_0 is replaced by an independent copy ε_0'.

Assuming $\mathbb{E}|Z_i|^\theta < \infty$ for some $\theta \geq 1$, the functional dependence measure is defined as

$$\delta_{i,\theta} = (\mathbb{E}|Z_i - Z_{i,0}|^\theta)^{1/\theta}, \quad \text{where} \quad Z_{i,0} = G(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_i). \tag{2.3.14}$$

Here, $Z_{i,0}$ represents the coupled version of Z_i, obtained by replacing the input ε_0 with an independent copy ε_0'. The functional dependence measure $\delta_{i,\theta}$ captures the θ-th moment difference between Z_i and its coupled version $Z_{i,0}$, providing a quantitative assessment of the dependence of Z_i on the input ε_0.

To characterize the cumulative dependence of the future observations $(Z_i)_{i\geq n}$ on the

input ε_0, we define the cumulative dependence measure as

$$\Theta_{n,\theta} = \sum_{i \geq n} \delta_{i,\theta}, \quad n \geq 0. \tag{2.3.15}$$

The cumulative dependence measure $\Theta_{n,\theta}$ aggregates the functional dependence measures $\delta_{i,\theta}$ over all future time points $i \geq n$, providing a comprehensive assessment of the overall dependence of the future observations on the input ε_0. The rate at which $\Theta_{n,\theta}$ decays as n increases characterizes the strength and persistence of the dependence in the noise process.

2.3.2 Testing Procedure

We now state results on the test described in Section 2.2.2. The first result provides the asymptotic distribution under the null hypothesis, while the second asserts asymptotic consistency under H_1, where $\tau = \tau_n$ and $d = d_n$ are allowed to depend on n.

Theorem 2.3.1. *Assume the short-range dependence condition*

$$\Theta_{0,2} = \sum_{i \geq 0} \delta_{i,2} < \infty. \tag{2.3.16}$$

(i) Under H_0, as $n \to \infty$,

$$\sup_{x \leq 0} |\mathbb{P}(\hat{T} \leq x) - e^{-2x^2}| \to 0. \tag{2.3.17}$$

(ii) Under H_1, if $(\tau_n/n)(1 - \tau_n/n) d_n \sqrt{n} \to \infty$, then $\hat{T} \to -\infty$ in probability as $n \to \infty$.

The proof is deferred to Section 2.6.1. Theorem 2.3.1(i) suggests using the α-quantile of the limit, $-(-0.5 \log \alpha)^{1/2}$, as the cutoff to test H_0 at level $\alpha \in (0, 1)$. Let $\mathbb{B}$ denote a standard Brownian motion and $\mathbb{B}_1(u) = \mathbb{B}(u) - u\mathbb{B}(1)$ be the Brownian bridge. For all $x \leq 0$,

$$\mathbb{P}\Big(\inf_{u \in [0,1]} \mathbb{B}_1(u) \leq x \Big) = e^{-2x^2}; \tag{2.3.18}$$

cf. equation (9.41) in Billingsley [1999]. For small n, a refined approximation of $\mathbb{P}(\hat{T} \leq x)$ is $\mathbb{P}(T^\circ \leq x)$, where $T^\circ = \min_{j\in\{1,2,\ldots,n\}} \mathbb{B}_1(j/n)$, with distribution obtained by simulation. The test based on the latter can have more accurate performance. Theorem 2.3.1(ii) implies that, for any $q \in \mathbb{R}$, $\mathbb{P}(\hat{T} \leq q) \to 1$.

2.3.3 Locating Algorithm

To establish a convergence theory for the estimated change points, we require the following assumption on temporal dependence.

Condition 2.3.1. *$(Z_i)_{i\in\mathbb{Z}}$ satisfies $H_\theta := (\mathbb{E}|Z_i|^\theta)^{1/\theta} < \infty$ for some $\theta > 2$, and any one of the following holds:*

- *$\theta > 4$ and $\Theta_{n,\theta} = \mathcal{O}\big(n^{-\gamma_\theta}(\log n)^{-A}\big)$ as $n \to \infty$, for $A > 2\,(1/\theta + 1 + \gamma_\theta)\,/3$, where $\gamma_\theta = (\theta^2 - 4 + (\theta-2)\sqrt{\theta^2 + 20\theta + 4})/(8\theta)$;*
- *$\theta = 4$ and $\Theta_{n,\theta} = \mathcal{O}\big(n^{-1}(\log n)^{-A}\big)$ as $n \to \infty$, with $A > 3/2$;*
- *$2 < \theta < 4$ and $\Theta_{n,\theta} = \mathcal{O}\big(n^{-1}(\log n)^{-1/\theta}\big)$ as $n \to \infty$.*

Condition 2.3.1 holds, for example, under the geometric moment contraction $\delta_{n,\theta} = \mathcal{O}(\rho^n)$ for some $\rho \in (0,1)$, which is satisfied for many nonlinear time series models; see, for example, Shao and Wu [2007] or Wu [2011]. It can be weaker, allowing polynomially decaying dependence measures. By Corollary 2.1 in Berkes et al. [2014], Condition 2.3.1 implies the following optimal Komlós–Major–Tusnády result: on a possibly richer probability space $(\Omega_c, \mathcal{A}_c, \mathrm{P}_c)$, there exists $(Z_i^c)_{i\in\mathbb{Z}} \overset{\mathcal{D}}{=} (Z_i)_{i\in\mathbb{Z}}$, and a standard Brownian motion $\mathbb{B}_c(\cdot)$ such that

$$\sum_{i=1}^{n} Z_i^c = \sigma_\infty \mathbb{B}_c(n) + o_{a.s.}(n^{1/\theta}). \tag{2.3.19}$$

The next result asserts the consistency of $\hat{\sigma}^2$, defined in (2.2.8), for estimating σ_∞^2. We write $a_n \ll b_n$ or $b_n \gg a_n$ to mean $a_n = o(b_n)$ as $n \to \infty$. The quantities $d = d_n$ and $\tau = \tau_n$

from (2.2.4) are allowed to depend on n without making this explicit in the notation, and k is the user-chosen block size; cf. Section 2.2.3. Further, $m := \lfloor n/k \rfloor$ and $\eta := \lfloor \tau/k \rfloor$, as before.

Theorem 2.3.2. *Assume Condition 2.3.1 with $\theta > 4$, $d \gg n^{-1/\theta}$, and $n^{2/\theta}\log(n) \ll k \ll \tau$. Then,*

$$\hat{\sigma}^2 = \sigma_\infty^2 + \mathcal{O}_{\mathbb{P}}\Big(\frac{1}{k} + \frac{1}{\eta}\Big), \quad \text{as } n \to \infty.$$

The proof is deferred to Section 2.6.2.

Remark 1. *(i) Theorem 2.3.2 limits the block size k from below by $n^{2/\theta}\log(n)$ for noise reduction (with a smaller lower bound for lighter tails) and from above by τ (implying $\eta \to \infty$). In practice, with unknown noise moment θ, the nonadaptive choice $k = \lceil n^{1/3} \rceil$ is simple yet effective, satisfying the condition if $\theta > 6$ Bühlmann and Künsch [1999].*

(ii) The gap d can vanish asymptotically if slower than $n^{-1/\theta}$. The conditions on d and k must be satisfied for the same θ, so if d decays slowly and Condition 2.3.1 holds for large θ, then k can be chosen smaller.

Our main result bounds the error of estimating τ by $\hat{\tau}$ in terms of the minimum gap to the signal averaged over sliding blocks. More precisely, defining

$$d_* := \min_{\substack{i=k(\eta+1)+1,\\ \ldots,n-k+1}} \frac{1}{k}\sum_{j=i}^{i+k-1}(\mu_j - \mu_1), \tag{2.3.20}$$

we have the following:

Theorem 2.3.3. *Assume Condition 2.3.1, $d \gg n^{-1/\theta}$, $n^{2/\theta}\log(n) \ll k \ll \tau$, $n - \tau \geq 2k$, and that there exists a constant $K > \rho$ with $d > Kd_*$, where ρ is the tuning parameter from the definition of $\hat{\tau}$. Let the estimator $\hat{\sigma}_\infty^2$, used in (2.2.9), be consistent for σ_∞^2; i.e., $\hat{\sigma}_\infty^2 = \sigma_\infty^2 + o_{\mathbb{P}}(1)$. Then,*

$$\hat{\tau} = \tau + \mathcal{O}_{\mathbb{P}}\big(d_*^{-\theta/(\theta-1)}\big),$$

as $n \to \infty$.

The proof is deferred to Section 2.6.3.

Corollary 2.3.1. *Under the conditions of Theorem 2.3.3, we have:*
(i) If d *is bounded away from zero (i.e., if there exists a constant* M *with* $0 < M < d$*), then* $\hat{\tau}_n = \tau + \mathcal{O}_{\mathbb{P}}(1)$*, as* $n \to \infty$.
(ii) If d *is unbounded (i.e.,* $d \to \infty$*, as* $n \to \infty$*), then* $\mathbb{P}(\hat{\tau}_n = \tau) \to 1$*, as* $n \to \infty$.

Remark 2. *(i) Under the general alternative* (2.2.4)*, we do not impose regularity beyond the one-sidedness of the change. The gap* $d \leq \min_{t=\tau,\dots,n}(\mu_t - \mu_1)$ *could be determined by a single noisy observation, insufficient for consistent estimation of* d *itself. We show (Lemma 2.6.5) that* $\hat{d}$ *consistently estimates* d_**, which provides an upper bound for* d*. The regularity condition* $d > Kd_*$ *for a constant* $K > \rho$ *requires* d_* *to also facilitate a lower bound, sufficient for estimating* τ *by* $\hat{\tau}$.

(ii) The following example illustrates a situation where $d > Kd_*$ *is satisfied, for a constant* $K > \rho$*: Say* $\mu_j - \mu_1 = m\big((j-\tau)/(n-\tau)\big)$ *for a function* $m : [0,1] \to (0,\infty)$ *of bounded total variation* $\|m\|_{\mathrm{TV}} < \infty$ *and let* $K := 1/\big(1 + \|m\|_{\mathrm{TV}}/\big(k \inf_{x\in[0,1]} m(x)\big)\big)$*. Note that* $K > \rho$ *if* $k > \rho\|m\|_{\mathrm{TV}}/\big((1-\rho)\inf_{x\in[0,1]} m(x)\big)$ *and*

$$d_* \leq \inf_{x\in[0,1]} m(x) + \frac{1}{k}\|m\|_{\mathrm{TV}} = K^{-1}\big(\inf_{x\in[0,1]} m(x)\big) \leq K^{-1} \min_{t\in\tau,\dots,n} m\Big(\frac{j-\tau}{n-\tau}\Big) =: K^{-1}d.$$

As an example, take a continuously differentiable function f *and add a finite number of jump discontinuities at distinct* $x_1, \dots, x_b$*: i. e.,* $m(x) := f(x) + \sum_{i=1}^{b} \delta_i I\{x \leq x_i\}$*. Then* m *is of bounded variation:* $\|m\|_{\mathrm{TV}} = \int_0^1 |f'(t)|\mathrm{d}t + \sum_{i=1}^{b} |\delta_i|$.

(iii) If the true d *were known, we could use the following estimate for* τ*:*

$$\tilde{\tau} := \arg\min_{j=2,\dots,n} \sum_{t=1}^{j-1} (X_t - \hat{\mu}_1 - \rho d). \tag{2.3.21}$$

Following the proof of Theorem 2.3.3, it can be shown that $\tilde{\tau} = \tau + \mathcal{O}_{\mathbb{P}}\big(d^{-\theta/(\theta-1)}\big)$ *for all* $\rho \in (0, 1)$*, without an assumption on* d_**. Note that* $\tilde{\tau}$ *is only available if* d *is known.*

(iv) The conditions on τ *allow for 'early' and 'late' changes, not requiring* $\tau \asymp n$*. The requirement* $k \ll \tau$ *ensures an increasing number of blocks before the change, while* $n - \tau \geq 2k$ *(slightly weaker) ensures at least one complete block after the change, needed to estimate* d_* *(see Lemma 2.6.5).*

2.4 Monte Carlo Studies

2.4.1 Models Considered

We assess the finite sample performance of both the testing procedure (Section 2.2.2) and the two-stage locating algorithm (Section 2.2.3). Our experiments employ data crafted via the signal plus noise model delineated in (2.2.1).

For the noise component, we utilize a threshold AR model Tong [1990]:

$$Z_i' = \theta \left(|Z_{i-1}'| + |Z_{i-2}'| \right) + \varepsilon_i, \tag{2.4.22}$$

where θ governs temporal dependence, and the i.i.d. innovations ε_i follow $\mathcal{N}(0, 0.5^2)$. A higher absolute value of θ indicates stronger temporal dependence. The process remains stationary provided $|\theta| < 0.5$. The noise process (Z_i) is obtained by centering Z_i as $Z_i := Z_i' - \mathbb{E}(Z_i')$. Table 2.1 provides three-digit approximations to the values used. For $\theta < 0$, we use that the expectation of the process for θ and $-\theta$ have the same long-run variance and the expectation differs only in sign. Further, for $\varepsilon_i \sim \mathcal{N}(0, \xi^2)$, we obtain expectation and long-run variances by multiplying the ones from Table 2.1 by ξ and ξ^2, respectively. For example, for $\theta = -0.2$ and $\varepsilon_i \sim \mathcal{N}(0, 0.5^2)$, we use $\mathbb{E}Z_i' = -0.343 \cdot 0.5$ and $\sigma_\infty^2 = 1.332 \cdot 0.5^2$.

Regarding the signal μ_t, we examine two scenarios: (i) Under H_0, as defined in (2.2.3), the signal remains constant at $\mu_1 = 0$. (ii) Under H_1, as defined in (2.2.4), the signal

Table 2.1: Simulated expectation $\mathbb{E}Z_i'$ and long-run variance σ_∞^2 of (Z_i'), defined in (2.4.22), for $\varepsilon_i \sim \mathcal{N}(0,1)$.

θ	$\mathbb{E}Z_i'$	σ_∞^2
0.2	0.343	1.332
0.3	0.577	2.104
0.4	0.988	5.782

generation model is:

$$\mu_t = \begin{cases} \mu_1 := 0 & \text{for } t = 1, \ldots, \tau - 1 \\ \mu_1 + s\left(\frac{2t-3\tau+\tau'}{\tau'-\tau}\right) & \text{for } t = \tau, \ldots, \tau' \\ \mu_1 + s\left(2 + \exp\left(\frac{2(t-\tau')}{\tau''-\tau'}\right)\right) & \text{for } t = \tau'+1, \ldots, \tau'' \\ \mu_1 + s\left(2 + \exp(2) \cdot \frac{2n-\tau''-t}{2n-2\tau''}\right) & \text{for } t = \tau''+1, \ldots, n, \end{cases} \tag{2.4.23}$$

In this model, s defines the magnitude of deviation from the baseline mean state ($\mu_1 = 0$) for $t < \tau$ to the varied mean state for $t \geq \tau$. This model reflects trends similar to those observed in the search engine index data depicted in Figure 2.1.

Figure **??**tFig:data_tstrates an example of the data (X_i) and its corresponding signal (μ_i). The figure clearly demonstrates the increase in signal strength following the initial change point at $\tau = 320$, where it surpasses the stable level of $\mu\_1 = 0$ by at least $s = 0.5$. Subsequent to τ, a second significant change manifests at $\tau'' = 640$, where the signal further escalates, attaining a minimum of $8s = 4$ above the initial $\mu\_1 = 0$ level. The dependence parameter for the noise process is set to $\theta = 0.4$. It is evident from the plot that obtaining a precise estimate for τ through visual inspection is exceedingly challenging due to the presence of noise. This pattern is consistent with our observations in real-world data.

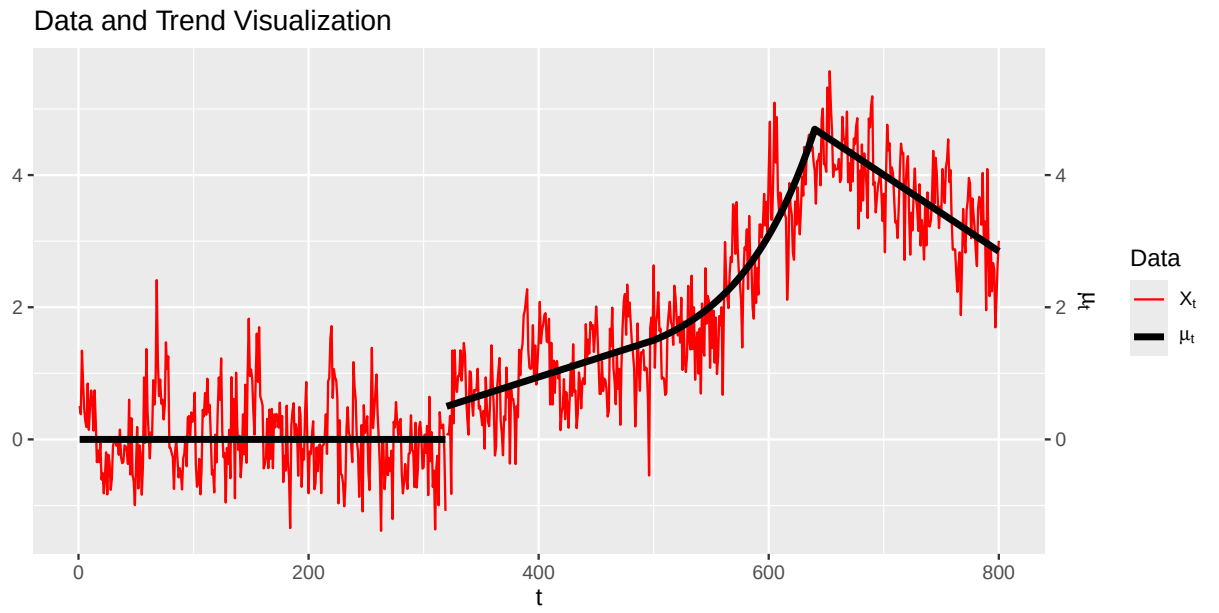

Figure 2.3: Illustration of (X_t) with $n = 800$, $\tau = 320$, $\tau' = 500$, $\tau'' = 640$, $s = 0.5$, and $\theta = 0.4$.

2.4.2 Synthetic Data under the Null Hypothesis

In this section, we illustrate that the testing procedure described in Section 2.2.2 has the correct size, asymptotically. We employ data structured as detailed in Section 2.4.1, operating under a constant signal (i.e., H_0).

We modulate the sample size, selecting n from $50, 100, 300, 500, 2000$, and adjust the dependence parameter θ from -0.4, -0.2, 0 (independence), 0.2, 0.4. The significance level remains fixed at $\alpha = 0.05$. We use the true long-run variance σ^2_∞ instead of $\hat{\sigma}^2_\infty$ in (2.2.5); cf. Table 2.1. The empirical sizes, derived from 100,000 replications, are summarized in Table 2.2.

Analyzing Table 2.2, it is evident that the rejection ratios—serving as proxies for type-one error—gravitate closer to the target significance level of $\alpha = 0.05$ as the sample size n expands and temporal dependence diminishes (absolute value of θ shrinks). This observation aligns seamlessly with our theoretical framework presented in Section 2.3.2. On juxtaposing the two methodologies, the finite-sample Gaussian approximation-based testing procedure

Table 2.2: Rejection ratios for change point testing procedure under the null hypothesis; cf. (2.2.3).

Approximation		θ				
Method	n	-0.4	-0.2	0	0.2	0.4
Asymptotic	50	0.0141	0.0270	0.0328	0.0318	0.0163
	100	0.0226	0.0340	0.0374	0.0369	0.0241
	300	0.0323	0.0392	0.0421	0.0408	0.0339
	500	0.0350	0.0417	0.0438	0.0430	0.0356
	2000	0.0418	0.0454	0.0454	0.0465	0.0423
Finite-sample	50	0.0210	0.0408	0.0500	0.0479	0.0244
	100	0.0296	0.0454	0.0500	0.0490	0.0314
	300	0.0380	0.0471	0.0497	0.0486	0.0392
	500	0.0398	0.0469	0.0501	0.0492	0.0404
	2000	0.0441	0.0485	0.0482	0.0495	0.0447

emerges superior in smaller datasets ($n = 50, 100, 300, 500$), compared to its asymptotic counterpart. However, the latter's performance converges with the finite-sample approach as the sample size surges to $n = 2000$. This implies that, for shorter datasets, the finite-sample Gaussian approximation can be advantageous. Conversely, for longer datasets, the more computationally economical asymptotic approach becomes viable.

2.4.3 Synthetic Data under Alternative Hypotheses

This section provides an in-depth analysis of our testing procedure's power and evaluates the efficacy of the algorithm used for locating the first change point, employing synthetic data. We adopt the data structure described in Section 2.4.1, with the signal defined as per (2.4.23) (i.e., H_1).

Our experimental setup is as follows: We vary the sample size, choosing n from the values $50, 100, 300, 500, 2000$. We select the dependence parameter θ from the values -0.4, -0.2, 0 (representing independence), 0.2, and 0.4. The gap parameter s ranges from 0 to 0.045, increasing in steps of 0.0006. It is important to note that the standard deviation of the innovation in the dependent process is fixed at 0.5, and we keep $\mu_1 = 0$. We standardize the ratios $\tau/n = 0.4$, $\tau'/n = 0.6$, $\tau''/n = 0.8$, maintaining $\mu_1 = 0$. For our testing methods, we

consistently set the significance level at $\alpha = 0.05$. Once the parameters for an experiment are established, we generate the trend (μ_i) using the aforementioned methodology. Subsequently, the additive noise process is simulated repeatedly, and this data is input into our testing and locating algorithms. For testing we use the true long-run variance σ_∞^2 instead of $\hat{\sigma}_\infty^2$ in (2.2.5); cf. Table 2.1. For the locating algorithm we use the long-run variance estimator defined in Section 2.2.3; i.e., $\hat{\sigma}_\infty^2 := \hat{\sigma}^2$; cf. (2.2.8) and (2.2.9). The results are derived from 100,000 independent simulations.

It is crucial to observe that testing for a change point remains challenging, even in scenarios with the largest gap parameter $s = 0.045$. This difficulty arises because, as indicated in Table 2.1, the gap $s = 0.045$ is considerably smaller than the noise levels, complicating the detection of the change point's presence significantly.

Initial observations indicate variations in the rejection ratio, an estimate of the true power, in relation to the gap parameter s. These variations are evident across different combinations of the dependence parameter θ and sample size n, as depicted in Figure 2.4. In each experiment, the rejection ratio progresses from the nominal level ($\alpha = 0.05$) to nearly 1 as s increases from 0 to 0.045. This trend suggests that as the task of detecting change points becomes less challenging, the power of our test approaches unity.

The graphical data highlight a marked increase in the rejection ratio with sample sizes expanding from 50 to 2000. This trend is in alignment with the theoretical insights presented in Theorem 2.3.1(ii). Additionally, it is noteworthy that despite a diminished test power under conditions of strong temporal dependence (with $|\theta| = 0.4$), the power can still approach unity given a sufficient sample size. This observation implies the efficacy of our testing procedure even under the influence of temporal dependent noise in the data.

Next, we showcase the absolute errors normalized by sample size $\mathbb{E}|\hat{\tau} - \tau|/n$ of our two-step locating algorithm across experiments with diverse parameters in Figure 2.5.

When the temporal dependence is moderate, error rates are smaller. As the disparity

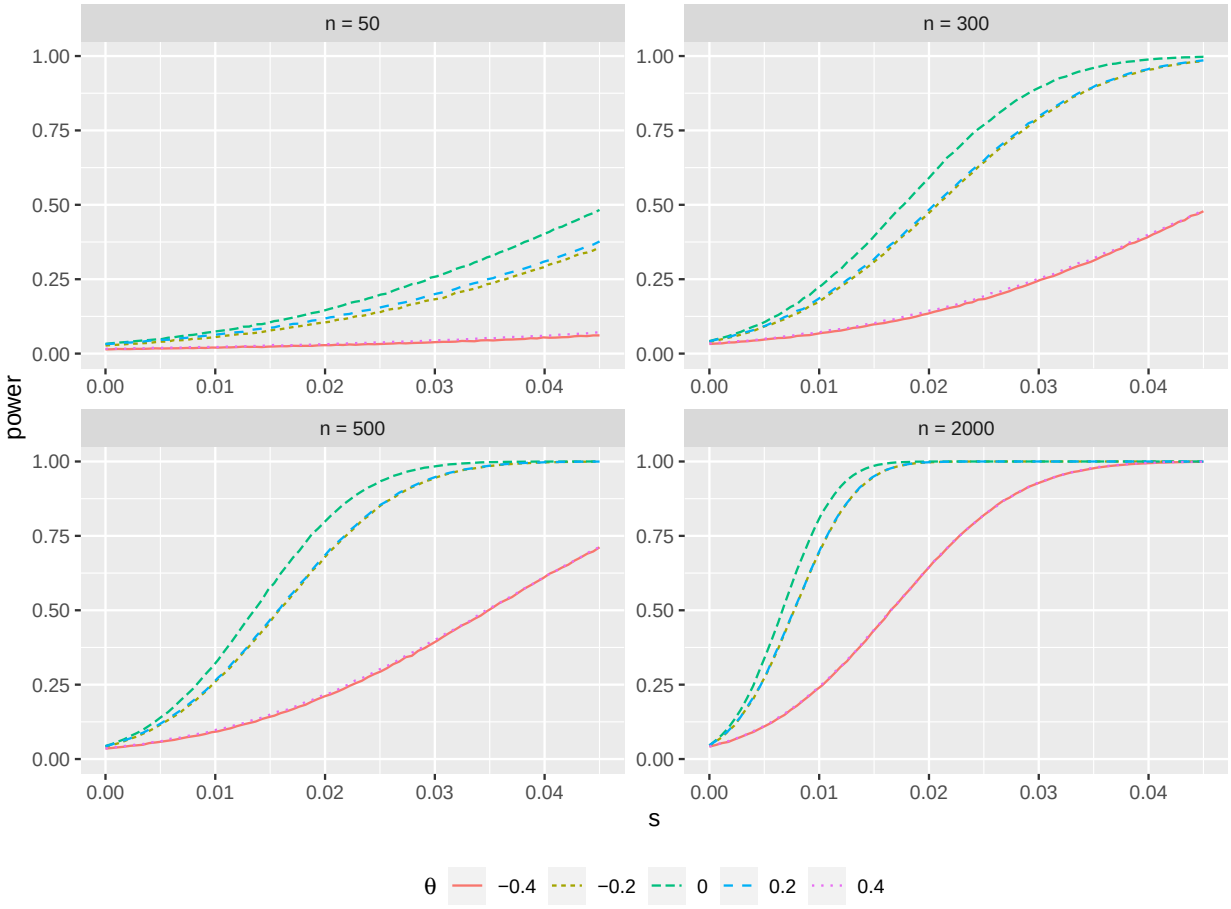

Figure 2.4: Rejection ratios for the testing procedure under the alternative hypothesis: $\mu_i = \mu_1$ for $i = 1, 2, \ldots, \tau - 1$; $\mu_i > \mu_1 + s$ for $i = \tau, \tau + 1, \tau + 2, \ldots, n$. The noise process is shaped by the dependence parameter θ. We adjust the gap parameter s over the set $\{0, 0.0006, 0.0012, \ldots, 0.045\}$, n over $\{50, 300, 500, 2000\}$, and θ over $\{-0.4, -0.2, 0, 0.2, 0.4\}$. Each data point represents 100,000 replications.

between the signal and non-signal segments grows from 0.8 to 2.5, the error rate progressively diminishes. As the sample size n expands from 50 to 2000 normalized error decreases. These findings resonate with Theorem 2.3.3, discussed in Section 2.3.3. For scenarios characterized by heightened dependence and minimal gap, error rates can be larger. Yet, in more favorable conditions, the error remains relatively stable or increases only marginally. This fact underscores the robustness of our methodology. We expanded our analysis to compare the performance of our locating algorithm with three established change point estimation techniques: a CUSUM-type method, a Maximum Likelihood Estimation-based

method, and an approach that employs binary segmentation. More precisely, we obtain $\arg\min_{j=2,3,\ldots,n+1} \sum_{i=1}^{j-1}(X_i - \bar{X}_n)$ and refer to it as CUSUM. This is related to our test statistic $\hat{T}$, defined in (2.2.5). Secondly, we apply the functions `cpt.mean`, with `method = "AMOC"`, and `cpts` from the `R` package `changepoint` Killick and Eckley [2014b], Killick et al. [2022] and refer to the obtained value as AMOC (at most one change). Thirdly, we apply the functions `sbs` and `changepoints` from the `R` package `wbs` Baranowski and Fryzlewicz [2019] and refer to the minimum of the obtained values as 1SBS (first time of change obtained with standard binary segmentation). We add $+1$ to AMOC and 1SBS to account for the fact that in our notation the change occurs from $\tau - 1$ to τ while there it occurs from τ to $\tau + 1$. Note that, while the first two methods estimate a single change point, the binary segmentation method estimates multiple change points of which we select the earliest. The outcome of this comparative study is detailed in Figure 2.5. It reveals that the errors associated with the three alternative methods are somewhat unstable and, depending on the scenario, can perform poorly. We also note that keeping the gap parameter s and the dependence parameter θ constant while increasing the sample size n from 50 to 2000 results in the mean absolute error (MAE) normalized by n for our method approach zero. This observation confirms our theory and previous numerical analysis that our method's error remains relatively constant with larger sample sizes. In contrast, for the other three classical methods, we see less stable behavior of the MAE $/n$, which either remains relatively unchanged as n increases, indicating that their errors grow with the sample size or behave reasonably under independence, but struggle in the presence of serial dependence. Furthermore, when θ and n are fixed and s is varied, our method shows a steady decline in error as s increases. This pattern of reducing error is not observed in the other methods. In fact, with these methods, increasing the gap parameter may lead to higher errors, questioning their stability and reliability in our experimental context. Overall, these results highlight the shortcomings of traditional methods in handling non-standard or complex data configurations, emphasizing the superiority and

versatility of our proposed method.

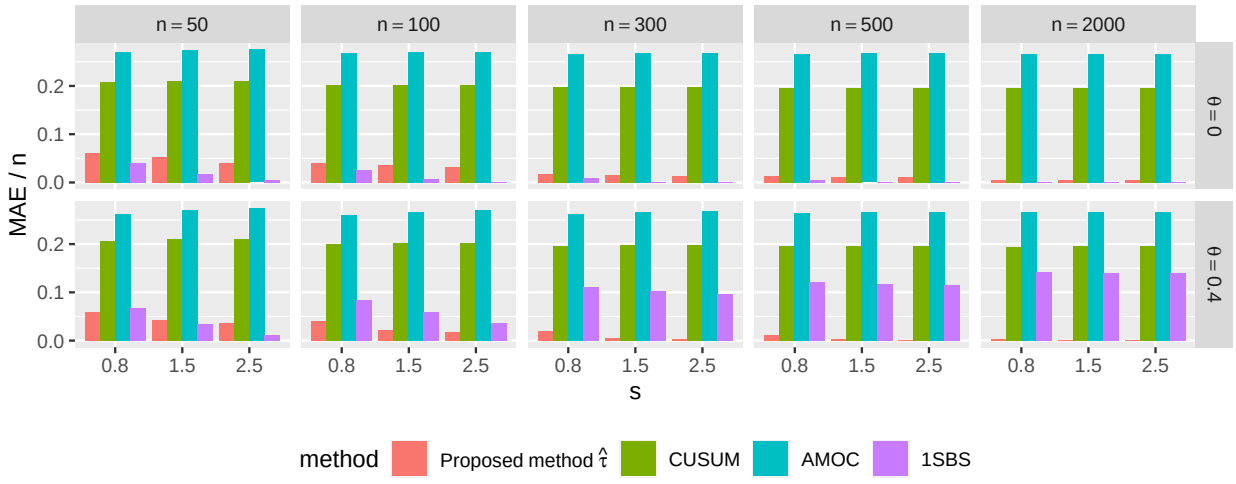

Figure 2.5: Expected absolute errors normalized by sample size across four change point detection methods. The red, green, blue and purple bars show $\mathbb{E}|\hat{\tau}-\tau|/n$ of our proposed method $\hat{\tau}$, CUSUM, likelihood-based method (AMOC), and earliest change point from standard binary segmentation (1SBS) method, respectively. The parameters varied in this study include the gap parameter s, the sample size n, and the parameter θ of the threshold autoregression noise process. Each bar in the graph represents the average result from $100,000$ replications.

2.5 Real Data analysis

2.5.1 Baidu Search Index for COVID-19 Related Symptoms

Numerous studies have endeavored to pinpoint the initial emergence of the SARS-CoV-2 virus among humans. The initial cases were likely linked to the Huanan Seafood Wholesale Market in late December 2019. However, this cluster is not believed to signify the pandemic's inception. To deduce the possible duration SARS-CoV-2 circulated in China before detection, we analyzed Baidu's search index (China's leading search engine) for COVID-19 symptom-related keywords between 1 October 2019 and 31 January 2020 in Hubei Province, China. We focused on the terms "fever" and "cough", aggregating searches from both desktop and mobile platforms. As depicted in Figure 2.6, the counts exhibit regular fluctuations until the series' end. Given the rapid transmission capability of COVID-19, the consistent mean assumption post-change point in conventional methods is inapplicable. Applying the test proposed in Section 2.2.2 for the null hypothesis H_0 of constant mean, defined in (2.2.3), against the alternative hypothesis H_1 of a one-sided upwards change, defined in (2.2.4), yields test statistics $\hat{T} < -10$ and $\hat{T} < -26$, for Baidu search indices "cough" and "fever", respectively. The p-values implied by Theorem 2.3.1(i) are essentially zero such that we reject the null hypothesis in both cases.

We continue the analysis by employing the two-stage locating method (Section 2.2.3). For the keyword "cough" (comprising $n = 123$ data points), the initial stage estimates the equilibrium data state's mean, μ_1, and the state gap parameter d, guiding the subsequent stage. We defined $k = \lceil n^{1/3} \rceil = 5$ for the batched mean length and computed

$$R_j = \frac{1}{k} \sum_{i=(j-1)k+1}^{jk} X_j, \quad j = 1, \ldots, m,$$

as defined in (2.2.6). We obtain $\hat{L} := \arg\min_{1 \le j \le m} R_j = 4$ and $\hat{\ell} := k\hat{L} = 20$, as defined

in (2.2.7). We find a pre-change sample mean of $\hat{\mu}_0 = 341.45$ obtained from the initial $\hat{\ell}$ data. The test statistics $\hat{D}_j = \sqrt{k}(R_j - \hat{\mu}_0)/\hat{\sigma}_\infty$ use $\hat{\sigma}_\infty \approx 51.4$ which is the square root of the estimated long-term variance from the initial $\hat{\ell}$ observations; cf. (2.2.8). Then, we obtain the test decisions $\hat{I}_j$, defined in (2.2.9), as

$$\hat{I}_j = \begin{cases} 1 & \text{if } \hat{D}_j \geq z_{1-1/m}, \\ 0 & \text{otherwise,} \end{cases}$$

where $z_{1-1/m}$ is the $1 - 1/m$ quantile of the standard normal distribution. We obtain

$$\hat{\eta} := \arg\min_t \sum_{j=1}^{m} \{I_j - 1_{[t+1,m]}(j)\}^2 = 14,$$

as defined in (2.2.10). A graphical representation of the test decisions and smoothing can be seen in Figure 2.2. The first-stage estimates thus are

$$\hat{\mu}_1 := \frac{1}{k\hat{\eta}} \sum_{i=1}^{k\hat{\eta}} X_i \approx 353.06, \quad \hat{d} := \min_{\substack{i=k(\hat{\eta}+1)+1, \\ \ldots, n-k+1}} \frac{1}{k} \sum_{j=i}^{i+k-1} (X_j - \hat{\mu}_1) \approx 21.9.$$

Setting $\rho = 0.5$, our refined change point estimate in the second phase is:

$$\hat{\tau} := \arg \min_{j=2,\ldots,n} \sum_{t=1}^{j-1} (X_t - \hat{\mu}_1 - \rho\hat{d}) = 69,$$

as defined in (2.2.12), which translates to 8 December 2019.

For comparative analysis, we evaluated the CUSUM, AMOC, and 1SBS estimates as discussed in Section 2.4.3. The CUSUM method identified 15 December 2019 as the change point, while the AMOC approach pinpointed 21 January 2020. Conversely, the Binary Segmentation method indicated 14 December 2019 as the initial outbreak date, slightly earlier than the other two methods but still later compared to our findings.

Remarkably, our analysis of the Baidu "fever" search index aligns with our results by also marking 8 December 2019 as the change point, consistent with the "cough" dataset. In contrast, the CUSUM, AMOC, and 1SBS methods suggested change points on 22 December 2019, 20 January 2020, and 3 December 2019, respectively. A graphical representation of these results is shown in Figure 2.6.

Reports, such as Worobey [2021], mention early COVID-19 cases, including a 41-year-old male showing symptoms on 16 December 2019 and a female seafood vendor with symptoms on 10 December 2019, aware of potential cases near Huanan Market from 11 December 2019. The CDC and other studies have highlighted early December 2019 as a critical period. These findings support the plausibility of our change point detection.

Notably, the Chinese government officially announced the outbreak on 20 January 2020, marking a significant tipping point. However, our focus is on identifying the initial outbreak, which our analysis suggests occurred before 1 January 2020. Classical methods may not be well-suited for detecting such early changes, possibly due to their reliance on sample means that can be influenced by later data points, leading to inaccuracies.

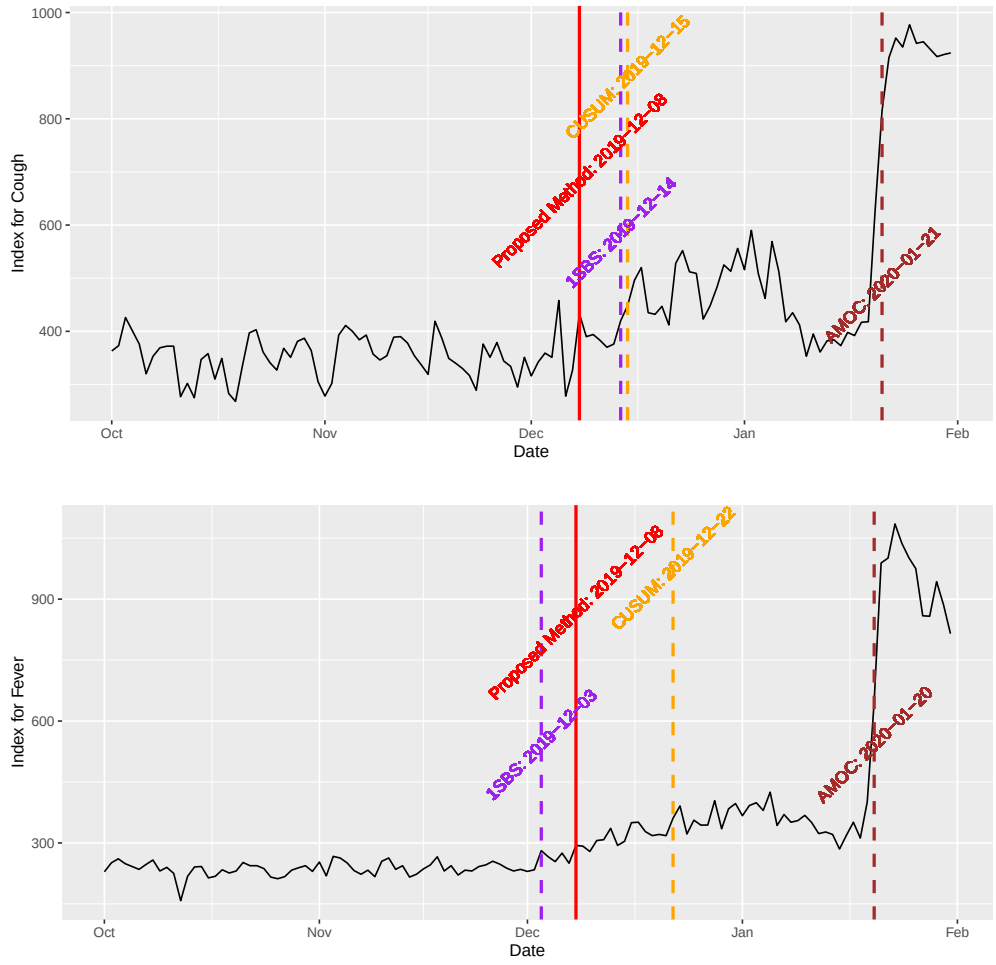

Figure 2.6: BAIDU search index for "fever" and "cough" from 1 October 2019 to 31 January 2020. The red solid vertical line indicates the change point detected by our method $\hat{\tau}$; the orange, brown, and purple dashed lines represent the change points detected by CUSUM, likelihood-based (AMOC), and earliest change detected by standard binary segmentation (1SBS) methods, respectively.

2.6 Proofs

2.6.1 Proof of Theorem 2.3.1

Let

$$T_n := \min_{j=1,2,\ldots,n} \frac{1}{\sqrt{n}\sigma_\infty} \sum_{i=1}^{j} (X_i - \bar{X}_n).$$

By $\hat{\sigma}^2 = \sigma_\infty^2 + o_{\mathbb{P}}(1)$ and Slutsky's lemma, it suffices to prove the result with $\hat{T}$ replaced by T_n. Note that

$$n^{-1/2} \sum_{i=1}^{j} (X_i - \bar{X}_n) = n^{-1/2} M_n(j) + n^{-1/2} \Big(S_j - \frac{j}{n} S_n \Big), \quad S_n = \sum_{i=1}^{n} Z_i, \tag{2.6.24}$$

where $M_n(j) := \sum_{i=1}^{j} (\mu_i - \bar{\mu}_n)$, $\bar{\mu}_n := n^{-1} \sum_{i=1}^{n} \mu_i$. By Theorem 3 in Wu [2005], under condition (2.3.16),

$$n^{-1/2} \{ S_{\lfloor nu \rfloor}, 0 \le u \le 1 \} \Rightarrow \{ \sigma_\infty \mathbb{B}(u), 0 \le u \le 1 \}.$$

Then $n^{-1/2} \{ S_{\lfloor nu \rfloor} - n^{-1} \lfloor nu \rfloor S_n, 0 \le u \le 1 \} \Rightarrow \{ \sigma_\infty \mathbb{B}_1(u), 0 \le u \le 1 \}$, and case (i) follows from the continuous mapping theorem and (2.3.18), as under H_0, $M_n(j) = 0$.

For (ii), under H_1, for $j, \tau \in \{1, \ldots, n\}$,

$$M_n(j) := \sum_{i=1}^{j} (\mu_i - \bar{\mu}_n) = \sum_{i=1}^{j} (\mu_i - \mu_1) - \frac{j}{n} \sum_{i=\tau}^{n} (\mu_i - \mu_1). \tag{2.6.25}$$

The representation in (2.6.25) has interesting consequences: (a) Since $(\mu_i - \mu_1) = 0$ for $i = 1, \ldots, \tau - 1$, the first term is non-negative, vanishing for $j < \tau$, and the second term decreases as j increases, implying $\arg\min_j M_n(j) \ge \tau - 1$. (b) We have $\min_{j=1,\ldots,n} M_n(j) \le$

$M_n(\tau-1) \le d(\frac{1-\tau}{n})(n-\tau+1)$. Thus,

$$n^{-1/2} \min_{j=1,\dots,n} M_n(j) \to -\infty, \quad \text{as } n^{-3/2}\tau(n-\tau)d \to \infty.$$

Therefore, (ii) follows from (2.6.24). □

2.6.2 Proof of Theorem 2.3.2

We introduce some notation and technical results. During the proofs, we refer to the following conditions, satisfied under the theorem's conditions: let the quantities n, d, τ, m, η (cf. comment before Theorem 2.3.2), and θ, as in Condition 2.3.1, satisfy, as $n \to \infty$,

(C0) $\eta \to \infty$,

(C1) $n^{1/\theta}\sqrt{\log(\eta)/k} = \mathcal{O}(1)$,

(C2) $(kd + \sqrt{k\log(\eta)})/n^{1/\theta} \to \infty$,

(C3) $\log(m) \ll kd^2$,

(C4) $m^{1/\theta}/k \ll k^{-1/2}$, as $n \to \infty$.

(C5) Let $\hat{\sigma}^2_\infty$ satisfy $\mathbb{P}\big(|\hat{\sigma}^2_\infty - \sigma^2_\infty| > \lambda\sigma^2_\infty\big) \to 0$, for some $\lambda \in (0,1)$.

By $\tau \le (\eta+1)k$, $k \ll \tau$ implies (C0). Further, $k \gg n^{2/\theta}\log(n)$ and $d \gg n^{-1/\theta}$ imply (C1)–(C4), and $\hat{\sigma}^2_\infty = \sigma^2_\infty + o_{\mathbb{P}}(1)$ implies (C5). If $\hat{\sigma}^2_\infty$ is our long-run variance estimate $\hat{\sigma}^2$ defined in (2.2.8) and Condition 2.3.1 is satisfied with $\theta > 4$, then (C5) holds.

Throughout Sections 2.6.2–2.6.4, we apply Theorem 2 from Wu and Wu [2016] providing a Nagaev-type inequality. The required condition is that the dependence adjusted norm, defined as

$$\Xi_{\theta,\alpha} := \sup_{i\ge 0}(i+1)^\alpha \Theta_{i,\theta}, \tag{2.6.26}$$

is finite, where the cumulative dependence measure $\Theta_{i,\theta}$ is as in (2.3.15).

More precisely, we require:

Condition 2.6.1. *$(Z_i)_{i\in\mathbb{Z}}$ satisfies $H_\theta := (\mathbb{E}|Z_i|^\theta)^{1/\theta} < \infty$ for some $\theta > 2$, and there exists $\alpha \geq 0$ such that $\Xi_{\theta,\alpha} < \infty$.*

Clearly, Condition 2.3.1 implies Condition 2.6.1 with $\alpha = 1$.

Lemma 2.6.1 asserts that, with probability tending to 1, $\hat{L}$ diverges and does not take values larger than $\eta - o(\eta)$.

Lemma 2.6.1. *Grant Conditions 2.3.1, (C1), (C2), and (C3). Then, for any sequence $t_n \in \mathbb{N}$ with $t_n \to \infty$ and $t_n = o(\eta)$,*

$$\mathbb{P}(t_n \leq \hat{L} \leq \eta - t_n) \to 1, \quad \text{as } n \to \infty. \tag{2.6.27}$$

The proof is deferred to Section 2.6.4.

Lemma 2.6.2. *Let $\delta_{i,2}$ be defined as in* (2.3.14)*. Then*

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| \leq 2\Big(\sum_{i=0}^{\infty} \delta_{i,2}\Big)^2, \quad \sum_{k=-\infty}^{\infty} |k||\gamma(k)| \leq 2\Big(\sum_{i=0}^{\infty} \delta_{i,2}\Big)\Big(\sum_{k=0}^{\infty} k\delta_{k,2}\Big). \tag{2.6.28}$$

In particular, Condition 2.3.1 with $\theta > 2$ implies $\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$ and with $\theta > 4$ implies $\sum_{k=-\infty}^{\infty} |k||\gamma(k)| < \infty$.

The proof is deferred to Section 2.6.4.

Lemma 2.6.3 concerns the consistency and rate of the estimate $\hat{\mu}_0$ used in the definition of the test statistic D_j.

Lemma 2.6.3. *Grant Conditions 2.3.1, (C1), (C2), and (C3). Then*

$$|\hat{\mu}_0 - \mu_1| = \mathcal{O}_{\mathbb{P}}\Big(\frac{1}{\sqrt{\eta k}}\Big), \quad \text{as } n \to \infty. \tag{2.6.29}$$

The proof is deferred to Section 2.6.4.

Now we prove Theorem 2.3.2. The following decomposition holds:

$$
\begin{aligned}
&\hat{\sigma}^2 - \sigma_\infty^2 \\
&= \hat{\sigma}^2 - \frac{1}{\hat{L}} \sum_{s=k}^{k\hat{L}} \Big(R_{s/k} - \mu_1 \Big)^2 && (2.6.30) \\
&\quad + \frac{k}{k\hat{L}} \sum_{s=k}^{k\hat{L}} \Big(R_{s/k} - \mu_1 \Big)^2 - \frac{k}{k\hat{L}} \sum_{s=k}^{k\hat{L}} \Big(R_{s/k} - \mathbb{E} R_{s/k} \Big)^2 && (2.6.31) \\
&\quad + \frac{k}{k\hat{L}} \sum_{s=k}^{k\hat{L}} \Big(R_{s/k} - \mathbb{E} R_{s/k} \Big)^2 - k\mathbb{E}\Big(R_1 - \mathbb{E} R_1 \Big)^2 && (2.6.32) \\
&\quad + k\mathbb{E}\Big(R_1 - \mathbb{E} R_1 \Big)^2 - \sigma_\infty^2 && (2.6.33)
\end{aligned}
$$

We will show that (2.6.30) $= \mathcal{O}_{\mathbb{P}}(1/\eta)$, (2.6.31) $= \mathcal{O}_{\mathbb{P}}\big((\eta k)^{-1/2}\big)$, (2.6.32) $= \mathcal{O}_{\mathbb{P}}\big((\eta k)^{-1/2}\big)$, and (2.6.33) $= \mathcal{O}(1/k)$.

To show (2.6.30) $= \mathcal{O}_{\mathbb{P}}(1/\eta)$, it suffices to show that (2.6.30) $= \mathcal{O}_{\mathbb{P}}(a_n/\eta)$ for any $a_n \to \infty$. Let $t_n := \eta/a_n$, then $t_n = o(\eta)$. For (2.6.30), note that

$$
\begin{aligned}
&\Big| \hat{\sigma}^2 - \frac{1}{\hat{L}} \sum_{s=k}^{k\hat{L}} \Big(R_{s/k} - \mu_1 \Big)^2 \Big| = \Big| \frac{1}{\hat{L}} \sum_{s=k}^{k\hat{L}} \Big(R_{s/k} - \hat{\mu}_0 \Big)^2 - \frac{1}{\hat{L}} \sum_{s=k}^{k\hat{L}} \Big(R_{s/k} - \mu_1 \Big)^2 \Big| \\
&= \Big| \frac{2}{\hat{L}} \sum_{s=k}^{k\hat{L}} (R_{s/k} - \mathbb{E} R_{s/k})(\mu_1 - \hat{\mu}_0) + \frac{1}{\hat{L}} \sum_{s=k}^{k\hat{L}} \Big((2\mathbb{E} R_{s/k})(\mu_1 - \hat{\mu}_0) + \hat{\mu}_0^2 - \mu_1^2 \Big) \Big| \\
&= \Big| \frac{2}{\hat{L}} \sum_{s=k}^{k\hat{L}} (R_{s/k} - \mathbb{E} R_{s/k})(\mu_1 - \hat{\mu}_0) + \frac{1}{\hat{L}} \sum_{s=k}^{k\hat{L}} \Big((\hat{\mu}_0 - \mu_1)^2 + 2(\mu_1 - \mathbb{E} R_{s/k})(\hat{\mu}_0 - \mu_1) \Big) \Big| \\
&\le 2k|\mu_1 - \hat{\mu}_0| \left(\Big| \frac{1}{k\hat{L}} \sum_{s=k}^{k\hat{L}} (R_{s/k} - \mathbb{E} R_{s/k}) \Big| + \frac{1}{2} |\hat{\mu}_0 - \mu_1| + \frac{1}{k\hat{L}} \sum_{s=k}^{k\hat{L}} |\mu_1 - \mathbb{E} R_{s/k}| \right) \\
&= \mathcal{O}_{\mathbb{P}} \left(k \frac{1}{(\eta k)^{1/2}} \Big(\frac{\eta/t_n}{(\eta k)^{1/2}} + \frac{1}{(\eta k)^{1/2}} + \frac{1}{(\eta k)^{1/2}} \Big) \right) = \mathcal{O}_{\mathbb{P}}\Big(\frac{a_n}{\eta} \Big),
\end{aligned}
$$

where we have used $|\hat{\mu}_0 - \mu_1| = \mathcal{O}_{\mathbb{P}}\big((\eta k)^{-1/2}\big)$ by Lemma 2.6.3, and which we will prove below:

$$\frac{1}{k\hat{L}} \sum_{s=k}^{k\hat{L}} (R_{s/k} - \mathbb{E}R_{s/k}) = \mathcal{O}_{\mathbb{P}}\Big(\frac{(\eta k)^{1/2}}{k t_n}\Big), \quad \forall t_n = o(\eta), \tag{2.6.34}$$

and

$$\frac{1}{k\hat{L}} \sum_{s=k}^{k\hat{L}} |\mu_1 - \mathbb{E}R_{s/k}| = \mathcal{O}_{\mathbb{P}}\big((\eta k)^{-1/2}\big). \tag{2.6.35}$$

For (2.6.34), first note that

$$\begin{aligned}
&\max_{t_n \le L \le \eta} \Big| \frac{1}{k} \sum_{s=k}^{kL} \sum_{i=s-k+1}^{s} Z_i \Big| = \max_{t_n \le L \le \eta} \Big| \frac{1}{k} \sum_{u=1}^{k} \sum_{i=u}^{k(L-1)+u} Z_i \Big| \\
&\le \max_{t_n \le L \le \eta} \max_{u=1,\dots,k} \Big| \sum_{i=u}^{k(L-1)+u} Z_i \Big| = \max_{t_n \le L \le \eta} \max_{u=1,\dots,k} \Big| \sum_{i=1}^{k(L-1)+u} Z_i - \sum_{i=1}^{u-1} Z_i \Big| \\
&\le 2 \max_{n=1,\dots,k\eta} \Big| \sum_{i=1}^{n} Z_i \Big|.
\end{aligned}$$

Thus, letting $r_n := 2(\eta k)^{1/2}/(k t_n)$,

$$\begin{aligned}
&\mathbb{P}\Big(\Big| \frac{1}{k\hat{L}} \sum_{s=k}^{k\hat{L}} (R_{s/k} - \mathbb{E}R_{s/k}) \Big| > M r_n \Big) \le \mathbb{P}\Big(\max_{t_n \le L \le \eta} \Big| \frac{1}{kL} \sum_{s=k}^{kL} \frac{1}{k} \sum_{i=s-k+1}^{s} Z_i \Big| > M r_n \Big) + o(1) \\
&\le \mathbb{P}\Big(\max_{1 \le n \le \eta k} \Big| \sum_{j=1}^{n} Z_j \Big| > M k t_n r_n / 2 \Big) + o(1) \\
&\le C_1 \frac{\eta k \Xi^{\theta}_{\theta,1}}{(M k t_n r_n/2)^{\theta}} + C_2 \exp\left(-\frac{C_3}{\Xi^2_{2,1}} \frac{(M k t_n r_n/2)^2}{\eta k} \right) + o(1) \\
&= C_1 \Xi^{\theta}_{\theta,1} \frac{1}{M^{\theta} (\eta k)^{(\theta-2)/2}} + C_2 \exp\left(-\frac{C_3}{\Xi^2_{2,1}} M^2 \right) + o(1),
\end{aligned}$$

which will be arbitrarily small for M large enough. For the $o(1)$ after the first inequality, we used Lemma 2.6.1. For the third inequality, note that Condition 2.6.1 holds. Therefore, we can apply Theorem 2 in Wu and Wu [2016], and hence, there exist positive constants C_1, C_2

and C_3, such that the third inequality holds.

For (2.6.35), we prove the following, more general statement: let $W_1, \ldots, W_n$ be any random variables such that $W_k = \ldots = W_{k\eta} = 0$ a.s., and $w_n > 0$ an arbitrary sequence signifying a rate. Then,

$$\sum_{i=k}^{k\hat{L}} W_i = o_{\mathbb{P}}(w_n). \tag{2.6.36}$$

To see this, note that for $\varepsilon > 0$

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{i=k}^{k\hat{L}} W_i\Big| > \varepsilon w_n\Big) &\leq \mathbb{P}\Big(\Big\{\Big|\sum_{i=k}^{k\hat{L}} W_i\Big| > \varepsilon w_n\Big\} \cap \Big\{\hat{L} \leq \eta\Big\}\Big) + \mathbb{P}(\hat{L} > \eta) \\ &\leq \mathbb{P}\Big(\sup_{L=1,\ldots,\eta} \Big|\sum_{i=k}^{kL} W_i\Big| > \varepsilon w_n\Big) + o(1) \to 0,\end{aligned}$$

due to $\sup_{L=1,\ldots,\eta} \sum_{i=k}^{kL} W_i = 0$ a.s., and $o(1)$ follows from Lemma 2.6.1. Thus, (2.6.35) follows from (2.6.36) with $W_i = |\mu_1 - \mathbb{E}R_{i/k}|/(k\hat{L})$ and $w_n = (\eta k)^{-1/2}$. We conclude (2.6.30) $= \mathcal{O}_{\mathbb{P}}(1/\eta)$.

For (2.6.31), note that

$$\begin{aligned}&\frac{k}{k\hat{L}} \sum_{s=k}^{k\hat{L}} \Big(R_{s/k} - \mu_1\Big)^2 - \frac{k}{k\hat{L}} \sum_{s=k}^{k\hat{L}} \Big(R_{s/k} - \mathbb{E}R_{s/k}\Big)^2 \\ &= \sum_{s=k}^{k\hat{L}} \frac{1}{\hat{L}} \Big(2R_{s/k}(\mathbb{E}R_{s/k} - \mu_1) + \mu_1^2 - (\mathbb{E}R_{s/k})^2\Big) = \mathcal{O}_{\mathbb{P}}\big((\eta k)^{-1/2}\big),\end{aligned}$$

with the rate following from (2.6.36).

For (2.6.33), note that

$$\begin{aligned}\sigma_\infty^2 - k\mathbb{E}\Big(R_1 - \mathbb{E}R_1\Big)^2 &= \sum_{u=-\infty}^{\infty} \gamma(u) - \sum_{|u|<k} (1 - |u|/k)\gamma(u) \\ &= \sum_{|u|\geq k} \gamma(u) + k^{-1} \sum_{|u|<k} |u|\gamma(u) = \mathcal{O}(1/k),\end{aligned}$$

where the $\mathcal{O}(1/k)$ rate follows from $\sum_{u=-\infty}^{\infty}|u\gamma(u)|<\infty$, which is satisfied due to Condition 2.3.1 with $\theta>4$; cf. Lemma 2.6.2.

Finally, for (2.6.32), denote

$$\begin{aligned}\tilde{Q}(u) &:= \sum_{t=k}^{u}\frac{1}{k}\Big(\sum_{s=t-k+1}^{t} G(\varepsilon_s,\varepsilon_{s-1},\ldots)\Big)^2 \\ &= \sum_{t=k}^{u}\frac{1}{k}\Big(\sum_{s=t-k+1}^{t} Z_s\Big)^2 = \sum_{t=k}^{u} k\Big(R_{t/k}-\mathbb{E}R_{t/k}\Big)^2.\end{aligned} \tag{2.6.37}$$

Now, to show (2.6.32) $=\mathcal{O}_{\mathbb{P}}\big((\eta k)^{-1/2}\big)$, for any $t_n=\eta/a_n=o(\eta)$, and for $M>0$, letting the rate $r_n:=(k\eta)^{1/2}/\big(k(t_n-1)+1\big)=O\big(a_n/(k\eta)^{1/2}\big)$, we have that

$$\begin{aligned}&\mathbb{P}\Big(\Big|\frac{k}{k(\hat{L}-1)+1}\sum_{s=k}^{k\hat{L}}\Big(R_{s/k}-\mathbb{E}R_{s/k}\Big)^2-k\mathbb{E}\Big(R_1-\mathbb{E}R_1\Big)^2\Big|>Mr_n\Big)\\ &\leq \mathbb{P}\Bigg(\max_{t_n\leq L\leq\eta}\Bigg|\sum_{s=k}^{kL}\Bigg(k\Big(R_{s/k}-\mathbb{E}R_{s/k}\Big)^2-k\mathbb{E}\Big(R_1-\mathbb{E}R_1\Big)^2\Bigg)\Bigg|>M(k\eta)^{1/2}\Bigg)+o(1)\\ &\leq \mathbb{P}\Bigg(\max_{k\leq u\leq k\eta}\Big|\tilde{Q}(u)-\mathbb{E}\tilde{Q}(u)\Big|>M(k\eta)^{1/2}\Bigg)+o(1)\\ &\leq \frac{Ck\eta\Theta_{0,2}^2}{(M(k\eta)^{1/2})^2}+o(1)=C\Big(\frac{\Theta_{0,2}}{M}\Big)^2+o(1),\end{aligned} \tag{2.6.38}$$

where the $o(1)$ after the first inequality is due to an application of Lemma 2.6.1 and, for the last inequality, we have used equation (A.16) from the proof of Theorem 5.1 in the supplement of Mies and Steland [2023]. Note that the condition (G.1) in Mies and Steland [2023] follow from Condition 2.3.1 and condition (G.2) is satisfied under stationarity. The right-hand side of the bound is arbitrarily small for all $n>N$ with N,M chosen large enough. Note that the above rate holds for all $a_n\to\infty$. Combining this with $\big(k(\hat{L}-1)+1\big)/(k\hat{L})=1+o_{\mathbb{P}}(1)$ and Slutzkly's lemma, we have thus shown (2.6.32) $=\mathcal{O}_{\mathbb{P}}\big((\eta k)^{-1/2}\big)$, which concludes the

proof of the theorem. □

2.6.3 Proof of Theorem 2.3.3

Our proof relies on the statistical properties of the estimates $\hat{\eta}$ defined in (2.2.10), $\hat{\mu}_1$ defined in (2.2.11), and $\hat{d}$ defined in (2.2.11), summarized in Theorem 2.6.1 and Lemmas 2.6.4 and 2.6.5 below.

Theorem 2.6.1. *Assume Conditions 2.3.1 and (C0)–(C5). Then,*

$$\mathbb{P}(\hat{\eta} = \eta) \to 1, \quad n \to \infty.$$

The proof of Theorem 2.6.1 is deferred to Section 2.6.4. For the proof of Theorem 2.3.3, we further employ the following lemmas:

Lemma 2.6.4. *Under the condition of Theorem 2.6.1, we have*

$$|\hat{\mu}_1 - \mu_1| = \mathcal{O}_{\mathbb{P}}\big((k\eta)^{-1/2}\big). \tag{2.6.39}$$

The proof of Lemma 2.6.4 is deferred to Section 2.6.4.

Lemma 2.6.5. *Under the condition of Theorem 2.6.1 and $n - \tau \geq 2k$, we have*

$$|\hat{d} - d_*| = \mathcal{O}_{\mathbb{P}}\left(\Big(\frac{\log(m - \eta + 1)}{k}\Big)^{1/2}\right). \tag{2.6.40}$$

The proof of Lemma 2.6.5 is deferred to Section 2.6.4.

Now, to prove $|\hat{\tau}_n - \tau| = \mathcal{O}_{\mathbb{P}}\big(d_*^{-\theta/(\theta-1)}\big)$, as $n \to \infty$, we need to show that for any $\varepsilon > 0$, there exists $\tilde{M}_\varepsilon \in \mathbb{N}$ and $N_\varepsilon \in \mathbb{N}$ such that

$$\mathbb{P}\left(|\hat{\tau}_n - \tau| \geq \tilde{M}_\varepsilon d_*^{-\theta/(\theta-1)}\right) < \varepsilon, \quad \forall n > N_\varepsilon.$$

We now derive a bound for $\mathbb{P}\left(|\hat{\tau}_n - \tau| \geq M_\varepsilon\right)$, where $M_\varepsilon := \lfloor \tilde{M}_\varepsilon d_*^{-\theta/(\theta-1)} \rfloor$. Note that from (2.2.12) we have that

$$\begin{aligned}
\Omega &= \left\{ \sum_{t=1}^{\tau-1} (X_t - \hat{\mu}_1 - \rho\hat{d}) \geq \sum_{t=1}^{\hat{\tau}_n - 1} (X_t - \hat{\mu}_1 - \rho\hat{d}) \right\} \\
&= \left(\left\{ \sum_{t=\hat{\tau}_n}^{\tau-1} (X_t - \hat{\mu}_1 - \rho\hat{d}) \geq 0 \right\} \cap \{\hat{\tau}_n < \tau\} \right) \\
&\quad \cup \left(\left\{ \sum_{t=\tau}^{\hat{\tau}_n - 1} (X_t - \hat{\mu}_1 - \rho\hat{d}) \leq 0 \right\} \cap \{\hat{\tau}_n > \tau\} \right) \cup \{\hat{\tau}_n = \tau\}.
\end{aligned}$$

Then for $n > N_\varepsilon \gg M_\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left(|\hat{\tau}_n - \tau| \geq M_\varepsilon\right) = \mathbb{P}\left(\bigcup_{\ell=M_\varepsilon}^{\tau-2} \{\hat{\tau}_n = \tau - \ell\} \cap \Omega \right) + \mathbb{P}\left(\bigcup_{\ell=M_\varepsilon}^{n-\tau} \{\hat{\tau}_n = \tau + \ell\} \cap \Omega \right) \\
&= \mathbb{P}\left(\bigcup_{\ell=M_\varepsilon}^{\tau-2} \left\{ \sum_{t=\tau-\ell}^{\tau-1} \left(X_t - \hat{\mu}_1 - \rho\hat{d} \right) \geq 0 \right\} \right) \\
&\quad + \mathbb{P}\left(\bigcup_{\ell=M_\varepsilon}^{n-\tau} \left\{ \sum_{t=\tau}^{\tau+\ell-1} \left(X_t - \hat{\mu}_1 - \rho\hat{d} \right) \leq 0 \right\} \right) \\
&\leq \mathbb{P}\left(\max_{\ell \geq M_\varepsilon} \sum_{t=\tau-\ell}^{\tau-1} \left(Z_t + \mu_t - \hat{\mu}_1 - \rho\hat{d} \right) \geq 0 \right) \qquad (2.6.41) \\
&\quad + \mathbb{P}\left(\max_{\ell \geq M_\varepsilon} \sum_{t=\tau}^{\tau+\ell-1} \left(\hat{\mu}_1 + \rho\hat{d} - Z_t - \mu_t \right) \geq 0 \right). \qquad (2.6.42)
\end{aligned}$$

We treat (2.6.41) and (2.6.42) separately.

For (2.6.41), note that $\mu_t = \mu_1$ for $t < \tau$. Then, for any ε with $0 < \varepsilon < \rho d_*/(1+\rho)$, we

argue as follows:

$$
\begin{aligned}
(2.6.41) &= \mathbb{P}\left(\max_{\ell\geq M_\varepsilon}\sum_{t=\tau-\ell}^{\tau-1} Z_t + \ell\Big(\mu_1-\hat{\mu}_1-\rho\hat{d}\Big)\geq 0\right)\\
&\leq \mathbb{P}\left(\max_{\ell\geq M_\varepsilon}\sum_{t=\tau-\ell}^{\tau-1} Z_t + \ell\Big((1+\rho)\varepsilon-\rho d_*\Big)\geq 0\right)+\mathbb{P}\Big(|\hat{d}-d_*|>\varepsilon\Big)+\mathbb{P}\Big(|\hat{\mu}_1-\mu_1|>\varepsilon\Big).
\end{aligned}
$$

Further, for the first term, we have

$$
\begin{aligned}
&\mathbb{P}\left(\max_{\ell\geq M_\varepsilon}\sum_{t=\tau-\ell}^{\tau-1} Z_t + \ell\Big((1+\rho)\varepsilon-\rho d_*\Big)\geq 0\right)\\
&\leq \sum_{k=1}^{\infty}\mathbb{P}\left(\max_{2^{k-1}M_\varepsilon\leq\ell\leq 2^kM_\varepsilon}\sum_{t=\tau-\ell}^{\tau-1} Z_t + \ell\Big((1+\rho)\varepsilon-\rho d_*\Big)\geq 0\right)\\
&\leq \sum_{k=1}^{\infty}\mathbb{P}\left(\max_{1\leq\ell\leq 2^kM_\varepsilon}\sum_{t=\tau-\ell}^{\tau-1} Z_t \geq 2^{k-1}M_\varepsilon\Big(\rho d_*-(1+\rho)\varepsilon\Big)\right)\\
&\leq \sum_{k=1}^{\infty} C_1\frac{2^kM_\varepsilon\Xi_{\theta,1}^{\theta}}{\Big(2^{k-1}M_\varepsilon r_1\Big)^{\theta}}+\sum_{k=1}^{\infty}C_2\exp\Big(-\frac{C_3\Big(2^{k-1}M_\varepsilon r_1\Big)^2}{2^kM_\varepsilon\Xi_{2,1}^2}\Big)\\
&= \frac{C_1\Xi_{\theta,1}^{\theta}}{\Big(M_\varepsilon r_1^{\theta/(\theta-1)}\Big)^{\theta-1}}\sum_{k=1}^{\infty}\frac{2^k}{2^{\theta(k-1)}}+\sum_{k=1}^{\infty}C_2\exp\Big(-\frac{C_3 2^{k-2}M_\varepsilon r_1^2}{\Xi_{2,1}^2}\Big),
\end{aligned}
$$

where $r_1 = r_1(d_*,\varepsilon) := \rho d_* - (1+\rho)\varepsilon$. For the third inequality, note that Condition 2.6.1 holds. Therefore, we can apply Theorem 2 in Wu and Wu [2016], and hence, there exist positive constants C_1, C_2 and C_3, such that the third inequality holds.

Recall the upper bound condition on d_*, by which we have that there exists $K > \rho$ with

$$
\min_{t\geq\tau}(\mu_t-\mu_1)\geq d > Kd_*.
$$

Thus, for $\varepsilon > 0$ small enough such that $(1+\rho)\varepsilon - (K-\rho)d_* < 0$,

$$
\begin{aligned}
(2.6.42) &= \mathbb{P}\left(\max_{\ell \geq M_\varepsilon} \sum_{t=\tau}^{\tau+\ell-1} (-Z_t) + \ell(\hat{\mu}_1 - \mu_1 + \rho\hat{d}) - \sum_{t=\tau}^{\tau+\ell-1} (\mu_t - \mu_1) \geq 0 \right) \\
&\leq \mathbb{P}\left(\max_{\ell \geq M_\varepsilon} \sum_{t=\tau}^{\tau+\ell-1} (-Z_t) + \ell(1+\rho)\varepsilon - \sum_{t=\tau}^{\tau+\ell-1} \Big((\mu_t - \mu_1) - \rho d_* \Big) \geq 0 \right) \\
&\qquad + \mathbb{P}\Big(|\hat{d} - d_*| > \varepsilon \Big) + \mathbb{P}\Big(|\hat{\mu}_1 - \mu_1| > \varepsilon \Big).
\end{aligned}
$$

Further, for the first term in the previous bound,

$$
\begin{aligned}
&\mathbb{P}\left(\max_{\ell \geq M_\varepsilon} \sum_{t=\tau}^{\tau+\ell-1} (-Z_t) + \ell(1+\rho)\varepsilon - \sum_{t=\tau}^{\tau+\ell-1} \Big((\mu_t - \mu_1) - \rho d_* \Big) \geq 0 \right) \\
&\leq \sum_{k=1}^{\infty} \mathbb{P}\left(\max_{2^{k-1} M_\varepsilon \leq \ell \leq 2^k M_\varepsilon} \sum_{t=\tau}^{\tau+\ell-1} (-Z_t) + \ell(1+\rho)\varepsilon - \sum_{t=\tau}^{\tau+\ell-1} \Big((\mu_t - \mu_1) - \rho d_* \Big) \geq 0 \right) \\
&\leq \sum_{k=1}^{\infty} \mathbb{P}\left(\max_{1 \leq \ell \leq 2^k M_\varepsilon} \sum_{t=\tau}^{\tau+\ell-1} (-Z_t) \geq 2^{k-1} M_\varepsilon \Big((K-\rho)d_* - (1+\rho)\varepsilon \Big) \right) \\
&\leq \frac{C_1 \Xi_{\theta,1}^{\theta}}{\Big(M_\varepsilon r_2^{\theta/(\theta-1)} \Big)^{\theta-1}} \sum_{k=1}^{\infty} \frac{2^k}{2^{\theta(k-1)}} + \sum_{k=1}^{\infty} C_2 \exp\Big(- \frac{C_3 2^{k-2} M_\varepsilon r_2^2}{\Xi_{2,1}^2} \Big),
\end{aligned}
$$

where $r_2 = r_2(d_*, \varepsilon) := (K-\rho)d_* - (1+\rho)\varepsilon$ and we have applied Theorem 2 from Wu and Wu [2016] again to obtain the third inequality. For c with $0 < c < \min\{\rho, K-\rho\}/(1+\rho)$ choose $\varepsilon = c d_*$. Then, let $C := \min\{\rho, K-\rho\}/(1+\rho) - c > 0$ and $M_\varepsilon = \tilde{M}_\varepsilon d_*^{-\theta/(\theta-1)}$. This yields

$$
\begin{aligned}
&\mathbb{P}\left(|\hat{\tau}_n - \tau| \geq \tilde{M}_\varepsilon d_*^{-\theta/(\theta-1)} \right) \\
&\leq 2 \frac{C_1 \Xi_{\theta,1}^{\theta}}{\Big(\tilde{M}_\varepsilon C^{\theta/(\theta-1)} \Big)^{\theta-1}} \sum_{k=1}^{\infty} \frac{2^k}{2^{\theta(k-1)}} + 2 \sum_{k=1}^{\infty} C_2 \exp\Big(- \frac{C_3 2^{k-2} \tilde{M}_\varepsilon C^2}{\Xi_{2,1}^2} \Big) + o(1),
\end{aligned} \tag{2.6.43}
$$

where the $o(1)$ follows from Lemmas 2.6.4 and 2.6.5 and the fact that $(\log(m)/k)^{1/2} = o(d_*)$

and $(k\eta)^{-1/2} = o(d_*)$, which can be seen from

$$0 \leq \frac{1}{k\eta d_*^2} \leq \frac{\log(m)}{kd_*^2} \leq \frac{\log(m)}{kd^2} = o(1),$$

where the first inequality holds for $\log(m)\eta \geq 1$, the second inequality is due to $d_* \geq d$ and the $o(1)$ is condition (C3). Noting that the bound in (2.6.43) will be arbitrarily small for all n large enough, if $\tilde{M}_\varepsilon$ is chosen large enough, which yields the bound of the theorem. □

2.6.4 Proof of the Lemmas in Sections 2.6.2–2.6.3

Proof of Lemma 2.6.1

For any $x \in \mathbb{R}$ and $s = 1, \ldots, m-1$, by definition of $\hat{L}$ in (2.2.7),

$$\Big\{ \min_{j \geq s+1} R_j \geq x \Big\} \cap \Big\{ \min_{j \leq s} R_j < x \Big\} \subset \{\hat{L} \leq s\}$$

and, for $s = 2, \ldots, m$,

$$\Big\{ \min_{j \leq s-1} R_j \geq x \Big\} \cap \Big\{ \min_{j \geq s} R_j < x \Big\} \subset \{\hat{L} \geq s\}.$$

Employing $A \cap B \subset C \Leftrightarrow C^c \subset A^c \cup B^c$ for events A, B, C with complements A^c, B^c, C^c, this implies that for any $x_n, x_n' \in \mathbb{R}$

$$\begin{aligned}
&\mathbb{P}(t_n \leq \hat{L} \leq \eta - t_n) \\
&= 1 - \mathbb{P}(\{t_n > \hat{L}\} \cup \{\hat{L} > \eta - t_n\}) \\
&\geq 1 - \mathbb{P}\Big(\min_{j \leq t_n - 1} \frac{\sqrt{k}(R_j - \mu_1)}{\sigma_\infty} < x_n \Big) - \mathbb{P}\Big(\min_{j \geq t_n} \frac{\sqrt{k}(R_j - \mu_1)}{\sigma_\infty} \geq x_n \Big) \\
&\quad - \mathbb{P}\Big(\min_{j \geq \eta - t_n + 1} \frac{\sqrt{k}(R_j - \mu_1)}{\sigma_\infty} < x_n' \Big) - \mathbb{P}\Big(\min_{j \leq \eta - t_n} \frac{\sqrt{k}(R_j - \mu_1)}{\sigma_\infty} \geq x_n' \Big).
\end{aligned}$$

Next, denote $\tilde{R}_j := \sqrt{k}(R_j - \mathbb{E}R_j)/\sigma_\infty$ and observe that $R_j - \mu_1 = R_j - \mathbb{E}R_j$ for $j \leq \eta$, $R_j - \mu_1 \geq R_j - \mathbb{E}R_j$ for $j > \eta$, and $R_j - \mu_1 \geq R_j - \mathbb{E}R_j + d$ for $j > \eta + 1$. Thus, for n large enough such that $t_n < \eta$,

$$\begin{aligned}\mathbb{P}(t_n \leq \hat{L} \leq \eta - t_n) \geq 1 &- \mathbb{P}\Big(\min_{j=1,\ldots,t_n} \tilde{R}_j < x_n\Big) - \mathbb{P}\Big(\min_{j=t_n,\ldots,\eta} \tilde{R}_j \geq x_n\Big) \\ &- \mathbb{P}\Big(\min_{\eta - t_n + 1 \leq j \leq \eta + 1} \tilde{R}_j < x_n'\Big) - \mathbb{P}\Big(\min_{j=1,\ldots,\eta - t_n} \tilde{R}_j \geq x_n'\Big) \\ &- \mathbb{P}\Big(\min_{j=\eta+2,\ldots,m} \tilde{R}_j + \sqrt{k}d/\sigma_\infty < x_n'\Big) I\{\eta + 1 < m\},\end{aligned} \tag{2.6.44}$$

where we have also employed that for sets $A \subset B$, $\min_{j \in B} \tilde{R}_j \leq \min_{j \in A} \tilde{R}_j$. Next, let $(Z_i^c)_{i \in \mathbb{Z}}$ and $\mathbb{B}_c(\cdot)$ be as in (2.3.19). Denote, for $i \in \mathbb{N}$,

$$\begin{aligned}W_i &= k^{-1/2}(\mathbb{B}_c(ik) - \mathbb{B}_c\big((i-1)k\big), \text{ and} \\ M_i &= k^{-1/2} \sum_{j=(i-1)k+1}^{ik} \Big(Z_j^c/\sigma_\infty - (\mathbb{B}_c(j) - \mathbb{B}_c(j-1))\Big).\end{aligned}$$

Further, denoting $R_i^c := k^{-1} \sum_{j=(i-1)k+1}^{ik} (Z_j^c + \mu_j)$, we have (jointly for all i)

$$\tilde{R}_i \stackrel{\mathcal{D}}{=} \tilde{R}_i^c := \sqrt{k}\frac{R_i^c - \mathbb{E}R_i^c}{\sigma_\infty} = W_i + M_i.$$

Note that W_i are i.i.d. standard normally distributed and that by Corollary 2.1 in Berkes et al. [2014], under Condition 2.3.1,

$$M^{(m)} := \max_{i=1,\ldots m} |M_i| = o_{a.s.}(n^{1/\theta} k^{-1/2}). \tag{2.6.45}$$

With the above notation,

$$(\min_{j \in J} W_i) - M^{(m)} \leq \min_{j \in J}(W_i + M_i) \leq (\min_{j \in J} W_i) + M^{(m)}, \text{ for any } J \subset \{1, \ldots, m\}.$$

Denote the cdf of the Gumbel distribution by $G(x) := \exp(-\exp(-x))$ and the scaling factor

$$\gamma_t = (2\log t - \log\log t - \log(4\pi))^{1/2} = (2\log t)^{1/2} - \frac{\log(4\pi\log(t))}{2(2\log t)^{1/2}} + o\big((\log t)^{-1/2}\big),$$

as $t \to \infty$. From the Fisher–Tippett–Gnedenko theorem, for any sequence $N \to \infty$,

$$\sup_x \Big|\mathbb{P}\Big(\gamma_N\Big(\max_{j=1,\dots,N} W_j - \gamma_N\Big) \le x\Big) - G(x)\Big| = o(1). \tag{2.6.46}$$

Thus, for $x_n = -(\gamma_{t_n} + \gamma_{\eta-t_n+1})/2$,

$$\begin{aligned}
&\mathbb{P}\Big(\min_{j=1,\dots,t_n} \tilde{R}_j < x_n\Big) \le \mathbb{P}\Big(\min_{j=1,\dots,t_n} W_j < x_n + M^{(m)}\Big)\\
&= \mathbb{P}\Big(\gamma_{t_n}\Big(\Big(\max_{j=1,\dots,t_n} -W_j\Big) - \gamma_{t_n}\Big) > \gamma_{t_n}((-\gamma_{t_n} + \gamma_{\eta-t_n+1})/2 - M^{(m)})\Big)\\
&\le \mathbb{P}\Big(\Big\{\gamma_{t_n}\Big(\Big(\max_{j=1,\dots,t_n} -W_j\Big) - \gamma_{t_n}\Big) > \gamma_{t_n}(\gamma_{\eta-t_n+1} - \gamma_{t_n})(1/2 - M^{(m)}/(\gamma_{\eta-t_n+1} - \gamma_{t_n}))\Big\}\\
&\quad \cap \Big\{M^{(m)}/(\gamma_{\eta-t_n+1} - \gamma_{t_n}) \le 1/4\Big\}\Big) + \mathbb{P}\Big(M^{(m)}/(\gamma_{\eta-t_n+1} - \gamma_{t_n}) > 1/4\Big)\\
&\le 1 - G\Big(\gamma_{t_n}(\gamma_{\eta-t_n+1} - \gamma_{t_n})/4\Big) + o(1) \to 0.
\end{aligned}$$

The $o(1)$ in the last line relates to two convergences: for the first probability, decrease the lower bound for the scaled maximum to $\gamma_{t_n}(\gamma_{\eta-t_n+1} - \gamma_{t_n})/4$ and then use (2.6.46); for the second probability to vanish, note that $1/(\gamma_{\eta-t_n+1} - \gamma_{t_n}) = o\big((\log\eta)^{1/2}\big)$, as $n \to \infty$, which together with (2.6.45) and (C1) implies $M^{(m)}/(\gamma_{\eta-t_n+1} - \gamma_{t_n}) = o_{\mathrm{a.\,s.}}(1)$. For the final convergence we used $\gamma_{t_n}(\gamma_{\eta-t_n+1} - \gamma_{t_n}) \to \infty$. For the second probability in the right-hand

side of (2.6.44),

$$\begin{aligned}
&\mathbb{P}\Big(\min_{j=t_n,\ldots,\eta}\tilde{R}_j \geq x_n\Big) \leq \mathbb{P}\Big(\min_{j=t_n,\ldots,\eta} W_j \geq x_n - M^{(m)}\Big) \\
&= \mathbb{P}\Big(\gamma_{\eta-t_n+1}\Big(\Big(\max_{j=t_n,\ldots,\eta} -W_j\Big) - \gamma_{\eta-t_n+1}\Big) \\
&\qquad\qquad \leq -\gamma_{\eta-t_n+1}((\gamma_{\eta-t_n+1}-\gamma_{t_n})/2 - M^{(m)})\Big) \\
&\leq G\Big(-\gamma_{\eta-t_n+1}(\gamma_{\eta-t_n+1}-\gamma_{t_n})/4\Big) + o(1) \to 0,
\end{aligned}$$

where $-\gamma_{\eta-t_n+1}(\gamma_{\eta-t_n+1}-\gamma_{t_n})/4 \to -\infty$ due to condition (C1).

The probabilities $\mathbb{P}\big(\min_{\eta-t_n+1\leq j\leq \eta+1}\tilde{R}_j < x'_n\big)$ and $\mathbb{P}\big(\min_{j=1,\ldots,\eta-t_n}\tilde{R}_j \geq x'_n\big)$ are treated analogously, with $x'_n = -(\gamma_{t_n}+\gamma_{\eta-t_n})/2$. It remains to bound the fifth one:

$$\begin{aligned}
&\mathbb{P}\Big(\min_{j=\eta+2,\ldots,m}\tilde{R}_j + \sqrt{k}d/\sigma_\infty < x'_n\Big) \leq (m-\eta-1)\mathbb{P}\Big(|\sum_{i=1}^{k} Z_i| > kd + \sigma_\infty|x'_n|\sqrt{k}\Big) \\
&\leq (m-\eta-1)\frac{C_1\Xi_{\theta,1}^{\theta}k}{(kd+\sigma_\infty|x'_n|\sqrt{k})^\theta} + C_2(m-\eta-1)\exp\left(-\frac{C_3(kd+\sigma_\infty|x'_n|\sqrt{k})^2}{k\Xi_{2,1}^2}\right),
\end{aligned} \tag{2.6.47}$$

where we used sub-additivity of $\mathbb{P}$ and the fact that $-x'_n = |x'_n|$ for the first inequality and applied Theorem 2 in Wu and Wu [2016] for the second inequality. Now, we see that the first term in the right-hand side of (2.6.47) is $o(1)$ by employing the fact that $m-(\eta+1) \leq (n-\tau)/k$ and

$$\frac{n-\tau}{(kd+\sigma_\infty|x'_n|\sqrt{k})^\theta} = \mathcal{O}\Big(\frac{n}{(kd+(k\log\eta)^{1/2})^\theta}\Big),$$

due to $|x'_n| \asymp (\log\eta)^{1/2}$. So the first term vanishes due to condition (C2). For the second term in the right-hand side of (2.6.47), note that $0 \leq \log(m-\eta-1) \leq \log(m) \ll kd^2 \to \infty$, due to condition (C3). □

Proof of Lemma 2.6.2

Denote $\xi_t := \sigma(\varepsilon_i : i \le t)$, $\mathcal{P}_t(Z) := \mathbb{E}(Z|\xi_t) - \mathbb{E}(Z|\xi_{t-1})$, $Z \in \mathcal{L}^\theta$. Then, for $k \in \mathbb{N}_0$,

$$\begin{aligned}|\gamma(k)| = |\gamma(-k)| = |\mathbb{E}(Z_k Z_0)| &= \Big|\mathbb{E}\bigg(\Big(\sum_{i\le 0}\mathcal{P}_i Z_k\Big)\Big(\sum_{i\le 0}\mathcal{P}_i Z_0\Big)\bigg)\Big| \\ &= \bigg|\sum_{i\le 0}\mathbb{E}\bigg(\Big(\mathcal{P}_i Z_k\Big)\Big(\mathcal{P}_i Z_0\Big)\bigg)\bigg| \le \sum_{i\le 0}\|\mathcal{P}_i Z_k\|_2 \|\mathcal{P}_i Z_0\|_2 \le \sum_{i\ge 0}\delta_{i+k,2}\delta_{i,2}\end{aligned}$$

This implies $\sum_{k=-\infty}^{\infty}|\gamma(k)| \le 2\Theta_{0,\theta}^2$ and

$$\begin{aligned}\sum_{k=-\infty}^{\infty}|k\gamma(k)| &\le 2\sum_{k=1}^{\infty}k\sum_{i\le 0}\delta_{i+k,2}\delta_{i,2} \\ &= 2\sum_{i=0}^{\infty}\delta_{i,2}\sum_{k=1}^{\infty}k\delta_{i+k,2} \le 2\Big(\sum_{i=0}^{\infty}\delta_{i,2}\Big)\Big(\sum_{k=1}^{\infty}k\delta_{k,2}\Big)\end{aligned}$$

It remains to show that the right-hand side is finite under Condition 2.3.1. To this end, note that $\delta_i = \Theta_{i,2} - \Theta_{i-1,2}$ and by Condition 2.3.1, if $2 < \theta \le 4$, then there exist $C, A > 0$ and $\gamma := 1$, such that $|\Theta_{n,2}| \le Cn^{-\gamma}(\log n)^{-A}$. If $\theta > 4$, then there exist $C > 0$, $A \ge 1$, $\gamma > 1$ such that $|\Theta_{n,2}| \le Cn^{-\gamma}(\log n)^{-A}$. In any of these cases,

$$\begin{aligned}|\delta_{n,2}| &\le C\Big(n^{-\gamma}(\log n)^{-A} - (n+1)^{-\gamma}(\log(n+1))^{-A}\Big) \\ &\le C(A+\gamma)n^{-(\gamma+1)}(\log n)^{-A}.\end{aligned}$$

Hence, $|\delta_{n,2}|$ is summable since $\gamma > 0$ and $|n\delta_{n,2}|$ is summable since $\gamma > 1$ for $\theta > 4$. □

Proof of Lemma 2.6.3

To show (2.6.29) holds, we prove an equivalent proposition:

$$|\hat{\mu}_0 - \mu_1| = \mathcal{O}_{\mathbb{P}}\Big(\frac{a_n}{\sqrt{\eta k}}\Big) \tag{2.6.48}$$

for any $\{a_n\}$, $a_n \to \infty$. Choose $r_n = a_n^{-1}$, then

$$\begin{aligned}
&\mathbb{P}\Big(|\hat{\mu}_0 - \mu_1| \geq \frac{a_n}{\sqrt{\eta k}} M\Big) \\
&\leq \mathbb{P}(\hat{L} < \eta r_n) + \mathbb{P}\left(\eta \geq \hat{L} \geq \eta r_n, \frac{|\sum_{i=1}^{\hat{L}k}(X_i - \mu_1)|}{\hat{L}k} \geq \frac{a_n}{\sqrt{\eta k}} M\right) + \mathbb{P}(\hat{L} > \eta) \\
&\leq 1 - \mathbb{P}\big(\lceil \eta r_n \rceil \leq \hat{L} \leq \eta - \lceil \eta r_n \rceil\big) + \mathbb{P}\left(\max_{j \geq \eta r_n k} \Big|\frac{1}{j}\sum_{i=1}^{j} Z_i\Big| \geq \frac{a_n}{\sqrt{\eta k}} M\right) = o(1),
\end{aligned}$$

which implies (2.6.48). For the $o(1)$ in the above, apply Lemma 2.6.1 with $t_n := \lceil \eta r_n \rceil = o(\eta)$ to the first probability and Lemma 2.6.6, below, with $g_n^{-1/2} = (\eta k r_n)^{-1/2} = a_n^{1/2}(\eta k)^{-1/2}$, to the second probability.

Lemma 2.6.6. *Grant condition 2.6.1 with $\alpha = 1$. Then, for any sequence $g_n \in \mathbb{N}, g_n \to \infty$,*

$$\max_{j \geq g_n} \left|\frac{1}{j}\sum_{i=1}^{j} Z_i\right| = \mathcal{O}_{\mathbb{P}}\Big(\frac{1}{\sqrt{g_n}}\Big). \tag{2.6.49}$$

Proof. For any $G > 0$, note that by Bonferroni's inequality,

$$\begin{aligned}
\mathbb{P}\left(\max_{j \geq g_n} \left|\frac{1}{j}\sum_{i=1}^{j} Z_i\right| \geq \frac{G}{\sqrt{g_n}}\right) &\leq \sum_{k=1}^{\infty} \mathbb{P}\left(\max_{2^{k-1} g_n \leq j \leq 2^k g_n} \left|\frac{1}{j}\sum_{i=1}^{j} Z_i\right| \geq \frac{G}{\sqrt{g_n}}\right) \\
&\leq \sum_{k=1}^{\infty} \mathbb{P}\left(\max_{2^{k-1} \leq j \leq 2^k g_n} \left|\frac{1}{2^{k-1} g_n}\sum_{i=1}^{j} Z_i\right| \geq \frac{G}{\sqrt{g_n}}\right) \\
&\leq \sum_{k=1}^{\infty} \mathbb{P}\left(\max_{1 \leq j \leq 2^k g_n} \left|\sum_{i=1}^{j} Z_i\right| \geq 2^{k-1} G \sqrt{g_n}\right).
\end{aligned}$$

As argued before, Condition 2.6.1 holds. Therefore, we can apply the Nagaev-type inequality under dependence from Theorem 2 in Wu and Wu [2016], and hence, there exist

positive constants C_1, C_2 and C_3, such that

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le j\le 2^k g_n}\left|\sum_{i=1}^{j} Z_i\right| \ge 2^k G\sqrt{g_n}/2\right)\\
&\le \frac{C_1\Xi^{\theta}_{\theta,1}(2^k g_n)}{\left(2^k G\sqrt{g_n}/2\right)^{\theta}} + C_2\exp\left(-\frac{C_3 2^{2k}G^2 g_n/4}{2^k g_n \Xi^2_{2,1}}\right)\\
&= C_1\Xi^{\theta}_{\theta,1}\frac{1}{2^{k(\theta-1)}}\frac{1}{g_n^{\theta/2-1}(G/2)^{\theta}} + C_2\exp\left(-\frac{C_3}{\Xi^2_{2,1}}2^k(G/2)^2\right).
\end{aligned}$$

Thus, employing $k \le 2^k$,

$$\mathbb{P}\left(\max_{j\ge g_n}\left|\frac{1}{j}\sum_{i=1}^{j} Z_i\right| \ge G/\sqrt{g_n}\right) \le C_1\Xi^{\theta}_{\theta,1}\frac{2^{\theta-1}}{1-2^{\theta-1}}\frac{1}{g_n^{\theta/2-1}(G/2)^{\theta}} + \frac{C_2}{\exp(\frac{C_3}{4\Xi^2_{2,1}}G^2)-1},$$

which is arbitrarily small for G large enough. The result follows. □

Proof of Theorem 2.6.1

We use approximations to $\hat{D}_j$ and R_j defined by

$$\tilde{D}_j := \sqrt{k}(\tilde{R}_j - \mu_1)/\sigma_\infty, \quad \tilde{R}_j := \mathbb{E}(R_j|\varepsilon_{(j-1)k+1},\ldots,\varepsilon_{jk}), \quad j = 1,2,\ldots. \tag{2.6.50}$$

Note that $\tilde{R}_1,\ldots,\tilde{R}_m$ are independent, as the ε_t are independent, and hence the $\tilde{D}_j$ are independent. We use the following result that asserts a rate for the approximation of R_j by the independent $\tilde{R}_j$.

Lemma 2.6.7. *Assume Condition 2.3.1 holds. Then,*

$$\max_{j=1,\ldots,m}\left|R_j - \tilde{R}_j\right| = \mathcal{O}_{\mathbb{P}}\Big(\frac{m^{1/\theta}}{k}\Big), \quad n\to\infty. \tag{2.6.51}$$

The proof of Lemma 2.6.7 is deferred to Section 2.6.4. Based on $\tilde{D}_j$, we also define

approximations $\tilde{I}_j$ to the test decisions I_j. Again, $\tilde{I}_j$ will be independent and we apply the following result, interesting in itself.

Lemma 2.6.8. *Let $I_1, I_2, \ldots$ be a sequence of independent Bernoulli-distributed random variables with $\mathbb{P}(I_i = 1) = p_i = 1 - \mathbb{P}(I_i = 0)$. Then,*

$$\mathbb{P}\Big(\max_{j\geq 1}\sum_{i=1}^{j}(2I_i - 1) \geq 0\Big) = \sum_{k=1}^{\infty}\sum_{(i_1,\ldots,i_{2k})\in A_k}\prod_{\ell=1}^{2k}\mathbb{P}(2I_\ell - 1 = i_\ell),$$

where

$$A_k = \big\{(i_1,\ldots,i_{2k}) \in \{-1,1\}^{2k} : \sum_{\ell=1}^{2k} i_\ell = 0 \text{ and } \sum_{\ell=1}^{j} i_\ell \leq 0, \forall j = 1,\ldots,2k\big\}.$$

The proof of Lemma 2.6.8 is deferred to Section 2.6.4.

Now we proceed with the proof of Theorem 2.6.1. Due to the assumed conditions (C1), (C2) and (C3), we can apply Lemma 2.6.3. Due to Condition 2.6.1, we can apply Lemma 2.6.7. Our aim is to prove

$$\mathbb{P}\left(|\hat{\eta} - \eta| \geq 1\right) \to 0, \text{ as } n \to \infty. \tag{2.6.52}$$

Note that

$$\begin{aligned}\Omega = &\left(\left\{\sum_{j=\hat{\eta}+1}^{\eta}\left(2\hat{I}_j - 1\right) \geq 0\right\} \cap \{\hat{\eta} < \eta\}\right) \\ &\cup \left(\left\{\sum_{j=\eta+1}^{\hat{\eta}}\left(2\hat{I}_j - 1\right) \leq 0\right\} \cap \{\hat{\eta} > \eta\}\right) \cup \{\hat{\eta} = \eta\}.\end{aligned}$$

Thus,

$$\mathbb{P}\left(|\hat{\eta}-\eta|\geq 1\right)=\mathbb{P}\left(\max_{\ell=1,\dots,\eta-1}\sum_{j=\eta-\ell+1}^{\eta}\left(2\hat{I}_j-1\right)\geq 0\right)$$
$$+\,\mathbb{P}\left(\max_{\ell=1,\dots,m-\eta-1}\sum_{j=\eta+1}^{\eta+\ell}\left(1-2\hat{I}_j\right)\geq 0\right).$$

Fix $c_1, c_2 > 0$ and $\lambda \in (0,1)$ as in (C5); recall the notation defined in (3.6.67), then

$$\mathbb{P}\left(\max_{\ell=1,\dots,\eta-1}\sum_{j=\eta-\ell+1}^{\eta}(2\hat{I}_j-1)\geq 0\right)\leq \mathbb{P}\Big(\max_{\ell=1,\dots,\eta-1}\sum_{j=\eta-\ell+1}^{\eta}(2\hat{I}_j-1)\geq 0,$$
$$|\hat{\mu}_0-\mu_1|<\frac{c_1}{\sqrt{k}},\forall j>1,\ |R_j-\tilde{R}_j|<\frac{c_2}{\sqrt{k}},\ (1-\lambda)^{1/2}\leq\frac{\hat{\sigma}_\infty}{\sigma_\infty}\leq(1+\lambda)^{1/2}\Big)$$
$$+\,\mathbb{P}\Big(|\hat{\mu}_0-\mu_1|>\frac{c_1}{\sqrt{k}}\Big)+\mathbb{P}\Big(\max_{j=1,\dots,m}|R_j-\tilde{R}_j|>\frac{c_2}{\sqrt{k}}\Big)+\mathbb{P}\Big(|\hat{\sigma}^2_\infty-\sigma^2_\infty|>\lambda\sigma^2_\infty\Big)$$
$$=: A_n+B_n+C_n+D_n.$$

By Lemma 2.6.3 and the fact that $(\eta k)^{-1/2}\ll k^{-1/2}$, by condition (C0), $B_n\to 0$. By Lemma 2.6.7 and condition (C4), $C_n\to 0$. By condition (C5), $D_n\to 0$, as $n\to\infty$. Now define

$$\tilde{I}_j=\mathbf{1}\Big\{\tilde{D}_j\geq\frac{z_{1-1/m}}{(1-\lambda)^{1/2}}-\frac{c_1+c_2}{\sigma_\infty}\Big\},\quad j=1,2,\dots\,.$$

Note that $\tilde{I}_1,\tilde{I}_2,\dots,\tilde{I}_\eta$ are i.i.d. Bernoulli-distributed with $\mathbb{P}(\tilde{I}_1=1)=:p\to 0$, as $n\to\infty$, due to $m\to\infty$ and $\tilde{D}_1=\mathcal{O}_{\mathbb{P}}(1)$. Let $\tilde{I}_{\eta+1},\dots,\tilde{I}_{\eta+2},\dots$ be independent and distributed as $\tilde{I}_1$. Then, Lemma 2.6.8 entails that

$$\mathbb{P}\Big(\max_{j\geq 1}\sum_{i=1}^{j}(2\tilde{I}_i-1)\geq 0\Big)=\sum_{k=1}^{\infty}|A_k|\big(p(1-p)\big)^k\leq\min\Big\{\frac{4p}{1-4p},1\Big\}\leq 8p,$$

where we have used $|A_k| \le 2^{2k}$ and $p(1-p) \le p$. On

$$\{|\hat{\mu}_0 - \mu_1| < c_1/\sqrt{k}\} \cap \{|R_j - \tilde{R}_j| < c_2/\sqrt{k}\} \cap \Big\{(1-\lambda)^{1/2} \le \frac{\hat{\sigma}_\infty}{\sigma_\infty} \le (1+\lambda)^{1/2}\Big\},$$

we have $\tilde{I}_j \ge \hat{I}_j$. Thus,

$$A_n \le \mathbb{P}\left(\max_{\ell=1,\dots,\eta-1} \sum_{j=\eta-\ell+1}^{\eta} (2\tilde{I}_j - 1) \ge 0\right) \le \mathbb{P}\Big(\max_{j\ge 1} \sum_{i=1}^{j} (2\tilde{I}_i - 1) \ge 0\Big) \to 0.$$

Similarly, using the notation $\hat{J}_j = 1 - \hat{I}_j$ and

$$\tilde{J}_j := \mathbf{1}\Big\{\tilde{D}_j < \frac{z_{1-1/m}}{(1+\lambda)^{1/2}} + \frac{c_1+c_2}{\sigma_\infty}\Big\}, \quad j = \eta+1, \dots, m$$

and adding independent $\tilde{J}_1, \dots, \tilde{J}_\eta$ and $\tilde{J}_{m+1}, \tilde{J}_{m+2}, \dots$ that have distribution as J_{m+1},

$$\bar{p} := \sup_{\substack{j\ge 1 \\ j\ne\eta+1}} \mathbb{P}(\tilde{J}_j = 1) \le \mathbb{P}\Big(\tilde{D}_j - \mathbb{E}\tilde{D}_j \le \frac{z_{1-1/m}}{(1+\lambda)^{1/2}} - (\sqrt{k}d + c_1 + c_2)/\sigma_\infty\Big) \to 0,$$

as $n \to \infty$, because $\tilde{D}_j - \mathbb{E}\tilde{D}_j$ are i.i.d. and

$$\frac{z_{1-1/m}}{(1+\lambda)^{1/2}} - \mathbb{E}\tilde{D}_j \le \frac{z_{1-1/m}}{(1+\lambda)^{1/2}} - \sqrt{k}d \to -\infty,$$

for $j \ne \eta+1$. The right hand side diverges, as $z_{1-1/m} \asymp (2\log(m))^{1/2} \ll \sqrt{k}d$, where the $\ll$ is condition (C3). Noting that $1 - \hat{I}_j \le \tilde{J}_j$,

$$\begin{aligned}\mathbb{P}\left(\max_{\ell=1,\dots,m-\eta-1} \sum_{j=\eta+1}^{\eta+\ell} \big(1 - 2I_j\big) \ge 0\right) &\le \mathbb{P}\Big(\max_{j\ge 1} \sum_{i=1}^{j} (2\tilde{J}_i - 1) \ge 0\Big) \\ &\le \min\Big\{ \sum_{k=1}^{\lfloor \eta/2 \rfloor} (4\bar{p})^k + \sum_{k=\lfloor \eta/2 \rfloor + 1}^{\infty} 4^k \bar{p}^{k-1}, 1\Big\} \to 0.\end{aligned}$$

This concludes the proof of Theorem 2.6.1. □

Proof of Lemma 2.6.7

By the sub-additivity of $\mathbb{P}$ and the stationarity of $R_j - \tilde{R}_j$, we have for $u > 0$, that

$$\mathbb{P}\Big(\max_{j=1,\ldots,m} \Big| R_j - \tilde{R}_j \Big| > u \Big) \leq m \cdot \mathbb{P}\left(|R_1 - \tilde{R}_1| > u \right).$$

Let $\Delta_j := \mathbb{E}(R_1 | \varepsilon_k, \ldots, \varepsilon_j) - \mathbb{E}(R_1 | \varepsilon_k, \ldots, \varepsilon_{j-1})$, then it follows that

$$\tilde{R}_1 - R_1 = \sum_{j=-\infty}^{1} \Delta_j.$$

Applying Markov's inequality, we have

$$\mathbb{P}\Big(\Big| \sum_{j=-\infty}^{1} \Delta_j \Big| > u \Big) \leq \frac{\mathbb{E}| \sum_{j=-\infty}^{1} \Delta_j |^\theta}{u^\theta} = \frac{1}{u^\theta} \Big\| \sum_{j=-\infty}^{1} \Delta_j \Big\|_\theta^\theta.$$

Since $\{\Delta_j\}$ is a martingale difference sequence, we may apply the Burkholder-Davis-Gundy inequality to obtain a bound for the right-hand side of the previous inequality

$$\Big\| \sum_{j=-\infty}^{1} \Delta_j \Big\|_\theta^2 \leq c_\theta \sum_{j=-\infty}^{1} \|\Delta_j\|_\theta^2.$$

Recall that $R_1 = \frac{1}{k} \sum_{i=1}^{k} X_i$. Therefore, by the triangle inequality and Jensen's inequality, we have

$$\begin{aligned} \|\Delta_j\|_\theta &\leq \frac{1}{k} \sum_{i=1}^{k} \big(\|\mathbb{E}(Z_i | \varepsilon_i, \ldots, \varepsilon_j) - \mathbb{E}(Z_i | \varepsilon_i, \ldots, \varepsilon_{j-1})\|_\theta \big) \\ &\leq \frac{1}{k} \sum_{i=1}^{k} \delta_{(i-j+1),\theta} \leq \frac{1}{k} \Theta_{2-j,\theta}. \end{aligned}$$

Combining the results above, it follows that

$$\mathbb{P}\Big(\max_{j=1,\dots,m}|R_j-\tilde{R}_j|>u\Big)\leq\frac{m}{(uk)^\theta}\left(c_\theta\sum_{j=0}^{\infty}\Theta_{j,\theta}^2\right)^{\theta/2},$$

where $\sum_{j=0}^{\infty}\Theta_{j,\theta}^2<\infty$ due to Condition 2.3.1. In other words:

$$\mathbb{P}\Big(\max_{j=1,\dots,m}|R_j-\tilde{R}_j|>M\frac{m^{1/\theta}}{k}\Big)=\mathcal{O}(M^{-\theta}),$$

will be arbitrarily small for M large enough. This concludes the proof of Lemma 2.6.7. □

Proof of Lemma 2.6.8

Note that $S_n=\sum_{i=1}^n(2I_i-1)=:\sum_{i=1}^n E_i$ is a random walk that starts at $S_0=0$. The first time $n>0$ that yields $S_n=0$ again is $t_0:=\inf\{n>0:S_n=0\}$ and

$$\mathbb{P}\Big(\max_{j\geq1}\sum_{i=1}^{j}(2I_i-1)>0\Big)=\sum_{k=1}^{\infty}\mathbb{P}(t_0=2k).$$

For $k>0$,

$$\mathbb{P}(t_0=2k)=\sum_{(i_1,\dots,i_{2k})\in A_k}\mathbb{P}(E_1=i_1,\dots,E_{2k}=i_{2k}).$$

The assertion follows from the assumed independence of $I_1,I_2,\dots$. □

Proof of Lemma 2.6.4

For the proof of (2.6.39), note that

$$\begin{aligned}
\mathbb{P}\Big(|\hat{\mu}_1 - \mu_1| > \varepsilon\Big) &\le \mathbb{P}\Big(\Big|\frac{1}{k\hat{\eta}}\sum_{i=1}^{k\hat{\eta}}(X_i - \mu_1)\Big| > \varepsilon, \hat{\eta} = \eta\Big) + \mathbb{P}\Big(\hat{\eta} \neq \eta\Big) \\
&\le \mathbb{P}\Big(\Big|\sum_{i=1}^{k\eta} Z_i\Big| > \varepsilon k\eta\Big) + o(1) \\
&\le C_1 \frac{k\eta \Xi_{\theta,1}^{\theta}}{(\varepsilon k\eta)^{\theta}} + C_2 \exp\Big(-\frac{C_3(\varepsilon k\eta)^2}{k\eta \Xi_{2,1}^2}\Big) + o(1) \\
&= C_1 \Xi_{\theta,1}^{\theta}\Big(\frac{1}{\varepsilon(k\eta)^{(\theta-1)/\theta}}\Big)^{\theta} + C_2 \exp\Big(-\frac{C_3}{\Xi_{2,1}^2}\big(\varepsilon(k\eta)^{1/2}\big)^2\Big) + o(1),
\end{aligned}$$

which implies (2.6.39), since $(k\eta)^{(\theta-1)/\theta} = (k\eta)^{1/2}(k\eta)^{(\theta-2)/(2\theta)}$ with $(\theta-2)/(2\theta) > 0$. □

Proof of Lemma 2.6.5

For the proof of (2.6.40), recall

$$\hat{d} := \min_{\substack{i=k(\hat{\eta}+1)+1,\\ \ldots, n-k+1}} \frac{1}{k}\sum_{j=i}^{i+k-1}(X_j - \hat{\mu}_1)$$

and d_* defined in (2.3.20) as

$$d_* := \min_{\substack{i=k(\eta+1)+1,\\ \ldots, n-k+1}} \frac{1}{k}\sum_{j=i}^{i+k-1}(\mu_j - \mu_1).$$

Due to $|\min_i x_i - \min_i y_i| \leq \max_i |x_i - y_i|$ (in the 2nd inequality),

$$
\begin{aligned}
&\mathbb{P}\Big(|\hat{d} - d_*| > \varepsilon\Big) \leq \mathbb{P}\Big(|\hat{d} - d_*| > \varepsilon, \hat{\eta} = \eta\Big) + \mathbb{P}\Big(\hat{\eta} \neq \eta\Big) \\
&\leq \mathbb{P}\Big(\min_{\substack{i=k(\eta+1)+1, \\ \ldots, n-k+1}} \Big|\frac{1}{k}\sum_{j=i}^{i+k-1}(X_j - \mu_j + \mu_1 - \hat{\mu}_1)\Big| > \varepsilon\Big) + o(1) \\
&\leq \mathbb{P}\Big(\min_{\substack{i=k(\eta+1)+1, \\ \ldots, n-k+1}} \Big|\frac{1}{k}\sum_{j=i}^{i+k-1} Z_j\Big| > \varepsilon/2, |\hat{\mu}_1 - \mu_1| \leq \varepsilon/2\Big) + \mathbb{P}\Big(|\hat{\mu}_1 - \mu_1| > \varepsilon/2\Big) + o(1) \\
&\leq (m - \eta + 1)\mathbb{P}\Big(\max_{i=1,\ldots,2k} \Big|\sum_{j=1}^{i} Z_j\Big| > k\varepsilon/4\Big) + \mathbb{P}\Big(|\hat{\mu}_1 - \mu_1| > \varepsilon/2\Big) + o(1) \\
&\leq (m - \eta + 1)C_1 \frac{2k\Xi_{\theta,1}^{\theta}}{(k\varepsilon/4)^{\theta}} + (m - \eta + 1)C_2 \exp\Big(-\frac{C_3(k\varepsilon/4)^2}{2k\Xi_{2,1}^2}\Big) \\
&\qquad + \mathbb{P}\Big(|\hat{\mu}_1 - \mu_1| > \varepsilon/2\Big) + o(1) \\
&\leq C_1 \frac{2\Xi_{\theta,1}^{\theta}}{(((m - \eta + 1)k^{1-\theta})^{-1/\theta}\varepsilon/4)^{\theta}} + C_2 \exp\Big(\log(m - \eta + 1) - k\varepsilon^2 \frac{C_3}{32\Xi_{2,1}^2}\Big) \\
&\qquad + \mathbb{P}\Big(|\hat{\mu}_1 - \mu_1| > \varepsilon/2\Big) + o(1).
\end{aligned}
$$

Choose $\varepsilon_M := M \max\{((m - \eta + 1)k^{1-\theta})^{1/\theta}, \tilde{C}(\log(m - \eta + 1)/k)^{1/2}\}$, for some $\tilde{C} \geq \big(64\Xi_{2,1}^2/C_3\big)^{1/2}$. By assumption, $n - \tau \geq 2k$ such that $m - \eta + 1 \geq 2$. Thus,

$$
\begin{aligned}
\mathbb{P}\Big(|\hat{d} - d_*| > \varepsilon_M\Big) &\leq C_1 \frac{2\Xi_{\theta,1}^{\theta}}{(M/4)^{\theta}} + C_2 \exp\Big(-M^2 \log(2)\Big) \\
&\qquad + \mathbb{P}\Big(|\hat{\mu}_1 - \mu_1| > M\tilde{C}(k\eta)^{-1/2}/2\Big) + o(1),
\end{aligned}
$$

where in the second line we used $(k\eta)^{-1/2} \leq (\log(m - \eta + 1)/k)^{1/2}$. Recalling Lemma 2.6.4, we can thus choose M large enough for the bound to be arbitrarily small. Therefore, we have proved

$$\hat{d} = d_* + \mathcal{O}_{\mathbb{P}}(\max\{(\log(m - \eta + 1)/k)^{1/2}, (mk^{1-\theta})^{1/\theta}\}).$$

Finally, by condition (C1),

$$m^{1/\theta}k^{(1-\theta)/\theta} = \frac{(mk)^{1/\theta}}{k^{1/2}k^{1/2}} \leq \frac{1}{k^{1/2}}\frac{n^{1/\theta}}{\sqrt{k}} = \frac{1}{k^{1/2}}\mathcal{O}(1),$$

which finishes the proof. □

CHAPTER 3
MULTIVARIATE CHANGE POINT ANALYSIS UNDER IRREGULAR SIGNALS

3.1 Introduction

In this chapter, we aim to extend our uni-variate change point analysis framework under irregular signals, introduced in Chapter 2, to a multivariate change point analysis framework. This extension is motivated by the application to COVID-19 data analysis discussed in Section 2.5.1.

In Section 2.5.1, we separately detected and located the first change point in search index time series for each COVID-19 related symptom. Interestingly, for both "fever" and "cough" keywords, we identified the first change point as being on the same date, December 08, 2019. This outcome leads us to consider whether aggregating these uni-variate time series into a single multivariate time series and conducting change point analysis might yield the same date, thus providing further evidence for an early outbreak of COVID-19. To explore this possibility, it becomes necessary to adapt our framework for change point analysis under irregular signals to a multivariate setting.

A substantial body of literature has been dedicated to the development of methodologies for multivariate change point analysis. Vert and Bleakley [2010] reformulated the multivariate change point analysis problem into the group lasso form and considered it within the framework of variable selection. Cho and Fryzlewicz [2015] proposed Sparse Binary Segmentation, employing a hard thresholding of the CUSUM matrix followed by an l_1 aggregation to mitigate the impact of irrelevant noise from high-dimensional problems. Lavielle and Teyssière [2006] introduced a procedure based on penalized Gaussian log-likelihood as a cost function, with the estimator computed via dynamic programming. Enikeeva and Harchaoui [2019] presented a test based on a linear statistic that considers all coordinates and

a scan statistic that uses l_2 aggregation on the CUSUM for all coordinates under sparsity assumptions.

In contrast to the existing literature, our objective is to conduct change point analysis on data for which the multivariate signals post-change point can be highly irregular, potentially leading to tipping points (as illustrated in Figure 3.2). Such scenarios can render CUSUM-based methods ineffective. To address these challenges, we propose a framework for multivariate change point analysis that works under these conditions. Our change point locating algorithm adopts a two-step estimation procedure similar to that in Chapter 2, where we conducted uni-variate change point locating under irregular signals. However, unlike Chapter 2, which only considers a one-sided case (signal is stronger after the change point), the multivariate signals exhibit a potentially more complex structure after the first change point. Consequently, for the initial estimation, we have more quantities to estimate.

The extension of our uni-variate framework to the multivariate setting enables a more comprehensive analysis of complex, high-dimensional data from various domains, including public health surveillance, financial market analysis, and climate change monitoring, where the identification of first change points is crucial for informed decision-making and timely interventions.

The chapter is structured as follows: Section 3.2 presents our methodology for hypothesis testing for the existence of change points and the two-step change point localization. Theoretical results are developed in Section 3.3, where we demonstrate the type-one error control and power analysis of the testing procedure and the convergence rate of the estimated change point to the true change point. The empirical efficacy of our method is evaluated through simulations in Section 3.4. In Section 3.5, we apply our methodology to the aggregated Baidu search indices for COVID-19 symptoms in 2019-2020 and compare the located first pandemic outbreak date with our result in Section 2.5.1. Comprehensive proofs are provided in Section 3.6.

3.2 Methodology

3.2.1 Model for the Data and the Problems Considered

Consider a p-dimensional sequential noisy model given by

$$\mathbf{X}_t = \boldsymbol{\mu}_t + \mathbf{Z}_t, \quad t = 1, 2, \ldots, n, \tag{3.2.1}$$

where $\boldsymbol{\mu}_t \in \mathbb{R}^p$ denotes the signal vector at time t, and the sequence $(\mathbf{Z}_t)_t$ represents p-dimensional stationary noise with a long-run covariance matrix defined as

$$\boldsymbol{\Sigma} := \sum_{k=-\infty}^{\infty} \mathrm{Cov}(\mathbf{Z}_t, \mathbf{Z}_{t+k}) \in \mathbb{R}^{p\times p}. \tag{3.2.2}$$

Let $(Q_t)_{t=1,2,\ldots,n}$ denote the sequence of squared Euclidean norms of the noise vectors $(\mathbf{Z}_t)_{t=1,2,\ldots,n}$:

$$Q_t = \mathbf{Z}_t^\top \mathbf{Z}_t, \quad t = 1, 2, \ldots, n. \tag{3.2.3}$$

Define $\gamma := \mathbb{E}(Q_t)$ as the mean of this sequence and ω^2 as its long-run variance:

$$\omega^2 := \sum_{k=-\infty}^{\infty} \mathrm{Cov}(Q_t, Q_{t+k}) < \infty \tag{3.2.4}$$

The null hypothesis is formulated as follows:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_n, \tag{3.2.5}$$

which posits that the signal remains constant over time. In contrast, the alternative hypothesis is given by:

$$H_1 : \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_{\tau-1} = \boldsymbol{\mu}_0, \quad \|\boldsymbol{\mu}_\tau - \boldsymbol{\mu}_1\|^2, \ldots, \|\boldsymbol{\mu}_n - \boldsymbol{\mu}_1\|^2 \geq d, \quad d > 0, \tag{3.2.6}$$

which suggests a shift in the signal level after a specific observation τ, with the change exceeding a minimum Euclidean distance $\sqrt{d}$. This model generalizes the uni-variate case discussed in Chapter 2 to a multivariate setting, enabling a more comprehensive analysis of changes across multiple dimensions.

Our goal is to develop a methodology for estimating the change point τ under the alternative hypothesis H_1. This approach is motivated by the analysis of the dataset introduced in Section 2.5.1, where instead of estimating change points for each time series corresponding to symptom-related keywords individually, we aim to aggregate these time series into a singular multivariate series to ascertain the collective change point, enhancing the robustness and interpretability of our findings.

3.2.2 Testing Procedure

We introduce a statistical test aimed at distinguishing between the null hypothesis H_0 and the alternative hypothesis H_1, as formally defined in equations (3.2.5) and (3.2.6), respectively. This test is developed in the context of observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ adhering to the distribution specified in (3.2.1). The proposed test statistic is given by:

$$\hat{T} := \min_{j=1,2,\ldots,n} \frac{1}{\sqrt{n}\hat{\omega}} \left(\sum_{i=1}^{j} \|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_0\|^2 - \frac{j}{n} \sum_{i=1}^{n} \|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_0\|^2 \right), \tag{3.2.7}$$

where $\hat{\boldsymbol{\mu}}_0$, and $\hat{\omega}^2$ are estimators of the non-signal mean $\boldsymbol{\mu}_0$, and the long-run variance ω^2, as described in (3.2.3). Section 3.2.3 introduces estimators for both ω^2 and $\boldsymbol{\mu}_0$ that are consistent under the assumptions of both H_0 and H_1.

In Section 3.3.2, we demonstrate that under the null hypothesis H_0, the distribution of the test statistic T_n converges, in a distributional sense, to that of the minimum of a standard Brownian bridge. This convergence allows for the derivation of critical values through either simulation techniques or asymptotic approximations. Under the alternative hypothesis H_1, it is shown that T_n converges in probability to $-\infty$ as $n \to \infty$. Consequently, the critical region is defined by $\{T_n < c\}$, where c is determined as the α-quantile of the minimum of a standard Brownian bridge distribution, as further elaborated in Section 3.3.2.

3.2.3 A Two-Step Locating Algorithm

Data Blocking

To reduce noise and emphasize the signal, we partition the data set into $\lfloor n/k \rfloor$ blocks of size k. An initial estimate for $\boldsymbol{\mu}_0$ is obtained using the sample mean of the first block:

$$\bar{\boldsymbol{\mu}}_0 := \frac{1}{k} \sum_{i=1}^{k} \mathbf{X}_i. \tag{3.2.8}$$

We then define the following quantities that measure the average squared Euclidean distance between our samples and the estimated pre-change mean:

$$R_{j,k} := \frac{1}{k} \sum_{i=(j-1)k+1}^{jk} \|\mathbf{X}_i - \bar{\boldsymbol{\mu}}_0\|^2, \quad j = 1, 2, \ldots, \lfloor n/k \rfloor. \tag{3.2.9}$$

The minimizer of these quantities is defined as $\hat{L}$:

$$\hat{L} := \underset{j \in \{2, \ldots, \lfloor n/k \rfloor\}}{\arg\min} R_{j,k}, \quad \hat{\ell} := k\hat{L}, \tag{3.2.10}$$

where $\hat{L}$ indicates a block likely to contain only pre-change observations and $\hat{\ell}$ is the index of the last observation in the $\hat{L}$th block.

Preliminary Estimates for Several Quantities

Using the first $\hat{L}$ blocks, we obtain refined estimates for $\boldsymbol{\mu}_0$ and γ:

$$\hat{\boldsymbol{\mu}}_0 := \frac{1}{\hat{L}k} \sum_{i=1}^{\hat{L}k} \mathbf{X}_i \tag{3.2.11}$$

and

$$\tilde{\gamma} := \frac{1}{\hat{L}k} \sum_{i=1}^{\hat{L}k} \|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_0\|^2. \tag{3.2.12}$$

We further estimate the long-run variance ω^2 of (Q_t) by

$$\hat{\omega}^2 := \frac{k}{\hat{\ell}} \sum_{s=k}^{\hat{\ell}} \left(R_{s/k} - \tilde{\gamma} \right)^2, \quad R_{s/k} := \frac{1}{k} \sum_{i=s-k+1}^{s} \|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_0\|^2, \tag{3.2.13}$$

where $R_{s/k}$ extends (3.2.9) to overlapping blocks. The motivation for this estimate is provided in Section 2.2.3.

Locating Algorithm: Step 1

Step 1 aims to refine the estimates of $\boldsymbol{\mu}_1$ and γ and provide an estimate of the change magnitude d. Using the block averages $R_{j,k}$, the preliminary estimates $\hat{\boldsymbol{\mu}}_0$, $\tilde{\gamma}$, and $\hat{\omega}^2$ from (3.2.11), (3.2.12), and (3.2.13), respectively, we compute the standardized 'test statistics' $\hat{D}_j$ and corresponding 'test decisions' $\hat{I}_j$:

$$\hat{D}_j := \frac{1}{\sqrt{k}} \cdot \frac{\sum_{i=(j-1)k+1}^{jk} \left(\|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_0\|^2 - \tilde{\gamma} \right)}{\hat{\omega}}, \quad \hat{I}_j = \begin{cases} 1 & \text{if } \hat{D}_j \geq z_{1-1/\lfloor n/k \rfloor}, \\ 0 & \text{otherwise,} \end{cases} \tag{3.2.14}$$

where z_α is the α-quantile of the standard normal distribution, and $m := \lfloor n/k \rfloor$. We then estimate the index η of the last block before the change:

$$\hat{\eta} := \underset{t=1,\ldots,\lfloor n/k \rfloor - 1}{\arg\min} \sum_{j=1}^{\lfloor n/k \rfloor} \left(\hat{I}_j - 1_{[t+1,\lfloor n/k \rfloor]}(j) \right)^2. \tag{3.2.15}$$

Finally, we obtain further refined estimates for $\boldsymbol{\mu}_1$ and γ, as well as a preliminary estimate for the change magnitude d:

$$\hat{\boldsymbol{\mu}}_1 := \frac{1}{k\hat{\eta}} \sum_{i=1}^{k\hat{\eta}} \mathbf{X}_i, \tag{3.2.16}$$

and

$$\hat{\gamma} := \frac{1}{\hat{\eta}k} \sum_{i=1}^{\hat{\eta}k} \|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_1\|^2, \tag{3.2.17}$$

and

$$\hat{d} := \min_{i \geq k\hat{\eta}+k+1} \frac{1}{k} \sum_{j=i}^{i+k-1} \left(\left\| \mathbf{X}_j - \hat{\boldsymbol{\mu}}_1 \right\|^2 - \hat{\gamma} \right). \tag{3.2.18}$$

Let $\eta := \lfloor \tau/k \rfloor$ be the index of the last block before the change point τ, ensuring $\eta k + 1 \leq \tau \leq (\eta+1)k$. For $j = 1, \ldots, \eta$, $\mathbb{E}R_j = \gamma$, while for $j = \eta+1, \ldots, m$, $\mathbb{E}R_j > \gamma$. Thus, $R_1, \ldots, R_\eta$ are close to γ, but $R_{\eta+1}, \ldots, R_m$ exceed γ, making $\hat{L}$ nearly uniform over $\{1, \ldots, \eta\}$. Consequently, $\hat{\boldsymbol{\mu}}_0$ averages about $k\eta/2$ pre-change observations, providing a $\sqrt{n}$-consistent estimate for $\boldsymbol{\mu}_0$ if τ scales with n.

To improve accuracy, $\hat{\boldsymbol{\mu}}_1$ uses $\hat{\eta}$ instead of $\hat{L}$, which is closer to the true η. The test decisions $\hat{I}_j$ differentiate between pre-change ($\mathbb{E}R_j = \gamma$) and post-change ($\mathbb{E}R_j > \gamma$) blocks. The sequence $\hat{I}_1, \ldots, \hat{I}_m$ resembles a step function with $I_j = 0$ for $j \leq \eta$ and $I_j = 1$ for $j > \eta$. $\hat{\eta}$ refines the estimate of η by fitting a step function that shifts from 0 to 1 at the change point.

Locating Algorithm: Step 2

The refined estimate for the change point τ as delineated in (3.2.6) is given by:

$$\tilde{\tau} := \arg \min_{j=1,\ldots,n} \left(\sum_{t=1}^{j} \left(\|\mathbf{X}_t - \hat{\boldsymbol{\mu}}_1\|^2 - \hat{\gamma} - \hat{c} \right) + 1 \right), \tag{3.2.19}$$

where $\hat{c} = \rho \hat{d}$, with $\rho \in (0, 1)$ serving as a tuning parameter. Theorem 3.3.2 provides theoretical support for the estimator $\tilde{\tau}$.

3.3 Theory

3.3.1 Assumptions on the Multivariate Stationary Noise Process

To derive meaningful insights into the statistical properties of our proposed methods, we make certain assumptions regarding the multivariate noise process $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ in model (3.2.1).

We adopt a framework of functional dependence measure inspired by Wu [2005]. Within this framework, the multivariate causal stationary process $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ is conceptualized as the output from a physical system, described by

$$\mathbf{Z}_t := \mathbf{G}(\ldots, \varepsilon_{t-1}, \varepsilon_t), \tag{3.3.20}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ represents the system's input information, assumed to be i.i.d. with mean zero and variance one. The function $\mathbf{G}$ is a measurable mapping acting as a filter or mechanism of the system. To assess the system's dependence, we examine the variation in outputs upon substituting the input at time $t = 0$ with an i.i.d. copy ε_0'. The functional dependence measure for the p-dimensional process $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ is defined as:

$$\delta_q(i) := \max_{j=1,\ldots,p} \left(\mathbb{E}\left|\mathbf{Z}_{i,j} - \mathbf{Z}'_{i,j}\right|^q\right)^{1/q}, \text{ where } \mathbf{Z}'_i = \mathbf{G}\left(\ldots, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_i\right). \tag{3.3.21}$$

This measure quantifies the maximum difference in the q-th moment between the original process and its coupled version, across all dimensions $j = 1, \ldots, p$, when the input at time 0 is replaced by an independent copy. It captures the sensitivity of the process to changes in the input information at a specific time point.

To fully characterize the temporal dependence of the time series, we introduce the cumulative dependence measure of $(\mathbf{Z}_i)_{i\geq t}$ on ε_0, defined as:

$$\Theta_{t,q} = \sum_{i \geq t} \delta_q(i), \quad t \geq 0. \tag{3.3.22}$$

This measure accumulates the functional dependence across all time points $i \geq t$, providing a comprehensive assessment of the process's dependence on the input at time 0. It takes into account the long-range dependence structure of the process.

Building upon this, we introduce the dependence adjusted norm (DAN), as discussed in Wu and Wu [2016] and Han and Tsay [2020]:

$$\Xi_{q,\alpha} = \sup_{i \geq 0} (i+1)^{\alpha} \Theta_{i,q}, \tag{3.3.23}$$

which incorporates a weight function $(i+1)^{\alpha}$ to adjust for the decay of dependence over time. It provides a unified measure of the process's dependence, taking into account both the temporal decay and the cumulative effect of the functional dependence. The parameter α controls the rate at which the weights decrease with increasing time lag.

In some cases, due to the presence of strong dependence, $\Xi_{q,\alpha}$ might be infinite while the q-th moment of the process remains finite:

$$H_q := \max_{j=1,2,\ldots,p} \left(\mathbb{E} \left[|\mathbf{Z}_{i,j}|^q \right] \right)^{1/q} < \infty. \tag{3.3.24}$$

This is not surprising, as the following inequality holds by stationarity:

$$\begin{aligned} H_q &\leq \max_{j=1,2,\ldots,p} \sum_{l=0}^{\infty} \left\| \mathbb{E} \left(\mathbf{Z}_{0,j} \mid \mathcal{F}_{-l} \right) - \mathbb{E} \left(\mathbf{Z}_{0,j} \mid \mathcal{F}_{l-1} \right) \right\|_q \\ &= \max_{j=1,2,\ldots,p} \sum_{l=0}^{\infty} \left\| \mathbb{E} \left(\mathbf{Z}_{l,j} - \mathbf{Z}'_{l,j} \mid \mathcal{F}_0 \right) \right\|_q \\ &\leq \max_{j=1,2,\ldots,p} \sum_{l=0}^{\infty} \left\| \mathbf{Z}_{l,j} - \mathbf{Z}'_{l,j} \right\|_q = \Xi_{q,0}, \end{aligned} \tag{3.3.25}$$

where $\mathcal{F}_l = \sigma(\ldots, \varepsilon_{l-1}, \varepsilon_l)$ denotes the σ-algebra generated by the input information up to time l. This inequality establishes a connection between the q-th moment of the process and the DAN.

In the special case where $\mathbf{Z}_i$, for $i \in \mathbb{Z}$, are i.i.d., the DAN $\Xi_{q,\alpha}$ and the q-th moment H_q are equivalent, satisfying the following bounds:

$$H_q \leq \Xi_{q,\alpha} \leq 2H_q. \tag{3.3.26}$$

This equivalence highlights the fact that, in the absence of dependence, the DAN reduces to the q-th moment of the process.

These assumptions and definitions provide a solid foundation for analyzing the statistical properties of our proposed methods in the context of multivariate time series with noise processes that exhibit temporal dependence. By carefully characterizing the dependence structure, we can derive theoretical properties of our estimators under mild conditions.

3.3.2 Testing Procedure

We now state results on the test described in Section 3.2.2. The first result provides the asymptotic distribution under the null hypothesis, and the second one asserts asymptotic consistency under H_1, where we will let τ and d depend on n.

Theorem 3.3.1. *Consider a sequence $\{\mathbf{Z}_i\}$ where each element follows the framework specified in* (3.3.20)*. It is assumed that the sequence maintains a finite second moment ($H_2 < \infty$) for each dimension and adheres to a short-range dependence criterion expressed as:*

$$\Theta_{0,4} = \sum_{i \geq 0} \delta_{i,4} < \infty. \tag{3.3.27}$$

(i) Under H_0, as $n \to \infty$, the convergence of the distribution of the test statistic is given by

$$\sup_{x \leq 0} |\mathbb{P}(\hat{T}_n \leq x) - e^{-2x^2}| \to 0, \tag{3.3.28}$$

indicative of the test statistic's asymptotic distribution aligning with a specific exponential decay function.

(ii) Under H_1, with the condition $(\tau_n/\sqrt{n})(1-\tau_n/n)d_n \to \infty$, it is demonstrated that as $n \to \infty$,

$$\hat{T}_n \to -\infty \quad \textit{in probability},$$

highlighting the test statistic's tendency to diverge under the alternative hypothesis.

The short-range dependence criterion, stated in (3.3.27), guarantees that the dependence between the elements of the sequence decays rapidly as the distance between them increases. This condition is crucial for establishing the asymptotic distribution of the test statistic under the null hypothesis.

Theorem 3.3.1(i) proves that, under the null hypothesis, the test statistic $\hat{T}_n$ converges in distribution to a random variable with a known distribution function, which can be used to determine the critical values for the test. Furthermore, Theorem 3.3.1(i) suggests using the α-quantile of the asymptotic limit, $-(-0.5 \log \alpha)^{1/2}$, as the critical value for testing H_0, given any significance level $\alpha \in (0, 1)$. This critical value is derived from the asymptotic distribution of the test statistic under the null hypothesis, which is given by the exponential decay function e^{-2x^2}. By employing this critical value, the test will have an asymptotic size equal to the desired significance level α.

The condition in Theorem 3.3.1(ii) ensures that the change point is not too close to the boundaries of the observed sequence and that the magnitude of the change, d_n, is sufficiently large relative to the sample size n. Under these conditions, the test statistic $\hat{T}_n$ diverges to negative infinity in probability, implying that the test will reject the null hypothesis with

high probability when the alternative hypothesis is true.

3.3.3 Locating Algorithm

To establish a convergence theory for the estimated change points, we require the following assumption on temporal dependence.

Condition 3.3.1. *Consider a sequence* $\{\mathbf{Z}_i\}$*, where each* $\mathbf{Z}i$ *is as defined in* (3.3.20)*. We assume that for every dimension, the sequence* $\mathbf{Z}_i$ *satisfies the condition that its* q*-th moment* H_q *is finite, for* $q > 2$*, and that* $\Theta_{m,2q} = O\left(m^{-\beta_{2q}}(\log m)^{-A}\right)$*, where* $A > \left(1/q + 2 + 2\beta_{2q}\right)/3$*, and*

$$\beta_x = \frac{x^2 - 4 + (x-2)\sqrt{x^2 + 20x + 4}}{8x}.$$

Additionally, any one of the following conditions must be satisfied:

- $q > 4$*, and it holds that* $\Theta_{m,q} = O\left(m^{-\beta_q}(\log m)^{-A}\right)$*, where* $A > 2\left(1/q + 1 + \beta_q\right)/3$*;*
- $q = 4$*, and it holds that* $\Theta_{m,q} = O\left(m^{-1}(\log m)^{-A}\right)$ *with* $A > \frac{3}{2}$*;*
- $2 < q < 4$*, and it holds that* $\Theta_{m,q} = O\left(m^{-1}(\log m)^{-1/q}\right)$*.*

Condition 3.3.1 imposes moment and dependence restrictions on the sequence $\{\mathbf{Z}_i\}$. The finiteness of the q-th moment, H_q, for $q > 2$, ensures that the sequence has well-behaved tail properties. The conditions on $\Theta_{m,2q}$ and $\Theta_{m,q}$ control the rate at which the dependence measure decays as the lag m increases. These conditions are crucial for establishing the convergence rates of the estimated change points. The parameter β_x is a function of the moment order x and appears in the conditions on $\Theta_{m,2q}$ and $\Theta_{m,q}$. It determines the rate of decay required for the dependence measure, with larger values of β_x corresponding to faster decay rates.

Our main result bounds the error of estimating τ by $\tilde{\tau}$ in terms of the minimum gap d to the signal. We have:

Theorem 3.3.2. *Assume Condition 3.3.1 holds. Then, depending on a regime for the gap parameter d, we have:*

(i) If d is fixed, then we have $|\tilde{\tau}_n - \tau_n| = \mathcal{O}_{\mathbb{P}}(1)$.

(ii) If $d_n \to \infty$, as $n \to \infty$, and $c = d/2$, in the estimator, then we have $\mathbb{P}(\tilde{\tau}_n = \tau_n) \to 1$.

(iii) If $0 < d_n \to 0$, as $n \to \infty$, then we have $|\tilde{\tau}_n - \tau_n| = \mathcal{O}_{\mathbb{P}}(d_n^{-2})$.

Theorem 3.3.2 provides convergence rates for the estimated change point $\tilde{\tau}_n$ under different regimes for the gap parameter d. When d is fixed, the estimation error is bounded in probability by a constant. When d_n tends to infinity, the estimated change point is consistent, i.e., it converges in probability to the true change point. When d_n tends to zero, the estimation error is bounded in probability by the inverse square of d_n. These results demonstrate the robustness of the proposed locating algorithm under various signal strengths. The proof is deferred to Section 3.6.2. Our second result of this section asserts that we may expect the statistical properties of the estimate $\hat{\eta}_n$, defined in (3.2.15), to be very good as well, which legitimises $\tilde{\tau}_n$, in (3.2.19), as our final estimate.

Theorem 3.3.3. *Assume Condition 3.3.1 holds, and* $\log(\eta_n) n^{1/q} k^{-1} \to 0$. *Then,*

$$\mathbb{P}\left(\hat{\eta}_n = \eta_n\right) \to 1, \qquad n \to \infty. \tag{3.3.29}$$

Theorem 3.3.3 establishes the consistency of the estimated index $\hat{\eta}_n$, which is used to construct the final estimate $\tilde{\tau}_n$. The condition $\log(\eta_n) n^{1/q} k^{-1} \to 0$ imposes a restriction on the growth rate of the block size k relative to the sample size n and the moment order q. This condition ensures that the bias introduced by the block-wise estimation is asymptotically negligible. The proof is deferred to Section 3.6.3. If $k_n = o(n)$, then the condition $\log(\eta_n) n^{1/q} k^{-1} \to 0$, from Theorem 3.3.3, is equivalent to $k_n^{-1} = o\left(\log(n)^{-1} n^{-1/q}\right)$. This can be seen from the definitions of τ and η, by which we have $\tau_n = \eta_n k_n + O\left(k_n\right) = cn + o(n)$. Generally, in practice, the moment of the noise q is unknown, we suggest to use $k_n = \lfloor \sqrt{n} \rfloor$,

which is sufficient if $q > 4$.

3.4 Monte Carlo Studies

3.4.1 Models Considered

We evaluate the finite sample efficacy of both the testing procedure (Section 3.2.2) and the two-stage locating algorithm (Section 3.2.3) through simulation experiments. These experiments involve data simulated according to the signal plus noise model outlined in (3.2.1).

For the noise component, a multivariate threshold autoregressive (AR) model is employed:

$$\mathbf{Z}'_{i,j} = \theta \left(|\mathbf{Z}'_{i-1,j}| + |\mathbf{Z}'_{i-2,j}| \right) + \varepsilon_{i,j}, \qquad j = 1, 2, \ldots, p, \tag{3.4.30}$$

where θ is the coefficient determining temporal dependence and the i.i.d. innovations $\varepsilon_{i,j}$ are distributed as $\mathcal{N}(0, 0.5^2)$. The strength of the temporal dependence increases with the absolute value of θ, and the process remains stationary for $|\theta| < 0.5$. The noise process $(\mathbf{Z}_i)$ is derived by centering $\mathbf{Z}_i$ as $\mathbf{Z}_i := \mathbf{Z}'_i - \mathbb{E}(\mathbf{Z}'_i)$.

Regarding the signal μ_t, two scenarios are considered:

(i) Under H_0, as defined in (3.2.5), the signal is constant, with $\boldsymbol{\mu}_1 = \mathbf{0}$.

(ii) Under H_1, as defined in (3.2.6), the signal evolves according to:

$$\boldsymbol{\mu}_t = \begin{cases} \boldsymbol{\mu}_1 := \mathbf{0} & \text{for } t = 1, \ldots, \tau - 1 \\ \boldsymbol{\mu}_1 + s \left(\frac{2t - 3\tau + \tau'}{\tau' - \tau} \right) \cdot \boldsymbol{\zeta}_t / \|\boldsymbol{\zeta}_t\| & \text{for } t = \tau, \ldots, \tau' \\ \boldsymbol{\mu}_1 + s \left(2 + \exp\left(\frac{2(t - \tau')}{\tau'' - \tau'} \right) \right) \cdot \boldsymbol{\zeta}_t / \|\boldsymbol{\zeta}_t\| & \text{for } t = \tau' + 1, \ldots, \tau'' \\ \boldsymbol{\mu}_1 + s \left(2 + \exp(2) \cdot \frac{2n - \tau'' - t}{2n - 2\tau''} \right) \cdot \boldsymbol{\zeta}_t / \|\boldsymbol{\zeta}_t\| & \text{for } t = \tau'' + 1, \ldots, n, \end{cases} \tag{3.4.31}$$

where s indicates the scale of deviation from the baseline mean state ($\boldsymbol{\mu}_1 = \mathbf{0}$) for $t < \tau$ to a variable mean state for $t \geq \tau$. The random vectors $\boldsymbol{\zeta}_t$ are i.i.d. and follow a standard p-dimensional Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$. This model effectively captures trends similar to

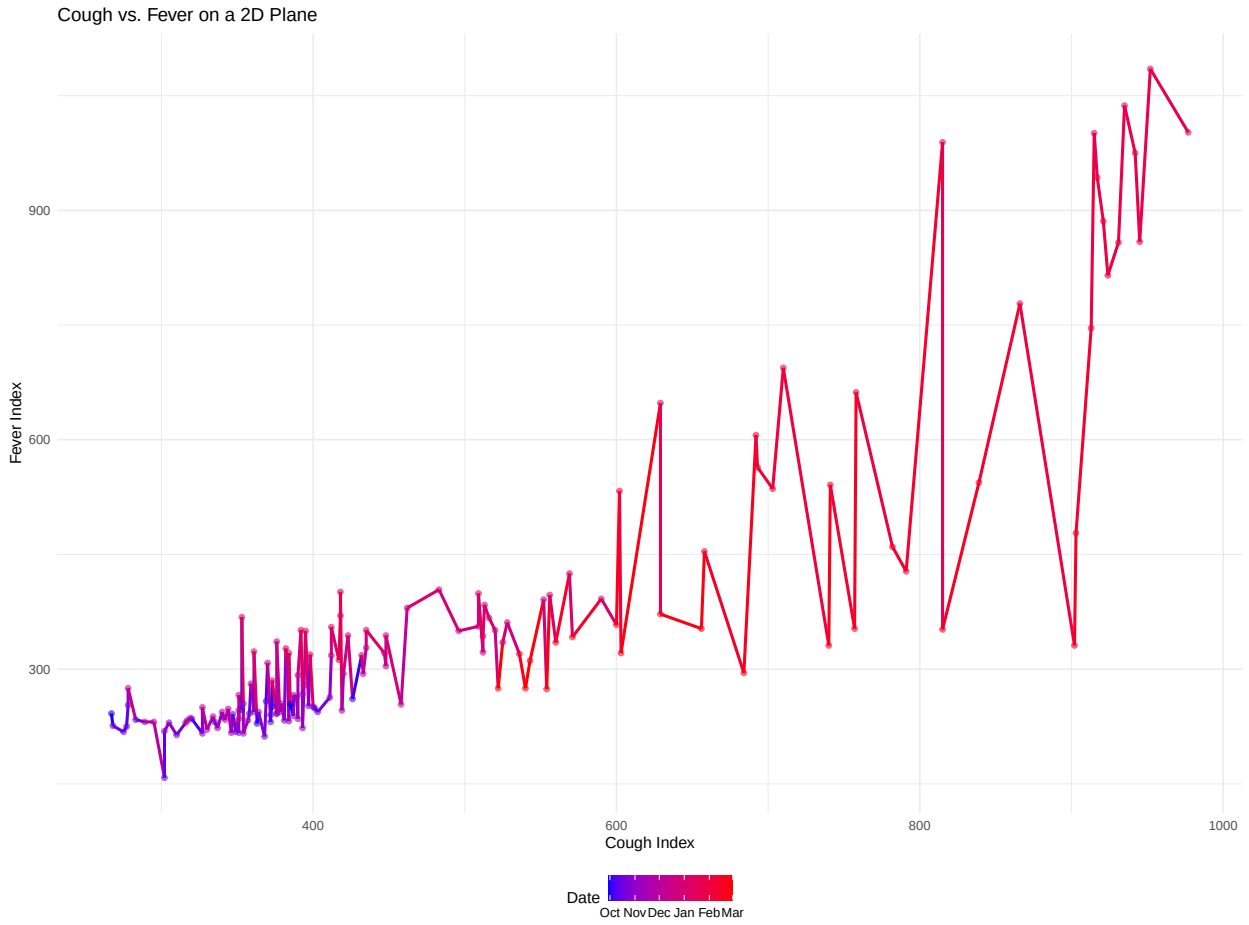

Figure 3.1: Daily Baidu search indices for COVID-related symptom keywords from October 1, 2019, to January 31, 2020. The indices for the keywords "cough" and "fever" are aggregated and displayed in a two-dimensional time series plot.

those observed in multivariate time series data, such as the search engine index data depicted in Figure 3.1, where indexes from various keywords are aggregated into a multivariate time series.

Figure 3.2 presents an example of the three-dimensional data series ($\mathbf{X}_i$), illustrating the dynamics of signal strength $\|\boldsymbol{\mu}_t\|$. Initially, the signal remains at a baseline level $\boldsymbol{\mu}_1 = \mathbf{0}$. At $\tau = 120$, there's a discernible increase in signal strength by at least $s = 0.5$. Subsequently, at $\tau' = 180$, the signal experiences a further elevation, surpassing $8s = 4$ above the original baseline. This depiction mirrors patterns observed within real-world datasets, demonstrating significant changes at specified points.

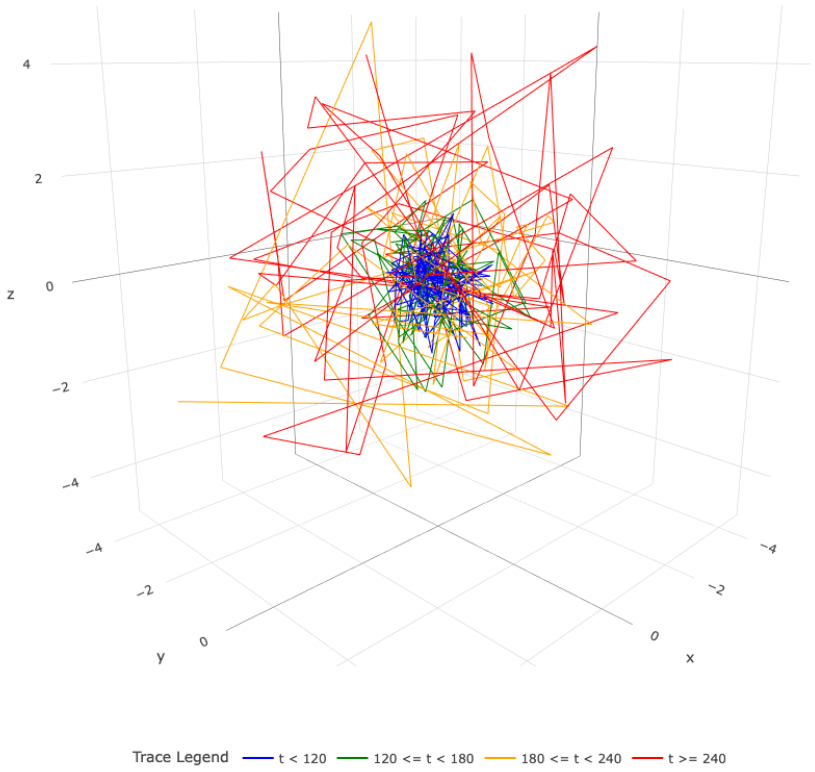

Figure 3.2: Visualization of $(\mathbf{X}_t)$ for $n = 300$, $p = 3$, marking significant shifts at $\tau = 120$, $\tau' = 180$, and $\tau'' = 240$, with an initial signal increase factor of $s = 0.5$.

3.4.2 Synthetic Data under the Null Hypothesis

In this section, we demonstrate that the testing procedure described in Section 3.2.2 has the correct size, asymptotically. We employ data structured as detailed in Section 3.4.1, operating under a constant signal.

We vary the sample size n, selecting values from the set $\{50, 100, 300, 500, 2000\}$, and choose the dimensionality p from $\{2, 3, 5\}$. Additionally, we adjust the dependence parameter θ across a range including 0 (indicating independence), 0.2, and 0.4. Throughout, we maintain a constant significance level of $\alpha = 0.05$. The empirical sizes, computed from 100,000 simulations, are concisely presented in Table 3.1.

Table 3.1: Rejection ratios for multivariate change point testing procedure under the null hypothesis; cf. (3.2.5).

n	p	θ		
		0.0	0.2	0.4
50	2	0.077	0.078	0.076
	3	0.076	0.073	0.075
	5	0.073	0.073	0.076
100	2	0.065	0.072	0.074
	3	0.061	0.062	0.066
	5	0.059	0.063	0.060
500	2	0.055	0.061	0.062
	3	0.057	0.059	0.063
	5	0.056	0.060	0.060
2000	2	0.053	0.059	0.061
	3	0.055	0.058	0.063
	5	0.052	0.058	0.063
5000	2	0.049	0.056	0.061
	3	0.051	0.055	0.059
	5	0.050	0.570	0.062

Upon reviewing the data in Table 3.1, we observe that the rejection ratios, which serve as indicators for the type-one error rate, progressively align more closely with the predetermined significance level ($\alpha = 0.05$) as the sample size n increases and the degree of temporal dependence θ decreases. These findings are in harmony with the theoretical results laid out in Section 3.3.2. Furthermore, the type-one error demonstrates negligible variance within our finite-dimensional framework ($p = 2, 3, 5$), reinforcing the suitability of this testing approach for the analysis of multivariate time series data, as depicted in scenarios described in Section 3.5.1.

3.4.3 Synthetic Data under Alternative Hypotheses

This section delves into the effectiveness of the algorithm for identifying the initial change point using synthetic data. We utilize the data framework outlined in Section 3.4.1, where the signal is specified according to (3.4.31) (that is, H_1).

The experimental design is as follows: The sample size, n, is varied among $50, 100, 300, 500, 2000$. The dimension, p, is selected from $2, 3, 5$. The dependence parameter, θ, includes 0 (indicating independence), 0.2, and 0.4. The gap parameter, s, encompasses $\{0.8, 1.2, 1.5, 2, 2.5\}$. It's important to note that the standard deviation of innovation in the dependent process is consistently 0.5. We fix the ratios $\tau/n = 0.4$, $\tau'/n = 0.6$, $\tau''/n = 0.8$, with $\boldsymbol{\mu}_1 = 0$. Following the establishment of experimental parameters, we generate the trend $(\boldsymbol{\mu}_i)$ using the described methodology. The additive noise process is then simulated multiple times, and this data is processed by our location algorithms. Outcomes are compiled from 100,000 independent simulations.

We now present the Normalized Root Mean Squared Errors (RMSE) for $\hat{\tau}$, derived from our two-step location algorithm, expressed as $\sqrt{\mathbb{E}(\hat{\tau} - \tau)^2}/n$, across various experimental settings as depicted in Figure 3.3.

Error rates generally decrease with moderate temporal dependence. Specifically, extending the gap between signal and non-signal segments from 0.8 to 2.5 results in a consistent reduction in error rates. Additionally, increasing the sample size from 50 to 2000 leads to lower normalized errors, whereas an increase in dimension from 2 to 5 induces a slightly higher normalized errors. These findings are consistent with the theoretical expectations of Theorem 3.3.2 discussed in Section 3.3.3 and its proof in Section 3.6.2.

In scenarios characterized by significant dependence and a narrow gap, error rates may increase. However, under more favorable conditions, the error rates remain largely stable or show only minor increases, underscoring the robustness and effectiveness of our proposed approach.

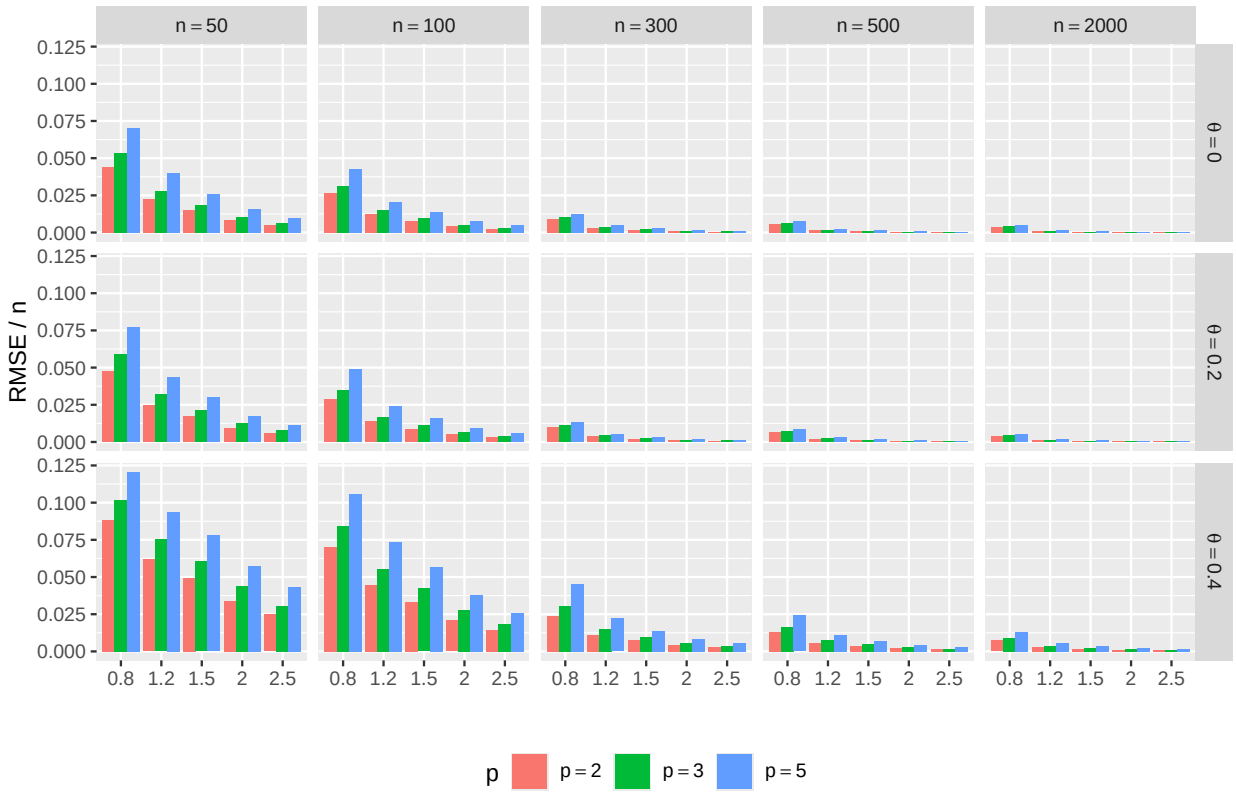

Figure 3.3: Normalized Root Mean Squared Errors (RMSE) for the proposed method $\hat{\tau}$, calculated as $\sqrt{\mathbb{E}(\hat{\tau}-\tau)^2}/n$. This analysis covers variations in parameters like the gap parameter s, sample size n, dimensionality p, and the θ parameter of the threshold autoregression noise model. Each bar graph represents the average outcome from $100,000$ replications.

3.5 Real Data Analysis

3.5.1 Baidu Search Index Analysis for COVID-19 Related Symptoms

In this analysis, we aggregate Baidu search index data for the keywords "fever" and "cough" from 1 October 2019 to 31 January 2020 in Hubei Province, China. Hence, this dataset is converted into a two-dimensional time series, which acts as the input for our multivariate change point analysis pipeline.

Employing the testing methodology described in Section 3.2.2, we evaluate the null hypothesis H_0 that the mean remains constant as detailed in Equation (3.2.5), against the alternative hypothesis H_1 which posits a variable mean following a change point, as depicted in Equation (3.2.6). The analysis yields test statistics $\hat{T} < -287.3842$ for the combined Baidu search indices of "cough" and "fever". With p-values from Theorem 3.3.1(i) approaching zero, we decisively reject the null hypothesis.

Further examination is carried out through a two-stage locating procedure (Section 3.2.3), focusing on the aggregated dataset for "fever" and "cough", which includes $n = 123$ data points. The initial phase calculates the mean of the equilibrium data state $\boldsymbol{\mu}_1$, the sum of the noise vector variances γ, and the state gap parameter d. This calculation aids the subsequent phase. The batch mean length is determined as $k = \lceil n^{1/3} \rceil = 5$, leading to

$$R_j := \frac{1}{k} \sum_{i=(j-1)k+1}^{jk} \|\mathbf{X}_i - \bar{\boldsymbol{\mu}}_0\|^2, \text{ for } j = 1, 2, \ldots, \lfloor n/k \rfloor,$$

as specified in Equation (3.2.9). This yields $\hat{L} := \arg\min_{1 \le j \le m} R_j = 9$ and $\hat{\ell} := k\hat{L} = 45$, aligning with Equation (3.2.10). The pre-change sample mean $\hat{\boldsymbol{\mu}}_0$ is approximated as $(235.31, 354.29)^\top$ from the initial $\hat{\ell}$ data points. Subsequently, the test statistics $\hat{D}_j$, using $\hat{\omega}_\infty \approx 2544.61$ (the square root of the estimated long-run variance from the initial $\hat{\ell}$ observations), are computed as per Equation (3.2.13), leading to the test decisions $\hat{I}_j$ out-

lined in Equation (3.2.14), with

$$\hat{I}_j = \begin{cases} 1 & \text{if } \hat{D}_j \geq z_{1-1/m} \\ 0 & \text{otherwise.} \end{cases}$$

We identify $\hat{\eta}$ as per Equation (3.2.15), resulting in first-stage estimates of

$$\hat{\boldsymbol{\mu}}_1 := \frac{1}{k\hat{\eta}} \sum_{i=1}^{k\hat{\eta}} \mathbf{X}_i \approx (237.32, 351.20)^\top, \quad \hat{\gamma} := \frac{1}{\hat{\eta}k} \sum_{i=1}^{\hat{\eta}k} \|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_1\|^2 \approx 1706.53,$$

and

$$\hat{d} := \min_{i \geq k\hat{\eta}+k+1} \frac{1}{k} \sum_{j=i}^{i+k-1} \left(\left\| \mathbf{X}_j - \hat{\boldsymbol{\mu}}_1 \right\|^2 - \hat{\gamma} \right) \approx 6951.74.$$

Setting $\rho = 0.5$, the refined second-phase change point estimate is obtained as follows:

$$\tilde{\tau} := \arg \min_{j=1,\ldots,n} \left(\sum_{t=1}^{j} \left(\|X_t - \hat{\mu}_1\|^2 - \hat{\gamma} - \hat{c} \right) + 1 \right) = 69,$$

as per Equation (3.2.19). This result identifies December 8, 2019, as the first change point date. In Figure 3.4, the observations before our estimated first change point date are colored in blue, while the observations after this date are colored in red. It is evident that detecting this first change point by visual inspection alone is almost impossible.

Remarkably, this date aligns with the findings presented in Section 2.5.1, further reinforcing the validity of our analytical approach.

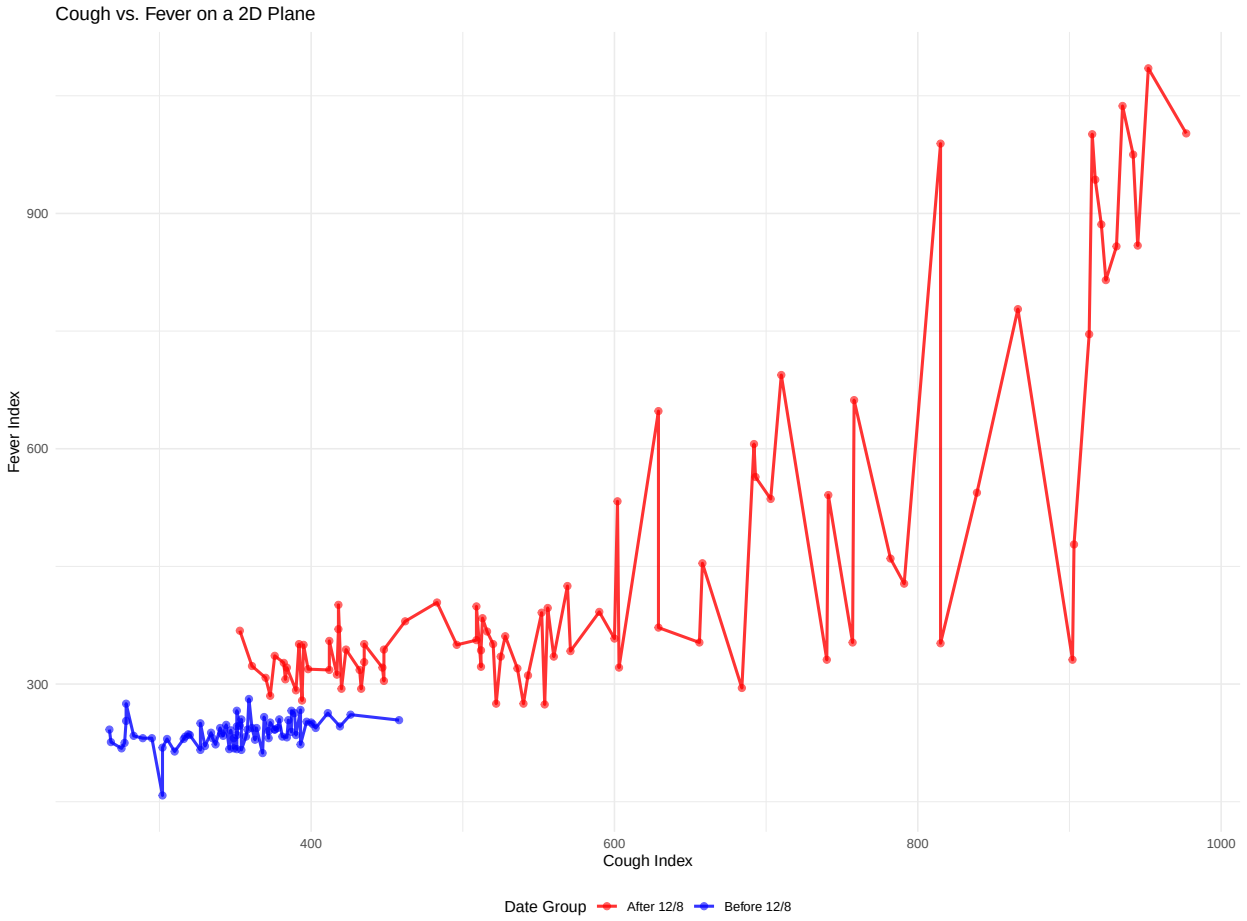

Figure 3.4: Baidu search index for "fever" and "cough" from October 1, 2019, to January 31, 2020. The blue points denote the observations before December 8, 2019, which is our estimated change point, while the red points denote the observations afterwards.

3.6 Proofs

This section outlines the formal proofs for the theorems introduced earlier in the chapter.

3.6.1 Proof of Theorem 3.3.1

We begin by defining the test statistic T_n as follows:

$$T_n := \min_{j=1,2,\ldots,n} \frac{1}{\sqrt{n}\omega} \left(\sum_{i=1}^{j} \|\mathbf{X}_i - \boldsymbol{\mu}_0\|^2 - \frac{j}{n} \sum_{i=1}^{n} \|\mathbf{X}_i - \boldsymbol{\mu}_0\|^2 \right), \tag{3.6.32}$$

where $\hat{\omega}^2 = \omega^2 + o_{\mathbb{P}}(1)$ and $\|\boldsymbol{\mu}_0 - \hat{\boldsymbol{\mu}}_0\| = o_{\mathbb{P}}(1)$. By applying Slutsky's lemma, we can replace $\hat{T}$ with T_n for the purpose of proving the theorem. To aid in our analysis, we delineate several auxiliary variables:

$$M_n(j) := \sum_{i=1}^{j} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_0\|^2 - \frac{j}{n} \sum_{i=1}^{n} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_0\|^2, \tag{3.6.33}$$

complemented by

$$N_n(j) := \sum_{i=1}^{j} \|\mathbf{Z}_i\|^2 - \frac{j}{n} \sum_{i=1}^{n} \|\mathbf{Z}_i\|^2, \tag{3.6.34}$$

and further,

$$C_n(j) := \sum_{i=1}^{j} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)^\top \mathbf{Z}_i - \frac{j}{n} \sum_{i=1}^{n} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)^\top \mathbf{Z}_i. \tag{3.6.35}$$

This allows us to express the difference as:

$$n^{-1/2} \left(\sum_{i=1}^{j} \|\mathbf{X}_i - \boldsymbol{\mu}_0\|^2 - \frac{j}{n} \sum_{i=1}^{n} \|\mathbf{X}_i - \boldsymbol{\mu}_0\|^2 \right) = n^{-1/2} M_n(j) + 2n^{-1/2} C_n(j) + n^{-1/2} N_n(j). \tag{3.6.36}$$

We further introduce S_i, the cumulative sum of centered squared norms of noise vectors, defined as:

$$S_i := \sum_{j=1}^{i} \left(\|\mathbf{Z}_j\|^2 - \mathbb{E}(\|\mathbf{Z}_j\|^2) \right). \tag{3.6.37}$$

Next, we define the delta measure $\tilde{\delta}_q(i)$, which quantifies the maximum expected difference in $\mathbf{Z}_i^2$ and its counterpart $\mathbf{Z}_i'^2$, across all dimensions $j = 1, \ldots, p$:

$$\tilde{\delta}_q(i) := \max_{j=1,\ldots,p} \left(\mathbb{E} \left| \mathbf{Z}_{i,j}^2 - \mathbf{Z}_{i,j}'^2 \right|^q \right)^{1/q}, \quad \text{where } \mathbf{Z}_i' = \mathbf{G}\left(\cdots, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_i \right). \tag{3.6.38}$$

Finally, we define the cumulative dependence measure $\tilde{\Theta}_{t,q}$ as the sum of $\tilde{\delta}_q(i)$ for all $i \geq t$:

$$\tilde{\Theta}_{t,q} = \sum_{i \geq t} \tilde{\delta}_q(i), \quad t \geq 0. \tag{3.6.39}$$

The subsequent analysis utilizes these definitions to rigorously prove the theorem. Before proceeding to prove our main theorem, it's crucial to introduce the following lemma:

Lemma 3.6.1. *For a sequence $\{\mathbf{Z}_i\}$, where each $\mathbf{Z}_i$ is defined as in* (3.3.20) *and assuming each dimension of the sequence $\{\mathbf{Z}_i\}$ has a finite qth moment H_q, then it can be shown that:*

$$\tilde{\Theta}_{t,q} \lesssim \Theta_{t,2q}. \tag{3.6.40}$$

The proof of Lemma 3.6.1 is detailed in Section 3.6.4. This lemma implies that, under the condition (3.3.27), we can establish that:

$$\tilde{\Theta}_{t,2} < \infty. \tag{3.6.41}$$

Consequently, applying Theorem 3 from Wu [2005] under the stipulated condition (3.6.41), we find:

$$n^{-1/2}\{S_{\lfloor nu \rfloor}, 0 \leq u \leq 1\} \Rightarrow \{\sigma_\infty \mathbb{B}(u), 0 \leq u \leq 1\},$$

leading to the conclusion that $n^{-1/2}\{S_{\lfloor nu \rfloor} - n^{-1}\lfloor nu \rfloor S_n, 0 \leq u \leq 1\} \Rightarrow \{\sigma_\infty \mathbb{B}_1(u), 0 \leq u \leq 1\}$. This follows from the continuous mapping theorem and (**??**), as under the null hypothesis H_0, both $M_n(j)$ and $C_n(j)$ are zero.

For part (ii), under the alternative hypothesis H_1, considering $j, \tau \in \{1, \ldots, n\}$, the expression detailed in (3.6.33) reveals that since $(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0) = \mathbf{0}$ for $i = 1, \ldots, \tau - 1$, the first component of $M_n(j)$ is non-negative and reduces to zero for $i < \tau$. Simultaneously, the second component decreases as i increments, suggesting that $\arg\min_j M_n(j) \geq \tau - 1$. Hence, it follows that $\min_{j=1,\ldots,n} M_n(j) \leq M_n(\tau - 1) \leq d \cdot (\frac{1-\tau}{n})(n - \tau + 1)$. As a result, we observe that

$$n^{-1/2} \min_{j=1,\ldots,n} M_n(j) \to -\infty, \text{ as } n^{-3/2}\tau(n - \tau) \cdot d \to \infty.$$

Accordingly, (ii) is concluded based on the premise outlined in (3.6.36).

□

3.6.2 Proof of Theorem 3.3.2 (i)–(iii)

Proof of Theorem 3.3.2 (i)

Fixing d, our objective is to establish that $|\tilde{\tau}_n - \tau_n| = \mathcal{O}_{\mathbb{P}}(1)$ as $n \to \infty$. Specifically, we aim to demonstrate that for any $\varepsilon > 0$, there exist $M_\varepsilon \in \mathbb{N}$ and $N_\varepsilon \in \mathbb{N}$ such that

$$\mathbb{P}\left(|\tilde{\tau}_n - \tau_n| \geq M_\varepsilon\right) < \varepsilon, \quad \forall n > N_\varepsilon.$$

From (3.2.19), we derive that

$$\begin{aligned}\Omega &= \left\{ \sum_{t=1}^{\tau_n-1} \left(\|\mathbf{X}_t - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{d}/2 \right) \geq \sum_{t=1}^{\tilde{\tau}_n-1} \left(\|\mathbf{X}_t - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{d}/2 \right) \right\} \\ &= \left(\left\{ \sum_{t=\tilde{\tau}_n}^{\tau_n-1} \left(\|\mathbf{X}_t - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{d}/2 \right) \geq 0 \right\} \cap \{\tilde{\tau}_n < \tau_n\} \right) \\ &\cup \left(\left\{ \sum_{j=\tau_n}^{\tilde{\tau}_n-1} \left(\|\mathbf{X}_t - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{d}/2 \right) \leq 0 \right\} \cap \{\tilde{\tau}_n > \tau_n\} \right) \cup \{\tilde{\tau}_n = \tau_n\}\end{aligned}$$

For $n > N_\varepsilon \gg M_\varepsilon > 0$, with $\hat{c} = \hat{d}/2$ and $c = d/2$, we obtain the following bound:

$$\begin{aligned}&\mathbb{P}\left(|\tilde{\tau}_n - \tau_n| \geq M_\varepsilon\right) \\ &= \mathbb{P}\left(\bigcup_{\ell=M_\varepsilon}^{\tau_n-1} \{\tilde{\tau}_n = \tau_n - \ell\} \cap \Omega \right) + \mathbb{P}\left(\bigcup_{\ell=M_\varepsilon}^{n-\tau_n} \{\tilde{\tau}_n = \tau_n + \ell\} \cap \Omega \right) \\ &= \mathbb{P}\left(\bigcup_{\ell=M_\varepsilon}^{\tau_n-1} \left\{ \sum_{j=\tau_n-\ell}^{\tau_n-1} \left(\|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{c} \right) \geq 0 \right\} \right) \\ &\quad + \mathbb{P}\left(\bigcup_{\ell=M_\varepsilon}^{n-\tau_n} \left\{ \sum_{j=\tau_n}^{\tau_n+\ell} \left(\|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{c} \right) \leq 0 \right\} \right) \\ &= \mathbb{P}\left(\max_{\ell \geq M_\varepsilon} \sum_{j=\tau_n-\ell}^{\tau_n-1} \left(\|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{c} \right) \geq 0 \right) \\ &\quad + \mathbb{P}\left(\max_{\ell \geq M_\varepsilon} \sum_{j=\tau_n}^{\tau_n+\ell} \left(\hat{\gamma} + \hat{c} - \|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2 \right) \geq 0 \right)\end{aligned}$$

Applying the Bonferroni inequality and Slutsky's lemma, we obtain

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{\ell \geq M\varepsilon} \sum_{j=\tau_n-\ell}^{\tau_n-1} \big(\|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{c} \big) \geq 0 \Big) \\
&\quad + \mathbb{P}\Big(\max_{\ell \geq M\varepsilon} \sum_{j=\tau_n}^{\tau_n+\ell} \big(\hat{\gamma} + \hat{c} - \|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2 \big) \geq 0 \Big) \\
&\lesssim \sum_{k=1}^{\infty} \mathbb{P}\Big(\max_{2^{k-1}M\varepsilon \leq \ell \leq 2^k M_\varepsilon} \sum_{j=\tau_n-\ell}^{\tau_n-1} \big(\sum_{u=1}^{p} \mathbf{Z}_{j,u}^2 - \gamma - c \big) \geq 0 \Big) \\
&\quad + \sum_{k=1}^{\infty} \mathbb{P}\Bigg(\max_{2^{k-1}M\varepsilon \leq \ell \leq 2^k M_\varepsilon} \sum_{j=\tau_n}^{\tau_n+\ell} \big(\gamma + c - \|\boldsymbol{\mu}_j - \boldsymbol{\mu}_0\|^2 \\
&\qquad - 2\sum_{u=1}^{p} (\boldsymbol{\mu}_{j,u} - \boldsymbol{\mu}_{0,u}) \mathbf{Z}_{j,u} - \sum_{u=1}^{p} \mathbf{Z}_{j,u}^2 \big) \geq 0 \Bigg) \\
&\leq \sum_{k=1}^{\infty} \mathbb{P}\Big(\max_{1 \leq \ell \leq 2^k M\varepsilon} \sum_{j=\tau_n-\ell}^{\tau_n-1} \big(\sum_{u=1}^{p} \mathbf{Z}_{j,u}^2 - \gamma \big) \geq 2^{k-1} M_\varepsilon c \Big) \\
&\quad + \sum_{k=1}^{\infty} \mathbb{P}\Bigg(\max_{1 \leq \ell \leq 2^k M\varepsilon} \sum_{j=\tau_n}^{\tau_n+\ell} \big(-2\sum_{u=1}^{p} (\boldsymbol{\mu}_{j,u} - \boldsymbol{\mu}_{0,u}) \mathbf{Z}_{j,u} \\
&\qquad + \gamma - \sum_{u=1}^{p} \mathbf{Z}_{j,u}^2 \big) \geq 2^{k-1} M_\varepsilon (d - c) \Bigg) \\
&= (I) + (II)
\end{aligned}
$$

Now we apply the Bonferroni inequality to bound (I) and (II).

$$
(I) \leq \sum_{k=1}^{\infty} \sum_{u=1}^{p} \mathbb{P}\left(\max_{1 \leq \ell \leq 2^k M_\varepsilon} \sum_{j=\tau_n-\ell}^{\tau_n-1} (\mathbf{Z}_{j,u}^2 - \mathbb{E}\mathbf{Z}_{j,u}^2) \geq 2^{k-1} M_\varepsilon c/p \right)
$$

and

$$(II) \leq 2\sum_{k=1}^{\infty}\sum_{u=1}^{p}\mathbb{P}\left(\max_{1\leq\ell\leq 2^k M_\varepsilon}\sum_{j=\tau_n}^{\tau_n+\ell}\left(-(\boldsymbol{\mu}_{j,u}-\boldsymbol{\mu}_{0,u})\mathbf{Z}_{j,u}\right)\geq 2^{k-1}M_\varepsilon(d-c)/(3p)\right)$$
$$+\sum_{k=1}^{\infty}\sum_{u=1}^{p}\mathbb{P}\left(\max_{1\leq\ell\leq 2^k M_\varepsilon}\sum_{j=\tau_n}^{\tau_n+\ell}\left(-(\mathbf{Z}_{j,u}^2-\mathbb{E}\mathbf{Z}_{j,u}^2)\right)\geq 2^{k-1}M_\varepsilon(d-c)/(3p)\right).$$

To further refine bounds for (I) and (II), we introduce the following lemma:

Lemma 3.6.2. *Assuming Condition 3.3.1 holds, then*

$$\tilde{\Xi}_{q,\alpha} := \sup_{t\geq 0}(t+1)^{\alpha}\sum_{i\geq t}\max_{j=1,\ldots,p}\left(\mathbb{E}\left|\mathbf{Z}_{i,j}^2-\mathbf{Z}_{i,j}'^2\right|^q\right)^{1/q}<\infty, \tag{3.6.42}$$

is also satisfied with $\alpha = 1$.

The proof of Lemma 3.6.2 is deferred to Section 3.6.4.

Given Lemma 3.6.2, consequently, to further bound (I) and (II), we apply Theorem 2 from Wu and Wu [2016].

This entails the existence of positive constants C_1, C_2, and C_3, such that

$$\mathbb{P}\left(\max_{1\leq\ell\leq 2^k M_\varepsilon}\sum_{j=\tau_n-\ell}^{\tau_n-1}(\mathbf{Z}_{j,u}^2-\mathbb{E}\mathbf{Z}_{j,u}^2)\geq 2^{k-1}M_\varepsilon c/p\right)\leq\frac{C_1 2^k M_\varepsilon\Xi_{2q,1}^{2q}}{(2^{k-1}M_\varepsilon c/p)^{2q}}$$
$$+C_2\exp\left(-\frac{C_3(2^{k-1}M_\varepsilon c/p)^2}{2^k M_\varepsilon\Xi_{2,1}^2}\right), \tag{3.6.43}$$

yielding that

$$(I)\leq\frac{C_1 p^{2q+1}\Xi_{2q,1}^{2q}}{M_\varepsilon^{2q-1}c^{2q}}\sum_{k=1}^{\infty}\frac{1}{2^{k(2q-1)-2q}}+C_2 p\sum_{k=1}^{\infty}\exp\left(-\frac{C_3 2^{k-2}M_\varepsilon(c/p)^2}{\Xi_{2,1}^2}\right). \tag{3.6.44}$$

As M_ε grows, (I) approach zero, confirming the asymptotic bounds under the prescribed

conditions.

To bound (II), define

$$w_{u,k} = \sqrt{\frac{2^k M_\varepsilon + 1}{\sum_{j=\tau_n}^{\tau_n + 2^k M_\varepsilon - 1} (\boldsymbol{\mu}_{j,u} - \boldsymbol{\mu}_{0,u})^2}}, \tag{3.6.45}$$

and construct

$$\mathbf{a}_{u,k} = (a_{u,k,\tau_n}, \cdots, a_{u,k,\tau_n + 2^k M_\varepsilon}), \tag{3.6.46}$$

where $a_{u,k,j} := -(\boldsymbol{\mu}_{j,u} - \boldsymbol{\mu}_{0,u}) \cdot w_{u,k}$.

There exist positive constants C_4, C_5, and C_6 such that

$$\begin{aligned}
&\mathbb{P}\left(\max_{1 \le \ell \le 2^k M_\varepsilon} \sum_{j=\tau_n}^{\tau_n + \ell} \left(-(\boldsymbol{\mu}_{j,u} - \boldsymbol{\mu}_{0,u}) \mathbf{Z}_{j,u} \right) \ge 2^{k-1} M_\varepsilon (d-c)/(3p) \right) \\
&= \mathbb{P}\left(\max_{1 \le \ell \le 2^k M_\varepsilon} \sum_{j=\tau_n}^{\tau_n + \ell} a_{u,k,j} \mathbf{Z}_{j,u} \ge 2^{k-1} M_\varepsilon w_{u,k} (d-c)/(3p) \right) \\
&\le C_4 \frac{\|\mathbf{a}_{u,k}\|_q^q \Xi_{q,1}^q}{\left(2^{k-1} M_\varepsilon w_{u,k} (d-c)/(3p)\right)^q} + C_5 \exp\left(- \frac{C_6 \left(2^{k-1} M_\varepsilon w_{u,k} (d-c)/(3p)\right)^2}{2^k M_\varepsilon \Xi_{2,1}^2} \right)
\end{aligned} \tag{3.6.47}$$

Additionally, positive constants C_7, C_8, and C_9 exist such that

$$\begin{aligned}
&\mathbb{P}\left(\max_{1 \le \ell \le 2^k M_\varepsilon} \sum_{j=\tau_n}^{\tau_n + \ell} \left(-(\mathbf{Z}_{j,u}^2 - \mathbb{E}\mathbf{Z}_{j,u}^2) \right) \ge 2^{k-1} M_\varepsilon (d-c)/(3p) \right) \\
&\le \frac{C_7 2^k M_\varepsilon \Xi_{2q,1}^{2q}}{(2^{k-1} M_\varepsilon (d-c)/(3p))^{2q}} + C_8 \exp\left(- \frac{C_9 (2^{k-1} M_\varepsilon (d-c)/(3p))^2}{2^k M_\varepsilon \Xi_{2,1}^2} \right),
\end{aligned} \tag{3.6.48}$$

It follows that

$$
\begin{aligned}
(II) &\lesssim \frac{C_4 \Xi_{q,1}^q p^{q+1}}{M_\varepsilon^{q-1}\left((d-c)/3\right)^q} \sum_{k=1}^{\infty} \frac{1}{2^{k(q-1)-q}} + C_5 \sum_{k=1}^{\infty} \exp\left(-\frac{C_6 2^{k-2} M_\varepsilon w_{u,k}^2 \left((d-c)/(3p)\right)^2}{\Xi_{2,1}^2}\right) \\
&\quad + \frac{C_7 \Xi_{2q,1}^{2q} p^{2q+1}}{M_\varepsilon^{2q-1}((d-c)/3)^{2q}} \sum_{k=1}^{\infty} \frac{1}{2^{k(2q-1)-2q}} + C_8 p \sum_{k=1}^{\infty} \exp\left(-\frac{C_9 2^{k-2} M_\varepsilon ((d-c)/(3p))^2}{\Xi_{2,1}^2}\right) \\
&\quad \to 0, \qquad \text{as } M_\varepsilon \to \infty.
\end{aligned}
\tag{3.6.49}
$$

This concludes the proof of Theorem 3.3.2 (i). □

Proof of Theorem 3.3.2 (ii)

Applying Slutsky's lemma to bound the probability

$$
\begin{aligned}
&\mathbb{P}\left(|\tilde{\tau}_n - \tau_n| \geq 1\right) \\
&= \mathbb{P}\left(\bigcup_{\ell=1}^{\tau_n - 1} \{\tilde{\tau}_n = \tau_n - \ell\} \cap \Omega\right) + \mathbb{P}\left(\bigcup_{\ell=1}^{n-\tau_n} \{\tilde{\tau}_n = \tau_n + \ell\} \cap \Omega\right) \\
&\lesssim \mathbb{P}\left(\max_{\ell \geq 1} \sum_{j=\tau_n - \ell}^{\tau_n - 1} \left(\sum_{u=1}^{p} Z_{j,u}^2 - \gamma - c\right) \geq 0\right) \\
&+ \mathbb{P}\left(\max_{1 \leq \ell \leq n - \tau_n} \sum_{j=\tau_n}^{\tau_n + \ell} \left(\gamma + c - \|\mu_j - \mu_0\|^2 - 2\sum_{u=1}^{p} (\mu_{j,u} - \mu_{0,u}) Z_{j,u} - \sum_{u=1}^{p} Z_{j,u}^2\right) \geq 0\right),
\end{aligned}
\tag{3.6.50}
$$

we can further bound it using the Bonferroni inequality:

$$
\begin{aligned}
&\mathbb{P}\left(|\tilde{\tau}_n - \tau_n| \geq 1\right) \\
&= \mathbb{P}\left(\bigcup_{\ell=1}^{\tau_n - 1} \{\tilde{\tau}_n = \tau_n - \ell\} \cap \Omega\right) + \mathbb{P}\left(\bigcup_{\ell=1}^{n-\tau_n} \{\tilde{\tau}_n = \tau_n + \ell\} \cap \Omega\right) \\
&\leq \sum_{u=1}^{p} \mathbb{P}\left(\max_{\ell \geq 1} \sum_{j=\tau_n - \ell}^{\tau_n - 1} \left(Z_{j,u}^2 - \mathbb{E}Z_{j,u}^2\right) \geq c/p\right) \\
&+ 2\sum_{u=1}^{p} \mathbb{P}\left(\max_{1 \leq \ell \leq n-\tau_n} \sum_{j=\tau_n}^{\tau_n + \ell} \left(-(\mu_{j,u} - \mu_{0,u}) Z_{j,u}\right) \geq (d-c)/(3p)\right) \\
&+ \sum_{u=1}^{p} \mathbb{P}\left(\max_{1 \leq \ell \leq n-\tau_n} \sum_{j=\tau_n}^{\tau_n + \ell} \left(-(Z_{j,u}^2 - \mathbb{E}Z_{j,u}^2)\right) \geq (d-c)/(3p)\right).
\end{aligned}
\tag{3.6.51}
$$

We begin by noting the equivalence $c \asymp d$. Leveraging the established Condition 3.3.1, as delineated in prior discussions, enables the application of Theorem 2 from Wu and Wu [2016]. Consequently, there exist positive constants C_1, C_2, and C_3, ensuring that

$$
\mathbb{P}\left(\max_{\ell \geq 1} \sum_{j=\tau_n - \ell}^{\tau_n - 1} \left(\mathbf{Z}_{j,u}^2 - \mathbb{E}\mathbf{Z}_{j,u}^2\right) \geq \frac{c}{p}\right) \leq \frac{C_1 \tau_n \Xi_{2q,1}^{2q}}{\left(\frac{c}{p}\right)^{2q}} + C_2 \exp\left(-\frac{C_3 \left(\frac{c}{p}\right)^2}{\tau_n \Xi_{2,1}^2}\right) \to 0, \tag{3.6.52}
$$

as $d \to \infty$.

Define the weight for adjustment as

$$
w_{u,k} = \sqrt{\frac{n - \tau_n + 1}{\sum_{j=\tau_n}^{n} (\boldsymbol{\mu}_{j,u} - \boldsymbol{\mu}_{0,u})^2}}, \tag{3.6.53}
$$

and introduce the vector

$$
\mathbf{a}_{u,k} = (a_{u,k,\tau_n}, \ldots, a_{u,k,n}), \tag{3.6.54}
$$

where each component is given by $a_{u,k,j} := -(\boldsymbol{\mu}_{j,u} - \boldsymbol{\mu}_{0,u}) \cdot w_{u,k}$.

Meanwhile, there exist positive constants C_4, C_5, and C_6, such that

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le\ell\le n-\tau_n}\sum_{j=\tau_n}^{\tau_n+\ell}\left(-(\boldsymbol{\mu}_{j,u}-\boldsymbol{\mu}_{0,u})\mathbf{Z}_{j,u}\right)\ge\frac{d-c}{3p}\right)\\
&=\mathbb{P}\left(\max_{1\le\ell\le n-\tau_n}\sum_{j=\tau_n}^{\tau_n+\ell}a_{u,k,j}\mathbf{Z}_{j,u}\ge\frac{w_{u,k}(d-c)}{3p}\right)\\
&\le C_4\frac{\|\mathbf{a}_{u,k}\|_q^q\Xi_{q,1}^q}{\left(\frac{w_{u,k}(d-c)}{3p}\right)^q}+C_5\exp\left(-\frac{C_6\left(\frac{w_{u,k}(d-c)}{3p}\right)^2}{(n-\tau_n)\Xi_{2,1}^2}\right)\to 0,
\end{aligned}\tag{3.6.55}$$

as $d\to\infty$.

Moreover, additional positive constants C_7, C_8, and C_9 exist such that

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le\ell\le n-\tau_n}\sum_{j=\tau_n}^{\tau_n+\ell}\left(-(\mathbf{Z}_{j,u}^2-\mathbb{E}\mathbf{Z}_{j,u}^2)\right)\ge\frac{d-c}{3p}\right)\\
&\le\frac{C_1(n-\tau_n)\Xi_{2q,1}^{2q}}{\left(\frac{d-c}{3p}\right)^{2q}}+C_2\exp\left(-\frac{C_3\left(\frac{d-c}{3p}\right)^2}{(n-\tau_n)\Xi_{2,1}^2}\right)\to 0,
\end{aligned}\tag{3.6.56}$$

as $d\to\infty$.

This concludes the proof of Theorem 3.3.2 (ii). □

Proof of Theorem 3.3.2 (iii)

Let $L_n=d_n^{-2}$. To establish that for any $\varepsilon>0$, there exist $M_\varepsilon,N_\varepsilon\in\mathbb{N}$ such that

$$\mathbb{P}\left(|\tilde{\tau}_n-\tau_n|\ge M_\varepsilon L_n\right)<\varepsilon,\quad\forall n>N_\varepsilon,$$

we consider $M_\varepsilon > 0$ and $n > N_\varepsilon$. The probability of interest is decomposed as follows:

$$\begin{aligned}
&\mathbb{P}\left(|\tilde{\tau}_n - \tau_n| \geq M_\varepsilon L_n\right) = \\
&\quad \mathbb{P}\left(\bigcup_{\ell=M_\varepsilon L_n}^{\tau_n - 1} \{\tilde{\tau}_n = \tau_n - \ell\} \cap \Omega\right) + \mathbb{P}\left(\bigcup_{\ell=M_\varepsilon L_n}^{n-\tau_n} \{\tilde{\tau}_n = \tau_n + \ell\} \cap \Omega\right) \\
&\quad = \mathbb{P}\left(\max_{\ell \geq M_\varepsilon L_n} \sum_{j=\tau_n-\ell}^{\tau_n - 1} \left(\|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{c}_n\right) \geq 0\right) \\
&\qquad + \mathbb{P}\left(\max_{\ell \geq M_\varepsilon L_n} \sum_{j=\tau_n}^{\tau_n + \ell} \left(\hat{\gamma} + \hat{c}_n - \|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2\right) \geq 0\right).
\end{aligned}$$

Applying Bonferroni inequality and Slutsky's lemma, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{\ell \geq M\varepsilon L_n} \sum_{j=\tau_n-\ell}^{\tau_n - 1} \left(\|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{c}_n\right) \geq 0\right) \\
&\quad + \mathbb{P}\left(\max_{\ell \geq M\varepsilon L_n} \sum_{j=\tau_n}^{\tau_n + \ell} \left(\hat{\gamma} + \hat{c}_n - \|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2\right) \geq 0\right) \\
&\leq \sum_{k=1}^{\infty} \mathbb{P}\left(\max_{2^{k-1} M\varepsilon L_n \leq \ell \leq 2^k M_\varepsilon L_n} \sum_{j=\tau_n-\ell}^{\tau_n - 1} \left(\|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2 - \hat{\gamma} - \hat{c}_n\right) \geq 0\right) \\
&\quad + \sum_{k=1}^{\infty} \mathbb{P}\left(\max_{2^{k-1} M\varepsilon L_n \leq \ell \leq 2^k M_\varepsilon L_n} \sum_{j=\tau_n}^{\tau_n + \ell} \left(\hat{\gamma} + \hat{c}_n - \|\mathbf{X}_j - \hat{\boldsymbol{\mu}}_0\|^2\right) \geq 0\right) \\
&\lesssim \sum_{k=1}^{\infty} \mathbb{P}\left(\max_{2^{k-1} M\varepsilon L_n \leq \ell \leq 2^k M_\varepsilon L_n} \sum_{j=\tau_n-\ell}^{\tau_n - 1} \left(\sum_{u=1}^{p} \mathbf{Z}_{j,u}^2 - \gamma - c_n\right) \geq 0\right) \\
&\quad + \sum_{k=1}^{\infty} \mathbb{P}\Bigg(\max_{2^{k-1} M\varepsilon L_n \leq \ell \leq 2^k M_\varepsilon L_n} \sum_{j=\tau_n}^{\tau_n + \ell} \Big(\gamma + c_n - \|\boldsymbol{\mu}_j - \boldsymbol{\mu}_0\|^2 \\
&\qquad - 2\sum_{u=1}^{p} (\boldsymbol{\mu}_{j,u} - \boldsymbol{\mu}_{0,u}) \mathbf{Z}_{j,u} - \sum_{u=1}^{p} \mathbf{Z}_{j,u}^2\Big) \geq 0\Bigg)
\end{aligned}$$

which can be further bounded as follows, using Bonferroni inequality:

$$
\begin{aligned}
&\sum_{k=1}^{\infty}\mathbb{P}\left(\max_{2^{k-1}M_\varepsilon L_n\le\ell\le 2^k M_\varepsilon L_n}\sum_{j=\tau_n-\ell}^{\tau_n-1}\left(\sum_{u=1}^{p}\mathbf{Z}_{j,u}^2-\gamma-c_n\right)\ge 0\right)\\
&\quad+\sum_{k=1}^{\infty}\mathbb{P}\Bigg(\max_{2^{k-1}M_\varepsilon L_n\le\ell\le 2^k M_\varepsilon L_n}\sum_{j=\tau_n}^{\tau_n+\ell}\Big(\gamma+c_n-\|\boldsymbol{\mu}_j-\boldsymbol{\mu}_0\|^2\\
&\qquad-2\sum_{u=1}^{p}(\boldsymbol{\mu}_{j,u}-\boldsymbol{\mu}_{0,u})\mathbf{Z}_{j,u}-\sum_{u=1}^{p}\mathbf{Z}_{j,u}^2\Big)\ge 0\Bigg)\\
&\le\sum_{k=1}^{\infty}\sum_{u=1}^{p}\mathbb{P}\left(\max_{1\le\ell\le 2^k M_\varepsilon L_n}\sum_{j=\tau_n-\ell}^{\tau_n-1}(\mathbf{Z}_{j,u}^2-\mathbb{E}\mathbf{Z}_{j,u}^2)\ge 2^{k-1}M_\varepsilon L_n c_n/p\right)\\
&\quad+2\sum_{k=1}^{\infty}\sum_{u=1}^{p}\mathbb{P}\Bigg(\max_{1\le\ell\le 2^k M_\varepsilon L_n}\sum_{j=\tau_n}^{\tau_n+\ell}\Big(-(\boldsymbol{\mu}_{j,u}-\boldsymbol{\mu}_{0,u})\mathbf{Z}_{j,u}\Big)\\
&\qquad\ge 2^{k-1}M_\varepsilon L_n(d_n-c_n)/(3p)\Bigg)\\
&\quad+\sum_{k=1}^{\infty}\sum_{u=1}^{p}\mathbb{P}\Bigg(\max_{1\le\ell\le 2^k M_\varepsilon L_n}\sum_{j=\tau_n}^{\tau_n+\ell}\Big(-(\mathbf{Z}_{j,u}^2-\mathbb{E}\mathbf{Z}_{j,u}^2)\Big)\\
&\qquad\ge 2^{k-1}M_\varepsilon L_n(d_n-c_n)/(3p)\Bigg)
\end{aligned}
$$

Given the equivalence $c_n \asymp d_n$ and leveraging Lemma 3.6.2, we apply Theorem 2 from Wu and Wu [2016]. Consequently, there exist positive constants C_1, C_2, and C_3 ensuring the following bound:

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\le\ell\le 2^k M_\varepsilon L_n}\sum_{j=\tau_n-\ell}^{\tau_n-1}(\mathbf{Z}_{j,u}^2-\mathbb{E}\mathbf{Z}_{j,u}^2)\ge 2^{k-1}M_\varepsilon L_n c_n/p\right)\\
&\le\frac{C_1 2^k M_\varepsilon L_n\Xi_{2q,1}^{2q}}{(2^{k-1}M_\varepsilon L_n c_n/p)^{2q}}+C_2\exp\left(-\frac{C_3(2^{k-1}M_\varepsilon L_n c_n/p)^2}{2^k M_\varepsilon L_n\Xi_{2,1}^2}\right)\\
&=\frac{C_1\Xi_{2q,1}^{2q}p^{2q}}{2^{k(2q-1)-2q}M_\varepsilon^{2q-1}d_n^{2-2q}}+C_2\exp\left(-\frac{C_3 2^{k-2}M_\varepsilon}{\Xi_{2,1}^2p^2d_n}\right),
\end{aligned}
\tag{3.6.57}
$$

which implies

$$\sum_{k=1}^{\infty}\sum_{u=1}^{p}\mathbb{P}\left(\max_{1\leq\ell\leq 2^k M_\varepsilon L_n}\sum_{j=\tau_n-\ell}^{\tau_n-1}(\mathbf{Z}_{j,u}^2-\mathbb{E}\mathbf{Z}_{j,u}^2)\geq 2^{k-1}M_\varepsilon L_n c_n/p\right)\to 0, \tag{3.6.58}$$

as $M_\varepsilon\to\infty$.

Defining $w_{u,k}$ and $\mathbf{a}_{u,k}$ as

$$w_{u,k}=\sqrt{\frac{2^k M_\varepsilon L_n+1}{\sum_{j=\tau_n}^{\tau_n+2^k M_\varepsilon L_n-1}(\boldsymbol{\mu}_{j,u}-\boldsymbol{\mu}_{0,u})^2}}, \tag{3.6.59}$$

$$\mathbf{a}_{u,k}=(a_{u,k,\tau_n},\ldots,a_{u,k,\tau_n+2^k M_\varepsilon L_n}), \tag{3.6.60}$$

with $a_{u,k,j}:=-(\boldsymbol{\mu}_{j,u}-\boldsymbol{\mu}_{0,u})\cdot w_{u,k}$, leads to the existence of constants C_4, C_5, C_6 that satisfy

$$\begin{aligned}&\mathbb{P}\left(\max_{1\leq\ell\leq 2^k M_\varepsilon L_n}\sum_{j=\tau_n}^{\tau_n+\ell}\left(-(\boldsymbol{\mu}_{j,u}-\boldsymbol{\mu}_{0,u})\mathbf{Z}_{j,u}\right)\geq 2^{k-1}M_\varepsilon L_n(d_n-c_n)/(3p)\right)\\&\lesssim C_4\frac{\Xi_{q,1}^q p^q}{2^{qk-q-k}M_\varepsilon^{q-1}d_n^{2-q}}+C_5\exp\left(-\frac{C_6 2^{k-2}M_\varepsilon w_{u,k}/9}{\Xi_{2,1}^2 d_n p^2}\right)\end{aligned} \tag{3.6.61}$$

demonstrating that

$$\sum_{k=1}^{\infty}\sum_{u=1}^{p}\mathbb{P}\left(\max_{1\leq\ell\leq 2^k M_\varepsilon L_n}\sum_{j=\tau_n}^{\tau_n+\ell}\left(-(\boldsymbol{\mu}_{j,u}-\boldsymbol{\mu}_{0,u})\mathbf{Z}_{j,u}\right)\geq 2^{k-1}M_\varepsilon L_n(d_n-c_n)/(3p)\right)\to 0, \tag{3.6.62}$$

as $M_\varepsilon\to\infty$.

Furthermore, positive constants C_7, C_8, C_9 exist such that

$$\begin{aligned}&\mathbb{P}\left(\max_{1\le\ell\le 2^k M_\varepsilon L_n}\sum_{j=\tau_n}^{\tau_n+\ell}\left(-(\mathbf{Z}_{j,u}^2-\mathbb{E}\mathbf{Z}_{j,u}^2)\right)\ge 2^{k-1}M_\varepsilon L_n(d_n-c_n)/(3p)\right)\\&\lesssim \frac{C_7\Xi_{2q,1}^{2q}p^{2q}}{2^{2qk-2q-k}M_\varepsilon^{2q-1}d_n^{2-2q}}+C_8\exp\left(-\frac{C_9 2^{k-2}M_\varepsilon/9}{\Xi_{2,1}^2 d_n p^2}\right),\end{aligned}\tag{3.6.63}$$

leading to

$$\sum_{k=1}^{\infty}\sum_{u=1}^{p}\mathbb{P}\left(\max_{1\le\ell\le 2^k M_\varepsilon L_n}\sum_{j=\tau_n}^{\tau_n+\ell}\left(-(\mathbf{Z}_{j,u}^2-\mathbb{E}\mathbf{Z}_{j,u}^2)\right)\ge 2^{k-1}M_\varepsilon L_n(d_n-c_n)/(3p)\right)\to 0,\tag{3.6.64}$$

as $M_\varepsilon\to\infty$.

This concludes the proof of Theorem 3.3.2 (iii). □

3.6.3 Proof of Theorem 3.3.3

We initiate the proof by delineating some preliminary notation and pivotal technical lemmas essential for our analysis. A key lemma pertains to the consistency and convergence rate of the estimator $\tilde{\gamma}$, which plays a crucial role in formulating the "test statistic" $\hat{D}_j$.

Lemma 3.6.3. *Under Condition 3.3.1, and provided that* $\log(t_n)n^{1/q}k^{-1}\to 0$ *as* $n\to\infty$, *it holds that*

$$|\tilde{\gamma}_n-\gamma_n|=\mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).\tag{3.6.65}$$

Define

$$U_{j,k}:=\frac{1}{k}\sum_{i=(j-1)k+1}^{jk}\|\mathbf{X}_i-\boldsymbol{\mu}_0\|^2.\tag{3.6.66}$$

For the purpose of approximating $\hat{D}_j$, we introduce

$$\tilde{D}_j:=\sqrt{k}(\tilde{U}_{j,k}-\gamma_n)/\sigma_\infty,\quad \tilde{U}_{j,k}:=\mathbb{E}(U_{j,k}|\varepsilon_{(j-1)k+1},\ldots,\varepsilon_{jk}),\quad j=1,2,\ldots.\tag{3.6.67}$$

It is crucial to note that $\tilde{U}_{1,k}, \ldots, \tilde{U}_{\lfloor n/k \rfloor,k}$ are independent due to the independence of the ε_t, making the $\tilde{D}_j$ likewise independent. The forthcoming lemma underscores the fidelity of approximating $U_{j,k}$ with $\tilde{U}_{j,k}$ under the framework of Theorem 3.3.3.

Lemma 3.6.4. *Let $s = \lfloor n/k \rfloor$ and assume Condition 3.3.1 is met. Then,*

$$\mathbb{P}\left(\max_{j \leq s} \left| U_{j,k} - \tilde{U}_{j,k} \right| > u\right) \to 0, \quad \text{as } s \to \infty. \tag{3.6.68}$$

In addition to $\tilde{D}_j$, we define $\tilde{I}_j$ to approximate the test decisions I_j. Since $\tilde{I}_j$ are independent, we leverage the following lemma, which offers valuable insights on its own.

Lemma 3.6.5. *Consider a sequence $\tilde{I}_1, \tilde{I}_2, \ldots$ of independent Bernoulli-distributed random variables with $\mathbb{P}(\tilde{I}_i = 1) = p_i = 1 - \mathbb{P}(\tilde{I}_i = 0)$. Then,*

$$\mathbb{P}\left(\max_{j \geq 1} \sum_{i=1}^{j} (2\tilde{I}_i - 1) \geq 0\right) = \sum_{k=1}^{\infty} \sum_{(i_1,\ldots,i_{2k}) \in A_k} \prod_{\ell=1}^{2k} \mathbb{P}(2\tilde{I}_\ell - 1 = i_\ell),$$

where

$$A_k = \{(i_1, \ldots, i_{2k}) \in \{-1, 1\}^{2k} : \sum_{\ell=1}^{2k} i_\ell = 0 \text{ and } \sum_{\ell=1}^{j} i_\ell \leq 0, \forall j = 1, \ldots, 2k\}.$$

Now we proceed with the proof of Theorem 3.3.3. Our goal is to demonstrate that

$$\mathbb{P}\left(|\hat{\eta}_n - \eta_n| \geq 1\right) \to 0, \text{ as } n \to \infty. \tag{3.6.69}$$

Consider the event Ω defined as

$$\begin{aligned}
\Omega &= \left\{ \sum_{j=1}^{\eta_n} I_j^2 + \sum_{j=\eta_n+1}^{\lfloor n/k \rfloor} \left(I_j - 1\right)^2 \geq \sum_{j=1}^{\hat{\eta}_n} I_j^2 + \sum_{j=\hat{\eta}_n+1}^{\lfloor n/k \rfloor} \left(I_j - 1\right)^2 \right\} \\
&= \left(\left\{ \sum_{j=\hat{\eta}_n+1}^{\eta_n} I_j - \sum_{j=\hat{\eta}_n+1}^{\eta_n} \left(1 - I_j\right) \geq 0 \right\} \cap \{\hat{\eta}_n < \eta_n\} \right) \\
&\quad \cup \left(\left\{ \sum_{j=\eta_n+1}^{\hat{\eta}_n} I_j - \sum_{j=\eta_n+1}^{\hat{\eta}_n} \left(1 - I_j\right) \leq 0 \right\} \cap \{\hat{\eta}_n > \eta_n\} \right) \cup \{\hat{\eta}_n = \eta_n\} \\
&= \left(\left\{ \sum_{j=\hat{\eta}_n+1}^{\eta_n} \left(2I_j - 1\right) \geq 0 \right\} \cap \{\hat{\eta}_n < \eta_n\} \right) \\
&\quad \cup \left(\left\{ \sum_{j=\eta_n+1}^{\hat{\eta}_n} \left(2I_j - 1\right) \leq 0 \right\} \cap \{\hat{\eta}_n > \eta_n\} \right) \cup \{\hat{\eta}_n = \eta_n\} .
\end{aligned}$$

Accordingly, the bound is given by

$$\begin{aligned}
&\mathbb{P}\left(|\hat{\eta}_n - \eta_n| \geq 1\right) \\
&= \mathbb{P}\left(\bigcup_{\ell=1}^{\infty} \{\hat{\eta}_n = \eta_n - \ell\} \cap \Omega \right) + \mathbb{P}\left(\bigcup_{\ell=1}^{\infty} \{\hat{\eta}_n = \eta_n + \ell\} \cap \Omega \right) \\
&= \mathbb{P}\left(\bigcup_{\ell=1}^{\eta_n} \left\{ \sum_{j=\eta_n-\ell+1}^{\eta_n} \left(2I_j - 1\right) \geq 0 \right\} \right) + \mathbb{P}\left(\bigcup_{\ell=1}^{\lfloor n/k \rfloor - \eta_n} \left\{ \sum_{j=\eta_n+1}^{\eta_n+\ell} \left(2I_j - 1\right) \leq 0 \right\} \right) \\
&= \mathbb{P}\left(\max_{\ell=1,\ldots,\eta_n} \sum_{j=\eta_n-\ell+1}^{\eta_n} \left(2I_j - 1\right) \geq 0 \right) + \mathbb{P}\left(\max_{\ell=1,\ldots,\lfloor n/k \rfloor - \eta_n} \sum_{j=\eta_n+1}^{\eta_n+\ell} \left(1 - 2I_j\right) \geq 0 \right) .
\end{aligned}$$

Fix $c_1, c_2 > 0$ and recall the notation from (3.6.67). We analyze the probability as follows:

$$\begin{aligned}
&\mathbb{P}\left(\max_{\ell=1,\dots,\eta_n} \sum_{j=\eta_n-\ell+1}^{\eta_n} (2I_j - 1) \geq 0\right) \\
&\leq \mathbb{P}\left(\max_{\ell=1,\dots,\eta_n} \sum_{j=\eta_n-\ell+1}^{\eta_n} (2I_j - 1) \geq 0, |\tilde{\gamma}_n - \gamma_n| < \frac{c_1}{\sqrt{k}}, \forall j > 1, |U_{j,k} - \tilde{U}_{j,k}| < \frac{c_2}{\sqrt{k}}\right) \\
&\quad + \mathbb{P}\left(|\tilde{\gamma}_n - \gamma_n| > \frac{c_1}{\sqrt{k}}\right) + \mathbb{P}\left(\max_{j=2,\dots,\lfloor n/k \rfloor} |U_{j,k} - \tilde{U}_{j,k}| > \frac{c_2}{\sqrt{k}}\right) =: A_n + B_n + C_n.
\end{aligned}$$

Invoking Lemma 3.6.3 and Lemma 3.6.4, it follows that $B_n \to 0$ and $C_n \to 0$ as $n \to \infty$.

Define

$$\tilde{I}_j = \mathbf{1}\left\{\tilde{D}_j > z_{1-1/\lfloor n/k \rfloor} - \frac{c_1 + c_2}{\sigma_\infty}\right\}, \quad j = 1, 2, \dots,$$

where $\tilde{I}_1, \tilde{I}_2, \dots, \tilde{I}_{\eta_n}$ are i.i.d. Bernoulli-distributed with $\mathbb{P}(\tilde{I}_1 = 1) := \kappa \to 0$ as $n \to \infty$. Extending $\tilde{I}_{\eta_n+1}, \tilde{I}_{\eta_n+2}, \dots$ to be independent and distributed as $\tilde{I}_1$, Lemma 3.6.5 yields

$$\mathbb{P}\left(\max_{j \geq 1} \sum_{i=1}^{j} (2\tilde{I}_i - 1) \geq 0\right) = \sum_{k=1}^{\infty} |A_k| \left(\kappa(1-\kappa)\right)^k \leq \min\left\{\frac{4\kappa}{1-4\kappa}, 1\right\} \leq 8\kappa,$$

where $|A_k| \leq 2^{2k}$ and $\kappa(1-\kappa) \leq \kappa$.

On the event $\{|\tilde{\gamma}_n - \gamma_n| < c_1/\sqrt{k}\} \cap \{|U_{j,k} - \tilde{U}_{j,k}| < c_2/\sqrt{k}\}$, we have $\tilde{I}_j \geq I_j$. Consequently,

$$A_n \leq \mathbb{P}\left(\max_{\ell=1,\dots,\eta_n} \sum_{j=\eta_n-\ell+1}^{\eta_n} (2\tilde{I}_j - 1) \geq 0\right) \leq \mathbb{P}\left(\max_{j \geq 1} \sum_{i=1}^{j} (2\tilde{I}_i - 1) \geq 0\right) \to 0.$$

Similarly, employing the notation $\tilde{J}_j = 1 - \tilde{I}_j$ for $j = \eta_n + 1, \dots, \lfloor n/k \rfloor$ and introducing independent $\tilde{J}_1, \dots, \tilde{J}_{\eta_n}$ and $\tilde{J}_{\lfloor n/k \rfloor + 1}, \tilde{J}_{\lfloor n/k \rfloor + 2}, \dots$, distributed as J_{η_n+1}, yields $\sup_{j \geq 1} \mathbb{P}(\tilde{I}_j = 1) := \bar{\kappa} \to 0$, as $n \to \infty$. This follows because $\tilde{D}_j - \mathbb{E}\tilde{D}_j$ are i.i.d. and

$z_{1-1/\lfloor n/k \rfloor} - \mathbb{E}\tilde{D}_j \leq z_{1-1/\lfloor n/k \rfloor} - \sqrt{k}d \to -\infty$ for $j \geq \eta_n + 1$. Thus,

$$\bar{\kappa} := \sup_{j \geq 1} \mathbb{P}(\tilde{J}_j = 1) \leq \mathbb{P}\left(\tilde{D}_j - \mathbb{E}\tilde{D}_j \leq z_{1-1/\lfloor n/k \rfloor} - \frac{\sqrt{k}d + c_1 + c_2}{\sigma_\infty} \right) \to 0.$$

Utilizing similar reasoning as before to approximate I_j with $\tilde{I}_j$, we deduce that

$$\begin{aligned} \mathbb{P}\left(\max_{\ell=1,\ldots,\lfloor n/k \rfloor - \eta_n} \sum_{j=\eta_n+1}^{\eta_n+\ell} \big(1 - 2I_j\big) \geq 0 \right) &= \mathbb{P}\Big(\max_{j \geq 1} \sum_{i=1}^{j} (2\tilde{J}_i - 1) \geq 0 \Big) + o(1) \\ &\leq \min\Big\{ \sum_{k=1}^{\infty} (4\bar{\kappa})^k, 1 \Big\} + o(1) \leq 8\bar{\kappa} + o(1) \to 0. \end{aligned}$$

This concludes the proof of Theorem 3.3.3. □

3.6.4 Proof of Lemma 3.6.1–3.6.5

Proof of Lemma 3.6.1

Utilizing the Cauchy-Schwarz inequality to establish a bound for (3.6.38), for each $j = 1, \ldots, p$, we derive

$$\begin{aligned} \Big(\mathbb{E} \Big| \mathbf{Z}_{i,j}^2 - \mathbf{Z}_{i,j}'^2 \Big|^q \Big)^{\frac{1}{q}} &= \Big(\mathbb{E} \left(|\mathbf{Z}_{i,j} - \mathbf{Z}'_{i,j}|^q \cdot |\mathbf{Z}_{i,j} + \mathbf{Z}'_{i,j}|^q \right) \Big)^{\frac{1}{q}} \\ &\leq \Big(\mathbb{E}(\mathbf{Z}_{i,j} - \mathbf{Z}'_{i,j})^{2q} \Big)^{\frac{1}{2q}} \cdot \Big(\mathbb{E}(\mathbf{Z}_{i,j} + \mathbf{Z}'_{i,j})^{2q} \Big)^{\frac{1}{2q}} \\ &\leq 2H_{2q} \cdot \Big(\mathbb{E}(\mathbf{Z}_{i,j} - \mathbf{Z}'_{i,j})^{2q} \Big)^{\frac{1}{2q}}. \end{aligned} \tag{3.6.70}$$

Given $H_{2q} \leq H_q < \infty$, the inequality (3.6.70) indicates that $\tilde{\delta}_q(i) \lesssim \delta_{2q}(i)$ for all $i = 1, \ldots, n$, thereby implying $\tilde{\Theta}_{t,q} \lesssim \Theta_{t,2q}$. □

Proof of Lemma 3.6.2

Under Condition 3.3.1, it is established that $\Xi_{q,\alpha} < \infty$ for $\alpha = 1$. The goal is to show that, under the same condition, $\tilde{\Xi}_{q,\alpha} < \infty$ is also true for $\alpha = 1$.

By Lemma 3.6.1, we have that $\tilde{\Theta}_{t,q} \lesssim \Theta_{t,2q}$. Consequently, Condition 3.3.1 implies

$$\tilde{\Theta}_{m,2q} = \mathcal{O}\left(m^{-\beta_{2q}}(\log m)^{-A}\right), \ \textit{for } A > \left(1/q + 2 + 2\beta_{2q}\right)/3, \tag{3.6.71}$$

thereby affirming that $\tilde{\Xi}_{q,\alpha} < \infty$ for $\alpha = 1$. □

Proof of Lemma 3.6.3

To demonstrate that (3.6.65) is satisfied, we establish an equivalent condition:

$$|\tilde{\gamma}_n - \gamma_n| = \mathcal{O}_{\mathbb{P}}\left(\frac{a_n}{\sqrt{n}}\right) \tag{3.6.72}$$

for any sequence $\{a_n\}$ where $a_n \to \infty$. Setting $r_n = a_n^{-1/2}$ yields

$$\begin{aligned}
&\mathbb{P}\left(|\tilde{\gamma}_n - \gamma_n| \geq \frac{a_n}{\sqrt{n}}\right) \\
&\lesssim \mathbb{P}\left(\hat{L} \leq \eta r_n, \frac{\left|\sum_{i=1}^{\hat{L}k}\left(\|\mathbf{X}_i - \boldsymbol{\mu}_0\|^2 - \gamma_n\right)\right|}{\hat{L}k} \geq \frac{a_n}{\sqrt{n}}\right) \\
&\quad + \mathbb{P}\left(\hat{L} > \eta r_n, \frac{\left|\sum_{i=1}^{\hat{L}k}\left(\|\mathbf{X}_i - \boldsymbol{\mu}_0\|^2 - \gamma_n\right)\right|}{\hat{L}k} \geq \frac{a_n}{\sqrt{n}}\right) \\
&\leq \mathbb{P}(\hat{L} \leq \eta r_n) \\
&\quad + \mathbb{P}\left(\max_{j > \eta r_n k} \frac{\left|\sum_{i=1}^{j}\left(\|\mathbf{X}_i - \boldsymbol{\mu}_0\|^2 - \gamma_n\right)\right|}{j} \geq \frac{a_n}{\sqrt{n}}\right) \to 0,
\end{aligned}$$

validating (3.6.72). To confirm this convergence to zero, we need to establish two key points:

First, for any sequence $t_n = o(\eta_n)$, it is shown that

$$\mathbb{P}(\hat{L} \geq t_n) \to 1, \quad \text{as } n \to \infty. \tag{3.6.73}$$

Second, for any sequence g_n, we demonstrate that

$$\max_{j \geq g_n} \left| \frac{\sum_{i=1}^{j} \left(\|\mathbf{X}_i - \boldsymbol{\mu}_0\|^2 - \gamma_n \right)}{j} \right| = \mathcal{O}_{\mathbb{P}} \left(\frac{1}{\sqrt{g_n}} \right). \tag{3.6.74}$$

These findings collectively culminate in the proof of Lemma 3.6.3 by establishing (3.6.73) and (3.6.74).

We commence with establishing (3.6.73). Initially, observe that

$$\mathbb{P}(\hat{L} \geq t_n) = \mathbb{P} \left(\max_{1 \leq i \leq \eta_n} k^{1/2} \omega_\infty^{-1} (R_{i,k} - \gamma_n) \geq \max_{1 \leq i \leq t_n} k^{1/2} \omega_\infty^{-1} (R_{i,k} - \gamma_n) \right).$$

Define $\theta_t = (2 \log t - \log \log t - \log(4\pi))^{1/2}$. We aim to demonstrate that

$$\begin{aligned} &\mathbb{P} \left(\max_{1 \leq i \leq \eta_n} k^{1/2} \omega_\infty^{-1} (R_{i,k} - \gamma_n) \geq \frac{\theta_{\eta_n} + \theta_{t_n}}{2} \right) \to 1, \\ &\mathbb{P} \left(\max_{1 \leq i \leq t_n} k^{1/2} \omega_\infty^{-1} (R_{i,k} - \gamma_n) \leq \frac{\theta_{\eta_n} + \theta_{t_n}}{2} \right) \to 1, \end{aligned} \tag{3.6.75}$$

as $n \to \infty$, which implies (3.6.73).

For a sequence t_n satisfying $\log(t_n) n^{1/q} k^{-1} \to 0$, as $n \to \infty$, it follows that

$$\sqrt{2 \log(t_n)} \left\{ \max_{1 \leq i \leq t_n} k^{1/2} \omega_\infty^{-1} (R_{i,k} - \gamma_n) - \theta_{t_n} \right\} \Rightarrow V, \tag{3.6.76}$$

where V follows the extreme value distribution $\mathbb{P}(V \leq x) = \exp\{-\exp(-x)\}$.

Given Condition 3.3.1 and (3.6.71), and applying Corollary 2.1 in Berkes et al. [2014], we find $Z_i^c \overset{\mathcal{D}}{=} \|\mathbf{Z}_i\|^2 - \gamma_n$, and a standard Brownian motion $\mathbb{B}_c(\cdot)$ on a space $(\Omega_c, \mathcal{A}_c, \mathrm{P}_c)$,

satisfying

$$\sum_{i=1}^{j} Z_i^c = \omega_\infty \mathbb{B}_c(j) + o_{a.s.}(j^{1/(2q)}).$$

Let $W_{i,k} = k^{-1/2}\left(\mathbb{B}_c(ik) - \mathbb{B}_c((i-1)k)\right)$ for $i = 1, 2, \ldots, t_n$, which are i.i.d. standard normally distributed, and $M_{i,k} = \sum_{j=(i-1)k+1}^{ik} \left(Z_j^c - (\mathbb{B}_c(j) - \mathbb{B}_c(j-1))\right)$, then

$$R_{i,k} - \gamma_n \stackrel{\mathcal{D}}{=} \omega_\infty k^{-1/2} W_{i,k} + k^{-1}\omega_\infty M_{i,k},$$

with $\max_{i \le t_n} M_{i,k} = o_{a.s.}(n^{1/(2q)})$. This setup leads to

$$k^{1/2}\omega_\infty^{-1}(R_{i,k} - \gamma_n) \stackrel{\mathcal{D}}{=} W_{i,k} + o_{a.s.}\left\{n^{1/(2q)}k^{-1/2}\right\} = W_{i,k} + o_{a.s.}\left\{1/\theta_{t_n}\right\}.$$

This equation, alongside the basic property that

$$\sqrt{2\log(t_n)}\left\{\max_{1 \le i \le t_n} (\mathbb{B}_c(i) - \mathbb{B}_c(i-1)) - \theta_{t_n}\right\} \Rightarrow V,$$

as $n \to \infty$, facilitates the proof of (3.6.73).

We now establish (3.6.74), where $\tilde{S}_i$ denotes $\sum_{j=1}^{i}\left(\|\mathbf{X}_j - \boldsymbol{\mu}_0\|^2 - \gamma_n\right)$. For any $G > 0$ and setting $u_{g_n} = G/\sqrt{g_n}$, the application of Bonferroni's inequality yields

$$\begin{aligned}
&\mathbb{P}\left(\max_{j \ge g_n}\left|\frac{\tilde{S}_j}{j}\right| \ge u_{g_n}\right) \\
&\le \sum_{k=1}^{\infty}\sum_{v=1}^{p}\mathbb{P}\left(\max_{2^{k-1}g_n \le i \le 2^k g_n}\left|\frac{\sum_{j=1}^{i}\left(\mathbf{Z}_{j,v}^2 - \mathbb{E}\mathbf{Z}_{j,v}^2\right)}{i}\right| \ge \frac{u_{g_n}}{p}\right) \\
&\le \sum_{k=1}^{\infty}\sum_{v=1}^{p}\mathbb{P}\left(\max_{1 \le i \le 2^k g_n}\left|\sum_{j=1}^{i}\left(\mathbf{Z}_{j,v}^2 - \mathbb{E}\mathbf{Z}_{j,v}^2\right)\right| \ge 2^{k-1}u_{g_n}g_n/p\right).
\end{aligned} \tag{3.6.77}$$

By applying a Nagaev-type inequality under dependence (as per Theorem 2 in Wu and

Wu [2016]) given Lemma 3.6.2, there exist positive constants C_1, C_2, and C_3 such that

$$\begin{aligned}&\mathbb{P}\left(\max_{1\le i\le 2^k g_n}\left|\sum_{j=1}^{i}\left(\mathbf{Z}_{j,v}^2-\mathbb{E}\mathbf{Z}_{j,v}^2\right)\right|\ge 2^{k-1}u_{g_n}g_n/p\right)\\&\le C_1\frac{\Xi_{2q,1}^{2q}p^q}{2^{k(q-1)-q}u_{g_n}^q g_n^{q-1}}+C_2\exp\left(-\frac{C_3 2^{k-2}u_{g_n}^2 g_n}{p^2\Xi_{2,1}^2}\right).\end{aligned}\tag{3.6.78}$$

Considering

$$\max\left\{\frac{\Xi_{2q,1}^{2q}}{u_{g_n}^q g_n^{q-1}},\frac{\Xi_{2,1}^2}{u_{g_n}^2 g_n}\right\}\to 0,\quad \text{as } g_n\to\infty,$$

it follows that

$$\sum_{k=1}^{\infty}\sum_{v=1}^{p}\mathbb{P}\left(\max_{1\le i\le 2^k g_n}\left|\sum_{j=1}^{i}\left(\mathbf{Z}_{j,v}^2-\mathbb{E}\mathbf{Z}_{j,v}^2\right)\right|\ge 2^{k-1}u_{g_n}g_n/p\right)\to 0,\quad \text{as } g_n\to\infty,$$

thereby concluding our result. □

Proof of Lemma 3.6.4

Utilizing the Bonferroni inequality, we obtain

$$\mathbb{P}\left(\max_{j\le s}\left|U_{j,k}-\tilde{U}_{j,k}\right|>u\right)\le s\cdot\mathbb{P}\left(|U_{1,k}-\tilde{U}_{1,k}|>u\right).\tag{3.6.79}$$

Define $\Delta_j:=\mathbb{E}(U_{1,k}|\varepsilon_k,\ldots,\varepsilon_j)-\mathbb{E}(U_{1,k}|\varepsilon_k,\ldots,\varepsilon_{j-1})$. It follows that

$$\tilde{U}_{1,k}-U_{1,k}=\sum_{j=-\infty}^{1}\Delta_j.$$

By applying the Markov inequality, we have

$$\mathbb{P}\left(\left|\sum_{j=-\infty}^{1}\Delta_j\right|>u\right)\leq\frac{\mathbb{E}\left|\sum_{j=-\infty}^{1}\Delta_j\right|^q}{u^q}.$$

Given $\{\Delta_j\}$ as a martingale difference sequence, the Burkholder-Davis-Gundy inequality allows us to bound $\mathbb{E}\left|\sum_{j=-\infty}^{1}\Delta_j\right|^q$, yielding

$$\left(\mathbb{E}\left|\sum_{j=-\infty}^{1}\Delta_j\right|^q\right)^{2/q}\leq c_q\sum_{j=-\infty}^{1}\left(\mathbb{E}\left|\Delta_j\right|^q\right)^{2/q}.$$

Considering $U_{1,k}=\frac{1}{k}\sum_{i=1}^{k}\|\mathbf{X}_i-\boldsymbol{\mu}_0\|^2$, the triangle and Jensen's inequalities, alongside Lemma 3.6.1, lead to

$$\begin{aligned}\left(\mathbb{E}\left|\Delta_j\right|^q\right)^{1/q}&\leq\frac{1}{k}\sum_{i=1}^{k}\left(\mathbb{E}\left|\mathbb{E}\left(\|\mathbf{Z}_i\|^2|\varepsilon_i,\ldots,\varepsilon_j\right)-\mathbb{E}\left(\|\mathbf{Z}_i\|^2|\varepsilon_i,\ldots,\varepsilon_{j-1}\right)\right|^q\right)^{1/q}\\&\leq\frac{p}{k}\sum_{i=1}^{k}\tilde{\delta}_{(i-j+1),\theta}\leq\frac{p}{k}\tilde{\Theta}_{2-j,q}\leq\frac{p}{k}\Theta_{2-j,2q}.\end{aligned}\tag{3.6.80}$$

Under Condition 3.3.1 and combining the aforementioned results, it follows that

$$\mathbb{P}\left(\max_{j\leq s}\left|U_{j,k}-\tilde{U}_{j,k}\right|>u\right)\lesssim sp^q(ku)^{-q},\tag{3.6.81}$$

where the constant in $\lesssim$ depends only on q.

This completes the proof of Lemma 3.6.4. □

Proof of Lemma 3.6.5

Consider $S_n = \sum_{i=1}^{n}(2\tilde{I}_i - 1) =: \sum_{i=1}^{n} X_i$, a random walk beginning at $S_0 = 0$. Define $t_0 := \inf\{n > 0 : S_n = 0\}$ as the first instance $n > 0$ for which S_n returns to zero, yielding

$$\mathbb{P}\Big(\max_{j\geq 1}\sum_{i=1}^{j}(2\tilde{I}_i - 1) > 0\Big) = \sum_{k=1}^{\infty}\mathbb{P}(t_0 = 2k).$$

For $k > 0$, we find

$$\mathbb{P}(t_0 = 2k) = \sum_{(i_1,\ldots,i_{2k})\in A_k}\mathbb{P}(X_1 = i_1, \ldots, X_{2k} = i_{2k}),$$

where the sum is taken over the set A_k of all sequences of $2k$ terms from $\{-1, 1\}$ that return to zero exactly at $2k$. This formulation directly results from the independence of $\tilde{I}_1, \tilde{I}_2, \ldots$, completing the proof. □

CHAPTER 4

NON-PARAMETRIC INFERENCE FOR CHANGE POINT UNDER NON-STATIONARY NOISE

4.1 Introduction

Chapter 2 and Chapter 3 focused on change point analysis under irregular signals. Although the discussed methods do not assume any parametric model and allow for temporal dependence in the noise sequence, they still require the stationarity of the noise process. However, in practice, the stationarity condition for the noise process may not be satisfied for data from financial market analysis or pandemic research. One approach is to apply methodologies that assume stationarity to analyze such datasets. However, as indicated by Mercurio and Spokoiny [2004], treating non-stationary data as though they are from a stationary process in data analysis is highly risky. Therefore, the next question we aim to address is whether we can extend our change point analysis to non-stationary scenarios and what the associated trade-offs are for this extension. The goal of this chapter is to develop a novel framework for non-parametric statistical inference of change points under non-stationary noise.

To the best of our knowledge, most of the current relevant literature adopts a parametric approach for statistical inference of change points under non-stationary noise. Davis et al. [2006] explore modeling non-stationary time series using piece-wise auto-regressive processes with unknown segment numbers, locations, and orders, employing the minimum description length principle and a genetic algorithm to optimize the model structure. Chowdhury et al. [2012] introduces a Bayesian online inference for spectral change point detection (BOSCPD) technique applied to online automatic speech recognition (ASR), significantly enhancing change point detection in non-stationary noise environments. Korkas and Pryzlewicz [2017] propose a technique utilizing Wild Binary Segmentation (WBS) combined with CUSUM

statistics from wavelet periodograms for consistent estimation of change-point numbers and locations in non-stationary time series. Safikhani and Shojaie [2022] introduce a three-stage procedure for detecting change points and estimating parameters in high-dimensional piecewise vector auto-regressive (VAR) models under non-stationary conditions, using penalized least squares with total variation penalty to handle the change point detection as a variable selection problem. Ma et al. [2020] present a method for identifying change points in non-stationary time series by modeling each segment as an auto-regressive process with distinct parameters, using a likelihood ratio scan for detection and spectral discrimination tests for validation. Ching et al. [2014] address the detection of multiple structural breaks in non-stationary economic time series using the Minimum Description Length (MDL) and Genetic Algorithm (GA) for model optimization.

The aforementioned studies provide valuable insights into change point detection under non-stationary conditions. However, they primarily rely on parametric models, which may not always be suitable for capturing the complex dynamics of time series with non-stationary noise. In this chapter, we aim to bridge this gap by developing a non-parametric framework for statistical inference of change points in non-stationary settings. We propose a novel test statistic based on local linear fit, which effectively captures change points in the signal even in the presence of strong noise distraction. Additionally, we introduce a new bootstrap procedure to approximate the distribution of this test statistic under the null hypothesis, enabling the calculation of p-values for the test. This innovative bootstrap algorithm ensures a robust distributional approximation even in the presence of complex temporal dependence and non-stationarity.

To relax the stationarity condition assumed in Chapters 2 and 3, a trade-off is necessary. Unlike these chapters, which allow for highly irregular signals after the change point, we impose a slightly stronger restriction on the signal part, requiring continuity after the change point to accommodate the non-stationarity of the noise. Despite this restriction, our

methodology relies on very weak conditions, as we do not assume piece-wise constant signals or i.i.d. Gaussian noise, which are common assumptions in many other works.

Under the introduced settings, we investigate the theoretical properties of our proposed methodology and explore its performance through extensive synthetic data analysis. This non-parametric framework contributes to the growing body of literature on change point analysis by providing a novel perspective on handling non-stationarity, expanding the toolkit available for researchers and practitioners dealing with complex, non-stationary time series data.

The rest of this chapter is organized as follows: Section 4.2 introduces our methodology for testing for change points in signals with non-stationary noise processes. In Section 4.3, we develop the theoretical foundations of our approach, demonstrating the type-one error control by showing that the distribution of our test statistic is well approximated via the introduced bootstrap mechanism under the null hypothesis. The empirical efficacy of our method is rigorously evaluated through extensive simulations in Section 4.4. Detailed proofs of the theoretical results are provided in Section 4.5.

4.2 Methodology

4.2.1 Data Model and Problem Formulation

Consider a data model infused with noise, formally articulated as:

$$X_t = \mu(t/n) + Z_t, \quad \forall t = 1, \ldots, n, \tag{4.2.1}$$

where $\mu(\cdot)$ is the mean function, and $(Z_t)_{t\in\mathbb{Z}}$ denotes a non-stationary process with a mean of zero.

We consider the following hypotheses:

- **Null Hypothesis** (H_0):

$$H_0 : \mu(\cdot) \text{ is continuous on } [0, 1], \tag{4.2.2}$$

- **Alternative Hypothesis** (H_1):

$$H_1 : \text{There exists } t_0 \in [0, 1] \text{ such that } \mu(\cdot) \text{ is discontinuous at } t_0. \tag{4.2.3}$$

Given the observations $X_1, \ldots, X_n$, our objective is to develop a testing procedure to distinguish between H_0 and H_1.

Our test statistic innovates upon the traditional change point detection scheme, known as MOSUM (cf. Eichinger and Kirch [2018]), by adapting it to the specifics of our model. Classical MOSUM computes the absolute differences between mean values of data across windows bifurcating at each timepoint t, selecting the maximum difference as the test statistic T. The rejection criterion under MOSUM is defined by $\{T > c\}$, where c denotes a critical threshold ascertainable through simulation or bootstrap techniques.

In deviation from the classic MOSUM framework, our method involves the application

of a local linear model on the data segments partitioned by the time point t on both sides. We compute the absolute disparity in predictions from these twin models at t, defining our test statistic as the apex of these disparities. The algorithmic embodiment of this procedure is documented in Algorithm 2. For p-value calculation, we advocate a bootstrap strategy as elaborated in Algorithm 1, aiming to simulate the distribution of our test statistic under the null hypothesis (4.2.2). Theorem 4.3.1 validates that the distribution of our test statistic is aptly approximated via the introduced bootstrap mechanism, showcasing its efficacy in scenarios characterized by non-stationarity and temporal dependency in the noise process.

4.2.2 Bootstrap for Non-stationary Processes

The bootstrap is a powerful resampling technique that allows for statistical inference in complex settings where analytical results may be difficult to obtain. In the context of non-stationary processes, the bootstrap can be particularly useful for estimating the sampling distribution of statistics of interest. Algorithm 1 outlines a novel bootstrap procedure specifically designed for non-stationary time series data.

The key idea behind this new bootstrap is to divide the observed time series into contiguous blocks of size m and then use data within these blocks to preserve the local structure of dependence and variance of the original data.

To achieve this, Algorithm 1 first computes the block sums $B_j = \sum_{t=(j-1)m+1}^{jm} X_t$ for each block $j = 1, 2, \ldots, \lfloor n/m \rfloor$. These block sums serve as a measure of the local level of the time series within each block. Next, for each time point $i = 1, 2, \ldots, n$, the algorithm computes the cumulative sum of squared block sums B_j^2 and the product of adjacent block sums $B_j B_{j+1}$, up to the current block $\eta = \lfloor i/m \rfloor$, plus an additional term R for the remaining observations within the current block. Here, B_j^2 captures the time-varying variance structure, and $B_j B_{j+1}$ captures the local dependence structure. These cumulative sums, denoted by T_i, capture the global non-stationarity of the time series.

The bootstrap samples are then constructed by first sorting and ranking the T_i values and then using these ranks to generate Gaussian processes with increments that match the increments of the original T_i process. Specifically, the algorithm generates b independent standard normal variables $Z_1, Z_2, \ldots, Z_n$ and then constructs the bootstrap samples $\mathbf{Y}_s = (Y_{s,1}, Y_{s,2}, \ldots, Y_{s,n})$ for $s = 1, 2, \ldots, b$ by setting $Y_{s,1} = S_{s,r_1}$ and $Y_{s,t} = S_{s,r_t} - S_{s,r_{t-1}}$ for $t = 2, \ldots, n$, where $S_{s,t}$ is a cumulative sum process constructed from the Z_t variables and the increments of the sorted T_i process. Depending on whether the minimum value of the sorted T_i process, H_1, is negative or non-negative, the construction of the $S_{s,t}$ process differs slightly to ensure that it remains non-negative.

The resulting bootstrap samples $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_b$ can then be used to estimate the sampling distribution of any statistic of interest, such as the mean, variance, or autocorrelation function of the time series. By resampling blocks of the original data in a way that preserves both the local dependence structure and the global non-stationarity, this new bootstrap procedure provides a flexible and robust approach to inference for non-stationary time series.

Algorithm 1 Bootstrap for non-stationary process

Input: Observed data $\{X_i\}_{i=1}^n$; b, the size of Bootstrap; m, the block size.

for all $j = 1, 2, \cdots, \lfloor n/m \rfloor$ **do**

$B_j \leftarrow \sum_{t=(j-1)m+1}^{jm} X_t$

end for

for all $i = 1, 2, \cdots, n$ **do**

Update $\eta \leftarrow \lfloor i/m \rfloor$

Update $R \leftarrow \sum_{t=\eta m+1}^{i} X_t$; $R \leftarrow 0$ if $i = \eta m$

Update $(C_1, \cdots, C_\eta, C_{\eta+1}) \leftarrow (B_1, \cdots, B_\eta, R)$

Store $T_i \leftarrow \sum_{t=1}^{\eta+1} C_t^2 + 2\sum_{t=1}^{\eta} C_t C_{t+1}$

end for

Store $(H_1, \cdots, H_n) \leftarrow \text{sort}\,(T_1, \cdots, T_n)$

Store $(r_1, \cdots, r_n) \leftarrow \text{rank}\,(T_1, \cdots, T_n)$

if $H_1 < 0$ **then**

Store $w \leftarrow \max\{t : H_t < 0\}$.

for all $s = 1, 2, \cdots, b$ **do**

Generate $Z_1, Z_2, \cdots, Z_n$ i.i.d. standard normal.

Update $S_{s,w} \leftarrow \sqrt{-H_w} Z_w$, $S_{s,t} \leftarrow S_{s,t+1} + \sqrt{H_{t+1} - H_t} Z_t$, $t = 1, \cdots, w-1$.

Update $S_{s,w+1} \leftarrow \sqrt{H_{w+1}} Z_{w+1}$, $S_{s,t} \leftarrow S_{s,t-1} + \sqrt{H_t - H_{t-1}} Z_t$, $t = w+2, \cdots, n$.

Update $Y_{s,1} \leftarrow S_{s,r_1}$, $Y_{s,t} \leftarrow S_{s,r_t} - S_{s,r_{t-1}}$, $t = 2, \cdots, n$.

Store $\mathbf{Y}_s \leftarrow (Y_{s,1}, Y_{s,2}, \cdots, Y_{s,n})$

end for

else

for all $s = 1, 2, \cdots, b$ **do**

Generate $Z_1, Z_2, \cdots, Z_n$ i.i.d. standard normal.

Update $S_{s,1} \leftarrow \sqrt{H_1} Z_1$, $S_{s,t} \leftarrow S_{s,t-1} + \sqrt{H_t - H_{t-1}} Z_t$, $t = 2, \cdots, n$.

Update $Y_{s,1} \leftarrow S_{s,r_1}$, $Y_{s,t} \leftarrow S_{s,r_t} - S_{s,r_{t-1}}$, $t = 2, \cdots, n$.

Store $\mathbf{Y}_s \leftarrow (Y_{s,1}, Y_{s,2}, \cdots, Y_{s,n})$

end for

end if

Output: Output the data frame $(\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_b)$.

4.2.3 Testing Based on Local Linear Fit

We propose a novel testing procedure for detecting break points in the mean of a non-stationary time series, as outlined in Algorithm 2. This procedure is based on fitting local linear models to the data within a sliding window, and comparing the fitted values at the center of the window on either side of each potential change point.

Specifically, for each time point $j \in \{k+1, \ldots, n-k\}$, where k is the window size and n is the total number of observations, we fit two linear models to the data within the windows $\{(j-i, X_i) : i = j-k+1, \ldots, j\}$ and $\{(i-j, X_i) : i = j, \ldots, j+k-1\}$, respectively. These models take the form $X_i = \beta_0 + \beta_1(j-i)$ for the left window and $X_i = \beta_0 + \beta_1(i-j)$ for the right window, where β_0 and β_1 are the intercept and slope parameters to be estimated. We denote the fitted intercepts from these models by $\hat{\beta}_{j,l}$ and $\hat{\beta}_{j,r}$, respectively.

The test statistic S_j is then defined as the absolute difference between these fitted intercepts, i.e., $S_j = |\hat{\beta}_{j,l} - \hat{\beta}_{j,r}|$. Intuitively, if there is a break point in the mean of the time series at time j, we would expect the fitted intercepts from the left and right windows to differ significantly, resulting in a large value of S_j.

To assess the significance of the observed test statistic, we employ a bootstrap procedure to approximate its distribution under the null hypothesis of no change point. This procedure, detailed in Algorithm 1, involves first estimating the residuals $\hat{r}_i = X_i - \hat{\mu}_i$ from the observed data, where $\hat{\mu}_i$ are the fitted signals. These residuals are then resampled using our introduced bootstrap approach to generate b bootstrap samples $\mathbf{W}_s = (W_{s,1}, W_{s,2}, \ldots, W_{s,n})$, $s = 1, 2, \ldots, b$.

For each bootstrap sample, we apply the same local linear fitting procedure as described above to compute the test statistic T_j at each time point j. The maximum value of T_j across all time points is then recorded as D_s for the s-th bootstrap sample. Finally, the p-value for the observed test statistic is estimated as $\hat{p} = \frac{1}{b+1}(1 + \sum_{s=1}^{b} \mathbf{1}\{\max_j S_j > D_s\})$, which represents the proportion of bootstrap samples for which the maximum test statistic exceeds

the observed maximum test statistic.

This testing procedure based on local linear fits is particularly well-suited for detecting break points in the mean of a time series, as it is able to capture local trends in the data while being robust to non-stationarity and temporal dependence in the noise process. The use of a bootstrap procedure to estimate the p-value further enhances the flexibility and applicability of this approach in a wide range of settings.

Algorithm 2 Testing for the change point of means.

Input: Sequential data $(X_i)_{i=1}^n$; Window size k; Size of Bootstrap sample b; Block size in Bootstrap, m.

for $j \in \{k+1, \cdots, n-k\}$ **do**
  With data $\left\{\big((j-i), X_i\big) : i = j-k+1, \cdots, j\right\}$ as input, we fit a linear model: $X_i = \beta_0 + \beta_1(j-i)$.
  $\hat{\beta}_{j,l} \leftarrow \hat{\beta}_0$.
  With data $\left\{\big((i-j), X_i\big) : i = j, \cdots, j+k-1\right\}$ as input, we fit a linear model: $X_i = \beta_0 + \beta_1(i-j)$.
  $\hat{\beta}_{j,r} \leftarrow \hat{\beta}_0$.
  $S_j \leftarrow |\hat{\beta}_{j,l} - \hat{\beta}_{j,r}|$.
end for
Let $\{\hat{\mu}_i\}_{i=1}^n$ be the fitted signals from the observed data.
for all $i = 1, 2, \cdots, n$ **do**
  $\hat{r}_i \leftarrow X_i - \hat{\mu}_i$
end for
for $s \in \{1, 2, \cdots, b\}$ **do**
  Apply Algorithm 1 with $\{\hat{r}_i\}$ and m as input to generate Bootstrap sample $\mathbf{W}_s = (W_{s,1}, W_{s,2}, \cdots, W_{s,n})$.
  With data $\left\{\big((j-i), W_{s,i}\big) : i = j-k+1, \cdots, j\right\}$ as input, we fit a linear model: $W_{s,i} = \gamma_0 + \gamma_1(j-i)$.
  $\hat{\gamma}_{j,l} \leftarrow \hat{\gamma}_0$.
  With data $\left\{\big((i-j), W_{s,i}\big) : i = j, \cdots, j+k-1\right\}$ as input, we fit a linear model: $W_{s,i} = \gamma_0 + \gamma_1(i-j)$.
  $\hat{\gamma}_{j,r} \leftarrow \hat{\gamma}_0$.
  $T_j \leftarrow |\hat{\gamma}_{j,l} - \hat{\gamma}_{j,r}|$.
  $D_s \leftarrow \max_{k+1 \le j \le n-k} T_j$
end for
$\hat{p} \leftarrow \frac{1}{b+1}(1 + \sum_{s=1}^b \mathbf{1}\{\max_j S_j > D_s\})$

Output: Estimated p value $\hat{p}$.

4.3 Theory

4.3.1 Assumptions on the Non-stationary Noise Processes

In this section, we introduce a framework for modeling non-stationary noise processes and establish the necessary assumptions for our change point analysis methodology. We consider the non-stationary process $(Z_t)_{t\in\mathbb{Z}}$ as the output of a physical system, represented by the following equation:

$$Z_t = g_t\left(\cdots, \varepsilon_{t-1}, \varepsilon_t\right) \tag{4.3.4}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ are i.i.d. inputs to the system and $(g_t)_{t\in\mathbb{Z}}$ is a set of $\mathbb{R}$-valued measurable functions. This representation allows for a flexible and general description of non-stationary processes, where the dependence structure and the functional form of the process can vary over time.

To quantify the dependence structure of the non-stationary process, we introduce the uniform functional dependence measure. Given a time lag $k \in \mathbb{Z}$, this measure assesses the sensitivity of the system's output Z_i to changes in the input information at time $i-k$. Specifically, we define:

$$\delta_p(k) := \sup_i \left(\mathbb{E}\left|Z_i - Z_{i,i-k}\right|^p\right)^{1/p}, \text{ where } Z_{i,i-k} = g_i\left(\cdots, \varepsilon_{i-k-1}, \varepsilon'_{i-k}, \varepsilon_{i-k+1}, \cdots, \varepsilon_i\right) \tag{4.3.5}$$

where $Z_{i,i-k}$ is a coupled version of Z_i, obtained by replacing the input ε_{i-k} with an i.i.d. copy ε'_{i-k}. The term $\left(\mathbb{E}\left|Z_i - Z_{i,i-k}\right|^p\right)^{1/p}$ captures the dependence of Z_i on ε_{i-k}. Since Z_i is a non-stationary process, the physical mechanism g_i is allowed to vary over time. Therefore, we define the functional dependence measure uniformly by taking the supremum over all i.

To express our dependence condition, we introduce the quantity:

$$\Theta_{i,p} = \sum_{k=i}^{\infty} \delta_p(k), \quad i \geq 0. \tag{4.3.6}$$

This quantity summarizes the cumulative dependence of the non-stationary process on its past inputs, starting from a given time lag i. The rate at which $\Theta_{i,p}$ decays as i increases will play a crucial role in determining the theoretical properties of our change point analysis methodology.

By modeling non-stationary noise processes through the lens of a physical system with time-varying functional mechanisms and quantifying the dependence structure using the uniform functional dependence measure, we establish a flexible and comprehensive framework for analyzing change points in the presence of non-stationarity. The assumptions outlined in this section provide the foundation for the theoretical developments presented in the subsequent sections of this chapter.

4.3.2 Main Theorem

Before presenting the main theorem, we introduce two essential conditions that are necessary for the theorem to hold. These conditions ensure that the dependence structure and moment properties of the underlying non-stationary noise sequence $\{Z_i\}$ are well-behaved.

Condition 4.3.1. *Consider a sequence $\{Z_i\}$ where each Z_i is expressed as in* (4.3.4)*. Assume that the p-norm of Z_i, denoted as $\|Z_i\|_p$, is uniformly bounded by a finite value for all i within $1 \leq i \leq n$ and for some $p > 2$. Further, suppose there exist constants $A > 1$ and $C > 0$ such that for all $i \geq 0$, the following inequality holds:*

$$\Theta_{i,p} = \sum_{k=i}^{\infty} \delta_p(k) \leq C(i+1)^{-A}. \tag{4.3.7}$$

This condition ensures that the dependence between the elements of the sequence $\{Z_i\}$ decays sufficiently fast as the distance between the indices increases. The parameter A controls the rate of decay, with larger values of A implying weaker dependence.

Condition 4.3.2. *Consider the series $(|Z_i|^p)$. It is required to satisfy the truncated uniform*

integrability condition, which can be stated as follows: For any fixed $a > 0$, the supremum of the expected value of $|Z_i|^p$ conditioned on $|Z_i|^p \geq an$, across all $1 \leq i \leq n$, tends to zero as $n \to \infty$. Formally,

$$\sup_{1 \leq i \leq n} \mathbb{E}\left(|Z_i|^p \mathbf{1}_{\{|Z_i|^p \geq an\}}\right) \to 0 \text{ as } n \to \infty. \tag{4.3.8}$$

This condition ensures that the moments of the noise sequence $\{Z_i\}$ are well-behaved and do not exhibit any extreme behavior. It is a form of uniform integrability condition that is commonly used in the study of dependent sequences.

Now we can introduce our main theorem. Let $Y_{n,k_n}(i) := \hat{\beta}_{i,l} - \hat{\beta}_{i,r}$, as in Algorithm 2. Define the function $\alpha_{n,k_n}(i,j)$ as follows:

$$\alpha_{n,k_n}(i,j) := \begin{cases} \frac{\left(\sum_{h=1}^{k_n-1} h^2\right) - \left(\sum_{h=1}^{k_n-1} h\right) \cdot (j-i)}{k_n\left(\sum_{h=1}^{k_n-1} h^2\right) - \left(\sum_{h=1}^{k_n-1} h\right)^2} & \text{for } j \in (i, i+k_n-1], \\ 0 & \text{for } j \in \{i\} \cup (0, i-k_n+1) \cup (i+k_n-1, n], \\ \frac{\left(\sum_{h=1}^{k_n-1} h\right) \cdot (i-j) - \left(\sum_{h=1}^{k_n-1} h^2\right)}{k_n\left(\sum_{h=1}^{k_n-1} h^2\right) - \left(\sum_{h=1}^{k_n-1} h\right)^2} & \text{for } j \in [i-k_n+1, i). \end{cases} \tag{4.3.9}$$

We also introduce the following notations:

$$B_a := \sum_{i=(a-1)m+1}^{am\wedge n} Z_i, \tag{4.3.10}$$

$$\tag{4.3.11}$$

$$T_j := \sum_{a=1}^{\lfloor j/m \rfloor} B_a^2 + 2 \sum_{a=1}^{\lfloor j/m \rfloor - 1} B_a B_{a+1}, \tag{4.3.12}$$

$$\tag{4.3.13}$$

$$R_j := I\{j/m \notin \mathbb{N}\} \cdot \sum_{i=\lfloor j/m \rfloor m+1}^{j} Z_i. \tag{4.3.14}$$

Using these notations, we have $S_j = \sum_{i=1}^{j} Z_i = \sum_{a=1}^{\lfloor j/m \rfloor} B_a + R_j$, and we define

$$\mathcal{T}_j := T_{\lfloor j/m \rfloor} + R_j^2 + 2B_{\lfloor j/m \rfloor} R_j. \tag{4.3.15}$$

Further, let

$$\hat{Y}_{n,k_n}(i) := \sum_{j=1}^{n} \alpha_{n,k_n}(i,j) \left\{ \mathbb{B}(\mathcal{T}_j) - \mathbb{B}(\mathcal{T}_{j-1}) \right\}, \tag{4.3.16}$$

where $\mathbb{B}(\cdot)$ is standard Brownian motion.

Theorem 4.3.1. *Consider the mean function $\mu(\cdot)$ as specified in (4.2.1), where $\mu \in L[0,1]$. Assume that for the sequence $(Z_t)_{t\geq 1}$, Condition 4.3.1 holds with*

$$A > A_0 = A_0(p) := \max\left\{ \frac{p^2 - p - 2 + (p-2)\sqrt{p^2 + 10p + 1}}{4p}, 1 \right\}. \tag{4.3.17}$$

Also, assume Condition 4.3.2. After taking $k_n = n^{(1+r_1)/3}(\log n)^{1/6}$, we then have

$$\max_{i \leq n} |Y_{n,k_n}(i) - \hat{Y}_{n,k_n}(i)| = \mathcal{O}_{\mathbb{P}}\left(n^{\frac{2r_1 - 1}{3}} (\log n)^{1/3} \right), \tag{4.3.18}$$

where $r_1 = \max\left\{\frac{1+A}{2(1+2A)}, \frac{1+4A/p}{2(1+2A)}\right\}$.

The rate of convergence in this theorem is given by $n^{\frac{2r_1-1}{3}}(\log n)^{1/3}$, which depends on the parameters A and p through r_1. The condition on A ensures that the dependence in the noise sequence $\{Z_i\}$ is not too strong.

Theorem 4.3.1 establishes the asymptotic equivalence between the difference of the local linear estimates $Y_{n,k_n}(i) := \hat{\beta}_{i,l} - \hat{\beta}_{i,r}$ and its Bootstrap approximation $\hat{Y}_{n,k_n}(i)$ under the null hypothesis (4.2.2). This equivalence is crucial as $\max_i |Y_{n,k_n}(i)|$, the maximum absolute difference, serves as the test statistic in Algorithm 2. Consequently, the theorem validates the Bootstrap procedure employed in Algorithm 2 for approximating the distribution of the test statistic under the null hypothesis.

The detailed proof of this theorem is provided in Section 4.5.1.

4.4 Monte Carlo Studies

4.4.1 Models Considered

We examine the finite sample performance of the testing procedure discussed in Section 4.2.3. The data for our experiments are generated based on the signal plus noise model as defined in (4.2.1). This model allows us to assess the effectiveness of the proposed testing procedure in detecting changes in the trend of a non-stationary time series.

For the noise component, we consider a causal non-stationary process as follows:

Let $w_t = \sin\left(\frac{4\pi t}{n}\right)$, and let $Z_t = \theta_t Z_{t-1} + \varepsilon_t$ where $\theta_t = \theta w_t$ and $Z_0 = 0$. In this causal process, the non-stationarity is evident as the variance of (Z_t) varies with respect to time t. The time-varying coefficient θ_t introduces a periodic structure in the dependence of the noise process, which is a common feature in many real-world applications, such as seasonal or cyclical patterns in economic or environmental data. The parameter θ determines the strength of the temporal dependence, with larger values of $|\theta|$ indicating stronger dependence. The innovations ε_t are assumed to be independent and identically distributed random variables with zero mean and finite variance.

For the signal component μ_t, we consider two scenarios:

(i) Under the null hypothesis H_0, as defined in (4.2.2), the signal function is continuous and is given by

$$\mu_t = \sin(\frac{2\pi t}{n}) \quad \text{for } t = 1, 2, \ldots, n. \tag{4.4.19}$$

(ii) Under the alternative hypothesis H_1, as defined in (4.2.3), the signal is generated according to

$$\mu_t = \begin{cases} \sin(\frac{2\pi t}{n}) + s & \text{for } t = 1, 2, \ldots, \tau - 1 \\ \sin(\frac{2\pi t}{n}) - s & \text{for } t = \tau, \tau + 1, \ldots, n, \end{cases} \tag{4.4.20}$$

where the parameter s governs the magnitude of the discontinuity in the signal function.

The choice of the signal function under the null hypothesis, as given in (4.4.19), represents a smooth, non-linear trend in the time series. The sinusoidal form captures a periodic pattern with a single cycle over the entire time range. The scaling factor $\frac{2\pi}{n}$ ensures that the period of the signal is equal to the sample size n.

Under the alternative hypothesis, the signal function in (4.4.20) introduces a change point at time τ, where the trend of the time series experiences an abrupt shift. The magnitude of the shift is determined by the parameter s, with larger values of s corresponding to more pronounced changes in the trend. The sinusoidal components on either side of the change point have the same frequency as the null hypothesis, with a period of n.

Figure 4.1 showcases an example of the data series (X_i) with sample size $n = 500$ under the alternative hypothesis H_1, featuring a distinct discontinuity at $\tau = 250$. The signal function μ_t transitions abruptly, causing a shift in magnitude of $2s = 1.0$, where $s = 0.5$. The noise process is rendered non-stationary and temporally dependent by setting $\theta = 0.5$.

It is important to note that the signal is generated only once and this identical signal is utilized across all replications. This approach ensures that the performance of the testing procedure is evaluated under a fixed signal structure, eliminating any variability that may arise from generating different signals for each replication. However, each replication involves adding a distinct realization of the noise process to the signal. This allows us to assess the robustness of the testing procedure to different noise scenarios and to obtain a reliable estimate of its performance metrics, such as size and power, under the specified signal plus noise model.

By considering these specific models for the noise and signal components, we aim to

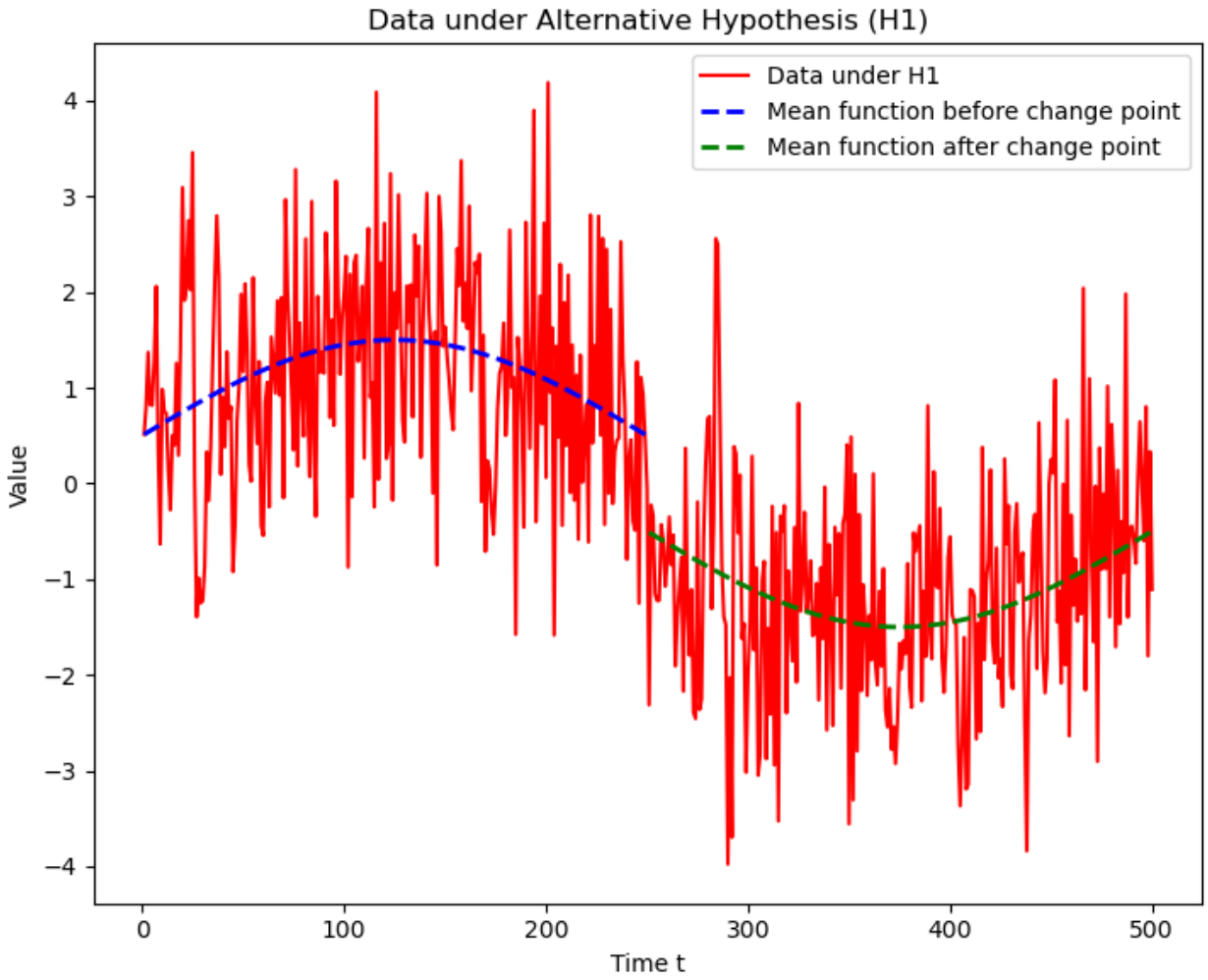

Figure 4.1: Visualization of (X_t) for $n = 500$, marking a change point at $\tau = 250$, with a signal decrease of magnitude $2s = 1.0$, and dependence parameter $\theta = 0.5$.

provide a comprehensive evaluation of the proposed testing procedure in detecting changes in the trend of a non-stationary time series. The causal noise process with time-varying coefficients and the sinusoidal signal functions offer a challenging and realistic setting to assess the effectiveness of the procedure in practical applications.

4.4.2 Synthetic Data under the Null Hypothesis

In this section, we investigate the finite-sample behavior of the testing procedure described in Section 4.2.3 under the null hypothesis. To assess the size of the test, we generate synthetic data according to the model detailed in Section 4.4.1, assuming a continuous signal function as defined in (4.4.19).

We conduct a series of experiments by varying the sample size n and the dependence parameter θ. The sample size n is selected from the set $50, 100, 300, 500, 2000$, covering a wide range of scenarios from small to large sample sizes. This allows us to examine the convergence of the empirical size to the nominal significance level as the sample size increases.

The dependence parameter θ is varied across a range of values, including -0.8, -0.4, 0 (indicating independence), 0.4, and 0.8. By considering both negative and positive values of θ, we can assess the performance of the testing procedure under different levels of temporal dependence in the noise process.

Throughout the experiments, we maintain a constant significance level of $\alpha = 0.05$, which is a commonly used value in hypothesis testing. The empirical sizes, defined as the proportion of simulations in which the null hypothesis is rejected, are computed based on 100,000 simulations for each combination of sample size and dependence parameter. This large number of simulations ensures that the empirical sizes are estimated with a high level of precision.

The results of the experiments are summarized in Table 4.1. The table presents the rejection ratios, which are the empirical sizes expressed as decimal values, for each combination of sample size n and dependence parameter θ.

Analyzing the results in Table 4.1, we observe several important patterns. Firstly, for a fixed value of θ, the rejection ratios tend to approach the nominal significance level of 0.05 as the sample size n increases. This convergence is evident across all values of θ, indicating that the testing procedure maintains the correct size asymptotically, regardless of the level

Table 4.1: Rejection ratios for change point testing procedure under the null hypothesis; cf. (4.2.2).

n	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$
50	0.0182	0.0349	0.0424	0.0399	0.0214
100	0.0261	0.0397	0.0437	0.0430	0.0278
300	0.0352	0.0432	0.0459	0.0447	0.0366
500	0.0374	0.0443	0.0470	0.0461	0.0380
2000	0.0436	0.0473	0.0488	0.0479	0.0445

of temporal dependence in the noise process.

Secondly, for a given sample size n, the rejection ratios are generally closer to the nominal level when the dependence parameter θ is closer to zero. This observation suggests that the testing procedure performs better in terms of size control when the temporal dependence is weaker. However, it is important to note that even for strong levels of dependence (e.g., $\theta = -0.8$ or 0.8), the rejection ratios still converge to the nominal level as the sample size increases.

In conclusion, the simulation results presented in Table 4.1 provide strong evidence that the proposed testing procedure achieves the correct size asymptotically under the null hypothesis, even in the presence of temporal dependence in the noise process. The convergence of the empirical sizes to the nominal significance level as the sample size increases supports the theoretical properties of the procedure. However, the finite-sample performance may be influenced by the level of temporal dependence, emphasizing the importance of considering the dependence structure when applying the procedure in practice. Overall, these findings contribute to the validation of the proposed testing procedure and its robustness to different data generating scenarios under the null hypothesis.

4.4.3 Synthetic Data under Alternative Hypotheses

This section provides an analysis of our testing procedure's power using synthetic data. We adopt the data structure described in Section 4.4.1, with the signal defined as per (4.4.20) (i.e., H_1).

Our experimental setup is as follows: We vary the sample size n, choosing from the values $50, 100, 300, 500, 2000$. This range allows us to examine the power of the test under different sample sizes, from small to large. The dependence parameter θ is selected from the values -0.8, -0.4, 0 (representing independence), 0.4, and 0.8. These values cover a wide range of dependence structures, from independence to strong dependence. The shift parameter s is selected from $\{0, 0.004, 0.01, 0.04, 0.06\}$, representing different magnitudes of the change in the signal.

It is important to note that the standard deviation of the innovation in the dependent process is fixed at 1.0. We standardize the ratios $\tau/n = 0.5$, which means that the change point occurs at the midpoint of the sample. For our testing methods, we consistently set the significance level at $\alpha = 0.05$. Once the parameters for an experiment are established, we generate the trend (μ_i) using the aforementioned methodology. Subsequently, the additive noise process is simulated repeatedly, and this data is input into our testing algorithm. The results are derived from 100,000 independent simulations, ensuring the reliability and robustness of our findings.

Figure 4.2 presents the rejection ratios (power) of our testing procedure under various combinations of sample size n, dependence parameter θ, and shift parameter s. Each subplot corresponds to a specific sample size n, and within each subplot, the rejection ratios are plotted against the shift parameter s for different values of θ.

Analyzing the power plot, we can make several important observations. Firstly, the rejection ratios increase as the shift parameter s increases, indicating that our testing procedure is more powerful in detecting larger changes in the signal. This behavior is consistent across all sample sizes and dependence structures. Secondly, for a fixed shift parameter s, the rejection ratios are generally higher for larger sample sizes n. This observation suggests that our testing procedure benefits from increased sample size, as it gains more power to detect changes in the signal.

Furthermore, the impact of the dependence parameter θ on the power of the test varies depending on the sample size and the magnitude of the shift. For smaller sample sizes (e.g., $n = 50, 100$), the power tends to be lower for strongly dependent cases ($\theta = -0.8, 0.8$) compared to the independent case ($\theta = 0$). However, as the sample size increases (e.g., $n = 300, 500, 2000$), the power becomes more robust to the dependence structure, and the differences between the rejection ratios for different values of θ diminish.

Lastly, for large sample sizes (e.g., $n = 2000$) and moderate to large shift parameters ($s \geq 0.04$), the rejection ratios approach 1, indicating that our testing procedure has excellent power in detecting changes in the signal under these conditions.

In conclusion, the power analysis conducted using synthetic data under the alternative hypothesis demonstrates the effectiveness of our testing procedure in detecting changes in the signal. The power of the test increases with the magnitude of the shift and the sample size, while being robust to different dependence structures, especially for larger sample sizes. These findings provide strong evidence for the practical applicability of our testing procedure in real-world scenarios, where the goal is to identify significant changes in the trend of time series data with non-stationary noises.

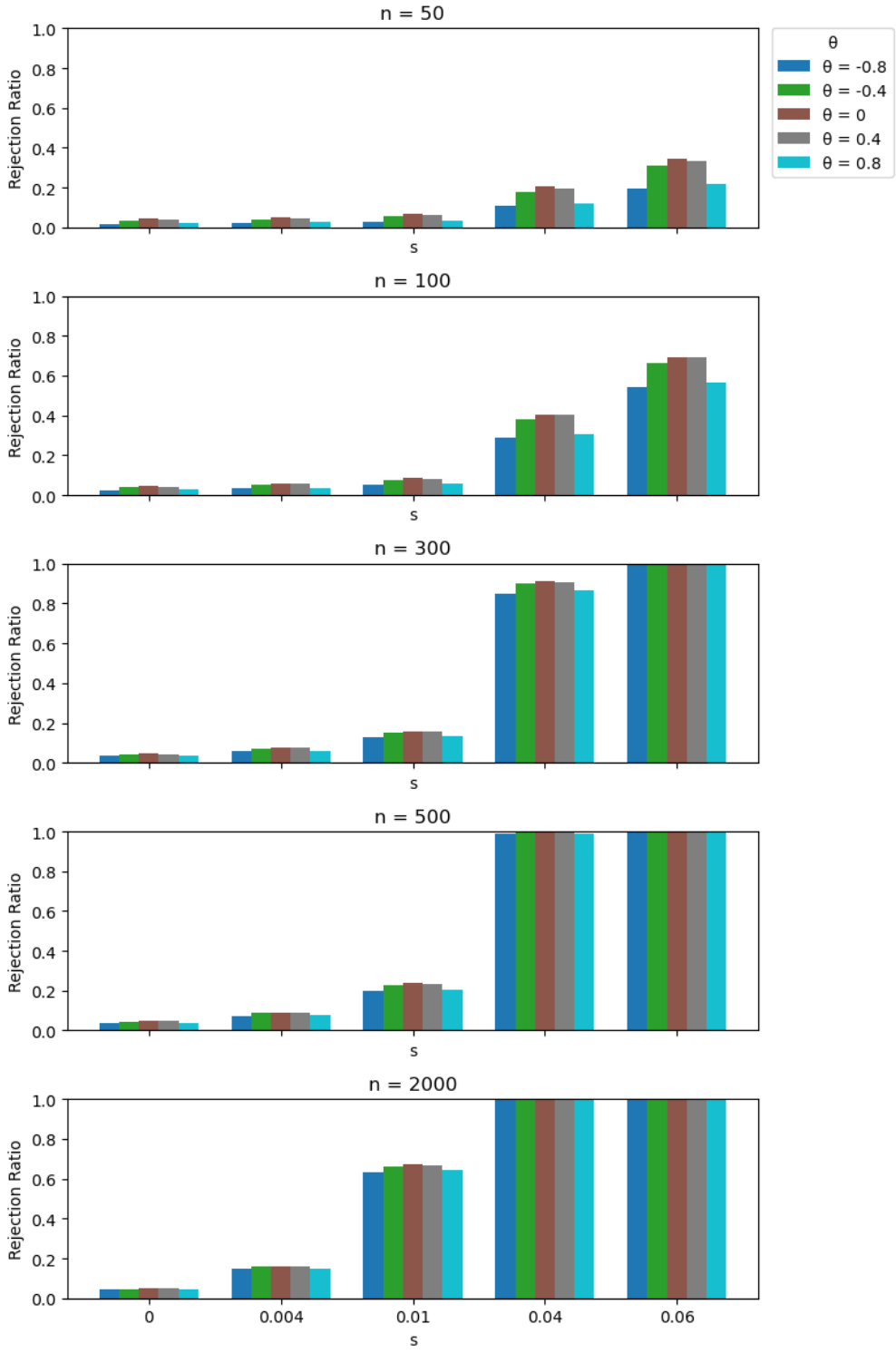

Figure 4.2: Rejection ratios for change point testing procedure under the alternative hypothesis; cf. (4.2.3). The noise process is shaped by the dependence parameter θ. We adjust the gap parameter s over the set $\{0, 0.004, 0.01, 0.04, 0.06\}$, n over $\{50, 100, 300, 500, 2000\}$, and θ over $\{-0.8, -0.4, 0, 0.4, 0.8\}$. Each data point represents 100,000 replications.

4.5 Proofs

4.5.1 Proof for Theorem 4.3.1

With a chosen window size k_n in Algorithm 2, the following equations are derived:

$$\hat{\beta}_{i,r} = \frac{1}{k_n \left(\sum_{h=1}^{k_n-1} h^2\right) - \left(\sum_{h=1}^{k_n-1} h\right)^2} \sum_{j=i}^{i+k_n-1} \left(\sum_{h=1}^{k_n-1} h^2 - (j-i) \sum_{h=1}^{k_n-1} h \right) X_j, \tag{4.5.21}$$

$$\hat{\beta}_{i,l} = \frac{1}{k_n \left(\sum_{h=1}^{k_n-1} h^2\right) - \left(\sum_{h=1}^{k_n-1} h\right)^2} \sum_{j=i-k_n+1}^{i} \left(\sum_{h=1}^{k_n-1} h^2 - (i-j) \sum_{h=1}^{k_n-1} h \right) X_j. \tag{4.5.22}$$

These equations represent the fitted intercepts from the local linear models on the right and left sides of time point i, respectively. The coefficients in these equations are determined by the window size k_n and the position of the observations within the window.

Then, it follows that

$$Y_{n,k_n}(i) = |\hat{\beta}_{i,l} - \hat{\beta}_{i,r}| = \sum_{j=1}^{n} \alpha_{n,k_n}(i,j)(\mu_j + Z_j) = \sum_{j=1}^{n} \alpha_{n,k_n}(i,j)\mu_j + \sum_{j=1}^{n} \alpha_{n,k_n}(i,j) Z_j, \tag{4.5.23}$$

which expresses the test statistic $Y_{n,k_n}(i)$ as a weighted sum of the true signal μ_j and the noise terms Z_j, where the weights are determined by the function $\alpha_{n,k_n}(i,j)$.

Define

$$L_{n,k_n}(i) := |\alpha_{n,k_n}(i,1)| + \sum_{j=2}^{n} |\alpha_{n,k_n}(i,j) - \alpha_{n,k_n}(i,j-1)|, \tag{4.5.24}$$

which represents the total variation of the weight function $\alpha_{n,k_n}(i,j)$ with respect to j for a fixed time point i. It measures the smoothness of the weight function and plays a crucial

role in controlling the approximation error.

By substituting the expression for $\alpha_{n,k_n}(i,j)$ and taking the maximum over i, we obtain

$$L_{n,k_n} := \max_{i \leqslant n} L_{n,k_n}(i) = M \max_i \left\{ |\alpha_{n,k_n}(i,1)| + \sum_{j=2}^{n} |\alpha_{n,k_n}(i,j) - \alpha_{n,k_n}(i,j-1)| \right\} \tag{4.5.25}$$

$$\leq M \frac{2k_n \sum_{h=1}^{k_n-1} h + 2\sum_{h=1}^{k_n-1} h^2}{k_n \left(\sum_{h=1}^{k_n-1} h^2\right) - \left(\sum_{h=1}^{k_n-1} h\right)^2} = \mathrm{O}\left(\frac{1}{k_n}\right). \tag{4.5.26}$$

This inequality provides an upper bound for the maximum total variation of the weight function $\alpha_{n,k_n}(i,j)$ over all time points i. The bound is of the order $\mathrm{O}\left(\frac{1}{k_n}\right)$, which suggests that the smoothness of the weight function improves as the window size k_n increases.

Applying **Corollary 3.1.** in Bonnerjee et al. [2024] under Conditions 4.3.1 and 4.3.2, we get

$$\max_{i \leqslant n} \left| \sum_{j=1}^{i} Z_j - \mathbb{B}(\mathcal{T}_i) \right| = \mathcal{O}_{\mathbb{P}}\left(n^{r_1}\sqrt{\log n}\right), \tag{4.5.27}$$

where

$$r_1 = \max\left\{ \frac{1+A}{2(1+2A)}, \frac{1+4A/p}{2(1+2A)} \right\}. \tag{4.5.28}$$

This result establishes the approximation of the partial sum process of the noise terms Z_j by a Brownian motion $\mathbb{B}(\mathcal{T}_i)$ evaluated at the time points $\mathcal{T}_i$. The rate of approximation is given by $n^{r_1}\sqrt{\log n}$, where r_1 depends on the parameters A and p from Conditions 4.3.1 and 4.3.2.

Then, using the triangle inequality, we have

$$|Y_{n,k_n}(i) - \hat{Y}_{n,k_n}(i)| \leq \left| \sum_{j=1}^{n} \alpha_{n,k_n}(i,j)\mu_j \right| + \left| \sum_{j=1}^{n} \alpha_{n,k_n}(i,j) \left(Z_j - \left\{ \mathbb{B}(\mathcal{T}_j) - \mathbb{B}(\mathcal{T}_{j-1}) \right\} \right) \right|$$

$$\leq (I) + (II) \tag{4.5.29}$$

where

$$(I) = \left| \sum_{j=1}^{n} \alpha_{n,k_n}(i,j)\mu_j \right| \leq O\left(\frac{k_n^2}{n} \right) \tag{4.5.30}$$

uniformly over $i = k_n, k_n + 1, \cdots, n - k_n$, because $\mu \in L[0,1]$, and

$$\begin{aligned} (II) &= \left| \sum_{j=1}^{n} \alpha_{n,k_n}(i,j) \left(Z_j - \left\{ \mathbb{B}(\mathcal{T}_j) - \mathbb{B}(\mathcal{T}_{j-1}) \right\} \right) \right| \\ &\leq L_{n,k_n}(i) \max_{i \leq n} \left| \sum_{j=1}^{i} Z_j - \mathbb{B}(\mathcal{T}_i) \right| = \mathcal{O}_{\mathbb{P}}\left(\frac{n^{r_1}\sqrt{\log n}}{k_n} \right). \end{aligned} \tag{4.5.31}$$

The terms (I) and (II) represent the approximation errors due to the signal and noise components, respectively. The term (I) is uniformly bounded by $O\left(\frac{k_n^2}{n}\right)$ because the signal μ is assumed to be in the function space $L[0,1]$. The term (II) is controlled by the product of the total variation of the weight function $L_{n,k_n}(i)$ and the maximum approximation error between the partial sum process of the noise terms and the Brownian motion. The rate of convergence for (II) is $\mathcal{O}_{\mathbb{P}}\left(\frac{n^{r_1}\sqrt{\log n}}{k_n}\right)$.

To balance (I) and (II), we take $k_n = n^{(1+r_1)/3}(\log n)^{1/6}$. It follows that

$$\max_{1\leq i\leq n} |Y_{n,k_n}(i) - \hat{Y}_{n,k_n}(i)| = \mathcal{O}_{\mathbb{P}}\left(n^{\frac{2r_1-1}{3}}(\log n)^{1/3}\right). \tag{4.5.32}$$

This choice of the window size k_n ensures that the approximation errors from the signal and noise components are of the same order, leading to the optimal rate of convergence $n^{\frac{2r_1-1}{3}}(\log n)^{1/3}$ for the maximum difference between the test statistic $Y_{n,k_n}(i)$ and its Bootstrap approximation $\hat{Y}_{n,k_n}(i)$. □

CHAPTER 5
MULTIVARIATE CHANGE POINT ANALYSIS UNDER NON-STATIONARY NOISE

5.1 Introduction

In Chapter 4, we introduced a non-parametric framework for testing change points in the mean function of time series with temporally dependent and non-stationary noise. This chapter aims to extend this framework to inference for change points in signals from data with multivariate non-stationary noise. The motivation for this extension stems from the increasing prevalence of applications dealing with multivariate or high-dimensional time series with non-stationary noise. Recent works have addressed change point detection for such data in various areas. Heo and Manuel [2022] develop the Greedy Copula Segmentation (GCS) algorithm for processing multivariate non-stationary climate data, demonstrating its potential for climate change adaptation and disaster risk reduction. Banerjee and Guhathakurta [2020] utilize sequential change-point detection in dynamic networks to analyze multivariate and non-stationary data from 13 global stock markets, detecting pre-crash changes and showcasing the method's potential as an early warning system for critical financial market regimes. Kirch et al. [2015] apply change point detection techniques to identify complex brain processes in multivariate non-stationary electroencephalograms during cognitive motor tasks. Schröder and Ombao [2018] focus on analyzing multivariate and non-stationary electroencephalogram data, identifying early disruptions and the progression of epileptic seizures across different brain regions.

As mentioned in Section 4.1, there is an extensive body of literature focusing on analyzing non-stationary time series data under parametric settings. However, literature addressing multivariate non-stationary time series data from a non-parametric perspective is relatively sparse. This scarcity is due to the complex nature of such data, with multivariate non-

stationary noise making the development of non-parametric statistical inference methods with valid theoretical foundations and good empirical performance extremely challenging. Sundararajan and Pourahmadi [2018] develop a non-parametric method for detecting multiple change points in multivariate non-stationary processes using a test statistic that measures differences in spectral density matrices. Park et al. [2023] introduce a two-step procedure that detects and estimates structural changes in multivariate non-stationary time series by initially assessing mean changes via a cumulative sum statistic and, if absent, examining second-order changes using a multivariate locally stationary wavelet process. Aue et al. [2009] introduce a non-parametric asymptotic test procedure for assessing the stability of volatilities and cross-volatilities in multivariate non-stationary time series models. Preuss et al. [2015] introduce MuBreD, a non-parametric method for detecting and estimating multiple structural breaks in the auto-covariance function of multivariate non-stationary processes.

Similar to our framework for uni-variate inference of change points in signals with non-stationary noise in Chapter 4, the multivariate inference methodology in this chapter allows for a time-varying signal after the change point, unlike the aforementioned literature that usually assumes a constant level after the change point. This setting significantly increases the complexity of our task, as the volatility of the data can originate from both the signal and multivariate noise components. Consequently, developing a reliable testing procedure becomes highly challenging, as it is difficult to obtain a good approximation for the distribution of the test statistic under the null hypothesis. We endeavor to tackle this issue by employing a bootstrap procedure proposed by Mies and Steland [2023]. However, as demonstrated in our Monte Carlo studies in Section 5.4.2, although our testing procedure exhibits good type-one error control when the sample size is sufficiently large (≥ 2000) and the dimensionality is relatively low (≤ 5), we encounter over-liberal issues when the sample size is small and the dimensionality is high. This can be attributed to the suboptimal finite-sample performance of the bootstrap procedure utilized in our algorithm.

Given that this work constitutes the initial endeavor to address this problem, we anticipate that future research efforts may yield further enhancements, potentially improving the performance of the testing procedure for small sample cases. One promising direction is the development of a bootstrap procedure that more effectively captures the intricate temporal dependence and non-stationarity in a multivariate setting. For instance, Bonnerjee et al. [2024] achieved a significant breakthrough in the uni-variate setting, and we posit that this work can be extended to a multivariate or even high-dimensional context, potentially leading to substantial improvements in the performance of our testing procedure for smaller sample sizes and higher dimensions.

The remainder of this chapter is structured as follows: Section 5.2 presents our methodology for testing change points in signals with multivariate non-stationary noise processes. In Section 5.3, we establish the theoretical foundations of our approach, demonstrating that the distribution of our test statistic can be approximated via the introduced bootstrap mechanism under the null hypothesis. The empirical efficacy of our method is assessed through simulations in Section 5.4. Comprehensive proofs of the theoretical results are provided in Section 5.5.

5.2 Methodology

5.2.1 Data Model and Problem Formulation

We consider a d-dimensional sequential noisy model given by

$$\mathbf{X}_t = \boldsymbol{\nu}(t/n) + \mathbf{Z}_t, \quad t = 1, 2, \ldots, n, \tag{5.2.1}$$

where $\boldsymbol{\nu}(\cdot)$ denotes the mean function defined on $\mathbb{R}^d$, and the sequence $(\mathbf{Z}_t)_t$ represents a d-dimensional non-stationary process with a mean of $\mathbf{0}$. This model provides a flexible framework for analyzing multivariate time series data with potential non-stationarity in both the mean function and the noise component.

The primary objective is to develop a testing procedure to discern between the following hypotheses:

- **Null Hypothesis** (H_0):

$$H_0 : \ \boldsymbol{\nu}(\cdot) \text{ is continuous across } [0, 1], \tag{5.2.2}$$

- **Alternative Hypothesis** (H_1):

$$H_1 : \ \text{There exists } t_0 \in [0, 1] \text{ at which } \boldsymbol{\nu}(\cdot) \text{ is discontinuous.} \tag{5.2.3}$$

The null hypothesis H_0 assumes that the mean function $\boldsymbol{\nu}(\cdot)$ is continuous throughout the interval $[0, 1]$, while the alternative hypothesis H_1 posits the existence of a change point t_0 at which $\boldsymbol{\nu}(\cdot)$ exhibits a discontinuity.

The test statistic we employ to distinguish between these hypotheses is:

$$S_{n,k}(\mathbf{X}_1, \ldots, \mathbf{X}_n) = \max_{i=k,\ldots,n-k} \left\| \sum_{j=i-k+1}^{i} \mathbf{X}_j - \sum_{j=i+1}^{i+k} \mathbf{X}_j \right\|, \tag{5.2.4}$$

where k is a window size parameter. This test statistic compares the sum of observations within two adjacent windows of size k and takes the maximum distance across all possible window locations. The intuition behind this test statistic is that if there exists a change point in the mean function, the distance between the sum of observations in the two adjacent windows will be large in the vicinity of the change point. By considering all possible window locations and taking the maximum distance, the test statistic aims to capture the most significant change in the mean function.

The choice of the window size k is crucial for the performance of the test statistic. A smaller window size allows for the detection of more localized changes but may be more sensitive to noise, while a larger window size provides a more stable estimate of the mean function but may miss smaller or more abrupt changes. In practice, the window size can be chosen based on prior knowledge about the expected scale of the changes or through the theoretical results provided in Theorem 5.3.1.

The null hypothesis H_0 is rejected for large values of $S_{n,k}$, indicating the presence of a significant break in the mean function. To calculate the p-values for the test, we need to approximate the distribution of $S_{n,k}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ under the null hypothesis H_0.

Unlike Chapter 3, where stationarity is assumed for the multivariate noise process, we are unable to derive an asymptotic distribution for $S_{n,k}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, such as what we did in Section 3.3.2, due to the non-parametric and non-stationary setting for the noise process. To address this challenge, in Section 5.2.2, we introduce Algorithm 3, a bootstrap procedure that works for multivariate temporally dependent non-stationary sequences. This procedure enables us to approximate the distribution of $S_{n,k}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ under H_0.

Building upon Algorithm 3, we propose our complete testing procedure for change points, as outlined in Algorithm 4 in Section 5.2.3. This procedure combines the multivariate non-stationary bootstrap with the maximum-type test statistic $S_{n,k}$ to detect change points in the mean function of a multivariate time series with non-stationary noise.

5.2.2 Bootstrap for Multivariate Non-Stationary Processes

Algorithm 3 introduces a novel bootstrap procedure specifically designed for multivariate non-stationary time series data.

In the multivariate setting, this new bootstrap aims to capture not only the local dependence within each block and the global non-stationarity across the entire time series but also the cross-sectional dependence between the different dimensions. This is achieved by introducing a novel resampling scheme that accounts for the covariance structure of the multivariate time series.

Unlike Algorithm 1, which is designed for uni-variate non-stationary time series, Algorithm 3 operates on the multivariate time series $\{\mathbf{X}_i\}_{i=1}^n$. It first computes the block sums $\mathbf{B}_t = \sum_{i=t-m+1}^{t} \mathbf{X}_i$ for each block ending at time point $t = m, m+1, \ldots, n$. These block sums capture the local level of the multivariate time series within each block.

Next, the algorithm computes the outer product of each block sum with itself, scaled by the block size m, to obtain the matrices $\hat{\mathbf{C}}_t = \mathbf{B}_t \mathbf{B}_t^\top / m$. These matrices represent the local covariance structure of the multivariate time series within each block. The Cholesky decomposition of each $\hat{\mathbf{C}}_t$ matrix is then computed to obtain the matrices $\hat{\mathbf{D}}_t = (\hat{\mathbf{C}}_t)^{1/2}$. These matrices capture the cross-sectional dependence between the different dimensions of the time series within each block.

The bootstrap samples are constructed by generating b independent standard multivariate normal random vectors $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$, and then setting $\mathbf{Y}_{s,t} = \hat{\mathbf{D}}_t \mathbf{Z}_t$ for $t = m, m+1, \ldots, n$. This construction ensures that the bootstrap samples have the same covariance structure as

the original multivariate time series within each block.

The resulting bootstrap samples $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_b$ can be used to estimate the sampling distribution of various statistics of interest in the multivariate setting, such as the mean vector, covariance matrix, or test statistics for change point detection. By resampling blocks of the original data in a way that preserves the local temporal dependence structure and cross-sectional dependence, Algorithm 3 provides a robust and flexible approach to inference in settings where the data exhibits complex non-stationarity structures across space and time.

Algorithm 3 Bootstrap for multivariate Non-Stationary Process

Input: Observed data $\{\mathbf{X}_i\}_{i=1}^n$; b, the number of Bootstrap samples; m, the block size.

for all $t = m, m+1, \ldots, n$ **do**
  Compute $\mathbf{B}_t \leftarrow \sum_{i=t-m+1}^{t} \mathbf{X}_i$
end for
for all $t = m, m+1, \ldots, n$ **do**
  Compute $\hat{\mathbf{C}}_t \leftarrow \frac{1}{m}\mathbf{B}_t\mathbf{B}_t^\top$
  Compute $\hat{\mathbf{D}}_t \leftarrow (\hat{\mathbf{C}}_t)^{1/2}$, where $(\cdot)^{1/2}$ denotes the Cholesky decomposition.
end for
for all $s = 1, 2, \ldots, b$ **do**
  Generate $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$ i.i.d. $N(\mathbf{0}, I_d)$.
  Compute $\mathbf{Y}_{s,t} \leftarrow \hat{\mathbf{D}}_t\mathbf{Z}_t$, for $t = m, m+1, \ldots, n$.
  Store $\mathbf{Y}_s \leftarrow (\mathbf{Y}_{s,m}, \mathbf{Y}_{s,m+1}, \ldots, \mathbf{Y}_{s,n})$.
end for

Output: Output the tensor $(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_b)$.

5.2.3 Testing for Change Point of Mean Function for multivariate Data with Non-Stationary Noise

In this section, we present a comprehensive testing procedure for detecting a change point in the mean function of a multivariate time series with non-stationary noise. The proposed methodology, outlined in Algorithm 4, combines the multivariate non-stationary bootstrap

introduced in Section 5.2.2 with the maximum-type test statistic $S_{n,k}$ defined in Section 5.2.1.

The testing procedure takes as input the sequential data $(\mathbf{X}_i)_{i=1}^n$, the window size k, the number of bootstrap samples b, and the block size m used in the bootstrap procedure. The first step is to compute the test statistic $T = S_{n,k}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ using the observed data. This test statistic captures the maximum distance between the sum of observations in two adjacent windows of size k across all possible window locations, as discussed in Section 5.2.1.

To approximate the distribution of the test statistic under the null hypothesis H_0, we employ a residual-based bootstrap approach. First, we estimate the mean function $\{\hat{\boldsymbol{\nu}}_i\}_{i=1}^n$ from the observed data using an appropriate non-parametric regression technique, such as local polynomial regression or spline regression. Then, we compute the residuals $\hat{\mathbf{r}}_i = \mathbf{X}_i - \hat{\boldsymbol{\nu}}_i$ for $i = 1, 2, \ldots, n$, which represent the estimated noise component of the time series.

Next, we apply the multivariate bootstrap procedure (Algorithm 3) to the residuals $\{\hat{\mathbf{r}}_i\}_{i=1}^n$ to generate b bootstrap samples $\mathbf{W}_s = (\mathbf{W}_{s,1}, \mathbf{W}_{s,2}, \ldots, \mathbf{W}_{s,n})$ for $s = 1, 2, \ldots, b$. This procedure takes into account the complex dependence structure and non-stationarity present in the noise process, as discussed in Section 5.2.2.

For each bootstrap sample $\mathbf{W}_s$, we compute the corresponding test statistic $D_s = S_{n,k}(\mathbf{W}_{s,1}, \mathbf{W}_{s,2}, \ldots, \mathbf{W}_{s,n})$. This step provides an approximation of the distribution of the test statistic under the null hypothesis, as the bootstrap samples are generated under the assumption of no change point in the mean function.

Finally, we estimate the p-value of the test as $\hat{p} = \frac{1}{b+1}(1 + \sum_{s=1}^b \mathbf{1}_{\{T > D_s\}})$, which represents the proportion of bootstrap test statistics that exceed the observed test statistic. The null hypothesis H_0 is rejected if $\hat{p} < \alpha$, where α is the desired significance level.

The choice of the window size k in the test statistic $S_{n,k}$ and the block size m used in the bootstrap procedure is crucial for the performance of the test. Theorem 5.3.1 provides theoretical guidance on the optimal choice of k and m to ensure the validity of the test.

Algorithm 4 Testing for Change Point of Covariance

Input: Sequential data $(\mathbf{X}_i)_{i=1}^n$; Window size k; Number of Bootstrap samples b; Block size in Bootstrap m.

Compute test statistic $T \leftarrow S_{n,k}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$.
Compute $\{\hat{\boldsymbol{\nu}}_i\}_{i=1}^n$, the fitted signals from the observed data $(\mathbf{X}_i)_{i=1}^n$.
for all $i = 1, 2, \ldots, n$ **do**
    Compute residuals $\hat{\mathbf{r}}_i \leftarrow \mathbf{X}_i - \hat{\boldsymbol{\nu}}_i$
end for
for $s = 1, 2, \ldots, b$ **do**
    Apply Algorithm 3 with inputs $\{\hat{\mathbf{r}}_i\}_{i=1}^n$ and m to generate the Bootstrap sample $\mathbf{W}_s = (\mathbf{W}_{s,1}, \mathbf{W}_{s,2}, \ldots, \mathbf{W}_{s,n})$.
    Compute $D_s \leftarrow S_{n,k}(\mathbf{W}_{s,1}, \mathbf{W}_{s,2}, \ldots, \mathbf{W}_{s,n})$
end for
Calculate $\hat{p} \leftarrow \frac{1}{b+1}(1 + \sum_{s=1}^b \mathbf{1}\{T > D_s\})$

Output: Output estimated p-value $\hat{p}$. Make a decision to reject H_0 if $\hat{p} < \alpha$.

5.3 Theory

5.3.1 Assumptions on the Multivariate Non-stationary Noise Processes

In this section, we extend the framework for modeling non-stationary noise processes introduced in Section 4.3.1 to the multivariate setting. We establish the necessary assumptions for our change point analysis methodology in the context of multivariate non-stationary processes. While the overall approach is similar to the one-dimensional case, there are key differences that arise due to the increased complexity and interdependence of multivariate data.

We consider the d-dimensional non-stationary process $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ as the output of a physical system described by:

$$\mathbf{Z}_t := \mathbf{G}_t(\cdots, \varepsilon_{t-1}, \varepsilon_t), \tag{5.3.5}$$

where $\mathbf{G}_t$, $t = 1, 2, \cdots, n$, are measurable mappings in $\mathbb{R}^d$. This representation extends the one-dimensional model in (4.3.4) to the multivariate setting, allowing for a flexible and general description of non-stationary processes, where the functional form of the process can vary over time and across dimensions.

To quantify the dependence structure of the multivariate non-stationary process, we introduce a functional dependence measure that takes into account the interdependence among the dimensions. Given a time lag $k \in \mathbb{Z}$ and a norm parameter r, we define:

$$\delta_{p,r}(i,k) := \left(\mathbb{E}\|\mathbf{Z}_i - \mathbf{Z}_{i,\{i-k\}}\|_r^p\right)^{1/p}, \text{ where } \mathbf{Z}_{i,\{i-k\}} = \mathbf{G}_i(\cdots, \varepsilon_{i-k-1}, \varepsilon'_{i-k}, \varepsilon_{i-k+1}, \cdots, \varepsilon_i) \tag{5.3.6}$$

This measure extends the uniform functional dependence measure in (4.3.5) to the multivariate setting by considering the L_r norm of the difference between the process and its coupled version. The parameters p and r control the moment and norm used in the defini-

tion, respectively. This additional flexibility allows for a more comprehensive assessment of the dependence structure in multivariate processes.

To express our dependence condition, we introduce the cumulative dependence measure:

$$\Theta_{j,p,r} = \sum_{k=j}^{\infty} \sup_i \delta_{p,r}(i,j), \quad j \geq 0, \tag{5.3.7}$$

which aggregates the functional dependence measures across all time points and dimensions, providing a comprehensive assessment of the overall dependence structure of the multivariate non-stationary process. The rate at which $\Theta_{j,p,r}$ decays as j increases will play a crucial role in determining the theoretical properties of our change point analysis methodology in the multivariate setting.

A key difference between the multivariate and one-dimensional settings is the presence of cross-sectional dependence among the dimensions. To capture this dependence, we introduce the concept of local long-run covariance matrices $(\Sigma_t)_{t\in\mathbb{Z}}$ for the d-dimensional non-stationary process $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$:

$$\Sigma_t := \sum_{h=-\infty}^{\infty} \operatorname{Cov}\left(\mathbf{G}_t(\cdots, \varepsilon_{-1}, \varepsilon_0), \mathbf{G}_t(\cdots, \varepsilon_{h-1}, \varepsilon_h)\right) \tag{5.3.8}$$

These local long-run covariance matrices extend the concept of long-run covariance from stationary processes to non-stationary processes, capturing the covariance structure of the multivariate process at each time point t. The existence and well-definedness of these matrices rely on the short-range dependence condition:

$$\Theta_{0,p,2} < \infty. \tag{5.3.9}$$

This condition ensures that the cumulative dependence measure decays sufficiently fast, allowing for the convergence of the series in (5.3.8).

By modeling multivariate non-stationary noise processes through the lens of a physi-

cal system with time-varying functional mechanisms, quantifying the dependence structure using the functional dependence measure, and introducing the concept of local long-run covariance matrices, we establish a comprehensive framework for analyzing change points in the presence of multivariate non-stationarity. The assumptions outlined in this section provide the foundation for the theoretical developments presented in the subsequent sections of this chapter.

5.3.2 Main Theorem

Before presenting the main theorem, we introduce two essential conditions that are necessary for the theorem to hold.

Condition 5.3.1. *Consider a d-dimensional non-stationary process $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ where each $\mathbf{Z}_i$ is expressed as in* (5.3.5)*. Assume that the p-norm of $\|\mathbf{Z}_i\|_2$, $(\mathbb{E}\|\mathbf{Z}_i\|_2^p)^{1/p}$, is uniformly bounded by a finite value η_n for all i within $1 \le i \le n$ and for some $p > 2$. Further, suppose there exist constants $A > 2$ such that for all $1 \le i \le n$, the following inequality holds:*

$$\delta_{p,2}(i,j) \le \eta_n \cdot (j \vee 1)^{-A}, \quad j \ge 0. \tag{5.3.10}$$

This condition imposes a moment bound on the non-stationary process $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ and requires the functional dependence measure $\delta_{p,2}(i,j)$ to decay at a polynomial rate with respect to the time lag j. The parameter A controls the rate of decay, with larger values of A indicating weaker dependence. The uniform bound η_n allows for the dependence structure to vary with the sample size n.

To perform asymptotic inference in our non-stationary setting, we further need to impose the following regularity condition on our physical system $(\mathbf{G}_t(\cdot))$:

Condition 5.3.2. *Assume for some $D_n \ge 1$,*

$$\sum_{t=2}^{n}\left(\mathbb{E}\left\|\mathbf{G}_t\left(\cdots,\varepsilon_{-1},\varepsilon_0\right)-\mathbf{G}_{t-1}\left(\cdots,\varepsilon_{-1},\varepsilon_0\right)\right\|_2^2\right)^{\frac{1}{2}}\leq\eta_n\cdot D_n. \tag{5.3.11}$$

This condition requires the physical system $(\mathbf{G}_t(\cdot))$ to exhibit a certain level of stability over time. It controls the cumulative differences between consecutive mappings $\mathbf{G}_t$ and $\mathbf{G}_{t-1}$, ensuring that the non-stationarity in the process is not too drastic. The parameter D_n quantifies the degree of non-stationarity, with larger values of D_n indicating more pronounced variations in the physical system over time.

Now let me present our main theorem. Recall the test statistic

$$S_{n,k}(\mathbf{X}_1,\ldots,\mathbf{X}_n)=\max_{i=k,\ldots,n-k}\left\|\sum_{j=i-k+1}^{i}\mathbf{X}_j-\sum_{j=i+1}^{i+k}\mathbf{X}_j\right\|$$

from Section 5.2.1, where $(\mathbf{X}_i)_{i=1}^n$ is the input data in Algorithm 4. Define

$$\mathbf{U}_{n,k_n}(i):=\frac{1}{k_n}\sum_{j=i-k_n+1}^{i}\mathbf{X}_j-\frac{1}{k_n}\sum_{j=i+1}^{i+k_n}\mathbf{X}_j, \tag{5.3.12}$$

which represents the difference between the average of observations in two adjacent windows of size k_n. Then, we have $S_{n,k_n}(\mathbf{X}_1,\ldots,\mathbf{X}_n)=k_n\max_{i=k_n,\ldots,n-k_n}\|\mathbf{U}_{n,k_n}(i)\|$.

Similarly, let

$$\hat{\mathbf{U}}_{n,k_n}(i):=\frac{1}{k_n}\sum_{j=i-k_n+1}^{i}\mathbf{W}_j-\frac{1}{k_n}\sum_{j=i+1}^{i+k_n}\mathbf{W}_j, \tag{5.3.13}$$

where $(\mathbf{W}_1,\cdots,\mathbf{W}_n)$ is a bootstrap sample generated by Algorithm 3 in the testing procedure described in Algorithm 4.

Now we present our main theorem:

Theorem 5.3.1. *Let $(\mathbf{X}_i)_{i=1}^n$ be the input data of Algorithm 4, following the form given in (5.2.1). We assume that the signal component $\boldsymbol{\nu}(\cdot)$ belongs to the function space $L[0,1]$. For the noise component $(\mathbf{Z}_t)_{t=1}^n$, we assume that Conditions 5.3.1 and 5.3.2 hold for some*

$p > 2$, with parameters $\eta_n > 0$, $A > 2$, and $D_n \geq 1$. Let the window size k_n in Algorithm 4 be chosen as

$$k_n = n^{\frac{1}{2}}(\log n)^{\frac{1}{2}}\eta_n\Bigg(D_n^{\frac{A-2}{2(A-1)}}\left(\frac{d}{n}\right)^{\lambda(p,A)} + D_n^{\frac{1}{4}}n^{\frac{1}{4}}m_n^{\frac{1}{8}} + n^{\frac{3}{8}}d^{\frac{1}{8}}m_n^{\frac{1}{8}} + n^{\frac{1}{2}}m_n^{-\frac{1}{4}} + n^{\frac{1}{2}}m_n^{\frac{1}{2}-\frac{A}{4}} + 1\Bigg), \tag{5.3.14}$$

where $m_n \asymp n^{\zeta}, \zeta \in (0,1)$ is the block size chosen in Algorithm 3, and

$$\lambda(p,A) = \begin{cases} \frac{p-2}{6p-4}, & A \geq 3, \\ \frac{(A-2)(p-2)}{(4A-6)p-4}, & \frac{3p+2}{p+2} < A < 3, \\ \frac{1}{2} - \frac{1}{A}, & 2 < A \leq \frac{3p+2}{p+2} \end{cases} \tag{5.3.15}$$

Then, the following asymptotic equivalence holds:

$$\max_{i\leq n}\|\mathbf{U}_{n,k_n}(i) - \hat{\mathbf{U}}_{n,k_n}(i)\| = \mathcal{O}_{\mathbb{P}}\Bigg(n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\eta_n\Bigg(D_n^{\frac{A-2}{2(A-1)}}\left(\frac{d}{n}\right)^{\lambda(p,A)} + D_n^{\frac{1}{4}}n^{\frac{1}{4}}m_n^{\frac{1}{8}} + n^{\frac{3}{8}}d^{\frac{1}{8}}m_n^{\frac{1}{8}} + n^{\frac{1}{2}}m_n^{-\frac{1}{4}} + n^{\frac{1}{2}}m_n^{\frac{1}{2}-\frac{A}{4}} + 1\Bigg)\Bigg). \tag{5.3.16}$$

This theorem establishes the asymptotic equivalence between the difference of averages $\mathbf{U}_{n,k_n}(i)$ and its bootstrap counterpart $\hat{\mathbf{U}}_{n,k_n}(i)$ under the null hypothesis H_0. By the relation $S_{n,k_n}(\mathbf{X}_1,\ldots,\mathbf{X}_n) = k_n \max_{i=k_n,\ldots,n-k_n}\|\mathbf{U}_{n,k_n}(i)\|$, this asymptotic equivalence result justifies the validity of the bootstrap procedure in Algorithm 4 for approximating the distribution of the test statistic $S_{n,k_n}(\mathbf{X}_1,\ldots,\mathbf{X}_n)$ under H_0.

The rate of convergence in (5.3.16) depends on several factors, including the sample size n, the dimension d, the moment order p, the dependence parameter A, the non-stationarity parameter D_n, and the block size m_n. The choice of the window size k_n is crucial for

achieving the optimal rate of convergence, and the theorem provides a specific choice of k_n that balances the various terms in the convergence rate.

The block size m_n is chosen to be of the order n^{ζ}, where $\zeta \in (0, 1)$. The choice of ζ affects the convergence rate through the terms involving m_n. In practice, ζ can be chosen based on the trade-off between capturing the dependence structure and the computational complexity of the bootstrap procedure in Algorithm 3.

This theorem provides theoretical justification for the validity of the bootstrap-based testing procedure in Algorithm 4 for detecting change points in the mean function of multivariate non-stationary time series. It also offers guidance on the choice of the window size k_n and the block size m_n to achieve the optimal convergence rate.

The proof of this theorem is deferred to Section 5.5.1.

5.4 Monte Carlo Studies

5.4.1 Models Considered

To comprehensively examine the finite sample performance of the testing procedure introduced in Section 5.2.3, we conduct an extensive set of Monte Carlo experiments. The data for these experiments are generated based on the signal plus noise model described in (5.2.1). This model provides a flexible framework for assessing the capability of the proposed testing procedure in detecting changes in the trend of a multivariate time series with non-stationary noise components.

The noise component is modeled as a multivariate causal non-stationary process, constructed as follows:

Define $w_{1,t} = 0.5 \times (-1)^{\lfloor 2t/n \rfloor}, t = 1, \ldots, n$, and let $\mathbf{Z}_t = \theta_t \mathbf{Z}_{t-1} + \boldsymbol{\varepsilon}_t$, with $\theta_t = \theta w_{1,t}$ and $\mathbf{Z}_0 = 0$. The innovations $\boldsymbol{\varepsilon}_t$ are drawn independently from a d-dimensional Gaussian distribution with mean $\mathbf{0}$ and a covariance matrix with diagonal elements equal to 1 and off-diagonal elements equal to 0.5.

The sequence $w_{1,t}$ imposes a periodic structure on the temporal dependence of the noise process, oscillating between 0.5 and -0.5 for each half of the time series. The strength of the temporal dependence is controlled by the parameter θ, with higher absolute values of θ signifying stronger dependence. The innovations $\boldsymbol{\varepsilon}_t$ are assumed to be independent and identically distributed random vectors following a d-dimensional Gaussian distribution with a specific covariance structure. The diagonal elements of the covariance matrix are set to 1, indicating unit variance for each dimension, while the off-diagonal elements are set to 0.5, introducing a moderate level of spatial correlation among the dimensions.

For the multivariate signal component $\boldsymbol{\nu}_t$, we consider two distinct scenarios:

The first scenario corresponds to the null hypothesis H_0, as stated in (5.2.2). Under this

hypothesis, the signal function is continuous and defined as:

$$\boldsymbol{\nu}_{t,j} = \sin\left(\frac{2(-1)^j \pi t}{n}\right) \quad \text{for } t = 1, 2, \ldots, n, \quad j = 1, \cdots, d \tag{5.4.17}$$

This specification of the signal function represents a smooth, non-linear trend in each dimension of the time series. The sinusoidal form captures a periodic pattern with a single cycle spanning the entire time range. The scaling factor $\frac{2\pi}{n}$ ensures that the signal's period matches the sample size n. The $(-1)^j$ term introduces alternating phases for odd and even dimensions, generating a more diverse and realistic multivariate signal.

The second scenario corresponds to the alternative hypothesis H_1, as described in (5.2.3). Under this hypothesis, the signal is generated according to:

$$\boldsymbol{\nu}_{t,j} = \begin{cases} \sin\left(\frac{2(-1)^j \pi t}{n}\right) + s & \text{for } t = 1, 2, \ldots, \tau - 1, \qquad j = 1, \cdots, d, \\ \sin\left(\frac{2(-1)^j \pi t}{n}\right) - s & \text{for } t = \tau, \tau + 1, \ldots, n, , \qquad j = 1, \cdots, d. \end{cases} \tag{5.4.18}$$

Here, the parameter s determines the magnitude of the discontinuity in the signal function. This formulation introduces a change point at time τ, where the trend of the time series undergoes a sudden shift. The severity of the shift is governed by the parameter s, with larger values of s indicating more substantial changes in the trend. The change point occurs simultaneously across all dimensions, reflecting a global structural change in the multivariate signal.

Figure 5.1 presents an illustrative example of a three-dimensional data series ($\mathbf{X}_i$) with sample size $n = 1000$ under the alternative hypothesis H_1, exhibiting a prominent discontinuity at $\tau = 500$. The signal function $\boldsymbol{\nu}_t$ undergoes an abrupt transition at the change point, with a magnitude of $s = 0.25$. The noise process is characterized by non-stationarity and temporal dependence, with the dependence parameter set to $\theta = 0.4$. Due to the presence of the multivariate non-stationary temporally dependent noise, visually detecting the change

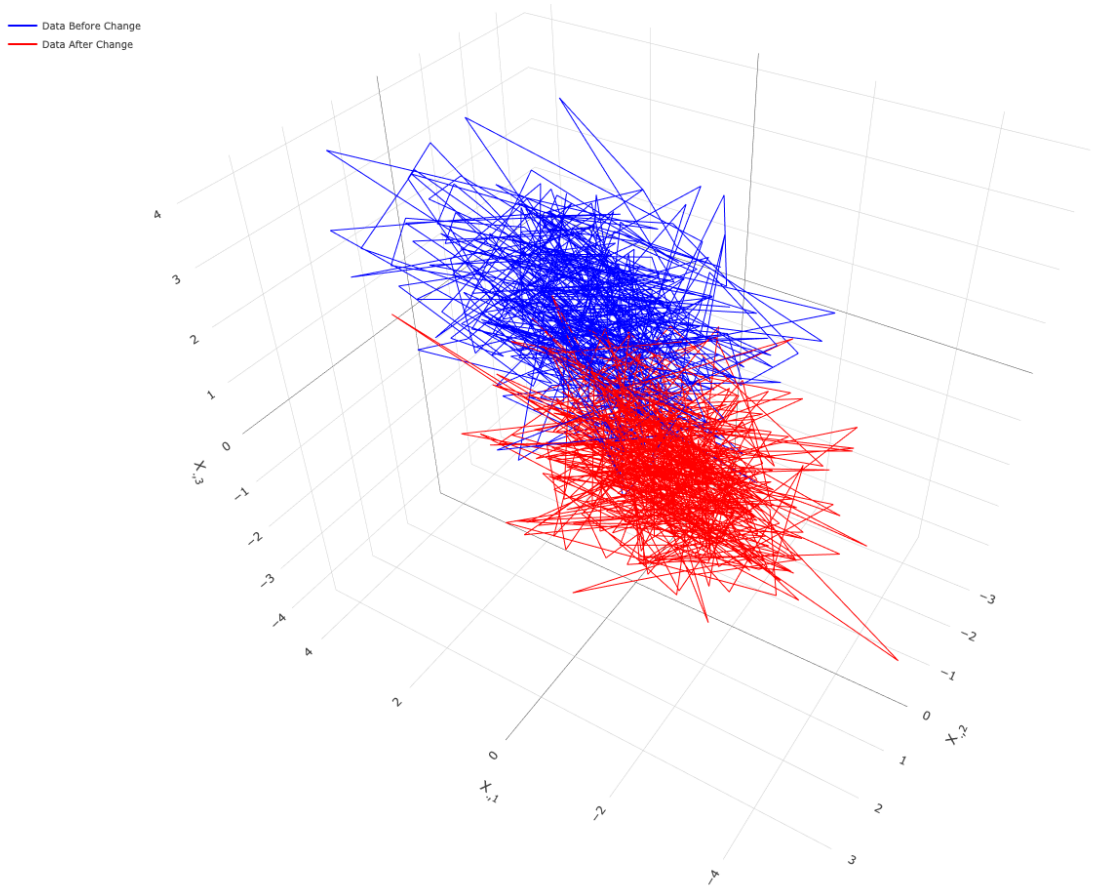

Figure 5.1: Visualization of $(\mathbf{X}_t)$ for $n = 1000$, marking a change point at $\tau = 500$, with a signal gap magnitude $s = 0.25$, and dependence parameter $\theta = 0.4$.

point becomes impossible for mortals.

By considering these specific models for the noise and signal components, we aim to provide a thorough evaluation of the proposed testing procedure's ability to detect changes in the trend of a multivariate non-stationary time series. The multivariate causal noise process with time-varying coefficients and the sinusoidal signal functions create a challenging and realistic setting to assess the procedure's effectiveness in practical applications. The Monte Carlo simulations presented in the subsequent sections will shed light on the finite sample

performance of the testing procedure under various parameter configurations and sample sizes, offering valuable insights into its robustness and reliability in real-world scenarios.

5.4.2 Synthetic Data under the Null Hypothesis

In this section, we investigate the finite-sample behavior of the testing procedure described in Section 5.2.3 under the null hypothesis. Synthetic data are generated according to the model detailed in Section 5.4.1, with the assumption of a continuous signal function as defined in (5.4.17).

A series of experiments are conducted by varying the sample size n, the dimensionality d, and the dependence parameter θ. The sample size n is selected from the set $\{50, 100, 300, 500, 2000\}$, spanning a comprehensive range from small to large datasets. The dimensionality d is chosen from $\{3, 5, 10\}$, exploring the performance of the test across low to moderate dimensional settings. The dependence parameter θ ranges from -0.8, -0.4, 0 (indicating independence), to 0.4, and 0.8, enabling an evaluation of the influence of varying degrees of temporal dependence in the noise process.

Throughout these experiments, a constant significance level of $\alpha = 0.05$ is maintained, aligning with common standards in hypothesis testing. The empirical sizes, defined as the proportion of simulations where the null hypothesis is incorrectly rejected, are computed based on 100,000 simulations for each configuration of sample size, dimensionality, and dependence parameter. This extensive number of simulations guarantees that the empirical sizes are estimated with high precision.

The results of the experiments are summarized in Table 5.1. The table presents the rejection ratios, which are the empirical sizes expressed as decimal values, for each combination of sample size n, dimensionality d, and dependence parameter θ.

Analyzing the results in Table 5.1, we observe several important patterns. Firstly, for a fixed value of d and θ, the rejection ratios tend to decrease as the sample size n increases. This

Table 5.1: Rejection ratios for multivariate change point testing procedure under the null hypothesis; cf. (5.4.17).

n	d	θ				
		-0.8	-0.4	0.0	0.4	0.8
50	3	0.1186	0.1136	0.1097	0.1157	0.1250
	5	0.1241	0.1195	0.1170	0.1195	0.1241
	10	0.1310	0.1217	0.1203	0.1263	0.1301
100	3	0.1138	0.1099	0.1034	0.1124	0.1157
	5	0.1202	0.1131	0.1094	0.1146	0.1163
	10	0.1255	0.1181	0.1157	0.1172	0.1265
300	3	0.1062	0.1002	0.0955	0.0980	0.1038
	5	0.1084	0.1052	0.1023	0.1040	0.1113
	10	0.1140	0.1104	0.1059	0.1110	0.1144
500	3	0.0954	0.0890	0.0739	0.0905	0.0933
	5	0.0991	0.0917	0.0794	0.0950	0.1011
	10	0.1078	0.0989	0.0882	0.1029	0.1028
2000	3	0.0595	0.0556	0.0501	0.0541	0.0555
	5	0.0611	0.0529	0.0590	0.0572	0.0598
	10	0.0617	0.0580	0.0603	0.0599	0.0617

observation suggests that the testing procedure becomes more conservative as the sample size grows, with the empirical size approaching the nominal significance level of 0.05 from above. The convergence to the nominal level is slower for higher dimensional settings and stronger levels of temporal dependence.

Secondly, for a given sample size n and dependence parameter θ, the rejection ratios generally increase as the dimensionality d increases. This pattern indicates that the testing procedure tends to reject the null hypothesis more frequently in higher dimensional settings, suggesting that the test may be more sensitive to the increased complexity of the data.

Thirdly, for a fixed sample size n and dimensionality d, the rejection ratios exhibit a U-shaped pattern with respect to the dependence parameter θ. The rejection ratios are generally higher for extreme values of θ (i.e., -0.8 and 0.8) compared to moderate values (i.e., -0.4, 0, and 0.4). This observation suggests that the testing procedure is more sensitive to strong levels of temporal dependence, both positive and negative, in the noise process.

In conclusion, the simulation results presented in Table 5.1 provide valuable insights

into the finite-sample behavior of the proposed testing procedure under the null hypothesis. The empirical sizes tend to be higher than the nominal significance level, especially for small sample sizes, high dimensional settings, and strong levels of temporal dependence. This over-liberal behavior may be related to the non-stationarity of the noise process, which introduces additional challenges in accurately assessing the significance of the test statistic. As the sample size increases, the empirical sizes gradually approach the nominal level, indicating that the testing procedure achieves the correct size asymptotically. However, the convergence rate appears to be slower compared to the case of stationary noise, highlighting the impact of non-stationarity on the finite-sample performance of the test.

The results emphasize the importance of considering the sample size, dimensionality, dependence structure, and the presence of non-stationarity when applying the testing procedure in practice. Researchers should be cautious when interpreting the results for small sample sizes and high dimensional settings, and may consider using more conservative critical values or adjusting the significance level to account for the finite-sample behavior of the test in the presence of non-stationary noise. Additionally, further research on developing refined testing procedures that are more robust to non-stationarity could be valuable in enhancing the reliability and applicability of change point detection methods in complex real-world scenarios.

5.4.3 Synthetic Data under Alternative Hypotheses

In this section, we conduct a comprehensive analysis of our testing procedure's power using synthetic data. We generate data according to the framework outlined in Section 5.4.1, with the signal specified as per (5.4.18), representing the alternative hypothesis H_1.

The experimental design encompasses a wide range of parameter settings to assess the performance of our testing procedure under various conditions. We vary the sample size n among $\{50, 100, 300, 500, 2000\}$ to examine the power of the test for different data sizes. The dimen-

sionality d is selected from $\{3, 5, 10\}$ to evaluate the impact of the number of dimensions on the test's performance. The dependence parameter θ is varied among $\{-0.8, -0.4, 0, 0.4, 0.8\}$ to assess the test's sensitivity to different levels of temporal dependence in the noise process. The gap parameter s, which governs the magnitude of the change in the signal, is chosen from $\{0, 0.008, 0.015, 0.035, 0.050, 0.075\}$ to investigate the test's ability to detect changes of varying sizes.

Throughout the experiments, we fix the ratio $\tau/n = 0.5$, indicating that the change point occurs at the midpoint of the observed time series. Our testing methodology employs a significance level of $\alpha = 0.05$ consistently across all settings. For each combination of experimental parameters, we generate the trend $(\boldsymbol{\nu}_i)$ according to the specified methodology. We then simulate the additive noise process multiple times and input the resulting data into our testing procedure. The outcomes are compiled from 100,000 independent simulations to ensure the reliability and robustness of our findings.

Figure 5.2 presents the rejection ratios, which represent the empirical power of our testing procedure, under various combinations of sample size n, dimensionality d, dependence parameter θ, and gap parameter s. Each subplot corresponds to a specific combination of n and d, and within each subplot, the rejection ratios are plotted against the gap parameter s for different values of θ.

Analyzing the results in Figure 5.2, we observe that the rejection ratios consistently increase as the gap parameter s increases, indicating that our testing procedure is more powerful in detecting larger changes in the signal. This behavior is intuitive, as larger changes are easier to distinguish from the null hypothesis of no change. Moreover, for a fixed gap parameter s, the rejection ratios generally improve as the sample size n increases, suggesting that our testing procedure benefits from larger sample sizes by gaining more power to detect changes in the signal. The increased sample size provides more information about the underlying process, enabling the test to make more accurate decisions.

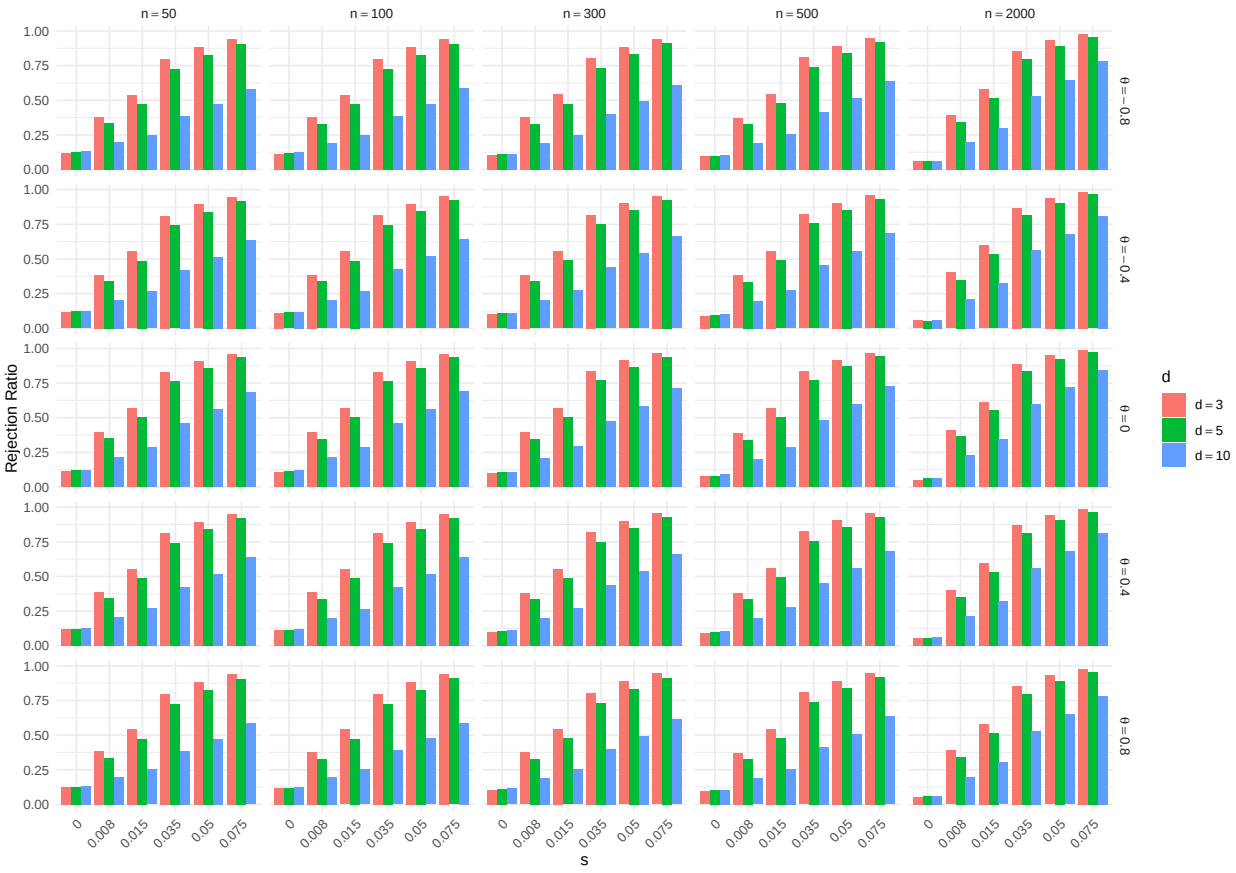

Figure 5.2: Rejection ratios for change point testing procedure under the alternative hypothesis; cf. (5.2.3). The noise process is shaped by the dependence parameter θ. We vary the gap parameter s over the set $\{0, 0.008, 0.015, 0.035, 0.050, 0.075\}$, the sample size n over $\{50, 100, 300, 500, 2000\}$, the dimensionality d over $\{3, 5, 10\}$, and the dependence parameter θ over $\{-0.8, -0.4, 0, 0.4, 0.8\}$. Each data point represents 100,000 replications.

The impact of the dimensionality d on the power of the test exhibits a more complex pattern. For smaller sample sizes (e.g., $n = 50, 100$), the rejection ratios tend to decrease as d increases, indicating that the test may struggle to detect changes in higher-dimensional settings when the sample size is limited. However, for larger sample sizes (e.g., $n = 300, 500, 2000$), the effect of dimensionality on power gradually diminishes, suggesting that the test becomes more robust to the curse of dimensionality as the sample size grows.

The dependence parameter θ also influences the power of the test, particularly for smaller sample sizes and gap parameters. Generally, the rejection ratios are lower for strongly depen-

dent cases ($\theta = -0.8, 0.8$) compared to the independent case ($\theta = 0$) when the sample size and gap parameter are small. This observation indicates that strong temporal dependence in the noise process can make it more challenging for the test to detect small changes in the signal. However, as the sample size and gap parameter increase, the impact of dependence on power diminishes, and the test becomes more robust to different levels of temporal dependence.

Notably, for large sample sizes (e.g., $n = 2000$) and moderate to large gap parameters ($s \geq 0.035$), the rejection ratios approach 1 across all dimensionalities and dependence levels, indicating that our testing procedure has excellent power in detecting changes in the signal under these conditions.

In conclusion, our extensive simulation study demonstrates the effectiveness and robustness of the proposed testing procedure in detecting changes in the mean of multivariate time series with non-stationary noise. The power of the test increases with the magnitude of the change and the sample size, while being sensitive to the dimensionality and temporal dependence, particularly for smaller sample sizes and change magnitudes. As the sample size grows, the test becomes more robust to the impact of dimensionality and dependence, exhibiting excellent power in detecting changes across a wide range of settings.

5.5 Proofs

5.5.1 Proof of Theorem 5.3.1

Let

$$\alpha_{n,k_n}(i,j) := \begin{cases} -1/k_n & \text{for } j \in (i, i+k_n], \\ 1/k_n & \text{for } j \in [i-k_n+1, i], \\ 0 & \text{for } j \in (0, i-k_n+1) \cup (i+k_n, n]. \end{cases} \tag{5.5.19}$$

Denote $\boldsymbol{\nu}_i = \mathbb{E}\mathbf{X}_i$. Therefore, we have

$$\mathbf{U}_{n,k_n}(i) := \sum_{j=1}^{n} \alpha_{n,k_n}(i,j)\mathbf{X}_j = \sum_{j=1}^{n} \alpha_{n,k_n}(i,j)(\boldsymbol{\nu}_j + \mathbf{Z}_j) \tag{5.5.20}$$

and

$$\hat{\mathbf{U}}_{n,k_n}(i) := \sum_{j=1}^{n} \alpha_{n,k_n}(i,j)\mathbf{W}_j \tag{5.5.21}$$

The function $\alpha_{n,k_n}(i,j)$ represents the weight assigned to the observation at time point j when computing the difference of averages $\mathbf{U}_{n,k_n}(i)$ and its bootstrap counterpart $\hat{\mathbf{U}}_{n,k_n}(i)$. The weights are chosen such that observations within the two adjacent windows of size k_n are assigned equal and opposite weights, while observations outside the windows are assigned zero weight.

The difference between $\mathbf{U}_{n,k_n}(i)$ and $\hat{\mathbf{U}}_{n,k_n}(i)$ can be bounded as follows:

$$\begin{aligned} \|\mathbf{U}_{n,k_n}(i) - \hat{\mathbf{U}}_{n,k_n}(i)\| &\leq \left\| \sum_{j=1}^{n} \alpha_{n,k_n}(i,j)\boldsymbol{\nu}_j \right\| + \left\| \sum_{j=1}^{n} \alpha_{n,k_n}(i,j)(\mathbf{W}_j - \mathbf{Z}_j) \right\| \\ &\leq (I) + (II) \end{aligned} \tag{5.5.22}$$

The inequality above decomposes the difference into two terms: (I) represents the contribution from the signal component $\boldsymbol{\nu}_j$, while (II) represents the contribution from the

difference between the bootstrap samples $\mathbf{W}_j$ and the noise component $\mathbf{Z}_j$.

Since we assume $\boldsymbol{\nu}(\cdot) \in L[0,1]$, it follows that $(I) = O(k_n/n)$ uniformly over $i = k_n, k_n + 1, \cdots, n - k_n$.

Let $(\mathbf{V}_j)$ be a sequence of independent, mean zero, Gaussian random vectors with $\mathbf{V}_j \sim N(\mathbf{0}, \Sigma_j)$, where Σ_j are the local long-run covariance matrices defined in (5.3.8). By Chebyshev's inequality and Theorem 3.1. in Mies and Steland [2023], under Condition 5.3.1 and Condition 5.3.2, for any $x > 0$, we have

$$\begin{aligned} \Pr\left(\max_{k\le n}\left\|\sum_{j=1}^{k}\mathbf{V}_j - \sum_{j=1}^{k}\mathbf{Z}_j\right\| \ge x\right) &\le \frac{\mathbb{E}\max_{k\le n}\left\|\sum_{j=1}^{k}\mathbf{V}_j - \sum_{j=1}^{k}\mathbf{Z}_j\right\|^2}{x^2} \\ &\le \frac{C\eta_n^2 D_n^{\frac{A-2}{A-1}}\log(n)\left(\frac{d}{n}\right)^{2\lambda(p,A)}}{x^2} \end{aligned} \tag{5.5.23}$$

where

$$\lambda(p,A) = \begin{cases} \frac{p-2}{6p-4}, & A \ge 3, \\ \frac{(A-2)(p-2)}{(4A-6)p-4}, & \frac{3p+2}{p+2} < A < 3, \\ \frac{1}{2} - \frac{1}{A}, & 2 < A \le \frac{3p+2}{p+2} \end{cases} \tag{5.5.24}$$

This bound on the difference between the partial sum processes of the Gaussian random vectors $\mathbf{V}_j$ and the noise component $\mathbf{Z}_j$ depends on the moment parameter p, the dependence parameter A, the non-stationarity parameter D_n, and the dimension d.

Consequently,

$$\max_{k\le n}\left\|\sum_{j=1}^{k}\mathbf{V}_j - \sum_{j=1}^{k}\mathbf{Z}_j\right\| = \mathcal{O}_{\mathbb{P}}\left(\eta_n D_n^{\frac{A-2}{2(A-1)}}\sqrt{\log(n)}\left(\frac{d}{n}\right)^{\lambda(p,A)}\right) \tag{5.5.25}$$

We define the cumulative process $\mathbf{Q}_t := \sum_{i=1}^{t}\Sigma_i$ and estimate it with $\hat{\mathbf{Q}}_t := \sum_{i=m}^{t}\hat{\mathbf{C}}_t$ in Algorithm 3. Let $\tilde{\mathbf{Q}}$ be some cumulative covariance process, and consider independent

Gaussian random vectors $\tilde{\mathbf{V}}_t \sim N(\mathbf{0}, \tilde{\mathbf{Q}}_t - \tilde{\mathbf{Q}}_{t-1})$, which are coupled with $(\mathbf{V}_t)$ such that

$$\mathbb{E} \max_{k=1,\ldots,n} \left\| \sum_{t=1}^{k} \mathbf{V}_t - \sum_{t=1}^{k} \tilde{\mathbf{V}}_t \right\|^2 \leq C \log(n)[\sqrt{n\omega_n\rho_n} + \rho_n] = \Delta_n \tag{5.5.26}$$

where

$$\omega_n := \max_{t=1,\ldots,n} \left\| \tilde{\mathbf{Q}}_t - \mathbf{Q}_t \right\|_{tr}, \quad \rho_n := \max_{t=1,\ldots,n} \|\mathbf{Q}_t\|_{tr}$$

by Proposition 5.2. in Mies and Steland [2023].

This coupling between the Gaussian random vectors $\mathbf{V}_t$ and $\tilde{\mathbf{V}}_t$ is constructed based on an estimated cumulative covariance process $\tilde{\mathbf{Q}}$. The term Δ_n bounds the expected maximum difference between the partial sum processes of $\mathbf{V}_t$ and $\tilde{\mathbf{V}}_t$, and it depends on the estimation error ω_n and the trace norm of the true cumulative covariance process ρ_n.

By Chebyshev's inequality,

$$\Pr\left(\max_{k=1,\ldots,n} \left\| \sum_{j=1}^{k} (\mathbf{V}_j - \tilde{\mathbf{V}}_j) \right\| > x \right) \leq \frac{\Delta_n}{x^2}. \tag{5.5.27}$$

Now, letting $\tilde{\mathbf{Q}}_t = \hat{\mathbf{Q}}_t$ and denoting the corresponding error (5.5.26) by $\hat{\Delta}_n$, since $\mathbf{W}_i \sim N(\mathbf{0}, \hat{\mathbf{Q}}_t - \hat{\mathbf{Q}}_{t-1})$, by Chebyshev's inequality, we obtain

$$\begin{aligned}
&\Pr\left(\max_{k\leq n} \left\| \sum_{j=1}^{k} \mathbf{W}_j - \sum_{j=1}^{k} \mathbf{V}_j \right\| > x \right) \\
&\leq \Pr\left(\max_{k\leq n} \left\| \sum_{j=1}^{k} \mathbf{W}_j - \sum_{j=1}^{k} \mathbf{V}_j \right\| > x, \hat{\Delta}_n < x \right) + \Pr\left(\hat{\Delta}_n \geq x \right) \\
&\leq \frac{\hat{\Delta}_n}{x^2} + \Pr\left(\hat{\Delta}_n \geq x \right) \\
&\leq \frac{1}{x} + \frac{\mathbb{E}|\hat{\Delta}_n|}{x},
\end{aligned} \tag{5.5.28}$$

which bounds the difference between the partial sum processes of the bootstrap samples $\mathbf{W}_j$

and the Gaussian random vectors $\mathbf{V}_j$ by conditioning on the event that the estimation error $\hat{\Delta}_n$ is smaller than x and using Chebyshev's inequality.

Note that by Theorem 5.1. in Mies and Steland [2023],

$$\hat{\Delta}_n = \mathcal{O}_{\mathbb{P}}\left(\log(n)\eta_n^2\left(D_n^{\frac{1}{2}} n^{\frac{1}{2}} m_n^{\frac{1}{4}} + n^{\frac{3}{4}} d^{\frac{1}{2}} m_n^{\frac{1}{4}} + n m_n^{-\frac{1}{2}} + n m_n^{1-\frac{A}{2}} + 1\right)\right) \tag{5.5.29}$$

Combining this result with (5.5.28) yields

$$\max_{k\leq n}\left\|\sum_{j=1}^{k}\mathbf{W}_j - \sum_{j=1}^{k}\mathbf{V}_j\right\| = \mathcal{O}_{\mathbb{P}}\left(\log(n)\eta_n^2\left(D_n^{\frac{1}{2}} n^{\frac{1}{2}} m_n^{\frac{1}{4}} + n^{\frac{3}{4}} d^{\frac{1}{2}} m_n^{\frac{1}{4}} + n m_n^{-\frac{1}{2}} + n m_n^{1-\frac{A}{2}} + 1\right)\right) \tag{5.5.30}$$

Since

$$\max_{k\leq n}\left\|\sum_{j=1}^{k}\mathbf{W}_j - \sum_{j=1}^{k}\mathbf{Z}_j\right\| \leq \max_{k\leq n}\left\|\sum_{j=1}^{k}\mathbf{W}_j - \sum_{j=1}^{k}\mathbf{V}_j\right\| + \max_{k\leq n}\left\|\sum_{j=1}^{k}\mathbf{V}_j - \sum_{j=1}^{k}\mathbf{Z}_j\right\|, \tag{5.5.31}$$

it follows that

$$\begin{aligned}\max_{k\leq n}\left\|\sum_{j=1}^{k}\mathbf{W}_j - \sum_{j=1}^{k}\mathbf{Z}_j\right\| = \mathcal{O}_{\mathbb{P}}\Bigg(\log(n)\eta_n^2\Bigg(&D_n^{\frac{A-2}{2(A-1)}}\left(\frac{d}{n}\right)^{\lambda(p,A)} \\ &+ D_n^{\frac{1}{2}} n^{\frac{1}{2}} m_n^{\frac{1}{4}} + n^{\frac{3}{4}} d^{\frac{1}{2}} m_n^{\frac{1}{4}} + n m_n^{-\frac{1}{2}} + n m_n^{1-\frac{A}{2}} + 1\Bigg)\Bigg),\end{aligned} \tag{5.5.32}$$

which combines the previous bounds to obtain a bound on the difference between the partial sum processes of the bootstrap samples $\mathbf{W}_j$ and the noise component $\mathbf{Z}_j$. The bound depends on various parameters, including the sample size n, the dimension d, the moment parameter p, the dependence parameter A, the non-stationarity parameter D_n, and the block size m_n used in the bootstrap procedure.

Define

$$L_{n,k_n}(i) := |\alpha_{n,k_n}(i,1)| + \sum_{j=2}^{n}|\alpha_{n,k_n}(i,j) - \alpha_{n,k_n}(i,j-1)|. \tag{5.5.33}$$

By substituting the expression for $\alpha_{n,k_n}(i,j)$ and taking the maximum over i, we obtain

$$L_{n,k_n} := \max_{i \leqslant n} L_{n,k_n}(i) = \mathcal{O}\left(\frac{1}{k_n}\right). \tag{5.5.34}$$

The term $L_{n,k_n}(i)$ represents the total variation of the weight function $\alpha_{n,k_n}(i,j)$ with respect to j for a fixed time point i. It measures the smoothness of the weight function and plays a crucial role in controlling the approximation error.

Using the smoothness of the weight function, as measured by $L_{n,k_n}(i)$, and the previously derived bound on the difference between the partial sum processes of $\mathbf{W}_j$ and $\mathbf{Z}_j$, we can bound the term (II) as follows:

$$\begin{aligned}
(II) &= \left\| \sum_{j=1}^{n} \alpha_{n,k_n}(i,j)(\mathbf{W}j - \mathbf{Z}_j) \right\| \\
&\leq Ln, k_n(i) \max_{k \leq n} \left\| \sum_{j=1}^{k} \mathbf{W}j - \sum j = 1^k \mathbf{Z}_j \right\| \\
&= \mathcal{O}_{\mathbb{P}}\left(\frac{\log(n)\eta_n^2}{k_n} \left(D_n^{\frac{A-2}{2(A-1)}} \left(\frac{d}{n}\right)^{\lambda(p,A)} + D_n^{\frac{1}{2}} n^{\frac{1}{2}} m_n^{\frac{1}{4}} \right.\right. \\
&\qquad \left.\left. + n^{\frac{3}{4}} d^{\frac{1}{2}} m_n^{\frac{1}{4}} + n m_n^{-\frac{1}{2}} + n m_n^{1-\frac{A}{2}} + 1 \right) \right).
\end{aligned} \tag{5.5.35}$$

To balance the contributions from the signal component (I) and the noise component (II), we choose the window size k_n as

$$k_n = n^{\frac{1}{2}} (\log n)^{\frac{1}{2}} \eta_n \left(D_n^{\frac{A-2}{2(A-1)}} \left(\frac{d}{n}\right)^{\lambda(p,A)} + D_n^{\frac{1}{4}} n^{\frac{1}{4}} m_n^{\frac{1}{8}} + n^{\frac{3}{8}} d^{\frac{1}{4}} m_n^{\frac{1}{8}} + n^{\frac{1}{2}} m_n^{-\frac{1}{4}} + n^{\frac{1}{2}} m_n^{\frac{1}{2}-\frac{A}{4}} + 1 \right). \tag{5.5.36}$$

This choice ensures that both terms are of the same order. The window size k_n depends on various parameters, including the sample size n, the dimension d, the moment parameter p, the dependence parameter A, the non-stationarity parameter D_n, and the block size m_n

used in the bootstrap procedure.

With this choice of k_n, we obtain the desired asymptotic equivalence between the difference of averages $\mathbf{U}_{n,k_n}(i)$ and its bootstrap counterpart $\hat{\mathbf{U}}_{n,k_n}(i)$:

$$\max_{i\leq n}\|\mathbf{U}_{n,k_n}(i)-\hat{\mathbf{U}}_{n,k_n}(i)\| = \mathcal{O}_{\mathbb{P}}\Bigg(n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\eta_n\Bigg(D_n^{\frac{A-2}{2(A-1)}}\left(\frac{d}{n}\right)^{\lambda(p,A)} + D_n^{\frac{1}{4}}n^{\frac{1}{4}}m_n^{\frac{1}{8}} + n^{\frac{3}{8}}d^{\frac{1}{4}}m_n^{\frac{1}{8}} + n^{\frac{1}{2}}m_n^{-\frac{1}{4}} + n^{\frac{1}{2}}m_n^{\frac{1}{2}-\frac{A}{4}} + 1\Bigg)\Bigg). \tag{5.5.37}$$

□

BIBLIOGRAPHY

S. Abosedra, A. Fakih, S. Ghosh, and K. Kanjilal. Financial development and business cycle volatility nexus in the uae: Evidence from non-linear regime-shift and asymmetric tests. *International Journal of Finance amp; Economics ; volume 28, issue 3, page 2729-2741 ; ISSN 1076-9307 1099-1158*, 2021.

A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046 – 4087, 2009. ISSN 00905364.

S. Banerjee and K. Guhathakurta. Change-point analysis in financial networks. *Stat*, 9(1), 2020. ISSN 20491573.

R. Baranowski and P. Fryzlewicz. *wbs: Wild Binary Segmentation for Multiple Change-Point Detection*, 2019. R package version 1.4.

R. Baranowski, Y. Chen, and P. Fryzlewicz. Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(3):649–672, 2019.

I. Berkes, W. Liu, and W. B. Wu. Komlos–Major–Tusnady approximation under dependence. *The Annals of Probability*, 42(2):794 – 817, 2014. ISSN 00911798.

P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 2 edition, 1999.

S. Bonnerjee, S. Karmakar, and W.B. Wu. Gaussian approximation for non-stationary time series with optimal rate and explicit construction. *Preprint*, 2024.

A. Bücher, H. Dette, and F. Heinrichs. Are deviations in a gradually varying mean relevant? a testing approach based on sup-norm estimators. *The Annals of Statistics*, 49(6):3583–3617, 2021.

P. Bühlmann and H. R. Künsch. Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis*, 31(3):295–310, 1999. ISSN 0167-9473. doi:https://doi.org/10.1016/S0167-9473(99)00014-6.

H. Cao and W.B. Wu. Changepoint estimation: Another look at multiple testing problems. *Biometrika*, 102(4):974–980, 2015. ISSN 14643510.

H. Cao and W.B. Wu. Testing and estimation for clustered signals. *Bernoulli*, 28(1):525–547, 2022.

Centre for Disease Control and Prevention. CDC museum COVID-19 timeline. *website*, 2022. Accessed: 2022-05-18.

Y. Chen, T. Wang, and R. J. Samworth. High-dimensional, multiscale online changepoint detection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1): 234–266, 2022.

T. S. Ching, N. A. Hamzah, and N. H. Moin. An introduction to multiple structural breaks estimation with minimum description length approach. *AIP Conference Proceedings*, 1605: 991 – 996, 2014. ISSN 0094243X.

H. Cho and P.Z. Fryzlewicz. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. 2, 2015.

M. F. R. Chowdhury, S.-A. Selouani, and D. O'Shaughnessy. Bayesian on-line spectral change point detection: a soft computing approach for on-line asr. *International Journal of Speech Technology*, 15(1):5 – 23, 2012. ISSN 1381-2416.

J. A. Dar and M. Asif. Is financial development good for carbon mitigation in india? a regime shift-based cointegration analysis. *Carbon Management*, 8(5-6):435 – 443, 2017. ISSN 1758-3004.

R. A. Davis, T. C. M. Lee, and G. A. Rodriguez-Yam. Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101 (473):223 – 239, 2006. ISSN 01621459.

H. Dette and W. Wu. Detecting relevant changes in the mean of nonstationary processes—a mass excess approach. *The Annals of Statistics*, 47(6):3578–3608, 2019.

H. Dette, T. Eckle, and M. Vetter. Multiscale change point detection for dependent data. *Scandinavian Journal of Statistics*, 47(4):1243–1274, 2020.

B. Eichinger and C. Kirch. A mosum procedure for the estimation of multiple random change points. *Bernoulli*, 24(1):526–564, 2018. ISSN 13507265.

F. Enikeeva and Z. Harchaoui. High-dimensional change-point detection under sparse alternatives. *The Annals of Statistics*, 47(4):2051 – 2079, 2019. ISSN 00905364.

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

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

P. Fryzlewicz. Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *The Annals of Statistics*, 46(6B):3390–3421, 2018.

Y. Guo, M. Gao, and X. Lu. Multivariate change point detection for heterogeneous series. *Neurocomputing*, 510:122 – 134, 2022. ISSN 0925-2312.

Y. Han and R. S. Tsay. High-dimensional linear regression for dependent data with applications to nowcasting. *Statistica Sinica*, 30(4):1797 – 1827, 2020. ISSN 10170405.

D. M. Hawkins. Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, 72(357):180–186, 1977.

F. Heinrichs and H. Dette. A distribution free test for changes in the trend function of locally stationary processes. *Electronic Journal of Statistics*, 15(2):3762–3797, 2021.

T. Heo and L. Manuel. Greedy copula segmentation of multivariate non-stationary time series for climate change adaptation. *Progress in Disaster Science*, 14(100221-), 2022. ISSN 2590-0617.

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

L. Horváth and P. Kokoszka. Change-point detection with non-parametric regression. *Statistics*, 36(1):9–31, 2002. doi:10.1080/02331880210930.

C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, and B. Cao. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet*, 395(10223): 497–506, 2020. doi:https://doi.org/10.1016/S0140-6736(20)30183-5.

R. Killick and I.A. Eckley. changepoint: An r package for changepoint analysis. *Journal of Statistical Software*, 58(1):1 – 19, 2014a. ISSN 1548-7660.

R. Killick and I.A. Eckley. changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, 58(3):1–19, 2014b.

R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590 – 1598, 2012. ISSN 01621459.

R. Killick, K. Haynes, and I.A. Eckley. *changepoint: An R package for changepoint analysis*, 2022. R package version 2.2.4.

C. Kirch, B. Muhsal, and H. Ombao. Detection of changes in multivariate time series with application to eeg data. *Journal of the American Statistical Association*, 110(511):1197 – 1216, 2015. ISSN 01621459.

K. K. Korkas and P. Pryzlewicz. Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica Sinica*, 27(1):287 – 311, 2017. ISSN 10170405.

S. N. Lahiri. Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404, 1999. doi:10.1214/aos/1018031117.

M. Lavielle and G. Teyssière. Detection of multiple change-points in multivariate time series. *LITHUANIAN MATHEMATICAL JOURNAL C/C OF LITOVSKII MATEMATICHESKII SBORNIK*, 46(3):287 – 306, 2006. ISSN 0363-1672.

K. Liu and X. Li. Bio-inspired multi-sensory pathway network for change detection. *Cognitive Computation*, 14(4):1421 – 1434, 2022. ISSN 1866-9956.

L. Ma, A. J. Grant, and G. Sofronov. Multiple change point detection and validation in autoregressive time series data. *Statistical Papers*, 61(4):1507 – 1528, 2020. ISSN 09325026.

A. Mallik, B. Sen, M. Banerjee, and G. Michailidis. Threshold estimation based on a p-value framework in dose-response and regression settings. *Biometrika*, 98(4):887–900, 2011. ISSN 00063444, 14643510.

A. Mallik, M. Banerjee, and B. Sen. Asymptotics for p-value based threshold estimation in regression settings. *Electronic Journal of Statistics*, 7(none):2477 – 2515, 2013. doi:10.1214/13-EJS845.

J. Markevičiūtė, A. Račkauskas, and C. Suquet. Testing the epidemic change in nearly non-stationary auto-regressive processes. *Nonlinear Analysis: Modelling and Control*, 19(1): 67–82, 2014.

D. Mercurio and V. Spokoiny. Statistical inference for time-inhomogeneous volatility models. *The Annals of Statistics*, 32(2):577 – 602, 2004. ISSN 00905364.

F. Mies and A. Steland. Sequential Gaussian approximation for nonstationary time series in high dimensions. *Bernoulli*, 29(4):3114 – 3140, 2023. doi:10.3150/22-BEJ1577.

H.-G. Muller. Change-points in non-parametric regression analysis. *The Annals of Statistics*, 20(2):737 – 761, 1992. doi:10.1214/aos/1176348654.

E. S. Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523 – 527, 1955a. ISSN 00063444.

E. S. Page. Control charts with warning lines. *Biometrika*, 42(1/2):243 – 257, 1955b. ISSN 00063444.

Y. Park, H. Im, and Y. Lim. Change points detection for nonstationary multivariate time series. *Communications for Statistical Applications and Methods*, 30(4):369–388 – 388, 2023. ISSN 23834757.

M. Peligrad and Q. M. Shao. Estimation of the variance of partial sums for ρ-mixing random variables. *Journal of Multivariate Analysis*, 52(1):140–157, 1995. ISSN 0047-259X. doi:https://doi.org/10.1006/jmva.1995.1008.

P. Preuss, R. Puchstein, and Ho. Dette. Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association*, 110(510):654 – 668, 2015. ISSN 01621459.

M. B. Priestley. *Nonlinear and nonstationary time series analysis*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, 1988. ISBN 0-12-564910-X.

A. Safikhani and A. Shojaie. Joint structural break detection and parameter estimation in high-dimensional non-stationary var models. *Journal of the American Statistical Association*, 117(537):251 – 264, 2022. ISSN 0162-1459.

A. L. Schröder and H. Ombao. Fresped: Frequency-specific change-point detection in epileptic seizure multi-channel eeg data. 2018. ISSN 0162-1459.

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

C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.

X. Shao and W. B. Wu. Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, 35(4):1773–1801, 2007. ISSN 0090-5364. doi:10.1214/009053606000001479.

H. Sheridan and A. L. Kleinsmith. Music reading expertise affects visual change detection: Evidence from a music-related flicker paradigm. *Quarterly Journal of Experimental Psychology*, 75(9):1643 – 1652, 2022. ISSN 17470218.

D. Siegmund. Confidence sets in change-point problems. *International Statistical Review/Revue Internationale de Statistique*, pages 31–48, 1988.

R.R. Sundararajan and M. Pourahmadi. Nonparametric change point detection in multivariate piecewise stationary time series. *Journal of Nonparametric Statistics*, 30(4):926–956, 2018. doi:10.1080/10485252.2018.1504943.

H. Tong. *Nonlinear time series*, volume 6 of *Oxford Statistical Science Series*. The Clarendon Press, Oxford University Press, New York, 1990. ISBN 0-19-852224-X. A dynamical system approach, With an appendix by K. S. Chan, Oxford Science Publications.

J.-P. Vert and K. Bleakley. Fast detection of multiple change-points shared by many signals using group lars. In *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*, number Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010, page 9p., Mines ParisTech CBIO, Institut Curie, INSERM U900, 2010.

A. Vexler. Guaranteed testing for epidemic changes of a linear regression model. *Journal of Statistical Planning and Inference*, 136(9):3101–3120, 2006. ISSN 0378-3758. doi:https://doi.org/10.1016/j.jspi.2004.11.010.

M. Vogt and H. Dette. Detecting gradual changes in locally stationary processes. *The Annals of Statistics*, 43(2):713 – 740, 2015. doi:10.1214/14-AOS1297.

J. R. Williams, M. M. Robinson, M. W. Schurgin, J. T. Wixted, and T. F. Brady. You cannot "count" how many items people remember in visual working memory: The importance of signal detection-based measures for understanding change detection performance.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY-HUMAN PERCEPTION AND PERFORMANCE, 48(12):1390 – 1409, 2022. ISSN 00961523.

M. Worobey. Dissecting the early covid-19 cases in wuhan: Elucidating the origin of the pandemic requires understanding of the wuhan outbreak. *Science*, 374(6572):1202, 2021. ISSN 0036-8075.

K. J. Worsley. Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, 73(1):91 – 104, 1986. ISSN 00063444.

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

W. B. Wu. Asymptotic theory for stationary processes. *Statistics and its Interface*, 4(2): 207–226, 2011.

W. B. Wu and Y.N. Wu. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, 10(1):352–379, 2016. ISSN 19357524.

Q. Yao. Tests for change-points with epidemic alternatives. *Biometrika*, 80(1):179–191, 1993. ISSN 00063444.

Z. Zhou. Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association*, 108(502):726–740, 2013. doi:10.1080/01621459.2013.787184.