Enhanced Backward Multiple Change-Point Detection

by

Shahab Pirnia

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2023

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Reza Modarres Professor, Dept. of Philosophy of Zoology, George Washington University
Supervisor(s):	Shoja'eddin Chenouri Professor, Dept. of Statistics and Actuarial Science, University of Waterloo
Internal Member:	Pengfei Li Professor, Dept. of Statistics and Actuarial Science, University of Waterloo
Internal Member:	Greg Rice Associate Professor, Dept. of Statistics and Actuarial Science, University of Waterloo
Internal-External Member:	Mikal Skuterud Professor, Dept. of Economics, University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Many statistical tools are built upon a specific set of assumptions on the distribution of the data at hand. However, the distribution of the observations in the dataset may not remain constant and may change due to some external events. For a sequence of observations, the points after which the distribution function has changed are commonly referred to as *change points*. Identifying such points can also be critical in gaining insights into the distributional behaviour of random variables and constructing statistical models. Thus, the change points analysis potentially applies to almost all data-driven disciplines, such as biology, finance, and public policies.

Change points analysis is categorized into online and offline analysis. The online change points analysis is designed to detect changes in the distribution of random variables as new observations are introduced. On the other hand, offline analysis is concerned with recovering change points within a historical dataset. In this thesis, we are only concerned with offline change point analysis; for simplicity, we refer to offline change points analysis as change points analysis.

Change point analysis was born 70 years ago from the quality control discipline Page (1954). Initially, the main focus of the change points literature was on the single change point scenario in which, at most, one change point exists within a sequence of random variables. However, with the advent of computers, the focus has switched to multiple change point detection problems. This shift does not imply that single change point detection methods are irrelevant. For instance, many multiple change point detection methods recover change points by conducting a single change point test locally. This class of change point detection methods is called *local search methods*.

One of the primary concerns of local search methods is the application of a single change point test statistic within the largest possible segment of the sequence of random variables with exactly one change point. Obtaining such intervals is a difficult task. For instance, wild binary segmentation Fryzlewicz et al. (2014) extracts the change points from intervals containing multiple change points. On the other hand, the narrowest over threshold Baranowski et al. (2019) estimates the change points within the narrowest intervals in which a predefined threshold is satisfied. Thus, the accuracy of the estimated locations of change points may suffer due to the shortness of these intervals. In this thesis, we propose two local search methods that attempt to infer locations of change points within the desirable intervals. The first method, *enhanced backward detection* (EBD), recovers the change points by eliminating unlikely candidates sequentially. The second method, i.e., *narrowest over threshold via interval selection with shortened exhaustive search* (NOT-IS.SES), estimates

the location of change points by following a top-down approach. That is, the change points are added to the active set sequentially. EBD and NOT-IS.SES are general procedures that can be applied to a wide range of change point problems by simply changing the underlying single change point test statistics.

Acknowledgements

I would like to thank many people who have helped me through the completion of this dissertation. First, I would like to express my sincere gratitude to my supervisor Professor Shojaeddin Chenouri, with whom I have had the luck to collaborate. I also wish to express my appreciation to Professor Pengfei Li, Professor Greg Rice, Professor Reza Modarres and Professor Mikal Skuterud for inspiring useful changes to my thesis. My appreciation also goes to all the members at the Department of Statistics and Actuarial Science, at the University of Waterloo, for their help and assistance during my PhD studies.

Special thanks to my family for their constant emotional support. Finally, my thanks also go to those who, directly or indirectly, helped me to complete my dissertation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction to Change Point Analysis

The statistical methods are constructed based on the assumptions on the probability distribution function from which the random variables are generated. If these assumptions are violated, then the statistical inference about the characteristics of the model based on the data at hand would be meaningless or, at best, misleading. For instance, in the introductory statistical courses, it is common to assume that random variables follow an identical probability distribution function. Subsequently, statistical inferences are conducted based on this assumption. Any significant change in the probability distribution function of the random variables leads to misleading conclusions and data analysis. Another important example is time series forecasting. For example, to build an ARMA model, the mean and variance of the data must be constant and independent of time. Otherwise, the ARMA model leads to erroneous forecasts.

Unfortunately, in many real-world scenarios, the probability distribution function does not remain constant and may change throughout data collection. The points that the distribution of random variables change are referred to as change points. Since the accuracy of the statistical inference often depends on the fulfilment of the assumptions on distribution functions of the random variables, disregarding the presence of change points results in misleading statistical inference. Thus, the change point analysis is an essential component of robust statistical data analysis. Providing a sound statistical inference is not the only application of the change point analysis. Often, detecting the points in which the distribution of random variables is altered is the problem of interest by itself. After recovering the locations of change points, one may have to interpret the events that may have influenced the distribution of the data.

One of the simplest change-point problems is the single normal mean change point scenario. In the single normal mean change-point problem, only one change occurs in the mean of the normally distributed random variables while the variance remains constant. In other words:

$$X_t \sim N(\mu_1, \sigma^2), \quad \text{for} \quad t = 1, \cdots, \tau,$$
$$X_t \sim N(\mu_2, \sigma^2), \quad \text{for} \quad t = \tau + 1, \cdots, T,$$

where $\mu_1 \neq \mu_2$ and $\sigma^2 < \infty$. This scenario can be demonstrated by the Nile river dataset in Figure 1.1, which was first investigated by Cobb (1978). The Nile dataset is the collection of the measurements on the Nile river's annual flow between the years 1871 to 1970. By simply observing the scatter plot of the Nile river data in 1.1, one can identify a change in mean around the year 1898. Ignoring the change in the mean of annual flow will lead to an inaccurate and erroneous inference. For instance, if the variance of the annual volume of the flow is estimated by the least-squares method while the mean change point is ignored, then the variance will be overestimated. Moreover, any further statistical inference on the Nile river data set, which is built upon the estimated variance, may lack sufficient accuracy and result in misleading inference on the Nile river data set. Note that, recovering the location of the change in the mean of annual flow of river Nile is also the problem of interest by itself and can be used to interpret the geographical events surrounding the Nile river.

The change point analysis is often conducted in scenarios where the data configuration is more complicated than the single normal mean change-point model. For instance, the change in distribution may occur in multiple instances concerning several parameters of the distribution function or even the form of the distribution function. Consider the motorcycle data set Silverman (1985) in Figure 1.2, which contains the accelerations of the head at different times during the simulated motorcycle accident. Unlike the Nile dataset containing only one change point, the motorcycle dataset indicates changes in the mean at several points. Moreover, the mean functions between the change points are not necessarily constant. In certain segments of the data, the mean is a polynomial function of time. As a result, the number of the parameters that are required to be estimated is larger. This observed behaviour adds to the computational and theoretical complexity of this change point problem compared to the Nile dataset. In both the Nile and Motorcycle datasets, the change points occur in the mean. The change in the mean is only the particular case of the change point problem. Other characteristics of the probability distribution function or the family of the probability distribution functions may change.

To demonstrate and compare the behaviour of different change-point models, we simulate multiple datasets from six distinct change-point models. The observations in each dataset are drawn independently from a normal distribution.

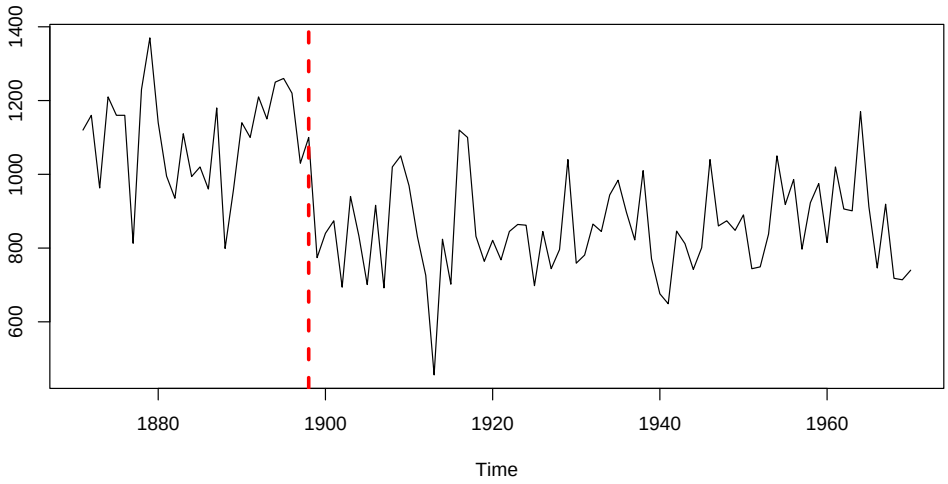

Figure 1.1: The vertical red dashed line represents the decrease in the volume of the annual flow of the Nile river.

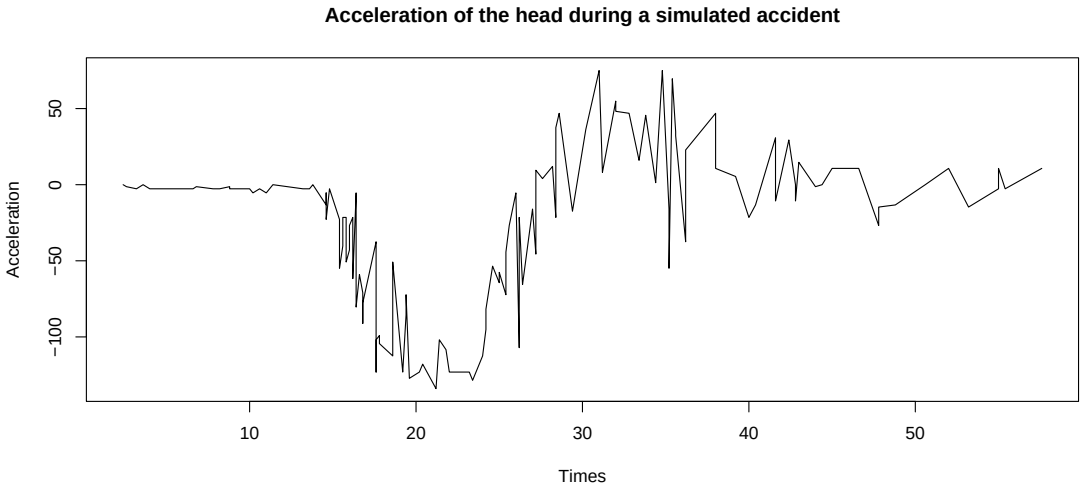

Figure 1.2: Unlike the Nile river dataset, the mean is a piecewise polynomial function of time in the motorcycle dataset.

Scenario 1. No change points present within the data set.

Scenario 2. Observations are generated from the normal distribution with a piecewise constant mean and a constant variance.

Scenario 3. The dataset is generated from a normal distribution with mean zero and a piecewise constant variance.

Scenario 4. Changes in the distribution are due to both mean and variance, and both mean and variance are piecewise constant functions.

Scenario 5. Changes occur in the mean, where the mean is a piecewise linear function of time while the variance remains constant.

Scenario 6. Changes in the distribution occur in both the mean and variance. The mean and variance are piecewise linear and constant functions of time, respectively.

In Scenarios 2 - 6, the locations of change points are at 50th, 100th, and 150th observations. We plotted the datasets for each scenario in Figure (1.3). The red dashed vertical lines represent the location of change points. In the case with no change point, the pattern of the dataset in different segments is very similar. In contrast, the change in the distribution of the random variables in other scenarios is evident.

As mentioned earlier, not only the change point analysis plays a vital role in statistical inference, but also, identification of the location of change points can be a problem of interest. The change point analysis was initially proposed by page Page (1954, 1955) to identify the point in which the quality of a product has deteriorated. Application of the change point analysis is not confined within the field of quality control. It has been extended to many other disciplines such as biology, economics, public policy. In biology, the change point methods has been applied to identify the copy number variation. Another application of the change point analysis is in financial data analysis. Many of the financial models on the behaviour of stock prices are constructed based on the variance of the return of the stock. Thus, discovering a point in time that the volatility of a stock price changes plays an important role. In terms of public policy, the change point methods are applied to verify whether a particularly desirable or undesirable change occurred after setting a new law into motion. For instance, a difference in car crash fatality rate can be investigated after introducing a new seat-belt law. Reduction of the rate of fatality in car crashes implies the effectiveness of a seat-belt law. A notable contemporary example of the application of change point analysis in public policy is assessing the effectiveness of the public health measures concerning the spread of the COVID-19 pandemic. Slow down in the rate of the

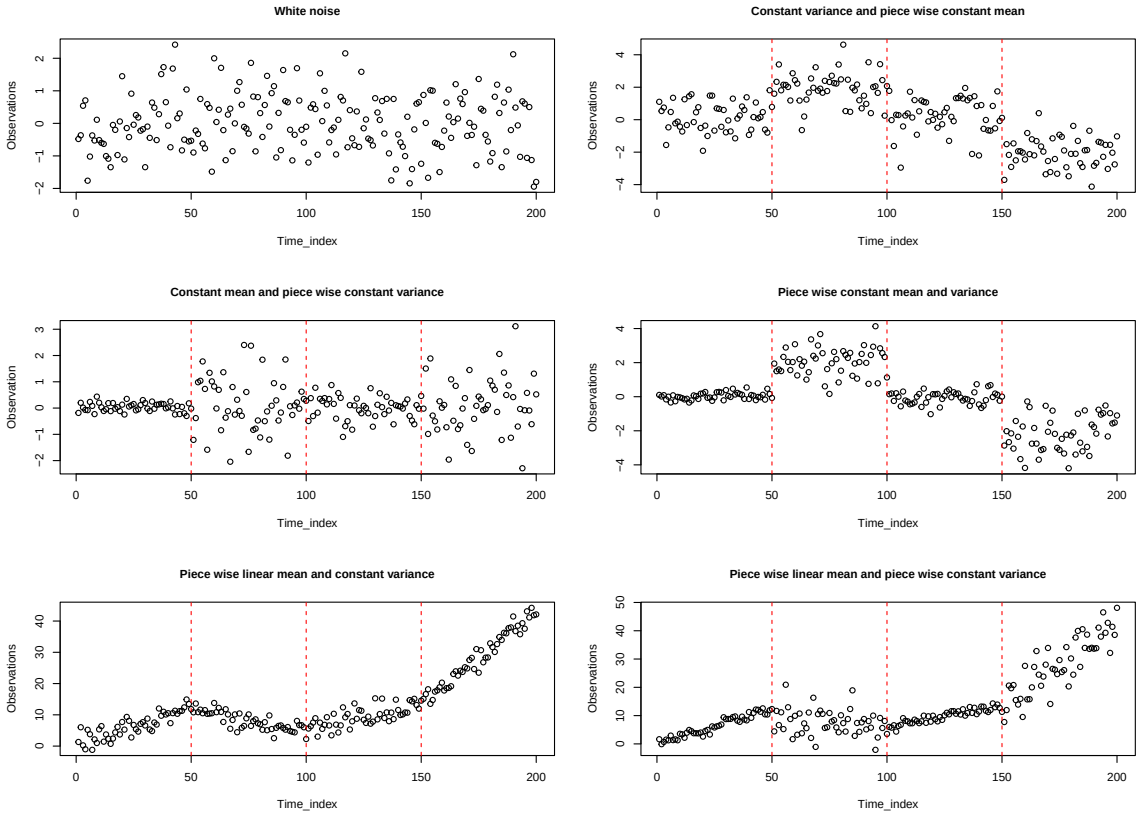

Figure 1.3: Change points in six simulated models are demonstrated by red vertical line.

new cases of COVID-19 implies the effectiveness of the public health measures that are designed to stop the spread of the disease. It is reasonable to assume that the daily new COVID 19 cases' mean function is a piecewise linear function based on Figure 1.4. The reduction in the slop of the linear models demonstrates the level of the effectiveness of the public health policies about the spread of COVID 19 disease.

The change points are the indexes of the observations such that the entire or some aspect of the underlying probability distribution which the random variables are generated from, changes afterwards. To clarify what we mean, consider the sequence of random variables $\{X_1, X_2, \cdots, X_T\}$ with the probability distribution function $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_T$, respectively. Change points are the subset of the indexes of the random variables, $\{1, 2, \ldots, T-1\}$ such that the distribution of the random variables change afterward. In other words, the set of

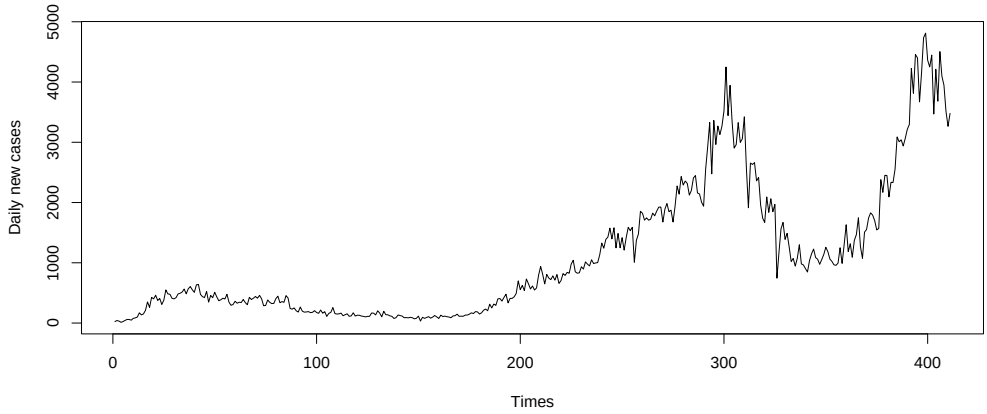

Figure 1.4: After the introduction of lockdown policies, the number of new daily cases and deaths decreased significantly.

indexes $\{\tau_1, \cdots, \tau_N\}$ of the sequence of random variables $X_1, \cdots, X_T$ are change points if

$$\mathcal{F}_1 = \cdots = \mathcal{F}_{\tau_1} \neq \mathcal{F}_{\tau_1+1} = \cdots = \mathcal{F}_{\tau_N} \neq \mathcal{F}_{\tau_{N+1}} \cdots = \mathcal{F}_T. \tag{1.1}$$

The change point definition in (1.1) is the most general from of the problem. The change point literature often assumes the random variables are generated from a common family of distribution functions and changes occur with respect to the parameters of the distribution function. In other words, the indexes $\tau_1, \cdots, \tau_N$ of the sequence of random variables $\{X_t\}_{t=1}^T$ with probability distribution function $\{\mathcal{F}_{\theta_t}\}_{t=1}^T$ (i.e, $X_t \sim \mathcal{F}_{\theta_t}$) are change points, if

$$\theta_1 = \cdots \theta_{\tau_1} \neq \theta_{\tau_1+1} = \cdots = \theta_{\tau_J} \neq \theta_{\tau_N+1} = \cdots = \theta_T. \tag{1.2}$$

One of the most popular versions of (1.2) is the normal mean change-points model. In this setting, the random variables are generated from Gaussian process with a piecewise constant mean and a constant variance. Even though, the assumptions of the normal mean change points model are too strong and seem to be unrealistic, this problem has been studied extensively. Despite this strong assumption, the normal mean change-point problem resembles some of the real-world applications of the change points analysis. One notable example is the copy number variation problem. Based on recent research, the copy number variation provides diagnostic tools and treatment regimens for cancer and other genetic diseases. Therefore, identifying the copy number variation is a crucial task. Many multiple change-point methods have been motivated by the copy number variation problem.

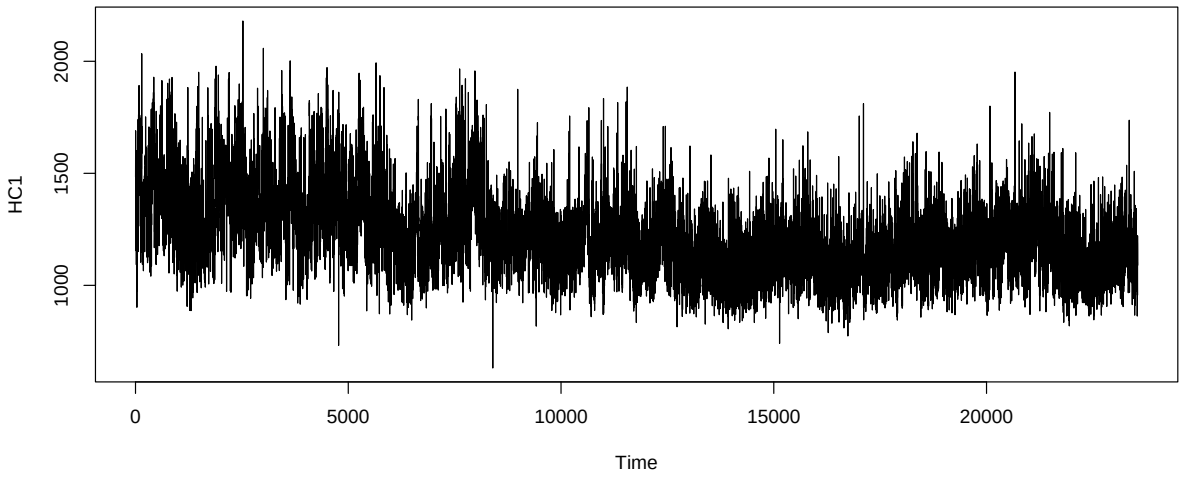

Figure 1.5: aCGH dataset of the chromosome one.

Moreover, the normal mean change point models can assess the effectiveness of the different change-point detection methods. Due to the simplicity of the normal mean change point model, if a change point method fails to provide a reasonable solution in the normal mean change point model, it would not be effective in other configurations. Moreover, many of the change point methods that were proposed for the normal mean change point model can be extended to more generalized settings with a simple modification in the algorithm.

As we previously mentioned, the change point analysis initially arose from the discipline of quality control and branched out to other applied disciplines such as biology, genomics, economy, and finance. The change point analysis can be applied to almost all data-driven fields. Due to its wide range of applications, much literature exists on the topic and several methods have been introduced over the past 70 years. With the advancements in computing, the change point problem attracted more attention. The methods that were deemed to be computationally infeasible are more manageable now. Despite of a significant progress in change point problem, many aspects of the change points analysis have remained untouched, and further investigation is required. In Chapter 2, some of the popular change point methods are introduced and reviewed. In Chapter 3, we introduce the sweeping process and enhanced backward selection and then evaluate our proposed methods under the assumption of the normal mean change-points model. In Chapter 4, we extend

the enhanced backward detection to a non-Gaussian setting using rank-based procedures. In Chapter 5, a different multiple change point detection method which we call narrowest over threshold with interval selection via shorten exhaustive search is introduced. Finally in Chapter 6, we introduce the hybrid enhanced backward detection which incorporates both enhanced backward detection and interval selection via shortest exhaustive search methods.

Chapter 2

A selective overview of off-line change point detection techniques

Early change point analysis literature assumed the presence of at most one change point within the sequence of random variables. Thus, many change point detection procedures such as CUSUM statistic Page (1954, 1955) and likelihood-based methods Hinkley (1970) were devised to tackle this particular problem. Despite the appeared simplicity of the single change-point problem, this problem is still an active area of research Horváth and Rice (2014), Horváth et al. (2020) and Barassi et al. (2020) . With the advances in computing, the focus has been shifted from a single change-point to more complicated multiple change-point problems. Although the multiple change-point research has dominated the change-point literature in recent years, single change-point literature maintains its relevance, since it is a building block of many multiple change point detection techniques.

Multiple change-point detection techniques are concerned with estimating the number of change points and their location. They consist of search methods that estimate locations of the change points and a stoppage rule that estimates the number of change points. One may categorize multiple change point detection methods to local and global search methods. Local search methods estimate each change point individually by conducting a single change point detection test concerning a segment of data. On the other hand, global search methods estimate change points simultaneously by minimizing a predefined loss function. Another concern of multiple change point detection techniques is the estimation of the number of change points. In global search methods, the number of change points is decided by adding a penalty function to the underlying loss function to avoid overestimation. While in local search methods, we set up a threshold based on the probability distribution of a single change point statistic such as CUSUM. If the magnitude of a change point test

statistic with respect to a certain interval exceeds a predefined threshold, then the presence of change point within the aforementioned interval is confirmed.

One of the primary concerns of local search change point detection techniques is conducting a change point test within a long interval of the random variables that contains precisely one change point. If the search domain of a single change point test is contaminated with more than one change points, the test may fail to detect any change in the distribution of the random variables. For instance, if the change points are too close to each other, they may offset each other and become invisible to single change point tests. Local search methods are grouped into top-down and bottom-up categories. Top-down change point detection techniques are initiated under the assumption that no change point exists. Then throughout the procedure, the change points are estimated sequentially by a single change point test. One of the early examples of top-down change point detection is binary segmentation Scott and Knott (1974). Binary segmentation conducts a single change point test on the entire sequence of random variables. If no change point is detected, then the procedure is stopped. Otherwise, the data is segmented with respect to the recently detected change point. A similar process is repeated within each segment. The binary segmentation procedure is continued until no new change point can be detected. In binary segmentation, a single change point detection procedure is conducted regardless of the number of change points in the search area. Thus, the binary segmentation fails to detect the change points that are too close to each other. One of the attractions of binary segmentation is its computational efficiency. Thus, the binary segmentation is still popular. Building upon the success of binary segmentation, different variations of binary segmentation such as circular binary segmentation Olshen et al. (2004), wild binary segmentation Fryzlewicz et al. (2014) and narrowest over threshold Baranowski et al. (2019) have been introduced over the years.

Unlike top-down methods, the bottom-up change point detection methods assume every single index of the random variables is a potential change point. Then change point candidates are eliminated sequentially until a desirable set of change point candidates is obtained.By desirable set, we mean a set of change points candidate which satisfied a predetermined criteria. At the early stages of the bottom-up methods, the change point candidates are assessed based on the relatively narrow segment of the data. Thus, the possibility of the removal of the true change points is diminished.

2.1 Single Change Point Test

As we mentioned previously, early change-point literature focused mainly on the single change-point scenarios in which at most one change-point occurs within the sequence of random variables. In other words, assuming the sequence of random variables $\{X_t\}_{t=1}^T$ are generated from the probability distribution function $\{F_t\}_{t=1}^T$, respectively, the single change-point detection problem aims to verify the existence of the point τ where

$$F_1 = \cdots = F_\tau \neq F_{\tau+1} = \cdots = F_T. \tag{2.1}$$

When the existence of a change point is confirmed, the location τ is estimated. The majority of the single change point literature narrows down the focus to the special case that the change in distribution occurs in terms of a parameter vector. That is, considering the sequence of random variables $\{X_t\}_{t=1}^T$ are generated from the common parametric family of the distribution F with parameter vectors $\theta_1, \cdots, \theta_T$, respectively. Then the change point tests are designed to conduct the following hypothesis testing:

$$\mathcal{H}_0 : \theta_1 = \cdots = \theta_T \quad \text{versus} \quad \mathcal{H}_1 : \theta_1 = \cdots = \theta_\tau \neq \theta_{\tau+1} = \cdots = \theta_T. \tag{2.2}$$

If the null hypothesis is rejected, then the location of τ is estimated. One of the most exhaustively studied variations of the problem (2.2) is the Gaussian mean change-point problem. In a single Gaussian mean change-point problem, the random variables are generated from Gaussian process with a constant known variance and mean function μ_t, where

$$\mu_1 = \cdots = \mu_\tau \neq \mu_{\tau+1} = \cdots = \mu_T. \tag{2.3}$$

With the advancements in computing, the change points literature has gradually shifted its attention to multiple change point problems. However, despite the recent popularities of the multiple change-point detection problems, the single change point analysis retains its importance. They are the main ingredient of many of the recent multiple change point detection techniques. Thus, single change point methods are still worth exploring.

2.1.1 CUSUM Statistic

As we mentioned, change-point analysis arose from the quality control discipline. The goal was to estimate the point of time in which the quality of the products started to deteriorate. Page (1954, 1955) proposed a CUSUM statistic to tackle this problem. The earliest variations of the CUSUM statistic were designed to detect a change-point in the

mean of the sequence of independently distributed random variables. Consider the sequence of independently and normally distributed random variables $X_1, \ldots, X_T$ with mean μ_t and a known common variance σ^2. Then the CUSUM statistic is calculated as follow:

$$\mathcal{D}(i,T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{i} \left(X_t - \frac{1}{T} \sum_{\ell=1}^{T} X_\ell \right), \tag{2.4}$$

Assuming exactly one change point lies within the sequence of random variables, the point in which the partial cumulative sum deviates the most from zero is highly likely to be the mean change point. In other words, if there is exactly one change point within the sequence of random variables $X_1, X_2, \ldots, X_T$ and

$$\tau = \underset{0<i<T}{\operatorname{argmax}} \mid \mathcal{D}(i,T) \mid, \tag{2.5}$$

then τ is assigned to be the estimated location of the change point. If the value of the CUSUM statistic deviates substantially from zero, then the CUSUM test confirms the change in the mean of the random variables. In other words, if the CUSUM statistic (2.5) exceeds a predefined threshold, then the change in the mean of the random variables is verified. A suitable threshold is obtained based on the probability distribution of CUSUM statistic under the assumption that no change point lies within the sequence of random variables. Assuming that a sequence of random variables is generated identically from a Gaussian distribution, the CUSUM statistic converges in distribution to a Brownian Bridge process. Thus, we can construct the threshold function accordingly.

CUSUM statistic has been investigated within more complicated change-point models. Brown et al. (1975) studied the change in parameters of a regression model under the assumption that the residuals are uncorrelated. Inclan and Tiao (1994) proposed applying the CUSUM statistic to detect a change in the variance of independently distributed random variables. Bai (1994), Lee et al. (2003), and Zhou and Liu (2009) adjusted the CUSUM statistic to be applied to a sequence of random variables with dependency. For more information, we refer the reader to Aue and Horváth (2013).

2.1.2 Likelihood Based Method

Suppose observations are generated from a member of a parametric family of distributions and a change in distribution function occurs only in the underlying parameter vector. A single change point detection test statistic can be derived from the generalized log-likelihood ratio test in this setting. Nevertheless, unlike the standard log-likelihood ratio

test for a two-sample test, the point at which the distribution of the observations changes is unknown. Therefore, the generalized log-likelihood ratio (GLR) statistic for change-point detection is more complicated than the standard log-likelihood ratio, both theoretically and computationally.

Consider a sequence of random variables that are generated from a common parametric probability distribution function $f(\cdot, \cdot)$ with parameter vectors $\theta_1, \cdots, \theta_T$. Suppose the problem of interest is testing whether the parameter vector changes after the τth observation or not. The standard log-likelihood ratio test is calculated as follow:

$$\mathcal{R}(\tau) = \mathcal{R}(\tau, \theta_a, \theta_b, \theta_0, X_1, \cdots, X_t) = \log\left(\frac{f(X_1, \cdots, X_\tau; \theta_a) f(X_{\tau+1}, \cdots, X_T; \theta_b)}{f(X_1, \cdots, X_T; \theta_0)}\right), \tag{2.6}$$

where θ_a and θ_b are the parameter vectors under the assumption of the alternative hypothesis (i.e., $\theta_a \neq \theta_b$) and θ_0 is the parameter vector under the null hypothesis. If the log-likelihood ratio is substantially large, the null hypothesis is rejected and the change in parameter vector is confirmed (i.e; $\theta_a \neq \theta_b$). In the context of the change point problem, the existence and location of change point are unknown. Thus, at first, the presence of change point is required to be verified; then, if the existence of change point is confirmed, the point which maximizes the log-likelihood ratio $\mathcal{R}(\cdot)$ in (2.6) is assigned to be the change point estimate. In other words, assuming the presence of exactly one change point within the sequence of random variables $X_1, \cdots, X_T$, the location of change point is estimated as follow:

$$\widehat{\tau} = \underset{d \leq t \leq T-d}{\operatorname{argmax}}\ \mathcal{R}(t), \tag{2.7}$$

where d is the dimension of the parameter vector (i.e; $\theta \in R^d$) and $\widehat{\tau}$ is the estimated change point. Recall that, each of the parameter vectors $\theta_0, \theta_a,$ and θ_b are estimated by the maximum likelihood method. In summary, the likelihood ratio $\mathcal{R}(t)$ is calculated for all the possible change point candidates ($t = d, \cdots, T-d$) and if the generalized log likelihood ratio test is substantially large, the point that is associated with the largest $\mathcal{R}(\cdot)$ is assigned to be the change point estimate.

To demonstrate an application of the generalized log-likelihood ratio test on the change point problem, we provide the following example. Assuming a sequence of random variables is generated independently from the normal mean change point model (2.3) ; one can show that the generalized log-likelihood ratio statistic (2.7) is equivalent to:

$$\tau = \underset{1 \leq i \leq n-1}{\operatorname{argmax}} \frac{\left|\overline{X}_{1:i} - \overline{X}_{(i+1):T}\right|}{\sqrt{\frac{1}{i} + \frac{1}{T-i}}}, \tag{2.8}$$

where $\overline{X}_{a:b}$ is the sample mean of $X_a, \cdots, X_b$ segment of the random variables. In the context of the mean change-point problem, test statistic (2.8) is the building block of many recent multiple change-point detection methods such as wild binary segmentation Fryzlewicz et al. (2014) Fryzlewicz (2020*a*), narrowest over threshold Baranowski et al. (2019) and isolated point detection Anastasiou and Fryzlewicz (2019) to name a few. In (2.8), the variance is assumed to be equal to one. In practice, the variance of the random variables must be estimated with the methods which are robust to change in the mean of the random variables. Examples of such methods are median absolute deviation Hampel (1974) and kernel-based estimators Eichinger et al. (2018). Note that, median absolute deviation can be applied if the random variables are independently distributed. An alternative approach for estimating the variance is kernel based method. One of the main drawbacks of kernel-based estimators is the difficulty in bandwidth selection. Shao and Zhang (2010) proposed a self normalized single change point test statistic for the location parameter change-point model. Unlike the traditional single change point tests, the self-normalized test of Shao and Zhang (2010) utilizes an inconsistent variance estimator. Since the inconsistent variance estimator is proportional to the true variance, then the nuisance parameter of the variance is cancelled out in the limiting distribution of the single change point test statistic. The self-normalized test was incorporated within the narrowest over the threshold by Jiang et al. (2020) to track the number of daily new COVID-19 cases. The self-normalized narrowest over threshold was designed to estimate the change in the mean function of daily new cases. This particular self-normalized change point test was designed under the assumption that the mean is a piecewise linear function of time.

To verify the existence of a change point within the sequence of random variables, the test statistic (2.6) must exceed a predefined threshold. In the ordinary likelihood ratio test, for two sample test, the test statistic (2.6) converges in distribution to χ^2_d, where d is the dimension of the parameter space. In the context of the change point problem, the asymptotic distribution of the standard likelihood ratio test is not valid , since the location at which the parameter vector changes is unknown. Hawkins (1977) obtained the exact distribution of the generalized log-likelihood ratio test statistic under the assumptions of normal mean change point model (2.3) and known variance. Worsley (1979) extended the results from Hawkins (1977) to the scenario in which the variance is unknown. Yao and Davis (1986) calculated the asymptotic distribution of generalized log likelihood ratio test statistic under the normal mean change point model (2.3). Jen and Gupta (1987) investigated the change in variance of the sequence of independently and normally distributed random variables with constant mean. They also calculated the exact distribution of the generalized log-likelihood ratio test under the null hypothesis of no change. Horváth (1993) investigated the behaviour of the generalized log-likelihood

ratio test under two distinct normal change point model:

i:) change occurs in the variance, but the mean remains the same throughout the data,

ii:) change occurs in the mean and variance simultaneously.

Horváth (1993) derived the asymptotic distribution of the generalized log-likelihood ratio test statistic under the null hypothesis for both cases. Single change point analysis is also investigated in the context of multivariate data analysis. The generalized log-likelihood ratio test and its properties have been investigated in the framework of the following change-point models:

i: change in the mean vector while the covariance matrix remains constant (Srivastava and Worsley (1986)),

ii: the mean vector is constant but the covariance matrix changes (Chen and Gupta (2004)),

iii: the mean vector and the covariance matrix change simultaneously (Chen and Gupta (2011)).

Davis et al. (1995) derived the single change point test statistic from the generalized log-likelihood for the $AR(p)$ process, constructed by the weakly stationary Gaussian white noise process with the finite fourth moment. This particular single change point test statistic is designed to identify a change in variance of white noise process and parameters of the AR(p) time series process. The asymptotic distribution of the obtained single change point test statistic is derived under the assumption of no change.

2.1.3 Information Criteria

Information Criteria can also be used as a single change point detection method. As before, consider the sequence of independent random variables $\{X_t\}_{t=1}^{T}$, assume only one change point lays within the sequence of random variables. A general information criteria is:

$$\mathcal{IC}(\lambda(T;d)) = \min_{d\leq\tau\leq T-d}\Big[-\sum_{t=1}^{\tau}\log(f(X_t;\theta_a)-\sum_{t=\tau+1}^{T}\log(f(X_t,\theta_b))+\lambda(T;d)\Big], \qquad (2.9)$$

where d is the dimension of the parameter vector and $\lambda(T;d)$ is the penalty function. Note that, θ_a and θ_b are estimated by the maximum likelihood method based on the

sub-sequences $X_1, \ldots, X_\tau$ and $X_{\tau+1}, \ldots, X_T$, respectively. Moreover, the location τ of the change point is also obtained by the maximum likelihood estimate. Calculating the information criteria under the assumption that the parameter vector has not changed through out the sequence of random variables is equivalent to maximizing the log likelihood function of the identically and independently distributed random variables. Information criteria under the assumption of no change point is denoted by $\mathcal{IC}(0)$, where $\lambda(T; d) = 0$. If the information criteria of the single change point model is less than the information criteria of the zero change point model (i.e., $\mathcal{IC}(\lambda(T; d)) < \mathcal{IC}(0)$), then the presence of the change point is verified by information criteria.

One of the main considerations in the information criteria is how to choose the penalty function. The penalty function must be large enough to avoid the false discovery of a change point and small enough to be sensitive to change in distribution. One of the most popular information criteria is the Schwarz information criteria or Bayesian information criteria (BIC) Schwarz et al. (1978). The penalty function in BIC is:

$$\lambda(T; d) = d \cdot \log(T).$$

Information criteria can also be applied in multiple change-point detection problems for estimating the number of change points. For a review on applications of information criteria in the context of single change point detection, we refer to Gupta and Chen Chen and Gupta (2011).

2.1.4 Rank-based single change point test

The accuracy of the information criteria and generalized log-likelihood ratio test-based single change point detection methods depends on whether the distribution of the random variables is specified accurately or not. Therefore, non-parametric approaches are attractive alternatives to parametric methods when the specification of the underlying distribution of the random variables is not possible. This thesis focuses on the location parameter (mean) change-point model where the random variables are generated independently.

Here, we develop a new variation of the rank-based single change point test statistic from Wilcoxon sum-rank or equivalently a two sample Kruskal and Wallis one-way analysis of variance. These particular tests are designed to verify whether multiple samples are generated from the same distribution or not. Suppose we want to verify whether the sub-samples $X_1, \ldots, X_\tau$ and $X_{\tau+1}, \cdots, X_T$ have common distribution or not. Before conducting the test, the random variables are ranked as follow:

$$r_t = \#\{X_\ell \mid X_\ell \leq X_t,\ \ell = 1, \cdots, T\}. \tag{2.10}$$

Then the Kruskal–Wallis statistic with respect to the index τ is calculated as follow:

$$\mathcal{H}(\tau) = (T-1)\,\frac{\tau(\overline{r}_{1:\tau} - \overline{r}_{1:T})^2 + (T-\tau)(\overline{r}_{(\tau+1):T} - \overline{r}_{1:T})^2}{\sum_{t=1}^{T}(r_t - \overline{r}_{1:T})^2}\,,$$

where $\overline{r}_{a:b}$ is the sample mean rank of the sub-sequence $\{X_t\}_{t=a}^{b}$; that is

$$\overline{r}_{a:b} = \frac{1}{b-a+1}\sum_{t=a}^{b} r_t\,.$$

A significantly large value of $\mathcal{H}$ implies that two sub-samples are generated from two distinct probability distribution functions. With simple algebraic manipulation, following can be shown:

$$\mathcal{R}(\tau) = \sqrt{\mathcal{H}(\tau)} = \frac{|\overline{r}_{1:\tau} - \overline{r}_{(\tau+1):T}|}{\sqrt{\sigma_r^2\left(\frac{1}{\tau} + \frac{1}{T-\tau}\right)}}, \quad \text{and} \quad \sigma_r^2 = \frac{1}{T-1}\sum(r_t - \overline{r}_{1:T})^2.$$

As mentioned earlier, the location of change is often unknown and required to be estimated. Assuming exactly one change point is located within the sequence of random variables $\{X_t\}_{t=1}^{T}$, the point in which $\mathcal{R}$ is maximized is the estimated location of the change point; that is

$$\tau = \underset{0<i<T}{\operatorname{argmax}}\,\mathcal{R}(i). \tag{2.11}$$

Unlike the parametric approaches that we mentioned earlier in this section, the rank-based change point test statistic (2.11) can detect the change in mean whether the variance remains constant throughout the sequence of random variables or not Chenouri et al. (2020).

2.2 Multiple Change Point Detection Techniques

2.2.1 Estimating the Number of Change Points

In recent years, most change point analysis literature has investigated the scenarios in which the number of change points is unknown. In the context of local search methods, the number of change points is sometimes estimated by a threshold function that is obtained based on the probability distribution of a single change point test statistic (Chen

and Gupta (2011) & Brodsky and Darkhovsky (1993)). Another popular approach is information criteria.Yao (1988) demonstrated the consistency of the estimation of the number of change points by the Bayesian information criterion. Bayesian information criteria are calculated as follow:

$$\mathcal{BIC}(\mathcal{M}(K);\theta) = -\log\big(L(\mathcal{M}(K);\theta)\big) + K\ d\ \log(T), \tag{2.12}$$

where $L(\mathcal{M}(K);\theta)$ is the likelihood function of the change point model $\mathcal{M}(K)$ with K change points and d is the dimension of the parameter vector θ. The set of change point candidates that minimizes (2.12) is assigned to be the set of estimated change points. Note that, to derive a consistent estimate of the number of change points by BIC, one must bound the number of change points from above. The specified upper bound for number of change points is much smaller than the number of observations. Since the regularity conditions of the likelihood function are not satisfied in context of the change point problem, BIC may not be an effective tool for estimating the number of change points. Therefore, Zhang and Siegmund (2007) proposed a new variation of the Bayesian information criteria referred to as modified Bayesian information criteria (mBIC), which is better suited for the change-point problem.

Fryzlewicz (2020*a*) argued that threshold-based methods and information criteria do not accurately estimate the number of change points in models with frequent change points. As mentioned earlier, information criteria do not provide a consistent estimation for the number of change points when changes in the distribution of the random variables occur frequently. In other words, the number of change points is required to be much less than the number of observations. Moreover, constructing a reliable threshold function in the context of a frequent change-point model is not practical. Recall that, to build a threshold function, the probability distribution function of the test statistic is required to be estimated accurately. However, such a task is not possible in the data configuration with frequent change points. Thus, Fryzlewicz (2020*a*) proposed the steepest drop to lowest level model selection in which threshold function plays a secondary role. Therefore, overestimation or underestimation of the threshold will not be troublesome. Later on, we will describe the steepest drop to lowest level model selection in more detail alongside wild binary segmentation 2.

2.2.2 Local Search Methods

Binary Segmentation

One of the earliest multiple change-point detection methods is binary segmentation (Scott and Knott (1974)). Binary segmentation initially searches the entire sequence of observations for a single change point. If no change point is detected, then the procedure stops. Otherwise, the dataset is divided into two segments based on the detected change point. For instance, if the index τ is assigned to be the estimated change point, then the data set is segmented into two intervals $\mathcal{I}_1 = (0, \tau]$ and $\mathcal{I}_2 = (\tau, T]$. A similar search is conducted within both of $\mathcal{I}_1$ and $\mathcal{I}_2$. This procedure is continued until no new change point can be obtained. The number of change points can be estimated by either information criteria or threshold-based approaches. In the threshold-based approach, binary segmentation is continued until none of the single change point tests statistic exceeds a predefined threshold. To estimate the number of change points by BIC, the upper bound for the number of change points $K_{\max}$ is required to be specified before the launch of the procedure. Unlike binary segmentation via thresholding, at most one change point candidate enters the active set at each stage of the binary segmentation via BIC. To be more precise, the change point candidate with the largest calculated single change point test statistic enters the active set at each stage of the procedure. This process continues until all $K_{\max}$ change point candidates are collected. So far, throughout the binary segmentation, $K_{\max} + 1$ active sets are obtained including the null model. The active set with the smallest BIC is assigned to be the set of estimated change points.

In threshold based search, the main ingredient of the binary segmentation approach is the single change point test statistic. The performance of the binary segmentation depends on the power of the single change point test. As long as the true change points are sufficiently far from each other, a single change point test retains its power in the region with more than one true change point. On the other hand, when the true change points are too close to each other, they may offset one another and become undetectable to a single change point test statistic. Therefore, a change point test should be conducted within the data segment that contains one change point and is bounded by other change points. Despite this particular drawbacks of binary segmentation, it remains popular due to its simplicity and computational efficiency. Many local search methods have been developed to build upon the desirable properties of binary segmentation while improving its accuracy. The main concern of these methods is detecting the individual change points within the longest intervals of the observations that contain precisely one change point and at the same time, avoid adding to the computational complexities significantly.

Circular Binary Segmentation

Throughout the binary segmentation, change point tests are performed regardless of their corresponding intervals. That is, many of the intervals which are defined for the tests contain more than one change points. Thus, binary segmentation may fail to recover the data segments which are created by closely located change points. One of the most investigated examples of such a scenario is the copy number variation in genomics. The copy number variation has become the motivation behind many of the multiple change point detection techniques such as circular binary segmentation ((Olshen et al. (2004))). Circular binary segmentation is adjusted in a way to recover the buried narrow changed segments of the data. Initially the sequence of random variables is clustered into the following two groups: $\{X_1, \ldots, X_i, X_{j+1}, \ldots, X_T\}$ and $\{X_{i+1}, \ldots, X_j\}$. Similar to the regular binary segmentation, a single change point test is conducted to verify whether the segment $X_{i+1}, \ldots, X_j$ has a different parameter vector from the rest of the observations or not. Assuming that the random variables are distributed normally and independently with unit variance, the single change point test statistic is calculated as follow:

$$Z_{i,j} = \frac{\left| \frac{S_T - S_j + S_i}{T-j+i} - \frac{S_j - S_i}{j-i} \right|}{\sqrt{\frac{1}{T-j+i} + \frac{1}{j-i}}}, \tag{2.13}$$

where $S_\ell = \sum_{t=1}^{\ell} X_t$. The largest value of the statistic is obtained and denoted by

$$Z_C = \max_{1 \le i < j \le T} Z_{i,j} \, .$$

If the observed Z_C exceeds a predefined threshold, then the change in the mean is confirmed. For instance, if $Z_{i,j} = Z_C$ and Z_C exceeds a pre-specified threshold, then i and j are signalled as mean change points. This procedure is conducted recursively to detect the remaining change points. When the random variables are generated from the normal mean change point model, we can calculate the threshold function with Monte Carlo simulations. To decrease the computational complexity of circular binary segmentation, Venkatraman and Olshen (2007) proposed a hybrid approach to construct the threshold function in almost linear time. They demonstrate that the new version of the circular binary segmentation is significantly faster and nearly as accurate as of the original circular binary segmentation.

Wild Binary Segmentation

One of the most popular variations of binary segmentation is wild binary segmentation (Fryzlewicz et al. (2014)). Wild binary segmentation is explicitly designed for mean change-point models where the variance is constant, and the mean is a piecewise constant function of time. This procedure is initiated with drawing a large number of randomized intervals within a sequence of random variables. Next, a single change point test statistic (CUSUM statistic (2.8)) is calculated concerning each of the random intervals. We refer to the points which maximize the weighted CUSUM statistic in their respective intervals as change point candidates. The change point candidate with the most significant CUSUM statistic in the respective interval is assigned to be the estimated change point. After adding the estimated change point to the active set, the randomized intervals that contain the estimated change point are discarded. This process is stopped when all of the randomly drawn data segments are discarded. Similar to binary segmentation, the number of change points can be estimated with some variations of the Schwarz information criterion (Schwarz et al. (1978)) or thresholding. Both approaches were studied by Fryzlewicz et al. (2014) thoroughly.

One of the primary considerations of wild binary segmentation is the choice of the number of randomized intervals. As the number of observations increases, the number of the required randomized intervals increases as well. The longer the sequence of observations becomes, the more computationally complex or inaccurate the wild binary segmentation method becomes.

At each stage of the wild binary segmentation, the change point candidate with the most prominent weighted CUSUM statistic enters the active set. The hope is that, the interval with the largest CUSUM statistic contains at most one change point. However, this wishful situation may not be fulfilled. Figure (2.1) demonstrates the data configuration in which the interval with the largest weighted CUSUM statistic may contain multiple true change points. This particular flaw of the binary segmentation and wild binary segmentation may not influence their performance in the context of the mean change-point model. However, Baranowski et al. (2019) argued against the use of binary segmentation and wild binary segmentation when the mean between the change points is the polynomial function of time.

Narrowest over threshold

In wild binary segmentation, the intervals that are used for estimating the locations of change points within data may contain more than one change point. Baranowski et al. (2019) demonstrated that the accuracy of generalized likelihood ratio tests is compromised

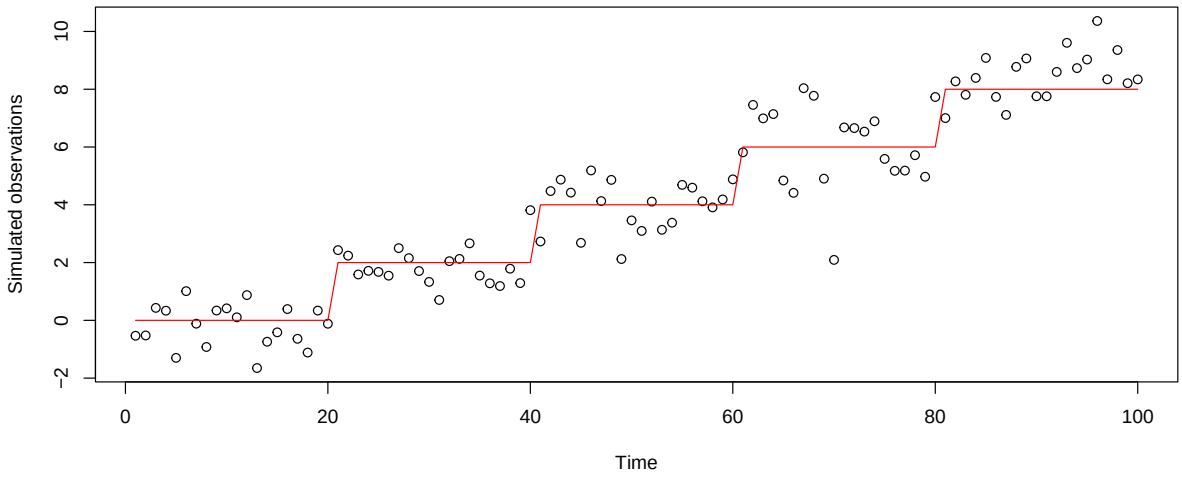

Figure 2.1: The conducted weighted CUSUM statistic concerning the entire dataset may be larger than the weighted CUSUM statistic calculated within other intervals of the data. Thus, the estimated change point is obtained from the contaminated interval.

when the search domain is contaminated with multiple change points and the mean is a piecewise polynomial function of time. Therefore, wild binary segmentation can not be extended beyond the simple piecewise constant mean change-point model. The narrowest over threshold (NOT) proposed by Baranowski et al. (2019) emphasizes extracting the estimated change points from narrow intervals.

Similar to wild binary segmentation, first, a large number of intervals are randomly drawn, and then the narrowest over threshold procedure is launched. A single change point test statistic is calculated concerning each of the randomized intervals. The intervals in which a predefined threshold is satisfied are kept, and the rest of the intervals are discarded. The narrowest interval among the recorded intervals is selected and the point that maximizes the single change point test statistic within the aforementioned interval is assigned to enter the active set. After the entrance of each estimated change point to the active set, the intervals that contain the estimated change points are eliminated. This procedure is continued until no intervals are left. Since the change points are extracted from the narrowest available interval, the possibility of inference on the location of change points within the intervals that are contaminated with multiple change points is diminished

significantly. Thus, the narrowest over threshold can be applied to a significantly larger spectrum of the change point problems in comparison to binary segmentation or wild binary segmentation. Baranowski et al. (2019) primarily investigated the performance of narrowest over threshold concerning several change point models under the assumption that random variables are generated independently. Jiang et al. (2020) conducted the narrowest over threshold to detect change in trends of the new COVID-19 cases. Due to the dependency of daily new cases, Jiang et al. (2020) replaced the generalized likelihood ratio test with a self-normalized test (Shao and Zhang (2010)). This particular variation of the narrowest over threshold is referred to as self-normalized narrowest over threshold. The threshold for self normalized narrowest over threshold is obtained by parametric bootstrap resampling as follow:

- A large number of the replications of the Gaussian white noise are generated ;
- For each of the generated replications,
  - The self normalized change point test statistic is calculated with respect to the sufficient number of the randomly drawn intervals;
  - The largest value of the calculated test statistic with respect to the randomized intervals in the replication is recorded;
- The $(1 - \alpha)$ percentile of the recorded test statistic values is assigned to be the threshold function.

The practitioners' value of α is chosen depending on the desired sensitivity to the false discovery of change points.

Even though the risk of estimating the change points within the interval that contains multiple change points is reduced significantly in narrowest over threshold, these intervals are still far from ideal. For instance, due to the short length of this interval, the estimated location of the change point may lack sufficient accuracy. Moreover, conducting the generalized likelihood ratio test in narrow intervals is prone to misdiagnosing the fluctuation of the observations with the change in the distribution of the random variables. The vulnerability of narrowest over threshold to the false discovery of change points is demonstrated in the following simulation study:

- M sequence of identically and independently distributed random variables with T observations are generated and each sequence is denoted with $\{X_t^m\}_{t=1}^T$ for $m = 1, \cdots, M$.

- For each of the M replications, do the following:
  - The (weighted) CUSUM test (2.8) is conducted for the entire sequence of observations. The value of the test statistic is recorded and referred to as a global test value.
  - B randomized intervals are drawn, and the single change point test (2.8) is conducted within each of the randomized intervals.
  - The largest value of underlying single change point statistic that is calculated within the randomized intervals is recorded and called the local test value.

At the end of this process, two separate histograms of the local CUSUM and global CUSUM statistics are plotted in Figure 2.2. Based on Figure 2.2, the global CUSUM statistic is less likely to pick the variation in random noises as the change in the distribution of the random variables in comparison to the local test. That is, the global test is less susceptible to false discovery of change points.

MOSUM and its variants

The primary ingredient of local search methods is the single change point test statistic that recovers the individual change points. Some of the multiple change-point methods, such as binary segmentation, estimate changes points by maximizing a single change point test statistic within data segments. An alternative approach among local search methods is the MOSUM (moving sum) statistic (Hušková and Slabỳ (2001)). For convenience, we describe an application of the MOSUM statistic in the normal mean change-point model. Recall that, the normal mean change-point model is the scenario in which a sequence of independent random variables is generated from a Gaussian process with piecewise constant mean and unit variance. Suppose we want to discover the points in which the mean of the random variables changed. The MOSUM procedure starts with guessing the minimum distances among the true change points. Next, the new parameter called bandwidth (h) is specified. The length of the bandwidth is ideally half of the minimum distance among true change points. The MOSUM statistic is calculated as follow:

$$\mathcal{D}(i,h) = \sqrt{\frac{h}{2}} \left| \overline{X}_{(i-h+1):i} - \overline{X}_{(i+1):h} \right| , \quad \text{for} \quad i = h, \ldots, T-h,$$

where $\overline{X}_{(a+1):b} = \frac{1}{b-a+1} \sum_{j=a+1}^{b} X_j$ and h is the bandwidth. All data points with MOSUM statistic that exceeds a predefined threshold are recorded. Then among the collected

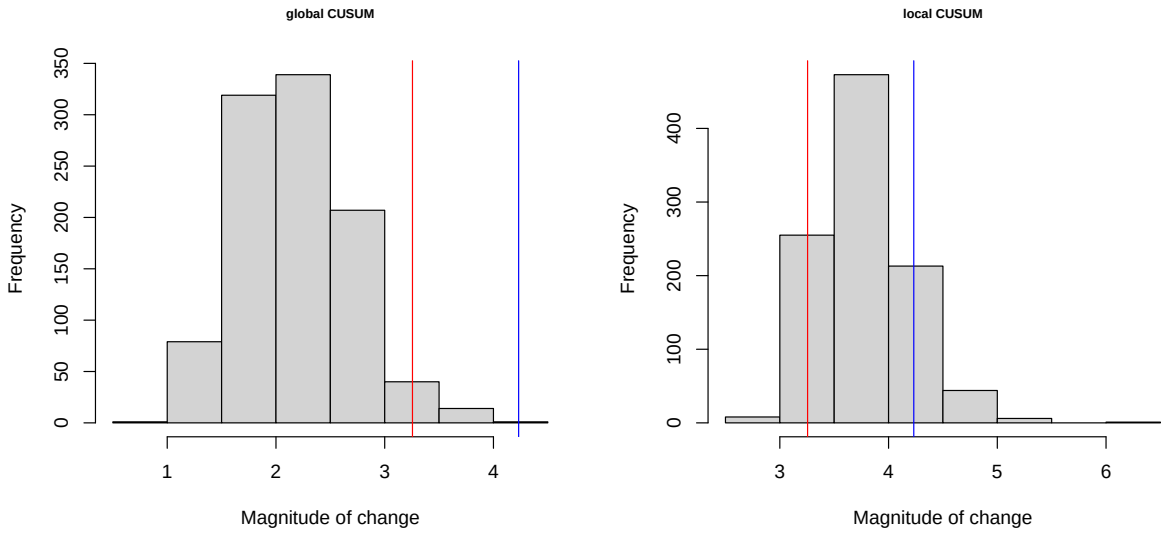

Figure 2.2: The global and local CUSUM statistics are plotted in two separates histograms. The blue and red lines represent the threshold functions $\lambda_1 = \sqrt{2\log(T)}$ and $\lambda_2 = 1.3\sqrt{2\log(T)}$, respectively. Based on these histograms, a larger portion of the calculated local CUSUM statistic exceeds both thresholds compared to the global CUSUM statistic. Thus, methods such as narrowest over threshold and wild binary segmentation are more likely to overestimate the number of change points.

points, the indexes such as τ is selected to be the change point if $\mathcal{D}(\tau, h) > \mathcal{D}(\eta, h)$ for every $\eta \in (\tau - h, \tau + h)$. Selection of the bandwidth h is very important. The MOSUM with a short bandwidth lacks the power to detect the change points. On the other hand, if the bandwidth is too wide, more than one change point may hide within the intervals defined for the MOSUM statistic. Subsequently, change points may offset or mask each other and become undetectable to the MOSUM statistic.

When the distance between the adjacent change points varies, a single bandwidth MOSUM statistic cannot recover all the change points. Thus, Multi-Filter algorithm (Messer et al. (2014)), and Multi-Scale Localized Pruning (Cho and Kirch (2019)) incorporate multiple single bandwidth mosum and aggregate the estimated change points. In the multi-filter algorithm, all change point candidates extracted by the MOSUM statistic with the shortest bandwidth are selected to be change points. Then all change point candidates that are within the bandwidth of the chosen change points are discarded. This process continues until no change point candidates left. Like narrowest over threshold algorithm, the multi-filter algorithm prioritizes the change point candidates that are extracted from shorter segments of data. Hence the estimated change points may not be accurate due to insufficient data points within segments. Moreover, when bandwidth is too short, random noise fluctuation may be mistaken with a change in mean as demonstrated in Figure 2.2.

Localized pruning was proposed as an alternative approach to combine several single bandwidth MOSUM statistics. At first, all the change point candidates alongside their corresponding bandwidths and magnitude of MOSUM statistic are recorded and sorted with respect to the magnitude of MOSUM statistic in the decreasing order. Notice that these are γ, h, $|\mathcal{D}(\gamma, h)|$, where γ is the index, h is the bandwidth, and $|\mathcal{D}(\gamma, h)|$ is the magnitude of the MOSUM statistic. Then the change point candidate with the largest MOSUM is selected and denoted by τ. All the change point candidates that are conflicted with τ are collected. Note that change point candidates η and τ with corresponding bandwidths g and h are conflicted if $\left|\eta - \tau\right| \leq \min(g, h)$. Suppose the points η_ℓ and η_r are the smallest and largest indexes that are conflicted with τ, respectively. Next, the change point candidate with the largest index to the left of η_ℓ and the change point candidate with the smallest index to the right of η_r are identified and denoted by η_L and η_R, respectively. The segment between η_L and η_R is referred to as a local environment of the change point candidate τ. This process is repeated for the remaining change point candidates. For each of the designated local environments, the subset of the change point candidates that minimizes the Bayesian information criteria is assigned to the active set. Recall that to estimate the number of change points with BIC consistently; we must specify the maximum number of change points. However, in the localized pruning algorithm, the upper bound for the number of change points is not required to be specified. Even though

the localized pruning seems to be more innovative and sophisticated compared to the multi-filter algorithm, this algorithm is extremely flawed. As the number of bandwidth increases, the computational time increases significantly and may even become infeasible. Moreover, the accuracy of the multiple change-point approaches decreases significantly if every single possible bandwidth is incorporated into the algorithm. The latter claim will be demonstrated in the simulation study section in the Chapter 3.

Wild Binary Segmentation 2 and Steepest Drop to Lowest Level

Recall that, in wild binary segmentation, a large number of randomized intervals are drawn prior to the launch of the procedure. This way, the computational complexity of the procedure remains independent from the number of change points. Unfortunately, this approach is not practical in frequent change-point models. Therefore, Fryzlewicz (2020*a*) proposed an alternative sampling approach that can be effective regardless of the frequency of the change points. Instead of drawing the randomized intervals prior to the start of the change point analysis, a smaller number of randomized intervals are drawn from each of the segments that are obtained throughout the procedure.

Consider the sequence of random variables $\{X_t\}_{t=1}^{T}$ that may contain change points. To locate the first change point candidate, we draw a relatively small number of randomized intervals. Next, we calculate the underlying single change point test statistic for each of the randomly drawn intervals. The point that is associated with the largest value of the single change point test statistic is declared as a change point candidate. Subsequently, the dataset is segmented into two partitions with respect to the recently obtained change point candidate. Then similar search is conducted concerning the newly constructed segments. This procedure is continued until no new sub-samples can be drawn. That is, there is no segment between the change point candidates with more than one data point. In other words, every observations is a change point candidate.

Wild binary segmentation 2 was designed to locate the undetected change points efficiently regardless of the frequency of the change points. Furthermore,Fryzlewicz (2020*a*) argued that the traditional tools for estimating the number of change points are not effective in data configuration with frequent change points. For instance, information criteria are proven to be too conservative in models with frequent change points. Moreover, obtaining a desirable threshold function is not practical in a dataset with many change points. For instance, a good quality variance estimate is critical to get a reliable threshold in the normal mean change-point model. In the data configuration with a low number of mean changes, the median absolute deviation can estimate the variance with a reasonable accuracy as long as the random variables are generated independently. However, variance

is often overestimated by median absolute deviation in the setting with frequent change points. Subsequently, the constructed threshold will become too large, and the single change point test becomes less sensitive to change in mean. Thus Fryzlewicz (2020*a*) proposed a new approach called Steepest Drop to Lowest Level. Fryzlewicz (2020*a*) claimed since the threshold function plays a secondary role in estimating the change points in this method, a poorly chosen threshold function will not decrease the accuracy of the estimated number of change points.

The change point candidates which were obtained by wild binary segmentation 2 are recorded alongside with their respective intervals that they are extracted from and the values of the test statistic. Then the change point candidates are sorted in the decreasing order of the test statistic values. Suppose b_k is the change point candidate with kth largest single change point test statistic value which is extracted from the interval $(s_k,\, e_k]$, and the quantity $\mathcal{U}(b_k|\, s_k,\ e_k))$ is the maximized change point test statistic with respect to the interval $(s_k,\, e_k]$, then b^k is recorded in the following row vector:

$$(b_k,\, s_k,\, e_k,\, \mathcal{U}(b_k|\, s_k,\, e_k))\,, \tag{2.14}$$

where $\mathcal{U}(b_1|\, s_1,\ e_1)) > \cdots > \mathcal{U}(b_{T-1}|\, s_{T-1},\ e_{T-1}))$. The difference of logarithm of the consecutive test statistic is calculated. Suppose wild binary segmentation 2 is conducted within the sequence of random variables $\{X_t\}_{t=1}^{T}$ and $T-1$ change point candidates are obtained. These change point candidates are recorded in the row vector (2.14) as mentioned earlier and then

$$\mathcal{Z}_K = \log\big(\mathcal{U}(b_k|\, s_k,\, e_k))\big) - \log\big(\mathcal{U}(b_{k+1}|\, s_{k+1},\, e_{k+1}))\big)\,.$$

Suppose Z_ℓ is associated with the largest difference. If $\mathcal{X}_{s_{l+1},e_{l+1}}^{b_{l+1}}$ is smaller than the pre-specified threshold, then set of estimated change points is $\{b_1,\, \ldots,\, b_\ell\}$. Otherwise, the second largest drop in the difference logs is identified, and the single change points test statistic associated with the drop is inspected. Suppose $\mathcal{Z}_d$ represents the largest drop, where $\mathcal{U}(b_{d+1}|\, s_{d+1},\, e_{d+1}))$ is smaller than a predefined threshold, then all the change point candidates such as b_i with the single change point test statistic $\mathcal{U}(b_i|\, s_i,\, e_i)) > \mathcal{U}(b_{d+1}|\, s_{d+1},\, e_{d+1}))$ are assigned to be the change points.

One of the main features of wild binary segmentation 2 is the re-sampling method which generates the randomized intervals. This modification can increase the accuracy of the estimated change points in comparison to wild binary segmentation. Note that the re-sampling process applied in wild binary segmentation 2 can also be incorporated in the narrowest over threshold method.

Detecting Multiple Change Points With Isolating Single Points

Isolated Detection methods byAnastasiou and Fryzlewicz (2019) is a novel approach that is computationally efficient and avoids the unpredictability of the randomized intervals in methods such as narrowest over threshold or wild binary segmentation. At first, the segments $(0,\ L)$ and $(T-L,\ T)$, where $0 < L < T$ are searched. Suppose no change point is detected, then similar search is repeated within $(0,\ 2L)$ and $(T-2L,\ T)$ intervals. Let say the first change point is identified within the interval $(0,\ 2L)$, then the next two segments that are searched for the change points are $(2L,\ 3L)$ and $(T-3L,\ T)$. This procedure is continued until the search areas from both sides coincide, and no further interval is available. Note that, the change point location is detected within the segments of the data that are not symmetrical, and one side of the true change point may contain more data points than the other side. Thus, the inference on the position of change point may not be accurate enough.

Narrowest Significant Pursuit

Narrowest Significant Pursuit (Fryzlewicz (2020*b*)) was proposed by Fryzlewicz to locate the confidence regions of the change points within the sequence of random variables. Similar to other local search methods, the single change point test is derived based on the assumed change-point model. In the same fashion as wild binary Segmentation 2, several randomized intervals are drawn at the start of each stage of the procedure. Then a single change point test statistic is calculated within each of the randomly drawn intervals. The narrowest randomized interval in which a single change point test indicates to presence of a change point is identified. Multiple sub-intervals are drawn with respect to the selected interval and the single change point test is conducted within each of the randomly drawn intervals. Among the sub-intervals in which the change point test confirms the alteration of the distribution, the narrowest one is assigned to be the confidence region. The confidence regions of change points may overlap with each other. The extent that confidence regions coincide with each other is allowed by parameters d_ℓ and d_r. Suppose the interval $(s,\ e)$ is assigned to be the first confidence region. Then similar search for the confidence regions is conducted within the interval $(0,\ s+d_\ell)$ and $(e-d_\ell,\ T)$. This procedure stops when no new confidence region is obtained.

Wild Energy Maximization with Gappy Schwarz Criterion

Another local search method that relies on randomly drawn intervals is Wild Energy Maximization with Gappy Schwarz (WEM.GSC) Cho and Fryzlewicz (2020). The main concern of WEM.GSC is to estimate the locations of mean change points within the sequence of the random variables with serial dependencies. That is, the goal of the WEM.GSC is to identify the indexes $\{\tau_1, \cdots, \tau_J\}$ within the sequence of the random variables that is generated from:

$$X_t = \mu_t + \epsilon_t, \tag{2.15}$$

where $\mu_{\tau_j} \neq \mu_{\tau_j+1}$ and ϵ_t is a mean zero $AR(p)$ stationary process. One of the main difficulty of detecting mean change points in model (2.15) is that the fluctuations of the random noise are likely to be mistaken by mean change points. Hence, the long range variance of the process (2.15) is required to be estimated. Two general approaches are the kernel based methods and self normalizing test. One of the challenges associated to kernalized variance estimator is selecting the length of the bandwidth. On the other hand, self normalized change point test is not concerned with determination of any nuisance parameter such as bandwidth. Unfortunately, the consistency of the self normalized change point test has not been investigated sufficiently outside the single change point setting.

The WEM.GCP consists of two components: (i) wild energy maximisation (WEM) which searches for the locations of mean change points and (ii) the gappy schwarz criterion (GSC) that estimates the number of change points. The intervals generation mechanism in the wild energy maximisation is similar to wild binary segmentation 2 (Fryzlewicz (2020*a*)). That is, relatively small number of the randomized intervals are generated and subsequently a weighted CUSUM statistic (2.8) is calculated with respect to each interval. The indexes which maximize the change point test statistic within the aforementioned intervals alongside their corresponding intervals and the maximized change point test statistic are recorded in the row vector. For instance, suppose the index τ maximizes the change point test statistic within the interval with lower bound α and upper bound β and $\mathcal{U}(\tau|\alpha, \beta)$ is the maximized test statistic with respect to the interval $(\alpha, \beta]$, then a row vector

$$[\tau,\ \alpha, \beta, \mathcal{U}(\tau|\alpha, \beta)]$$

is recorded. Suppose the maximized change point test statistic which corresponds to the index τ is larger than rest of the recorded change point test statistic in the recorded row vectors, then the sequence of the random variables is divided to two segments which are jointed by the index τ. Subsequently , similar procedure continues within the sub-intervals $X_1, \cdots, X_\tau$ and $X_{\tau+1}, \cdots, X_T$. That is, with respect to each of these segments:

- Relatively small number of randomized intervals are drawn. These intervals are denoted by $\mathcal{I}$;
- The change point test statistic (2.8) is calculated for each of the randomized intervals in the set $\mathcal{I}$;
- Suppose the change point test statistic $\mathcal{U}(\tau^*|\alpha^*, \beta^*)$ is larger than all of the maximized change point test statistic within the interval set $\mathcal{I}$, then the row vector

$$[\tau^*,\ \alpha^*,\ \beta^*,\ \mathcal{U}(\tau^*|\alpha^*, \beta^*)] \tag{2.16}$$

  is collected;
- The entire interval is segmented with respect to the index τ^*;

The WEM.GSC continues until no more randomized intervals with more than one observation can be drawn.

Similar to many of the local search methods, the number of change points is estimated by Schwartz information criterion in WEM.GSC. Recall that, in order to estimate the number change points consistently by information criteria, the upper bound for the number of change points denoted by $K_{\max}$ is required to be specified by the practitioner. Given the time series dependencies in the change point model (2.15), calculating the information criteria for all of the $K_{\max}+1$ is an infeasible task. Fortunately, Cho and Fryzlewicz (2020) proposed a modification to the Schwarz Criterion that reduces the change point model space. This variation of Schwarz criterion is named gappy Schwarz criterion (GSP).

In order to implement the Gappy Schwarz Criterion, the recorded row vectors (2.16) are collected in the matrix $\mathcal{W}$ in the manner that the row vectors with the larger change point test statistic are located in the top rows of the matrix $\mathcal{W}$. Furthermore, the elements of the matrix $\mathcal{W}$ are relabeled with respect to the magnitude of the recorded change point test statistic . That is, the i-th row of the matrix $\mathcal{W}$ is

$$[\tau^{(i)},\ \alpha^{(i)},\ \beta^{(i)},\ \mathcal{U}(\tau^{(i)}|\alpha^{(i)}, \beta^{(i)})],$$

where $\tau^{(i)}$ maximizes the change point test statistic within the sub-sequence of random variables $X_{\alpha^{(i)}+1}, \cdots, X_{\beta^{(i)}}$ and $\mathcal{U}(\tau^{(i)}|\alpha^{(i)}, \beta^{i})$ is the ith largest recorded change point test statistic. Then, the change point test statistic is replaced with the log of the change point test statistic (2.8); That is ,

$$\mathcal{Y}(\tau^{(t)}|\alpha^{(t)}, \beta^{(t)}) = \log\left[\mathcal{U}(\tau^{(t)}|\alpha^{(t)}, \beta^{(t)})\right], \quad \text{for} \quad t = 1, \cdots, T. \tag{2.17}$$

The recorded indexes in the matrix $\mathcal{W}$ are referred to as change point candidates. Either double CUSUM statistic (DC) (Cho (2016)) or the largest difference (LD)(Fryzlewicz (2020*a*)) are applied to decide which of the candidates set to be in the model space. The aim of the LD and DC is to identify the indexes in the matrix $\mathcal{W}$ after which the magnitude of the change point test statistic has dropped the most. Next, these sets are assigned to be the candidate sets in the model space to be defined for GSC. Note that, the number of the indexes in the candidate sets are required to be smaller than the predefined upper bound of the number of change points. To calculate the Schwartz Criterion, the $AR(p)$ time series model is fitted, where the order p of the autoregressive process is unknown. The parameter of the AR process are estimated with respect to the longest segment in the proposed change point model. Among the available change point candidate sets in the model space and the null model, the one with the smallest Schwartz criterion is assigned to be the estimated change point set.

Since the wild energy maximisation prioritize change point candidate with the largest change point test statistic value, the possibility of the false discovery change points is reduced significantly. Furthermore, by estimating the number of change points with gappy Schwartz criteria, the long range variance is not required to be estimated. In other words, the problem is reduced to obtaining the variance of the white noise which constructs the linear process.

Seeded Binary Segmentation

The common theme of local search methods is to obtain intervals of sequence of random variables which are relatively long and contains exactly one change point. For instance, wild binary segmentation and narrowest over threshold attempt to obtain such intervals by constructing large number of randomized intervals. Seeded binary segmentation (Kovács et al. (2020)) proposes an alternative approach to search for intervals in a deterministic manner. This approach is generic and can be applied to wide spectrum of change point models. One of the main attraction of the seeded binary segmentation is that the computational complexity of the method is independent of the data configuration. This property of the Seeded Binary Segmentation method does not hold true for methods which utilize the randomized intervals. For instance, the wild binary segmentation requires a significantly larger number of the randomized intervals in the frequent change points datasets.

The seeded binary segmentation constructs the search intervals in multiple layers, where the number of layers is defined by decay parameter a. The collection of the intervals in the k-th layer ($\mathcal{I}_k$), where

$$1 \leq k \leq \lceil \log_{\frac{1}{a}}(T) \rceil,$$

is generated as follow:

$$\mathcal{I}_k = \bigcup_{i=1}^{n_k} \{(\lfloor (i-1)s_k \rfloor, \lceil (i-1)s_k + l_k \rceil]\}, \tag{2.18}$$

where $n_k = 2\ \lceil \left(\frac{1}{a}\right)^{k-1} \rceil - 1$, $l_k = T\ a^{k-1}$ and $s_k = \frac{T - l_k}{n_k - 1}$. Moreover, a is selected by the practitioner from the interval $[\frac{1}{2}, 1)$. The overall interval collection is

$$\mathcal{I} = \bigcup_{k=1}^{\lceil \log_{\frac{1}{a}}(T) \rceil} \mathcal{I}_k. \tag{2.19}$$

Note that, $\lceil m \rceil$ are the ceiling value of m and $\lfloor m \rfloor$ is the floor value of m. As the value of the decay parameter increases, the number of constructed intervals and the computational complexity of the seeded binary segmentation increase. From the practical perspective, the decay parameter is recommended to be $\frac{1}{\sqrt{2}}$.

Rank based Wild Binary Segmentation

So far, we have discussed the wild binary segmentation method (Fryzlewicz et al. (2014)) and its variations in a parametric setting (1.2). In this problem, the sequence of the random variables is assumed to be generated from a common and known parametric family of distribution functions. However, in many real world scenarios, the distribution function is unknown and hard to approximate. Thus, rank based non-parametric approaches seems to be attractive alternatives. Ross (2021) proposed alteration of the generalized log-likelihood ratio test statistic (2.7) with a Lepage (Lepage (1971)) like statistic which combine the Mann-Whitney (Mann and Whitney (1947)) and Mood statistic Mood (1954).

Similar to wild binary segmentation (Fryzlewicz et al. (2014)) or wild binary segmentation 2, a single change point test statistic is calculated within intervals of the random variables. In other words, a rank based single change point statistic is calculated locally. To do this, we first rank the observations X_{p+1}, X_{p+2}, ..., X_q in the interval $(p,\ q]$; That is, rank of X_i with respect to the interval $(\ p,\ q]$ is:

$$r_{_i} = \sum_{j=p+1}^{q} 1(X_i \geq X_j).$$

After ranking the sequence of observations in the interval $(p,\ q]$, we compute the value of the rank change point statistic. One of the advantages of a rank based change point

detection methods is that they are non-parametric. Thus, the threshold function can be constructed in a distribution free manner. Ross (2021) constructed the threshold function by parametric bootstrap re-sampling similar to Jun Shin et al. (2020) and Jiang et al. (2020). Another variation of rank based wild binary segmentation was proposed by Padilla et al. (2019) which differs from the aforementioned rank based wild binary segmentation Ross (2021) with respect to the rank based change point test statistic and the constructed threshold.

Bottom Up Multiple Change Point Detection Methods

A common feature of all of the change point detection methods which has been discussed so far is that they are initiated with no change points and then change points are recovered sequentially by conducting a change point test. This category of the local search method is known as the top-down approach. The earliest top-down technique is binary segmentation. One of the main drawbacks of binary segmentation is searching for the individual change points within the data segments with multiple change points. Thus, binary segmentation may fail to detect change points when they are too close to each other. This particular drawback of binary segmentation has been one of the main motivations behind many more sophisticated top-down methods. The primary concern of the more recent top-down approaches has been circumventing the interference of the undetected change points with each other. These attempts lead to a rigorous search for the longest possible observation segments that contain precisely one change point. Thus the computational cost of the top-down change point detection techniques increases in comparison to binary segmentation.

Another category of local search methods is the bottom-up change point detection techniques. Bottom-up change point detection techniques initially regard every single index of observations as a potential change point candidate. Then change point candidates are sequentially eliminated from consideration until the desirable active set is obtained. Unlike the top-down approach, very few bottom-up change point detection techniques have been proposed over the years. The two examples that we know of are backward selection (Jun Shin et al. (2020)), and tail-greedy-bottom (Fryzlewicz et al. (2018)). Both of these methods follow a very similar scheme.

In this chapter, backward selection and bottom-up-tail-greedy change point techniques are investigated. The backward selection method was introduced in the context of the copy number variation. Thus, it was constructed under the assumptions of the normal mean change-point model. To decide which change point candidates to be removed from an active set, the importance of each candidate is measured by a quantity called scaled

jump size (2.20). Suppose the current set of change point candidates is $\{\tau_1, \ldots, \tau_N\}$, then the scaled jump size is calculated as follow:

$$\mathcal{U}(\tau_i|\tau_{i-1}, \tau_{i+1}) = \frac{\left|\overline{X}_{(\tau_{i-1}+1):\tau_i} - \overline{X}_{(\tau_i+1):\tau_{i+1}}\right|}{\sqrt{\frac{1}{\tau_{i+1}-\tau_i} + \frac{1}{\tau_i-\tau_{i-1}}}} \quad \text{for} \quad i = 1, \cdots, N, \tag{2.20}$$

where $\tau_0 = 0$ and $\tau_{N+1} = T$. At each run, the change point candidate with the smallest scaled jump size is removed from the active set. Subsequently, the scaled jump size of the change point candidates adjacent to the eliminated candidates are updated according to the renewed active set. Backward selection stops when the scaled jump size of the remaining change point candidates exceeds a predefined threshold. The threshold is constructed by estimating the probability distribution of the scaled jump size with a bootstrap re-sampling. That is, multiple replications of Gaussian process with constant mean and variance are generated. The mean of the random variables is estimated by maximum likelihood methods. However, the variance must be estimated by methods which are robust to change in mean of the random variables. The examples of such methods are kernel methods and median absolute deviation. Then multiple replications of the random variables are generated from Gaussian distribution with the estimated mean and variance. Backward selection is executed for each of the replicated sequences of random variables, and the largest calculated scaled jump size is recorded. Among the recorded scaled jump size, $(1 - \alpha) \times 100$th percentile of the scaled jump size is assigned to be the threshold. The value of $\alpha \in (0, 1)$ is decided by the practitioner.

Another variation of the bottom-up search methods is Tail Greedy Bottom Up, in which, unlike backward selection, more than one change points may be eliminated after each run as long as each segment is merged at most once. Prior to launch of the algorithm, maximum number of change point candidates which are eliminated in each run is specified and denoted by $r_{\max}$. The set of change point candidates which are obtained after the jth run is denoted by $\mathcal{M}(N(j)) = \{\tau_1^{(j)}, \cdots, \tau_{N(j)}^{(j)}\}$, where $N(j)$ is the number of change point candidates after the jth run. For example, every single observations is initially regarded as a potential change point candidate; thus, the designated active set prior to launch of the algorithm is $\mathcal{M}(N(0)) = \{1, \cdots, T-1\} = \{\tau_1^{(0)}, \ldots, \tau_{N(0)}^{(0)}\}$, where $N(0) = T - 1$. One major difference between the backward selection and bottom-up-tail-greedy is the way each change point candidate is assessed. Suppose the current active set is $\mathcal{M}(N(j))$ and $\tau_\ell^{(j)}$ is getting inspected. The importance of $\tau_\ell^{(j)}$ is inspected by the scaled jump sizes which were calculated within the sub-intervals of $(\tau_{\ell-1}^{(j)}, \tau_{\ell+1}^{(j)}]$ in the previous runs. Then the largest scaled jump size is recorded and denoted by $\mathcal{U}(\tau_\ell^{(j)})$. That is:

$$\mathcal{U}(\tau_\ell^{(j)}) = \max_{i=0,\ldots,j} \mathcal{U}(\tau_\ell^{(j)}|\tau_a^{(i)}, \tau_b^{(i)}), \tag{2.21}$$

where $\tau_a^{(i)}$ and $\tau_b^{(i)}$ were adjacent change point candidates to $\tau_\ell^{(j)}$ in the active set that was obtained after the ith run, for $i = 0, \ldots, j$. At each run of of the bottom-up-tail-greedy approach, up to $r_{\max}$ change point candidates with the smallest scaled jump size are eliminated from the active set. Moreover, test statistic (2.21) of all the eliminated change point candidates are smaller than a predefined threshold and none of them were adjacent with each other in the previous active set. The mth run would be the last run, if either

$$\mathcal{U}(\tau_\ell^{(m)}) \geq \lambda \quad \text{for} \quad \ell = 1, \ldots, N(m),$$

for a predefined threshold λ or no change point candidate left in the active set.

Both backward selection and bottom-up-tail-greedy follow the same scheme more or less with some minor differences. One of the most critical concerns of both methods is deciding on the threshold function. The threshold function is constructed based on the estimated probability distribution of the scaled jump sizes. If the selected threshold is oversized, then the backward selection is more likely to underestimate the number of change points compared to bottom up tail greedy. Assuming the scaled jump size of a true change point is less than a predefined threshold, the true change point is eliminated from the active set according to the backward selection procedure. Unfortunately, the search domains defined for the true change points adjacent to the eliminated true change point are contaminated with multiple change points. Therefore, the scale jump size may become less pronounced and subsequently backward procedure may eliminate more true change points. Thus, bottom up tail greedy method is more robust to oversized threshold. On the other hand, since the bottom-up-tail-greedy applies local change point test, it is prone to misdiagnosis of fluctuation of the randomized noise with change in the distribution of the random variables. As demonstrated in (2.2), the narrowness of the segment in which the scaled jump size is calculated leads to an overestimation of the scaled jump size and even a false discovery of the change points.

2.2.3 Global Search Methods

In global search methods, the locations of change points are estimated simultaneously and often by finding an optimal solution for a minimization problem. Note that the objective function of the optimization problem is devised based on the assumed change-point model. Here we categorize global search methods into exact search methods and LASSO-type estimator. In the exact search methods, the set of change point candidates that minimizes an objective function is assigned to be the set of estimated change points. The search for the optimal solution is ideally conducted by comparing all possible sets of change point

candidates. The candidate change point set, which minimizes the objective function, is assigned to be the set of estimated change points. This process is proven to be highly time-consuming and even infeasible for large datasets. Thus, exact search methods utilize the dynamic programming algorithm to reduced the computational complexity. The prominent examples of such a method are segmented neighbourhood search algorithm (Auger and Lawrence (1989)), and optimal partitioning algorithm (Jackson et al. (2005)). Another approach among the global search methods is the application of the LASSO type estimators called fused lasso. This approach will be discussed later on in this section.

Exact Search Methods

Consider the sequence of random variables $\{X_t\}_{t=1}^{T}$ that are generated from a common family of distribution with p.d.f. f and the m-dimensional parameter vector $\theta_1, \cdots, \theta_T$. To estimate locations of the possible changes in the distribution, the following optimization problem is conducted:

$$(\tau_1, \cdots, \tau_N)^t = \underset{0=\eta_0<\eta_1<\cdots<\eta_N<\eta_{N+1}=T}{\operatorname{argmin}} \mathcal{L}(0, \eta_1, \cdots, \eta_N, T) + N\, m\, \lambda_T, \tag{2.22}$$

$$\mathcal{L}(\eta_1, \cdots, \eta_N) = \sum_{\ell=0}^{N} \mathcal{C}(\eta_\ell + 1, \eta_{\ell+1}), \tag{2.23}$$

where $\mathcal{C}(\eta_\ell + 1, \eta_{\ell+1})$ is the cost function associated to the random variables $\{X_t\}_{t=\eta_\ell+1}^{\eta_{\ell+1}}$. Moreover, N is the number of change points, m is the dimension of the parameter vector and λ_T is the function of the number of observation. One of the most popular choices for the penalty function is $\lambda = \log(T)$ that is the penalty function for Bayesian information criterion. A data driven approach for choice of the penalty function was proposed by Haynes et al. (2017). Another consideration is the choice of the loss function $\mathcal{L}(\cdot)$ which is selected according to the assumptions of the change point model. In parametric settings, the loss function is derived from the likelihood function. For instance, in the normal mean change point model with unit variance, the cost function based on the likelihood function is given by mean of square errors, i.e:

$$\mathcal{C}(\eta_\ell + 1, \eta_{\ell+1}) = \sum_{t=\eta_\ell+1}^{\eta_{\ell+1}} \left(X_t - \overline{X}_{(\eta_\ell+1):\eta_{\ell+1}}\right)^2.$$

Fortunately, a unique solution exists for problem (2.22). The optimization problem (2.22) can be defined as either a constrained minimization problem or a penalised minimization

problem. In the constrained minimization problem, the upper bound for the number of change points denoted by $K_{\max}$ is specified. Next, the loss function is minimized for the given number of change points ,N, for $N = 0, 1, \ldots, K_{\max}$. The set of indexes that minimizes the function in (2.22) is assigned to be the estimated locations of change points. The other version of the optimization problem (2.22) is the penalized minimization problem. In the penalized minimization problem, the penalty function is the linear function of the number of segments in (2.22). The optimal partitioning (Jackson et al. (2005)) search method is applied to solve the penalized minimization problem.

Segment neighbourhood search algorithm (Auger and Lawrence (1989)) is designed to solve the constrained minimization problem. In this algorithm, first, the upper bound for the number of change points ,$K_{\max}$, is specified. Next, all of the sub-intervals of the sequence of random variables such as $\{X_s\}_{s=1}^{t}$ for $t = 2, \cdots, T$ are partitioned into two non-overlapping segments in a way that the loss function $\mathcal{L}$ is minimized. Then all of the sub-intervals such as $\{X_s\}_{s=1}^{t}$ for $t = 3, \ldots, T$ are partitioned to three segments with the help of segmentation in the previous stages. This procedure continues until all of the possible sub-intervals are segmented to $K_{\max} + 1$ segments in a manner that the loss function $\mathcal{L}$ is minimized. Then among the obtained segmentation, the one that minimizes the objective function in (2.22) is assigned to be the optimal solution for the constrained minimization problem. Consider a sub-interval of random variables $\{X_s\}_{s=1}^{t}$ with N change points and denote the minimum of the loss function with $N+1$ segmentation by $\mathcal{L}_{N,t}$. Due to the additivity of the loss function, $\mathcal{L}_{N,\,t}$ can be obtained from the previous segmentation in the following

$$\mathcal{L}_{N,\,t} = \min_{N \le s \le t-1} \big\{ \mathcal{L}_{N-1,s} + \mathcal{C}(s+1,\, t) \big\}. \tag{2.24}$$

The segment neighbourhood search is computationally expensive. To reduce the computational cost of the algorithm, Rigaill (2010) introduces a novel method called pruned dynamic programming for pruning the candidate change point sets in segment neighbourhood search.

In penalized minimization problem, the new segment and the penalty function are added to the penalized cost function simultaneously. As mentioned earlier, the optimal partitioning search algorithm is designed to obtain the optimal solution for a penalized minimization problem. The function which is required to be minimized by the optimal partitioning search is called a penalized cost function. The minimum penalized cost function by the iteration t is denoted by $F(t)$:

$$F(t) = \min_{\eta_1, \cdots, \eta_N} \Big\{ \sum_{\ell=0}^{N} \big[\mathcal{C}(\eta_\ell, \eta_{\ell+1}) + \ \lambda_T \big] - m \ \lambda_T \Big\} \tag{2.25}$$

where $\eta_{N+1} = t$ and t is called the endpoint. In optimal partition, the global minimum in problem (2.25) is obtained recursively for all the possible end points such as t for $t = 0, \ldots, T$. The problem (2.25) can be rewritten as follow:

$$F(t) = \min_{0 \leq \tau \leq T-1} \Big(F(\tau) + \mathcal{C}(\tau + 1,\, t) + m\, \lambda_T \Big)\,, \tag{2.26}$$

where $F(0) = -m\, \lambda_T$ and τ is referred to as the last change point. The minimum penalized cost function is calculated for all possible endpoints. Therefore, the optimal partitioning search algorithm can be computationally expensive for the long sequence of random variables. To reduce the computational complexity of the optimal partitioning search algorithm, Killick et al. (2012) proposed a pruning algorithm called PELT which eliminates some of the indexes of random variables from any consideration in (2.26). Assuming the following is true for some κ

$$\mathcal{C}(s + 1,\, t) + \mathcal{C}(t + 1,\, T) + \kappa \leq \mathcal{C}(s + 1,\, T) \quad \text{for} \quad s < t < T, \tag{2.27}$$

where κ is a positive constant, then if

$$F(s) + \mathcal{C}(s + 1,\, t) + \kappa > F(t)\,, \tag{2.28}$$

s can not be the last change points for any end point larger than t. Thus, by incorporating (2.28) into the optimal partitioning algorithm, many of the indexes that are not likely to be the last change points are identified and discarded in the optimization problem (2.26). Therefore, computational cost of PELT is significantly better than the optimal partitioning search algorithm. Moreover, under certain configuration, the computational cost of PELT is a linear function of time.

LASSO-Type Change Point Detection Methods

Another variation of the global search method is the LASSO-type methods. The main difference between LASSO-type change point detection methods and exact search methods is the penalty function which is relaxed from ℓ_0 to ℓ_1 norm penalization in the LASSO type methods. In this section, we mainly focus on the use of LASSO on the normal mean change-point model. Consider the sequence of random variables $\{X_t\}_{t=1}^{T}$ with piecewise constant mean and unit variance. The LASSO is launched to locate the time in which the mean of the random variables changes afterwards. Suppose the mean of the random variable at time t is μ_t. Then the mean change points are obtained by solving the following optimization

problem:

$$(\mu_1, \cdots, \mu_T)^t = \underset{\mu_1,\cdots,\mu_T}{\operatorname{argmin}} \sum_{t=1}^{T}(X_t - \mu_t)^2 + \lambda \sum_{t=1}^{T-1} \left|\mu_{t+1} - \mu_t\right|. \tag{2.29}$$

Application of LASSO on the change-point problem was initially proposed by Huang et al. (2005) in the context of copy number variation. They obtained the solution for (2.29) by modifying the optimization problem (2.29) to the standard model selection with LASSO (Tibshirani (1996)).Harchaoui and Lévy-Leduc (2010) investigated the multiple change-point detection methods via LASSO. There are several other literature on this topic, such as Harchaoui and Lévy-Leduc (2010) and Bleakley and Vert (2011) to name a few.

Rank Based Global Search Methods

In the global search methods, the locations of change points are often estimated by minimizing a loss function which is determined based on the underlying parametric distribution of the random variables. One of the most notable examples of the parametric penalized loss function is Shwartz information criterion which is closely related to log-likelihood function. Therefore, to use this procedure, the underlying parametric distribution is required to be specified accurately. To avoid the difficult task of determining the distribution of the random variables, Wang et al. (2020) proposed a rank based Schwartz information criterion with respect to the mean change point problem. Recall that in mean change point model, the sequence of random variables $\{X_t\}_{t=1}^{T}$ are generated from a common but unknown family of distribution function as follow:

$$X_t = \mu_t + \epsilon_t, \quad \text{where} \quad \mu_{\tau_j} \neq \mu_{\tau_j+1},$$

and $\tau_1, \ldots, \tau_N$ are the location of changes in the mean. Moreover, $\epsilon_1, \ldots, \epsilon_T$ are independently and identically distributed noises with mean zero and constant variance.

The main role of the rank based Schwartz criterion in Wang et al. (2020) is to determine the number of change points. Prior to start of the procedure, the sequence of the random variables are ranked as follow:

$$R_t = \sum_{i=1}^{T} I(X_i < X_t) + \sum_{i=1}^{T} I(X_i = X_t) + 0.5,$$

where $I(A)$ is the indicator function of A. Similar to the other variation information criteria, the rank based method consists of a loss function and a penalty term that circumvents overestimation of the number of change points. That is, the rank based Schwarz

information criterion with respect to the candidate set $\{\tau_1, \tau_2, \ldots, \tau_K\}$ is calculated as follow:

$$rSIC(\tau_1, \tau_2, \ldots, \tau_K) = \mathcal{L}_R(\tau_1, \tau_2, \ldots, \tau_K) + K\, C\, T^2 \log(T). \tag{2.30}$$

Note that $\mathcal{L}_R$ is calculated based on the designated rank of the random variables and C is a fixed constant. Wang et al. (2020) proposed several rank based loss functions. One example is:

$$\mathcal{L}_{R;\, LAD} = \sum_{j=0}^{K} \sum_{t=\tau_j+1}^{\tau_{j+1}} \left| R_t - \widetilde{R}_{(\tau_j+1):\tau_{j+1}} \right|,$$

where $\widetilde{R}_{(\tau_j+1):\tau_{j+1}}$ is the median of the rank $R_{\tau_j+1}, \ldots, R_{\tau_{j+1}}$. Another consideration is setting up a penalty term or more specifically determining the constant C in (2.30). They determined C by conducting an extensive simulation study and assigned C to be 0.1.

As mentioned earlier in this chapter, multiple change point detection methods consist of a search method and a mechanism which determines the number of change points. The rank based Schwartz criteria can be applied in order to determine the number of mean change points regardless of the underlying distribution family of the random variables. Moreover, rank based Schwarz criteria can be implemented in local or global search methods. For instance, Wang et al. (2020) proposed application of the rank based Schwarz criteria into PELT algorithm (Killick et al. (2012)).

2.3 Comparison of Global and Local Search Methods

Single change point tests are constructed under the assumption that at most one change point exists within the sequence of random variables. If an interval which is defined for a change point test contains multiple change points, then the power of the test may diminish. In other words, if a change point test is performed within the sequence of random variables with multiple change points, then the test may fail to recover any change points. An alternative approach to single change point methods is multiple change point methods which have been investigated extensively in the recent years. As mentioned earlier, to perform a multiple change point detection analysis, local search methods conduct a single change point test within a region of sequence of random variables. Thus, local search methods may also fail when the aforementioned assumption in the implemented single change point test is not satisfied (existence of at most one change point within local interval). One local search method which manages to diminish the risk of the interference of change points

with each others is the narrowest over threshold (Baranowski et al. (2019)). Note that, as a consequence of drawing insufficient number of intervals, the narrowest over threshold may also fail to detect a change in distribution of the random variables in some segment of the dataset.

As noted, the primary component of the local search methods is a single change point test. In fact, the accuracy of a local search method depends highly on the performance of the single change point test which is implemented into the method. Single change point tests are sometimes constructed from some sort of a loss function. In other words,

$$\mathcal{U}(\tau|\alpha,\beta) = \mathcal{L}(\alpha,\beta) - \mathcal{L}(\alpha,\tau,\beta), \tag{2.31}$$

where $\mathcal{U}(\tau|\alpha,\beta)$ is the single change point test statistic which measures the change in distribution of the random variables at the point τ within the sub-sequence of the random variables $X_{\alpha+1},\cdots,X_{\beta}$ and $\mathcal{L}()$ is some sort of a loss function. Note that, the corresponding loss function with respect to the candidate set $\{\tau_1,\cdots,\tau_N\}$ ($\mathcal{L}(0,\tau_1,\cdots,\tau_N,T)$) is calculated as follow:

$$\mathcal{L}(0,\tau_1,\cdots,\tau_N,T) = \sum_{l=0}^{N} \mathcal{C}((\tau_l+1):\tau_{l+1}).$$

where $\mathcal{C}((a+1):b)$ is the cost function between indexes a and b. One of the most prominent examples of such a single change point test statistic is the generalized log-likelihood ratio test statistic. Since single change point test statistics and the loss functions are related here, one can conclude that the global search methods are much more vulnerable to underestimation of the number of change points in comparison to their local search counterparts. To demonstrate this claim, the following scenario is constructed. Consider the candidate sets $\{\eta\}$ and $\{\gamma_1,\gamma_2\}$ such that

$$\begin{aligned} \{\eta\} &= \underset{i\in\{1,\cdots,T-1\}}{\arg\min}\ \mathcal{L}(0,i,T) \\ \{\gamma_1,\gamma_2\} &= \underset{(\alpha,\beta)\in\{(i,j)|\ d\le i<T-d\ \&\ i+d\le j<T\}}{\arg\min}\ \mathcal{L}(0,\alpha,\beta,T). \end{aligned} \tag{2.32}$$

As a result of equations (2.31) and (2.32),

$$\mathcal{U}(\gamma_2|\gamma_1,T) = \mathcal{L}(0,\gamma_1,T) - \mathcal{L}(0,\gamma_1,\gamma_2,T) \ge \mathcal{L}(0,\eta,T) - \mathcal{L}(0,\gamma_1,\gamma_2,T),$$

where d is the dimension of the changing parameter vector. Since the single change point test statistic (2.31) is larger than the difference of the two loss function, local search

methods are more likely to accept the possibility of a change in distribution of the random variables in comparison to global search methods. Moreover, if some of the true change points get eliminated due to the oversized penalty function, then the remaining change points may get offset by eliminated ones. To demonstrate this observation, let us return to the earlier scenario. Suppose the global search method is conducted with respect to the sequence of the random variables with *two change points*. Moreover, the upper bound for the number of change points is assigned to be two, and the penalty function is set to be the function of number change points (i.e, $\lambda(N)$, where N is the number of change points). Suppose, equation (2.32) holds in the aforementioned sequence of random variables. Since the maximum number of change points is assumed to be two, according to (2.32), the estimated change point set is selected among the sets $\{\}$, $\{\eta\}$ and $\{\gamma_1, \gamma_2\}$ (2.32). Suppose $\mathcal{U}(\gamma_2|\gamma_1, T) < \lambda(2) - \lambda(1)$, then $\mathcal{L}(0, \eta, T) + \lambda(1) < \mathcal{L}(0, \gamma_1, \gamma_2, T) + \lambda(2)$. Therefore, the set $\{\gamma_1, \gamma_2\}$ is rejected in favour of the set $\{\eta\}$. Since the candidate set $\{\gamma_1, \gamma_2\}$ is removed from any further considerations, then the estimated change point set is selected between the sets $\{\}$ and $\{\eta\}$. The comparison between these two candidate sets is similar to conducting a single change point test (2.31). Recall that, a single change point test may fail to detect a change in distribution of random variables, if an interval which is defined for the test is contaminated with closely located change points. Thus, the global search methods may fail to locate any change points in this scenario. In other words, as a result of the oversized penalty function and interference of the change points with each other, global search methods may fail to detect any change points regardless of the value of $\mathcal{U}(\gamma_1|0, \gamma_2)$. On the other hand, local search methods such as wild binary segmentation or narrowest over threshold are more likely to recover a change point, since they are more likely to conduct a change point test in an interval of the random variables with exactly one change point. In summary, the local search methods such as wild binary segmentation or narrowest over threshold are more robust to a large threshold than the global search methods to an oversized penalty function.

Chapter 3

Enhanced Backward Detection

3.1 Introduction

As we discussed in Section 2.2.2, throughout the bottom-up methods, the change point candidates are eliminated sequentially. In parametric settings, the change point candidates are often assessed with the log-likelihood ratio statistics. Consider the sequence of random variables $\{X_t\}_{t=1}^{T}$ that are generated from a common family of distributions with p.d.f. f and d-dimensional parameter vectors $\theta_1, \cdots, \theta_T$. Suppose the log-likelihood ratio test is conducted to decide on the elimination of the change point candidate τ. Moreover, let assume the change point candidates which are adjacent to τ are the indexes a and b. Thus, the log-likelihood ratio test is conducted within the sub-sequence of random variables $\{X_t\}_{t=a+1}^{b}$ where $0 \leq a < b \leq T$ for testing a change at τ. In this case, the log-likelihood ratio test is calculated as follow:

$$\mathcal{U}(\tau|a,b) = \log\left(\frac{f(X_{a+1},\cdots,X_\tau;\theta_a)f(X_{\tau+1},\cdots,X_b;\theta_b)}{f(X_{a+1},\cdots,X_b;\theta_0)}\right), \tag{3.1}$$

where θ_a, θ_b, and θ_0 are estimated by the maximum likelihood method within the intervals of random variables $\{X_t\}_{t=a+1}^{\tau}$, $\{X_t\}_{\tau+1}^{b}$ and $\{X_t\}_{t=a+1}^{b}$, respectively. The change point candidates with the smaller test statistic (3.1) are less likely to be true change points.

The common feature of the bottom-up methods is the sequential elimination of the change point candidates throughout the procedure. To decide on removing a change point candidate, we evaluate all of the change point candidates in the current active set based on the test statistic (3.1). Then one or multiple change-point candidates with the smallest value of the test statistic (3.1) are discarded. The collection of the change point candidates

that are retained after removing the unlikely change point candidates make up the new active set. Throughout the execution of the bottom-up change point detection methods, multiple active sets are obtained. Among the active sets, the most desirable one is assigned to be the set of estimated change points. Note that, the elements of an active set may not maximize the log-likelihood ratio statistic (3.1) within their corresponding intervals. In this chapter, we introduce a novel bottom up change point detection method called enhanced backward detection (EBD). Unlike other variations of the bottom-up change point methods, EBD ensures that all of the elements of the active sets are the solutions to the optimization problem (2.7) within their corresponding intervals. In other words, the set of change point candidates $\{\tau_1, \ldots, \tau_N\}$ is an active set in context of EBD, if

$$\tau_j = \underset{\tau_{j-1}<t<\tau_{j+1}}{\operatorname{argmax}} \mathcal{U}(t|\tau_{j-1}, \tau_{j+1}) \quad \text{for} \quad j = 1, \cdots, N, \tag{3.2}$$

where $\tau_0 = 0$ and $\tau_{N+1} = T$. The condition (3.2) is referred to as a *sweeping requirement*. The procedure that ensures the satisfaction of sweeping requirement (3.2) by all of the change point candidates in an active set is called the *sweeping process*. Throughout this chapter, the active set with N change point candidates is denoted by $\mathcal{M}(N)$. After eliminating a change point candidate with the smallest value of the generalized log-likelihood test statistic from the active set, we launch the sweeping process to obtain another active set. Similar to other variations of the bottom-up change point detection methods, the active sets that are obtained during EBD are recorded. Finally, the estimated change point set is selected among the recorded active sets by the help of Bayesian Information Criteria.

The motivation for sweeping process is that the generalized log-likelihood ratio test is much more accurate than the backward detection in the single change point detection problem. To demonstrate the superior performance of the generalized log likelihood ratio test over the backward detection, we constructed the following simulation study in the single normal mean change point setting. Consider a sequence of independently normally distributed random variables $\{X\}_{t=1}^{40}$ with unit variance and a piecewise constant mean μ_t where $\mu_t = 0$ for $t = 1, \ldots, 20$ and $\mu_t = 2$ for $t = 21, \ldots, 40$. Note that, the number of change points is known and assumed to be one in this particular simulation study. Thus, the only concern is estimating the location of the mean change point. For the backward detection method, the change point candidates are eliminated based on the magnitude of the scaled jump size quantity (2.20), until one change point candidate remains. This point is the estimated location of the change point by the backward detection. To assess the accuracy of the backward detection, we subtract the true change point from the estimated location of the change in the mean by backward detection. A similar procedure is repeated for the generalized log-likelihood ratio test. The number of trails in this simulation study

	≤ -3	-2	-1	0	1	2	≥ 3
Generalized log-likelihood ratio test Error $\tau_c - \tau$	360	366	1139	6201	1134	413	387
Backward detection Error $\tau_{bws} - \tau$	787	484	1056	5499	1060	459	655

Table 3.1: Estimates of the true change point i.e., τ are denoted with τ_c for the CUSUM method and with τ_{bws} for the backward detection.

is 10000. Based on the simulation results in Table (3.1), the generalized log-likelihood ratio test outperforms the backward detection significantly. This flaw of the backward detection is more problematic in multiple change-point detection cases. Thus, the sweeping process is necessary to be added to backward detection to improve the accuracy of this method.

In the next section (3.2), the sweeping process is described more precisely. Moreover, the conditions by which the sweeping process is guaranteed to converge are introduced. Next, we present EBD in Section (3.3). The performance of EBD is compared to some of the most well-known multiple change-point detection methods in Section (3.4). Furthermore, the role of the sweeping process in boosting the accuracy of EBD is demonstrated by comparing the precision of EBD before and after adding the sweeping process.

3.2 Sweeping Process

EBD is differentiated from other bottom-up methods by the sweeping process. The hope is that, the accuracy of EBD improves compared to other bottom-up change point methods by adding this procedure. In this section, the sweeping process is described more precisely. Later on, the condition in which the sweeping process is guaranteed to converge is presented.

To clarify the steps of the sweeping process, we construct the following example. Suppose the current active set is $\mathcal{M}(K) = \{\tau_1, \ldots, \tau_K\}$, where $\tau_1 < \cdots < \tau_K$. Recall that the set of change point candidates is considered to be an active set in the context of EBD, if all of the change point candidates satisfy the sweeping requirement (3.2). Suppose the change point candidate τ_j has the smallest value of the test statistic (3.1) among the members of the active set $\mathcal{M}(K)$. Subsequently, τ_j is eliminated from the active set $\mathcal{M}(K)$. The remaining change point candidates may not satisfy the sweeping requirement (3.2). Thus, the sweeping process is launched to obtain the next active set. To simplify the explanation

of the sweeping process in the context of this example, we relabel the remaining elements of the set $\mathcal{M}(K)$ as follow:

$$\begin{aligned} \tau_\ell^* &= \tau_\ell \qquad \text{for} \qquad \ell = 1, \ldots, j-1, \\ \tau_\ell^* &= \tau_{\ell+1} \qquad \text{for} \qquad \ell = j, \ldots, K-1. \end{aligned}$$

Moreover, the current change point candidates set is denoted by $\mathcal{M}^*(K-1)$. Note that, after removing τ_j, only portion of the remaining change point candidates are required to be verified whether they satisfy the sweeping requirement(3.1) or not. In this example, τ_{j-1}^* and τ_j^* are the only change point candidates that may violate the sweeping requirement and must be investigated by the sweeping process. Note that, since the sweeping requirement is satisfied by the members of $\mathcal{M}(K)$, then the change point candidates which are not adjacent to τ_j will maximize the change point test statistic within their corresponding intervals after removal of the change point candidate τ_j. The current and last change point candidates which are required to be verified are referred to as the count $c = \max(j-1, 1)$ and the maximum count $m.c = \min(K-1, c+1))$, respectively. Note that, the count and the maximum count variables are subject to change throughout the process. For example, if τ_c does not satisfy the sweeping requirement, then the count and maximum count are updated as follow:

$$\begin{aligned} c &= \max(c-1, 1), \\ m.c &= \min(K-1, \max(c+2, m.c)). \end{aligned}$$

Otherwise, the count is added by one ($c = c+1$). The sweeping process is continued until the count exceeds the maximum count (i.e., $c > m.c$). The steps of the sweeping process are presented in the algorithm (1). The sweeping process is denoted by $\mathcal{SP}(\{X_t\}_t, \mathcal{M}^*(K-1), c, m.c)$, where $\mathcal{M}^*(K-1)$ is the set which consists of the change point candidates before the launch of the sweeping process.

By adding the sweeping process to backward detection, the accuracy of the backward detection improves significantly while the computational complexity does not increase by much. One concern is that, the sweeping process may keep oscillating between multiple set of change point candidates and subsequently does not converge. Here, this scenario is explored in a very simple example. Consider the indexes of the random variables τ_1, τ_2, τ_1^*, τ_2^*,

Result: $\mathcal{M}(K-1)$
INITIALIZATION c, $m.c$ and $\mathcal{M}^*(K-1)$;
while $c \leq m.c$ **do**
  Obtain $\tau = \underset{\tau_{c-1}<t<\tau_{c+1}}{\operatorname{argmax}} \mathcal{U}(t|\tau_{c-1}, \tau_{c+1})$ where $\tau_0 = 0$ &$\tau_K = T$;
  if $\tau \neq \tau_c$ **then**
    $\tau_c = \tau$;
    $c = \max(1, c-1)$;
    $m.c = \min(K-1, \max(c+2, m.c))$;
  else
    $c = c+1$;
  end
end

Algorithm 1: Sweeping process.

such that

$$\begin{aligned}
\tau_1 &= \underset{0<i<\tau_2}{\operatorname{argmax}} U(i \mid 0,\, \tau_2), \\
\tau_2 &= \underset{\tau_1^*<i<T}{\operatorname{argmax}} U(i \mid \tau_1^*,\, T), \\
\tau_1^* &= \underset{0<i<\tau_2^*}{\operatorname{argmax}} U(i \mid 0,\, \tau_2^*), \\
\tau_2^* &= \underset{\tau_1<i<T}{\operatorname{argmax}} U(i \mid \tau_1,\, T).
\end{aligned} \tag{3.3}$$

Suppose the sweeping process is launched on the set $\{\tau_1, \tau_2\}$, where count is set to 1 and maximum count is 2. Since based on the assumptions in (3.3), τ_1 maximizes the generalized log-likelihood ratio test statistic within the interval $(0, \tau_2)$, then count c is added by 1 and the maximum count $m.c$ remains 2. Now the point τ_2 is getting inspected. Since $\mathcal{U}(\tau_2^* \mid \tau_1, T)$ is larger than $\mathcal{U}(\tau_2 \mid \tau_1, T)$, the τ_2 is replaced by τ_2^*. Subsequently, the count and maximum count are updated ($c = 1$ & $m.c = 2$). However, τ_1^* maximizes the generalized log-likelihood test statistic within the interval $(0, \tau_2^*)$, so τ_1^* replaces τ_1. Since, $\mathcal{U}(\tau_2 \mid \tau_1^*, T)$ is larger than $\mathcal{U}(\tau_2^* \mid \tau_1^*, T)$, then τ_2^* is replaced with τ_2. Since the point τ_1 maximizes the generalized log-likelihood ration test statistic in the segment $(0, \tau_2)$, then τ_1^* is replaced with τ_1. Thus, if the assumptions in (3.3) hold, then the sweeping process keeps oscillating between multiple sets and subsequently algorithm fails to converge.

Recall that, the loss function in (2.22) can be derived from the maximum likelihood function, given the random variables are generated from a common and known parametric

family of distribution. The likelihood based loss function of the change point candidates set $\{\tau_1, \cdots, \tau_N\}$ is calculated as follow:

$$\begin{aligned} \mathcal{L}(0, \tau_1, \cdots, \tau_N, T) &= \sum_{\ell=0}^{N} \mathcal{C}((\tau_\ell + 1) : \tau_{l+1}), \\ \mathcal{C}((\tau_\ell + 1) : \tau_{\ell+1}) &= \log\Big[f(X_{\tau_\ell+1}, \ldots, X_{\tau_{\ell+1}})\Big], \end{aligned} \tag{3.4}$$

where θ_ℓ is estimated by maximum likelihood estimator within the sequence of random variables $\{X_t\}_{t=\tau_\ell+1}^{\tau_{\ell+1}}$. The log-likelihood ratio test for the change point candidate τ_ℓ with respect to the sub-sequence of the random variables $\{X_t\}_{t=\tau_{\ell-1}+1}^{\tau_{\ell+1}}$ can be represented as follow:

$$\mathcal{U}(\tau_\ell \mid \tau_{\ell-1}, \tau_{\ell+1}) = \mathcal{C}((\tau_{\ell-1} + 1) : \tau_{\ell+1}) - \big[\mathcal{C}((\tau_{\ell-1} + 1) : \tau_\ell) + \mathcal{C}((\tau_\ell + 1) : \tau_{\ell+1})\big], \tag{3.5}$$

for $\ell = 1, \cdots, T$. The relationship between the cost function (3.4) and the single change point test statistic (3.1) is instrumental for the convergence of the sweeping process. Suppose the set $\{\tau_1, \tau_2\}$ is the change point candidates set within the sequence of the random variables $\{X_t\}_{t=1}^{T}$, then

$$\begin{aligned} \mathcal{L}(0, \tau_1, \tau_2, T) &= \mathcal{L}(0, \tau_2, T) - \mathcal{U}(\tau_1 \mid 0, \tau_2) = \mathcal{C}(1 : \tau_2) + \mathcal{C}((\tau_2 + 1 : T)) - \mathcal{U}(\tau_1 | 0, \tau_2)\,, \\ \mathcal{L}(0, \tau_1, \tau_2, T) &= \mathcal{L}(0, \tau_1, T) - \mathcal{U}(\tau_2 \mid \tau_1, T) = \mathcal{C}(1 : \tau_1) + \mathcal{C}((\tau_1 + 1 : T)) - \mathcal{U}(\tau_2 \mid \tau_1, T)\,. \end{aligned} \tag{3.6}$$

Let us return to the example in which the sweeping process did not converge. Assuming the equation (3.3) is true, the sweeping process keep oscillating between multiple change point candidates sets. Suppose the sweeping process is launched on the change point candidate set $\{\tau_1, \tau_2\}$, then the candidates sets rotate in the following manner:

$$\{\tau_1, \tau_2\} \Rightarrow \{\tau_1, \tau_2^*\} \Rightarrow \{\tau_1^*, \tau_2^*\} \Rightarrow \{\tau_1^*, \tau_2\} \Rightarrow \{\tau_1, \tau_2\} \Rightarrow \cdots\,. \tag{3.7}$$

In this example, the algebraic relationship between the loss function of the different change point candidates sets is as follow:

$$\begin{aligned} \mathcal{L}(0, \tau_1, \tau_2, T) &= \mathcal{C}(1 : \tau_1) + \mathcal{C}((\tau_1 + 1) : \tau_2) + \mathcal{C}((\tau_2 + 1) : T) \\ &= \mathcal{C}(1 : \tau_1) + \mathcal{C}((\tau_1 + 1) : T) - \mathcal{U}(\tau_2 \mid \tau_1, T) \\ &= \mathcal{C}(1 : \tau_1) + \mathcal{C}((\tau_1 + 1) : \tau_2^*) + \mathcal{C}((\tau_2^* + 1) : T) \\ &\quad + \mathcal{U}(\tau_2^* \mid \tau_1, T) - \mathcal{U}(\tau_2 \mid \tau_1, T) \\ &= \mathcal{L}(0, \tau_1, \tau_2^*, T) + \mathcal{U}(\tau_2^* \mid \tau_1, T) - \mathcal{U}(\tau_2 \mid \tau_1, T). \end{aligned} \tag{3.8}$$

Similarly, the following can be shown:

$$\begin{aligned}
\mathcal{L}(0,\, \tau_1,\, \tau_2^*,\, T) &= \mathcal{L}(0, \tau_1^*,\, \tau_2^*,\, T) + U(\tau_1^* \mid 0,\, \tau_2^*) - U(\tau_1 \mid 0,\, \tau_2^*)\,, \\
\mathcal{L}(0,\, \tau_1^*,\, \tau_2^*,\, T) &= \mathcal{L}(0, \tau_1^*,\, \tau_2,\, T) + U(\tau_2 \mid \tau_1^*,\, T) - U(\tau_2^* \mid \tau_1^*,\, T)\,, \\
\mathcal{L}(0,\, \tau_1^*,\, \tau_2,\, T) &= \mathcal{L}(0,\, \tau_1,\, \tau_2,\, T) + U(\tau_1 \mid 0,\, \tau_2) - U(\tau_1^* \mid 0,\, \tau_2)\,.
\end{aligned} \tag{3.9}$$

Based on the assumptions in the example (3.3) and results obtained in (3.8) and (3.9), the following holds:

$$\mathcal{L}(0,\, \tau_1,\, \tau_2,\, T) > \mathcal{L}(0,\, \tau_1,\, \tau_2^*,\, T) > \mathcal{L}(0,\, \tau_1^*,\, \tau_2^*,\, T) > \mathcal{L}(0,\, \tau_1^*,\, \tau_2,\, T) > \mathcal{L}(0,\, \tau_1,\, \tau_2,\, T). \tag{3.10}$$

The result in the equation (3.10) points out that assumptions in the equation (3.3) are impossible to be fulfilled, and the sweeping process will eventually converge as long as the single change point statistic is derived from some sort of a loss function. This example can be extended to more general scenarios with larger change point candidates set.

3.3 Enhanced backward detection via BIC

Enhanced Backward Detection (EBD) starts with an active set with large number of change point candidates. Then, the change point candidates with the smallest test statistic (3.1) is identified and eliminated from the active set. To ensure that the sweeping requirement (3.2) is satisfied by all the remaining change point candidates, the sweeping process is launched. Throughout the procedure, multiple active sets are obtained. Finally, the estimated set of change points is selected among the recorded active sets. In the context of EBD, selecting the most desirable active set is equivalent to estimating the number of change points.

Two of the most frequently used approaches for estimating the number of change points are the information criteria and threshold based methods. In this chapter, the number of change points is estimated by Bayesian information criteria. In order to estimate the number of change points consistently, the upper bound for the number of change points , $N_{\max}$, is required to be specified. Consider the sequence of random variables $\{X_1,\, \ldots,\, X_T\}$ and the active set $\mathcal{M}(N) = \{\tau_1,\, \ldots,\, \tau_N\}$, then the corresponding Bayesian information criterion is calculated by:

$$\mathrm{BIC}(\mathcal{M}(N)) = -\sum_{j=1}^{N+1} \log\Big[f(X_{\tau_{j-1}+1},\, \ldots,\, X_{\tau_j})\Big] + N \cdot d \cdot \log(T)\,, \tag{3.11}$$

where d is the dimension of the parameter vectors. During the execution of EBD, $N_{\max}+1$ active sets are recorded. The active set which minimizes (3.11) is assigned to be the estimated change points set and denoted by $\mathcal{M}$.

Result: $\mathcal{M}$
INPUTS: $X_1, \cdots, X_T$ and $\mathcal{M}(N_{\max}) = \{\tau_1, \cdots, \tau_{N_{\max}}\}$;
$\mathcal{M} = \mathcal{M}(N) \leftarrow \mathcal{SP}(\{X_t\}_{t=1}^T, \mathcal{M}(N_{\max}), c, m.c)$;
$BIC \leftarrow BIC(\mathcal{M}(N_{\max})))$;
$N \leftarrow N_{\max}$;
while $N \geq 0$ **do**
  $j \leftarrow \arg\min_{\tau \in \mathcal{M}(N)} \mathcal{U}(\tau_j | \tau_{j-1}, \tau_{j+1})$, where $\tau_0 = 0$ and $\tau_{N+1} = T$;
  $\mathcal{M}^*(N-1) \leftarrow \{\tau_1^*, \cdots, \tau_{N-1}^*\}$ where

$$\tau_l^* = \tau_l \quad \text{for} \quad l = 1, \cdots, j-1;$$
$$\tau_l^* = \tau_{l+1} \quad \text{for} \quad l = j, \cdots, N-1;$$

  $N \leftarrow N - 1$;
  $c \leftarrow \max(2, j-1)$;
  $m.c \leftarrow \min(N, \max(c+1, m.c))$;
  $\mathcal{M}(N) \leftarrow \mathcal{SP}(\{X_t\}_{t=1}^T, \mathcal{M}^*(N), c, m.c)$;
  if $BIC(\mathcal{M}(N)) < BIC$ **then**
    $BIC \leftarrow BIC(\mathcal{M}(N))$;
    $\mathcal{M} \leftarrow \mathcal{M}(N)$;
  end
end

Algorithm 2: Enhanced backward detection via BIC.

3.4 Simulation Study

In this section, performance of EBD is compared to several popular change points detection methods under normal mean change points model. The normal mean change points model can be rewritten as follow:

$$
\begin{aligned}
&X_t = \mu_t + \epsilon_t, \\
&\mu_1 = \cdots = \mu_{\tau_1} \neq \mu_{\tau_1+1} = \cdots = \mu_{\tau_2} \neq \mu_{\tau_2+1} = \cdots = \mu_{\tau_N} \neq \mu_{\tau_N+1} = \cdots = \mu_{\tau_T}, \\
&\epsilon_t \sim N(0, \sigma^2) \quad \text{for} \quad t = 1, \cdots, T,
\end{aligned}
$$

where the mean of the random variables is a piecewise constant function of time and variance is constant. Two classes of simulation studies are presented here:

- In class 1, the locations of the change points and the mean between the change points are randomized and the sequence of random variables are generated as follow :
  - The number of observations T is randomly drawn from the set
    $$\{100, 200, 500, 1000, 2000\}.$$
  - The maximum number of change points is set to $N_{max} = \lfloor \frac{T}{10} \rfloor$.
  - Number of change points is selected randomly from the set $\{0, 1, 2, \ldots, N_{max}\}$ and denoted by N.
  - Locations of change points are obtained by randomly drawing N points from the set $\{1, 2, \ldots, T-1\}$.
  - Difference in the mean before and after each change point is drawn from the normal distribution with mean zero and standard deviation σ_{j_s}, where σ_{j_s} is sampled from the set $\{1, 2, 5\}$.
  - Finally, random noises are added to the data sequence, where the randomized errors are generated independently and identically from the standard normal distribution;
  - The randomized change point model is referred to as random scenario and is denoted by R.
- In the second class, the parameters of the change point models remain constant in each trial. This class of the simulation study consists of the following change point models:

– In the normal mean change point Model 1 (M1), 2048 data points are generated. The locations of the mean change points are

$$(205,\ 267,\ 308,\ 472,\ 512,\ 820,\ 902,\ 1332,\ 1557,\ 1598,\ 1659)\,.$$

The mean between the change points are

$$(0,\ 14.64,\ -3.66,\ 7.32,\ -7.32,\ 10.98,\ -4.39,\ 3.29,\ 19.03,\ 7.68,\ 15.37,\ 0)\,.$$

The standard deviation is set to $\sigma = 10$.

– In the normal mean change point Model 2 (M2), 497 data points are generated. The locations of the mean change points are

$$(0,\ 139,\ 226,\ 243,\ 300,\ 309,\ 333,\ 497)\,.$$

The mean between the change points are

$$(-0.18,\ 0.08,\ 1.07,\ -0.53,\ 0.16,\ -0.69,\ -0.16)\,.$$

The standard deviation is set to $\sigma = 0.4$.

– In the normal mean change point Model 3 (M3), 560 data points are generated. The locations of the mean change points are

$$(0,\ 11,\ 21,\ 41,\ 61,\ 91,\ 121,\ 161,\ 201,\ 251,\ 301,\ 361,\ 421,\ 491,\ 560)\,.$$

The mean between the change points are

$$(7,\ -7,\ 6,\ -6,\ 5,\ -5,\ 4,\ -4,\ 3,\ -3,\ 2,\ -2,\ 1,\ -1)\,.$$

The standard deviation is set to $\sigma = 4$.

– In the normal mean change point Model 4 (M4), 140 data points are generated. The locations of the mean change points are

$$(0,\ 11,\ 21,\ 31,\ 41,\ 51,\ 61,\ 71,\ 81,\ 91,\ 101,\ 111,\ 121,\ 131,\ 140)\,.$$

The mean between the change points are

$$(0,\ 1,\ 0,\ 1,\ 0,\ 1,\ 0,\ 1,\ 0,\ 1,\ 0,\ 1,\ 0,\ 1)\,.$$

The standard deviation is set to $\sigma = 0.4$.

– In the normal mean change point Model 5 (M5), 150 data points are generated. The locations of the mean change points are

$$(0,\ 11,\ 21,\ 31,\ 41,\ 51,\ 61,\ 71,\ 81,\ 91,\ 101,\ 111,\ 121,\ 131,\ 141,\ 150)\,.$$

The mean between the change points are

$$(1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 12,\ 13,\ 14,\ 15)\,.$$

The standard deviation is set to $\sigma = 0.3$.

The accuracy of the fitted change points models is assessed by the mean square error:

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left(\mu_t - \widehat{\mu}_t\right)^2, \tag{3.12}$$

where μ_t and $\widehat{\mu}_t$ are the true mean and the estimated mean of the random variable X_t, respectively.

The performance of EBD is compared with other multiple change points detection techniques such as simple binary segmentation (SBS) (Scott and Knott (1974)), wild binary segmentation (WBS) (Fryzlewicz et al. (2014)), isolated detection (ID) (Anastasiou and Fryzlewicz (2019)), MOSUM with localized pruning (MLP) (Cho and Kirch (2019)), tail greedy bottom-up (TGUH) (Fryzlewicz et al. (2018)), SMUCE (Frick et al. (2014)), PELT (Killick et al. (2012)) and pruned dynamic programming (DPP) (Rigaill (2010)). For simple binary segmentation and PELT, the `R` package `change point` is used. In order to decide the number of change points, the Bayesian information criteria is applied for both PELT and binary segmentation. The `R` package `breakfast` is used for wild binary segmentation and tail greedy bottom-up. The number of change points for wild binary segmentation is also estimated with BIC. The package `not` (Baranowski et al. (2019)) is used for narrowest over threshold. Similar to wild binary segmentation, the number of change points is estimated by BIC. The package `MOSUM` (Cho and Kirch (2019)) is used for multi-scaled moving sum with localized pruning. For SMUCE (Frick et al. (2014)), the package `Step FDRseg` is applied. Finally, the package `Segmentor3IsBack` (Rigaill (2010)) is used for pruned dynamic programming, and the number of change points is estimated with Bayesian information criteria. The maximum number of change points is selected to be 20 for change point models with fixed parameters and N_{max} in the randomized scenario for dynamic pruning programming.

Based on the simulation studies in Table (3.2), the dynamic pruned programming is more accurate than other methods. However, dynamic pruned programming is proven to

		M1	M2	M3	M4	M5	R
MSE	EBD	2.4125	0.0035620	1.5105	0.059196	0.15413	0.15413
	PELT	**2.3850**	0.0036211	1.5346	0.070008	**0.021162**	0.15678
	ID	2.6038	0.0037935	1.5698	0.061695	**0.021162**	0.17614
	DDP	**2.4036**	0.0037782	1.5219	**0.052067**	**0.022073**	**0.14925**
	WBS	2.5418	0.0037327	**1.4901**	0.059667	0.02471	0.1786
	TGUG	3.215	0.0046856	1.8711	0.064325	0.025101	0.17753
	NOT	2.5641	0.0036122	1.5627	0.064230	**0.021640**	0.36156
	SMUCE	4.86	0.0104288	4.0255	0.22824	0.177337	0.2651
	MLP	2.5852	0.0047585	**1.4430**	**0.0444**	**0.018839**	0.23184
	SBS	3.6345	0.0082395	2.3657	0.17298	0.025064	0.17985

Table 3.2: M1, M2, M3, M4 and M5 are the models 1, 2, 3, 4 and 5 for scenarios with fixed parameters. R represents the result for the randomized scenario. In this table, EBD stands for our proposed enhanced backward detection procedure.

be too time-consuming. Therefore, the dynamic pruned programming approach is ideally conducted in scenarios with a small number of change points. According to Table 3.2, EBD is the second most accurate change point detection technique. Thus, EBD can be a reliable approach when the dynamic pruned programming is not feasible. Moreover, the adaptability of EBD has made it more applicable to a more complicated range of change point scenarios.

3.4.1 The Performance of EBD with respect to Zero Change Point Models

In this section, we investigate the susceptibility of EBD to false discovery of change point. EBD is conducted with respect to the sequence of Gaussian white noise with unit variables and $T = 100, 200, 500, 1000, 2000$ observations. For each of the designated number of observations T, one thousand observations are generated. Then we applied EBD with respect to each generated sequence of random variables in order to estimate number of change points. Note that, since the generated models contain zero change point, the estimated number of change points ideally should be zero. The results of the simulation study is summarized in Table 3.3. Based on the result in Table 3.3, as the number of observations has increased, the risk of false discovery of change points is reduced.

	T=100	T=200	T=500	T=1000	T=2000
$\widehat{\tau} = 0$	912	946	969	989	989
$\widehat{\tau} = 1$	53	36	21	11	9
$\widehat{\tau} = 2$	32	14	6	0	2
$\widehat{\tau} \geq 3$	3	4	4	0	0

Table 3.3: As the length of sequence of observations increases, the rate of false discovery of change points decreases.

	Mean square error (MSE)				
Change point model	M1	M2	M3	M4	M5
Enhanced backward detection	2.935	0.0035896	1.5319	0.05681	0.021956
backward detection	3.147	0.0045520	1.8796	0.064495	0.025162

Table 3.4: The mean square error for both enhanced backward detection and backward detection is estimated. As it is seen, adding sweeping process to backward detection improves the accuracy of the method significantly.

3.4.2 Does adding the sweeping process improves the accuracy of the backward detection?

EBD is distinguished from other bottom-up change point detection techniques by a sweeping process. To demonstrate the effectiveness of the sweeping process, we compared the performance of EBD with backward detection in the scenarios M1, M2, M3, M4, and M5. To evaluate these methods under identical circumstances, we estimated the number of change points by BIC for both methods. The results of these simulation studies are exhibited in Table 3.4 which indicate the superiority of EBD over ordinary backward detection. In all of the simulated change point scenarios, EBD is proven to be more accurate than backward detection.

3.4.3 Performance of the Multi-Scale Localized Pruning

Based on our simulation studies in Table 3.2, multi-scale localized pruning performs the best for some of the change points models and unperforms for the rest of the scenarios. In this section, we conduct simulation studies with respect to the M4 and M5 change point models in order to compare the performance of the multi-scaled localized pruning with the default set of bandwidths in the `mosum` (Cho and Kirch (2019)) package versus all possible

	Mean of the Square Error	
	M4	M5
Defult Bandwidth in mosum Package	0.04540	0.01870
All Possible Bandwidth	0.06142	0.03306

Table 3.5: As the number of bandwidths has increased, the mean of the square error increases as well.

bandwidths. Note that, all possible bandwidths for mosum change points detection method are

$$h = 2, \cdots, \left\lfloor \frac{T}{2} \right\rfloor - 1.$$

Intuitively, as the number of bandwidths increases, the accuracy of the estimated change points must increase as well. However, in terms of the mean of the square error, the multi-scale localized pruning with default set of bandwidths is overwhelmingly superior according to the results in Table 3.5. That means, by increasing the number of bandwidths and the computational complexity of the localized pruning, the accuracy of the methods would not increase necessarily. Therefore, the multiple-scale localised pruning is not suitable for combining the results that are obtained from multiple single bandwidth mosum statistic. Moreover, by increasing the number of bandwidths in multi-scale localized pruning, the computational time increases significantly and even the method became infeasible in M1, M2 and M3 scenarios.

3.4.4 Frequency of the Estimated Change Points by Enhanced Backward Detection

In this section, we conducted simulation studies in order to obtain the distribution of the estimated change points with respect to the models M1 to M5, which were introduced in the previous simulation studies. For each change point model

- One thousand replications are generated according to the aforementioned change point models;
- The enhanced backward detection is conducted with respect to each of the generated data sets;
- Estimated change points of all of the generated data sets are collected;

- The histogram of the estimated change points are plotted in Figure (3.1);

As we expected, most of the estimated change points are clustered near the true change points. In the other words, the enhanced backward detection is proven to be robust to false discovery of change points and yet sensitive enough to identify the change in mean.

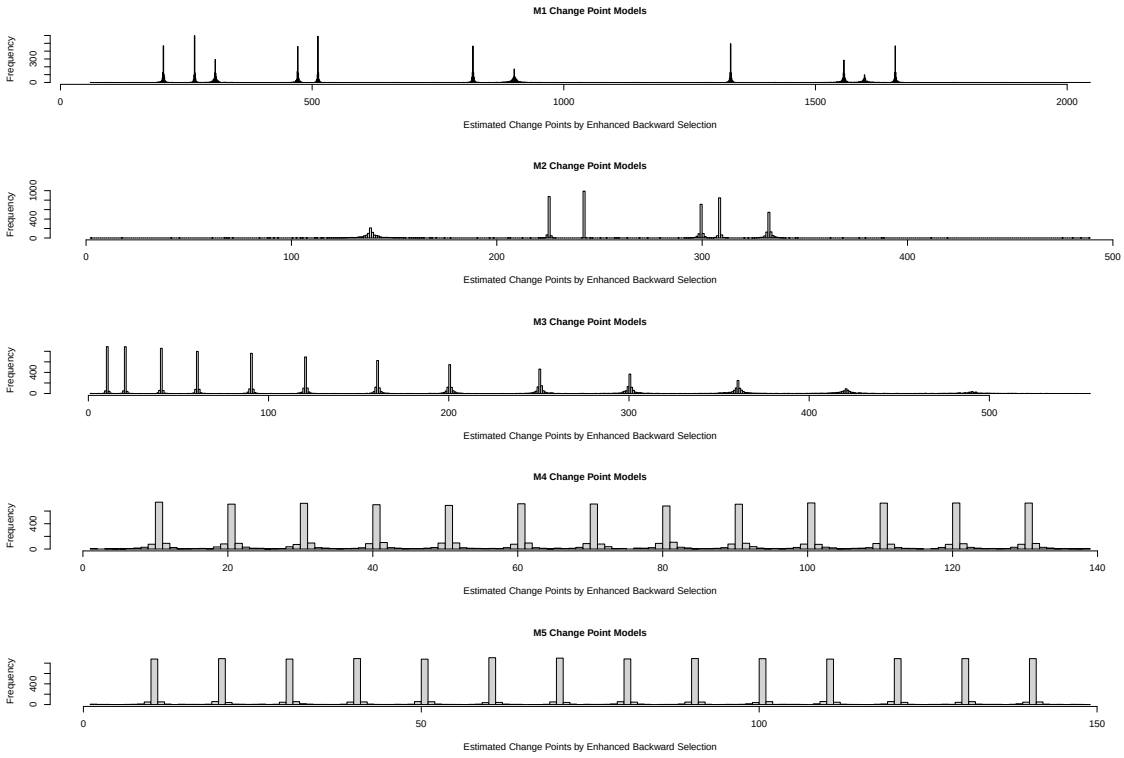

Figure 3.1: The frequency of the estimated locations of the change points are presented here by Histogram.

3.4.5 Computational Complexity of EBD

The computational complexity of algorithms is often calculated under the assumptions of the worst case scenarios. That is, in the worst case scenario, the sweeping process investigates all current change point candidates multiple times at each run of EBD. In other words, after elimination of each change point candidate, the change point test statistic (3.1) is re-calculated with respect to remaining candidates multiple times. However, this scenario

$O(\sqrt{T})$	$O(T)$	$O(T\log(T))$	$O(T^2)$	$O(T^3)$
1	91	670	223	15

Table 3.6: Empirical computational complexity of the enhanced backward detection suggest $O(T\log(T))$.

is unlikely to occur in our experience. Therefore, assessing the computational complexity of EBD by traditional approaches is not insightful. In this section, the computational complexity of EBD is obtained with the help of the R package `GuessCompx` (Agenis-Nevers et al. (2021)). For this purpose, several replications of the random variables are generated. Then EBD is conducted with respect to each generated dataset and the computational complexity of EBD is guessed by the function `CompEst` in the package `GuessCompx` (Agenis-Nevers et al. (2021)).

The configuration of random variables may influence the computational complexity of EBD. Therefore, each generated dataset is constructed from the randomized model in which the number and locations of change points and mean between the change points are determined randomly. More precisely, the randomized change point model are constructed as follow:

- The number of observations is set to be 1000.
- The number of change points is selected from a Poisson distribution, where the rate is sampled from the set $\{4, 8\}$.
- Locations of change points are drawn randomly and uniformly from the set $\{1, 2 \ldots, T-1\}$.
- The mean difference after occurrence of change points are selected from Gaussian distribution with mean zero and variance which is drawn randomly from the set $\{1, 3, 5\}$.
- Gaussian white noise with unit variance is added to the constructed mean function.

The result of the simulation study is recorded in the Table (3.6). based on the aforementioned simulation study, the computational complexity of EBD is most likely to be $O(T\log(T))$.

3.5 Data Analysis

In this section, the change point analysis is conducted for various datasets. We assume that, the random variables are normally and independently distributed. The mean is either constant, piece-wise constant or piece-wise linear function of time and variance is either constant or piece-wise constant. The problem of interest is to figure out the instances after which mean or variance functions change. Since the datasets are assumed to be normally distributed, the underlying change point test statistics are constructed from generalized log-likelihood ratio test.

3.5.1 Housing Price Index (HPI)

Since the housing expenditure such as rents and mortgage covers a large portion of the household expenses, the change in the housing price can affect the other economical activities. Thus, investigating the change in housing price is beneficial for understanding the effect of economical events on the households and society. In this section, we conduct a change point analysis via enhanced backward detection on the housing price indexes of the three boroughs of the city of London and then we link the estimated change points to the events that might have influenced the housing price.

For each dataset, a change point model is fitted based on the estimated locations after which mean or variance of the random variables has changed. To assess the goodness of the fits of the proposed change point model:

- First, we compute the residuals of the estimated change point models by

$$\widehat{r}_t = \frac{X_t - \widehat{\mu}_t}{\widehat{\sigma}_t}.$$

- Then, we verify the normality and stationarity of the calculated residuals via QQ-plot and time series plot, respectively.

We assume the HPI datasets are normally and independently distributed. Through out our analysis, We consider the following scenarios:

S1: The mean is piecewise constant and variance is constant.

S2: The mean is constant and variance is piecewise constant.

S3: The mean and variance are piecewise constant and both mean and variance change simultaneously after each change point.

Notice that, assuming the normality of HPI datasets, a single change point test statistic can be derived from the corresponding generalized log-likelihood ratio test (3.1). The change point analysis is conducted with respect to the datasets from the borough of Hackney, Newham and Tower Hamlets. For more information on these datasets, we refer the reader to data analysis section of Baranowski et al. (2019).

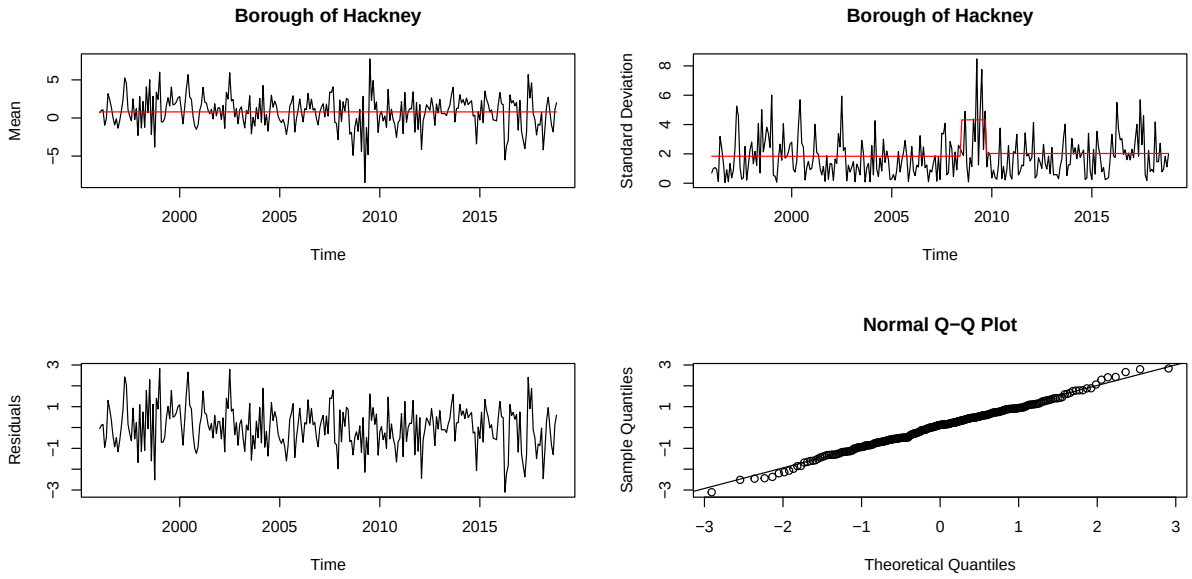

Figure 3.2: The increase of the variance of the change in the housing price index in the borough of the Hackney is visible between the year 2009 and 2010, while the mean remains constant throughout the dataset. The increase of the variance can be explained by world-wide financial crises during 2007-2009.

In the Borough of Hackney, the changes in variance is detected by enhanced backward detection. Based on the result demonstrated in Figure 3.2, the variance within the period between the years 2009-2010 increased significantly. This particular change in the housing price overlapped imperfectly with the global financial crisis period. Note that, the enhanced backward detection on the borough of Hackney is conducted under the assumption of the S2 change point model. Based on the QQ-plot and time series plot of the estimated residuals

by S2 based enhanced backward detection, our presumed assumptions on this data set seem to be reasonable.

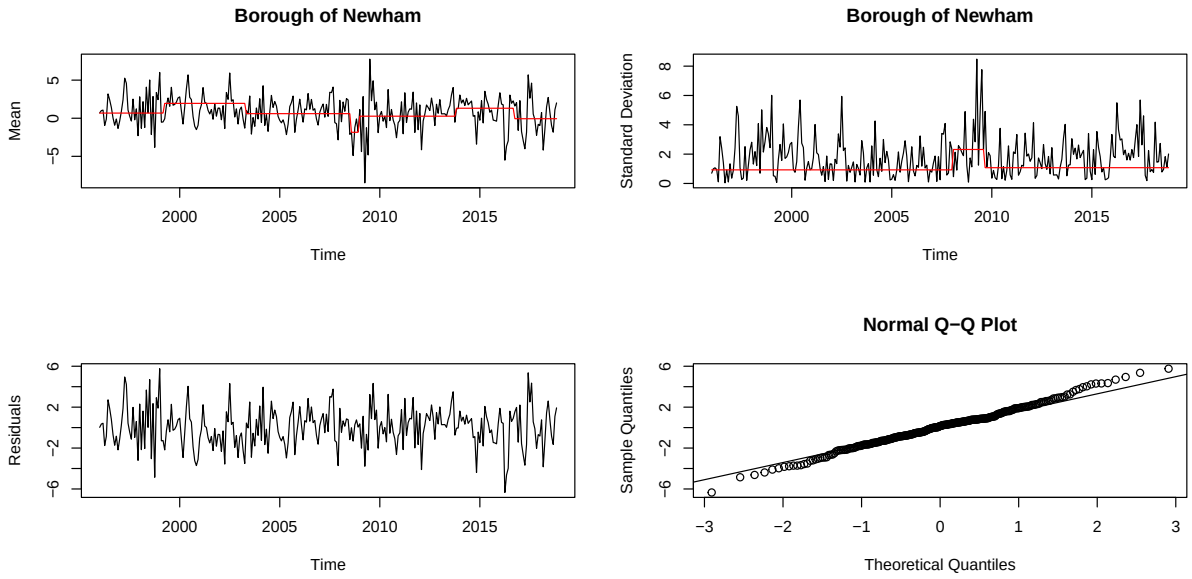

Figure 3.3: The sudden rise of the variance of the change in the housing price index for the borough of Newham is visible between the year 2009 and 2010. The change in the variance can be explained by world wide financial crises during 2007-2009. In other change points locations, the change in distribution occurs only with respect to mean and variance has remained constant.

To perform a change point analysis on Newham data set, enhanced backward detection is conducted under the assumptions of S1 and S3 change point models. The changes in mean and variance are detected with respect to 2008-2010. As mentioned earlier, the sharp increase in the variance of the random variables may be explained with 2007-2009 financial crisis. Also, the results in Figure 3.3 demonstrate the sharp fall in the housing prices of that period. Note that, the other estimated change points are only associated with change in the mean while the variance remains constant. Note that, the time series plots in Figure 3.3 points out to stationarity of the estimated residual. Furthermore. QQ-plot in Figure 3.3 confirm the normality assumption. Thus, our proposed change point model regarding the rate of change in the HPI of the borough of Newham seems to be reasonable.

The change point analysis is conducted on housing price index of the borough of the

Tower Hamlets by S3 based enhanced backward detection. Similar to previous two boroughs, the significant rise of the variance is detected within the period 2008-2010 which may be attributed to the 2007-2009 financial crisis. Fortunately, based on the Figure 3.4, the normality assumption is justifiable.

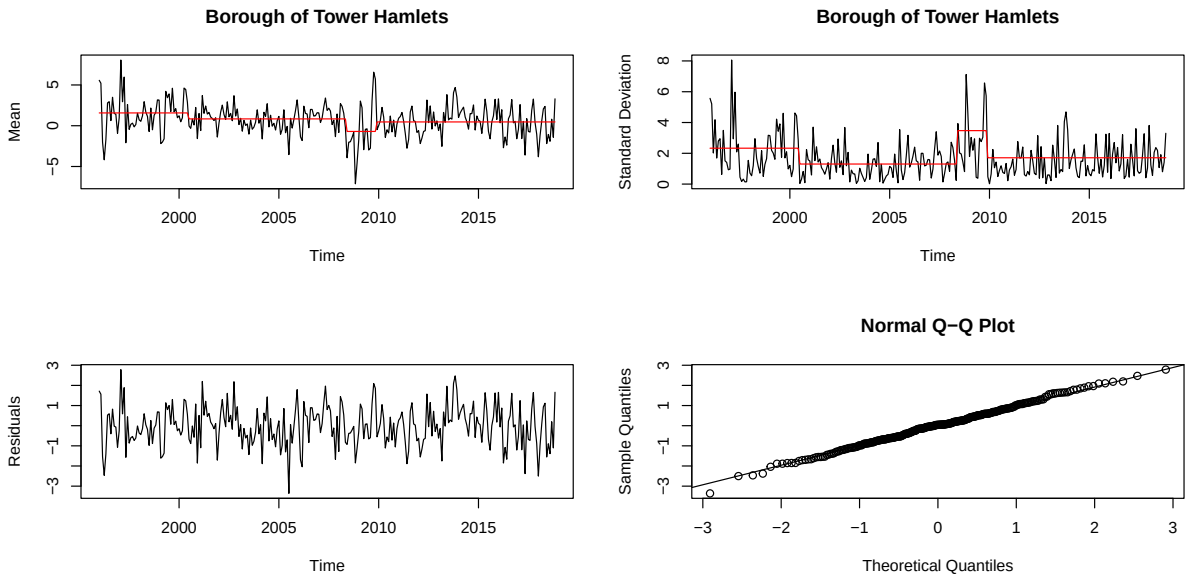

Figure 3.4: Enhanced backward detection is applied on the Tower Hamlets dataset via the assumptions of the change point model S3.

3.5.2 Global Surface Temperature

This data set consists of the yearly average of the monthly temperature anomalies recorded from January 1880 to June 2018. The anomalies here are defined as the difference between the global temperature average in a given month and the baseline value. The baseline value is obtained by averaging the temperature data collected over the 30 years starting from 1951 to 1980. By simply looking at Figure 3.5, one can realize the change in distribution occurs only with respect to the mean which is a piecewise linear function of time. To conduct a change point analysis in this data set, we assumed the random variable are generated independently and normally. Moreover, the mean is assumed to be piecewise linear and

variance is constant function of time. We referred to this change points model and the variation of the enhanced backward detection which is devised based on the aforementioned change point model as S4 change point model and S4 based enhanced backward detection, respectively.

The estimated change points are located in the years 1903, 1946 and 1964. Based on Figure 3.5, the surface temperature anomalies remain constant from 1880 to 1903. Another notable observation in Figure 3.5 is the consistent rise of surface temperature anomalies since 1964. Moreover, the global surface temperature rose at a much sharper rate in this period in comparison to other periods which are designated by enhanced backward detection.

Similar to HPI datasets, the accuracy of the fitted change point model is assessed via the estimated residuals. The recorded residuals are inspected in order to check whether they follow a normal distribution or not. Note that the normality of the data set is verified by the help of Shapiro test and QQ-plot. Both methods indicate that the sequence of residuals is normally distributed. The normality of the residuals is evident to the goodness of the fitted model by enhanced backward detection.

3.5.3 Covid-19 Daily Cases

One of the main events of 2020 was the worldwide spread of the Covid-19 which is attributed to the high contingency of the disease. Due to high transmission rate, the public health resources have been become overwhelmed. In order to mitigate the risk of the depletion of the public health resources, the local and central governments devised extreme public policies such as public lockdown. Economic lockdown is proven to put heavy tolls on the economical and mental health of the society. Thus, the effectiveness of these policies are required to be evaluated. This assessment can be done by estimating the change in the trends of the daily new cases. Obviously, the decreasing trends imply the effectiveness of the policies and increasing trends points out otherwise. In this section, we conducted the enhanced backward detection on the daily Covid-19 cases from March/31/2020 to Jun/02/2021 in Ontario.

We assumed the random variables are generated normally and independently with piecewise linear mean and piecewise constant variance. The parameter vector of the mean and variance functions changes simultaneously after occurrence of each change point. We call this change point model and the variation of the enhanced backward detection which is devised based on the presumed distribution of the daily new cases, S5 change point model and S5 based enhanced change point model, respectively. The mean and variance function

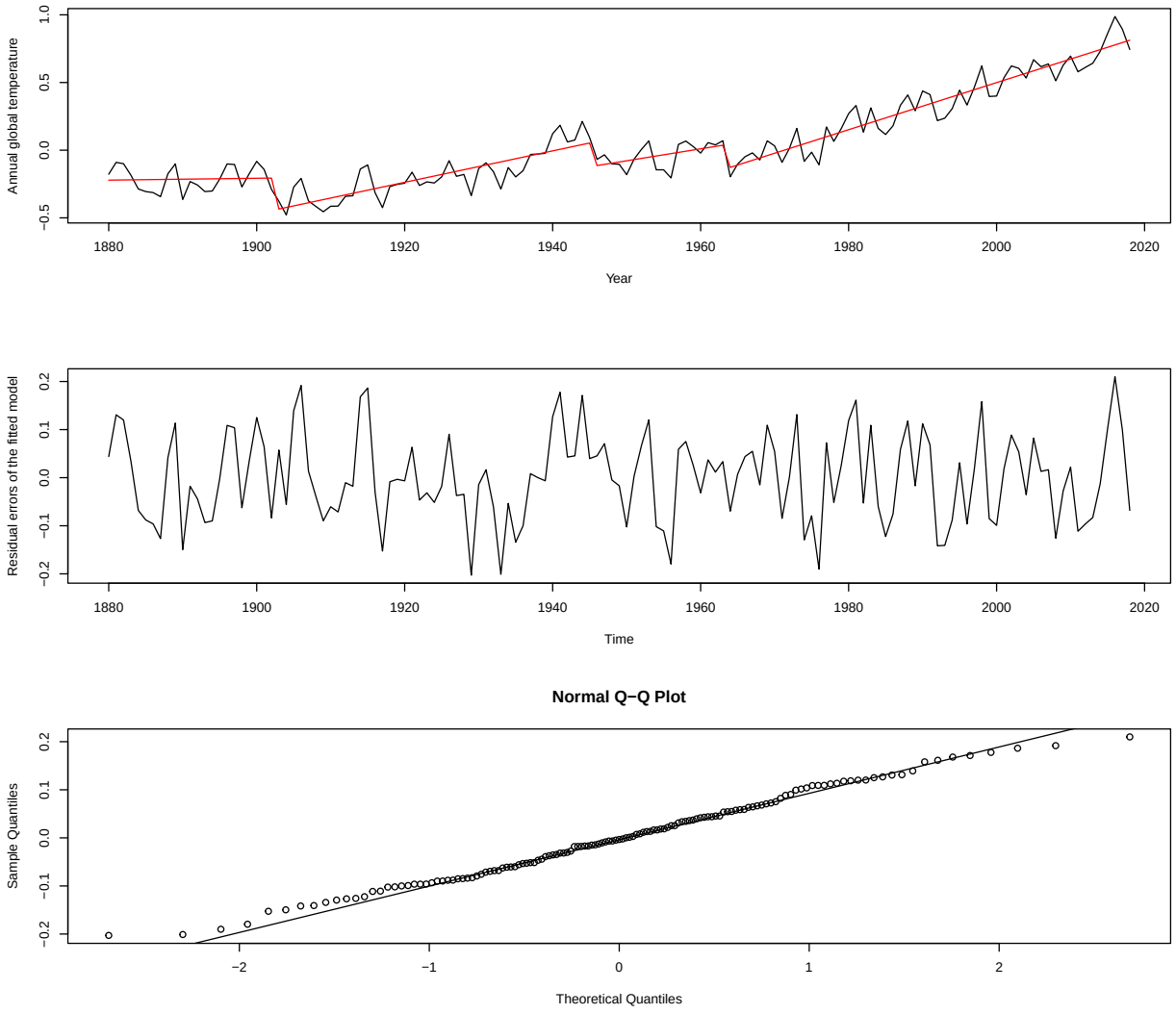

Figure 3.5: The simple linear model is fitted between the estimated change points for the global surface temperature dataset. One notable observation is that the temperature has been risen in a sharper rate since mid 1960s.

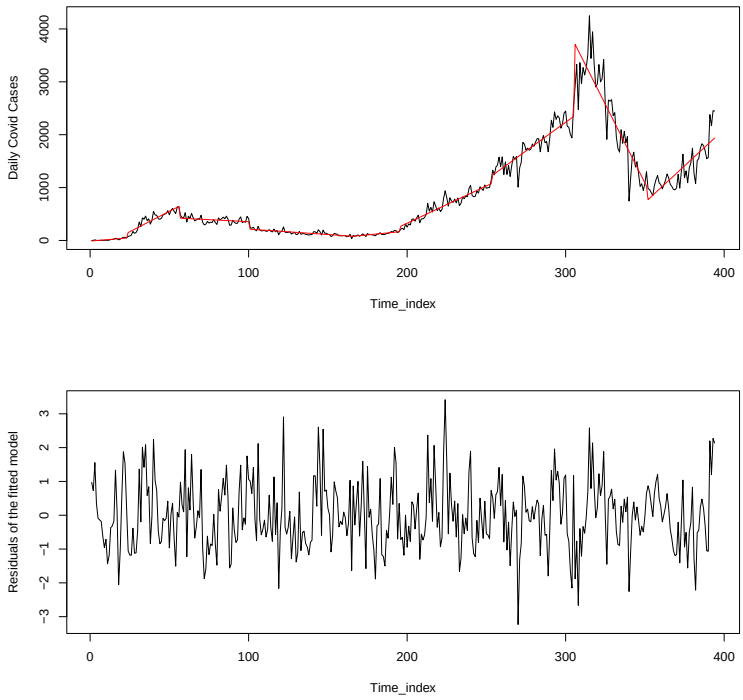

Figure 3.6: The trend of the daily new Covid-19 cases changes after the introduction of the public health policies.

are fitted according to the estimated change points by enhanced backward detection. Based on the Figure 3.6, after introducing each lock down in the province of Ontario, the trend in the daily new cases had decreased significantly. This observation implies the effective of these public health measures in controlling the spread of Covid-19.

Chapter 4

Rank Based enhanced backward detection

4.1 Introduction

In Chapter 3, the enhanced backward detection method (EBD) was introduced in the parametric setting. EBD estimates the locations of change points by using the generalized log-likelihood ratio test statistic locally. Throughout the procedure, multiple active sets are obtained. The active set which minimizes the Bayesian information criteria is assigned to be the estimated change points set. To perform a generalized log-likelihood ratio test effectively, a parametric family of distributions for the random variables must be specified with a reasonable accuracy. Otherwise, the change point analysis via EBD would not be sound. Therefore, non-parametric change point tests are reasonable alternatives to the generalized log-likelihood ratio test, when the distribution family of the random variables can not be determined with a reasonable accuracy.

This chapter mainly focuses on detecting changes in location and/or scale parameters of independently distributed random variables. In other words, considering the random variables $X_1, X_2, \ldots, X_T$ are generated independently from distribution functions $F_1, F_2, \ldots, F_T$ with location parameters $\mu_1, \mu_2, \ldots, \mu_T$ and scale parameters $\sigma_1^2, \sigma_2^2, \ldots, \sigma_T^2$, respectively. The problem of interest is to recover the *location change points* $\tau_1, \tau_2, \ldots, \tau_N$

and *scale change points* $\eta_1, \eta_2, \ldots, \eta_M$ such that:

$$\begin{aligned} &X_t = \mu_t + \sigma_t \epsilon_t \quad \text{for} \quad t = 1, \cdots, T, \quad \text{where} \\ &\mu_1 = \cdots = \mu_{\tau_1} \neq \mu_{\tau_1+1} = \cdots = \mu_{\tau_2} \neq \cdots = \mu_{\tau_N} \neq \mu_{\tau_N+1} = \cdots = \mu_T, \\ &\sigma_1^2 = \cdots = \sigma_{\eta_1}^2 \neq \sigma_{\eta_1+1}^2 = \cdots = \sigma_{\eta_2}^2 \neq \cdots = \sigma_{\eta_M}^2 \neq \sigma_{\eta_M+1}^2 = \cdots = \sigma_T^2, \quad \text{and} \end{aligned} \tag{4.1}$$

the sequence of random variables $\epsilon_1, \epsilon_2, \ldots, \epsilon_T$ are independently and identically distributed with location zero and scale 1. Note that, the location and scale change points are not identical, that is generally:

$$\{\tau_1, \tau_2, \ldots, \tau_N\} \neq \{\eta_1, \eta_2, \ldots, \eta_M\}.$$

To identify the location and scale change points in model 4.1, we propose substituting the generalized log-likelihood ratio test with a suitable rank based single change point test statistic in the EBD. We refer to this variation of EBD as *rank based enhanced backward detection* (REBD). To ensure the convergence of REBD, the underlying change point test statistic is required to be derived from a loss function as demonstrated in (3.5). The implemented test statistic in REBD algorithm will be elaborated more precisely in Section 4.2. Then REBD will be introduced formally in Section 4.3. Finally, change point analysis is conducted with respect to real world datasets via REBD algorithm. Moreover, performance of the rank based and parametric enhanced backward detection are compared in term of accuracy.

4.2 Rank Based Change Point Test

Suppose the sequence of random variables $X_1, X_2, \ldots, X_T$ is segmented into $N+1$ non-overlapping partitions as follow:

$$\{X_t\}_{t=1}^{\tau_1}, \ \{X_t\}_{t=\tau_1+1}^{\tau_2}, \ \cdots, \ \{X_t\}_{t=\tau_N+1}^{T},$$

where the random variables in each segment have a common location parameter. Note that, the location parameter of the adjacent segments may be equal or unequal with each other. To verify, whether the location parameter remains constant throughout the sequence of random variables or not, the following test is conducted:

$$\mathcal{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_T \quad \text{versus} \quad \mathcal{H}_1 : \text{otherwise}.$$

This investigation can be conducted by the help of Kruskal–Wallis test statistic (Kruskal and Wallis (1952)). Note that, Kruskall-Wallis test is a ranked based test. Thus, prior to calculating the Kruskall-Wallis test statistic, the observations are ranked as follow:

$$r_t = \sum_{s=1}^{T} 1_{\{X_s<X_t\}} + 0.5 \sum_{s=1}^{T} 1_{\{X_t=X_s\}} + 0.5 \quad \text{for} \quad t = 1,\, 2,\, \ldots,\, T. \tag{4.2}$$

Then, Kruskall-Walls test statistic is obtained as follow:

$$H(0, \tau_1, \cdots, \tau_N, T) = \sum_{i=1}^{N+1} (\tau_i - \tau_{i-1})(\overline{r}_i - \overline{r})^2, \tag{4.3}$$

where

$$\overline{r}_i = \frac{1}{\tau_i - \tau_{i-1}} \Big(\sum_{t=\tau_{i-1}+1}^{\tau_i} r_t \Big) \quad \text{and} \quad \overline{r} = \frac{1}{T} \sum_{t=1}^{T} r_t.$$

The large value of the test statistic H implies that some location parameters or center of sub-population are different from one another.

Now we switch our attention to the special case in which the dataset is partitioned into two segments: $\{X_\ell\}_{\ell=0}^{t}$ and $\{X_\ell\}_{\ell=t+1}^{T}$. In this scenario, the Kruskal–Wallis test can be reduced to:

$$H(0, t, T) = \left(\frac{|\overline{r}_1 - \overline{r}_2|}{\sqrt{\left(\frac{1}{t} + \frac{1}{T-t}\right)}} \right)^2 = \left[\sum_{\ell=1}^{T} (r_\ell - \overline{r})^2 - \Big(\sum_{\ell=1}^{t} (r_\ell - \overline{r}_1)^2 + \sum_{\ell=t+1}^{T} (r_\ell - \overline{r}_2)^2 \Big) \right]. \tag{4.4}$$

Suppose the index after which the location parameter of the random variables changes is unknown. The point which maximizes the test statistic (4.4) is the most likely candidate for the location parameter change point. That is, the location change point within the sequence of the independently distributed random variables X_1, X_2, …, X_T is estimated as follow:

$$\tau = \underset{0<t<t}{\operatorname{argmax}}\, \mathcal{U}(t \mid 0,\, T), \quad \text{where} \quad \mathcal{U}(t \mid 0, T) = \sqrt{H(0,\, t,\, T)}. \tag{4.5}$$

The test statistic in (4.5) is referred to as *rank CUSUM statistic* in this chapter. Note that, since rank CUSUM statistic is derived from sum of the square error of the rank of observations (aka, a loss function), we can conclude the sweeping process in REBD will terminates.

In the context of multiple change point detection problem, we compute the rank CUSUM statistic (4.5) locally and with respect to intervals of the random variables. The local rank CUSUM statistic with respect to the sub-sequence of the random variables $X_{e+1}, \ldots, X_s$ is calculated as follow:

$$\tau = \arg\max_{s<t<r} \left| \frac{\overline{r}_{_1} - \overline{r}_{_2}}{\sqrt{\frac{1}{t-e} + \frac{1}{s-t}}} \right|, \tag{4.6}$$

where

$$\overline{r}_{_1} = \frac{1}{t-e} \sum_{i=e+1}^{t} r_{_i} \quad \text{and} \quad \overline{r}_{_2} = \frac{1}{s-t} \sum_{i=t+1}^{s} r_{_i}.$$

Note that, the sequence of random variables $\{X_t\}_{t=1}^{T}$ is ranked globally as shown in (4.2). In other words, to conduct the rank CUSUM statistic with respect to the interval $(e, t]$, the random variables $X_{e+1}, X_{e+2}, \ldots, X_s$ are ranked as follow:

$$r_{_t} = \sum_{\ell=1}^{T} 1_{\{X_\ell < X_t\}} + 0.5 \sum_{\ell=1}^{T} 1_{\{X_t = X_\ell\}} + 0.5 \quad \text{for} \quad t = e+1, 2, \ldots, s.$$

If the maximized change point test statistic (4.6) is sufficiently large, then the index which maximizes the change point test statistic (4.6) within the interval $(e, s]$ is assigned to be the location change point. In this setting, the underlying probability distribution function of the random variables is not required to be known. However, since the rank CUSUM statistic (4.6) is derived from Kruskal-Wallis test, the random variables in the adjacent segments are assumed to have the same variability or shape.

4.3 Rank based enhanced backward detection

The rank based enhanced backward detection (REBD) utilizes the rank CUSUM statistic (4.5) to recover the location change points. At first, the sequence of the observations are ranked globally as shown in Equation (4.3). Then EBD is launched with respect to the rank of the observations. Throughout the execution of the rank based enhanced backward detection, multiple active sets are obtained. In the context of REBD, an active set is referred to the set of change point candidates $\{\tau_1, \ldots, \tau_N\}$ such that

$$\tau_{_j} = \underset{\tau_{j-1}<t<\tau_{j+1}}{\operatorname{argmax}} \ \mathcal{U}(t \mid \tau_{_{j-1}}, \tau_{_{j+1}}), \tag{4.7}$$

where $\tau_1 < \cdots < \tau_N$, $\tau_0 = 0$, $\tau_{N+1} = T$ and the change point test $\mathcal{U}$ is the rank CUSUM statistic in equation (4.6). An active set with N change point candidates is denoted by $\mathcal{M}(N)$.

Prior to the launch of REBD, the upper bound for the number of change points is specified and denoted by $N_{\max}$. Next, the set of change point candidates with at most $N_{\max}$ elements is constructed in a manner that the distance between change point candidates is at least 2 (i.e, $\tau_j - \tau_{j-1} \geq 2$ for $j = 1, 2, \ldots, N_{\max} + 1$). Then the sweeping process is launched with respect to the global ranks of random variables and the aforementioned set of change point candidates. After run of the sweeping process, the first active set (i.e., $\mathcal{M}(N_{\max})$) is obtained. At each run of REBD, the change point candidate with the smallest rank CUSUM statistic is eliminated and subsequently the sweeping process is launched to obtain a new active set. This procedure continues until a null active set is obtained. After the execution of REBD algorithm, $N_{\max}+1$ distinct active sets are obtained. The penalized loss function in equation (4.8) with respect to each active set is calculated. The penalized loss function with respect to the active set $\mathcal{M}(N) = \{\tau_1, \ldots, \tau_N\}$ is obtained by:

$$\mathcal{L}(\mathcal{M}(N)) = \Big(\sum_{j=1}^{N+1} \sum_{t=\tau_{j-1}+1}^{\tau_j} (r_t - \overline{r}_j)^2 \Big) + 0.1\, N\, T^2\, \log(T), \tag{4.8}$$

where $\overline{r}_j$ is the mean of $\{r_{\tau_{j-1}+1}, \cdots, r_{\tau_j}\}$. The loss function in (4.8) was proposed as a rank based variation of Schwartz Information Criteria (Wang et al. (2020)). Among the recorded active sets, the one with the smallest penalized loss function (4.8) is assigned to be the estimated change points set. The estimated change points set by the rank based enhanced backward detection is denoted by $\mathcal{M}$. The REBD algorithm is demonstrated in Algorithm 3.

The REBD algorithm is primarily designed to recover the potential change points in location of the distribution. More precisely, the rank CUSUM statistic (4.6) is not sensitive to change in scale or variability of the random variables. Fortunately, the ranked based enhanced backward detection can recover scale change points by modifying the ranking procedure. One way to recover the scale change points by REBD is as follow:

Step1: The location change points model is estimated by REBD.

Step2: The residuals of the estimated change points model is calculated:

$$R_t = X_t - \widehat{\mu}_t.$$

Result: $\mathcal{M}$
INPUTS: $X_1, \cdots, X_T$ and $\mathcal{M}(N_{\max}) = \{\tau_1, \cdots, \tau_{N_{\max}}\}$;
$\{r_t\}_{t=1}^T \leftarrow \text{rank}(\{X_t\}_{t=1}^T)$;
$\mathcal{M} = \mathcal{M}(N) \leftarrow \mathcal{SP}(\{r_t\}_{t=1}^T, \mathcal{M}(N_{\max}), 1, N_{\max})$;
$\mathcal{L} \leftarrow \mathcal{L}(\mathcal{M})$;
$N \leftarrow N_{\max}$;
while $N \geq 0$ **do**

$j \leftarrow \underset{\tau_\ell \in \mathcal{M}(N)}{\text{argmin}} \left[\mathcal{U}(\tau_\ell | \tau_{\ell-1}, \tau_{\ell+1})\right]$, where $\tau_0 = 0$ and $\tau_{N+1} = T$;

$\mathcal{M}^*(N-1) \leftarrow \{\tau_1^*, \cdots, \tau_{N-1}^*\}$ where

$$\tau_l^* = \tau_l \quad \text{for} \quad l = 1, \cdots, j-1;$$
$$\tau_l^* = \tau_{l+1} \quad \text{for} \quad l = j, \cdots, N-1;$$

$N \leftarrow N - 1$;
$c \leftarrow \max(2, j-1)$;
$m.c \leftarrow \min(N,\, c+1)$;
$\mathcal{M}(N) \leftarrow \mathcal{SP}(\{r_t\}_{t=1}^T, \mathcal{M}^*(N), c, m.c)$;
if $\mathcal{L}(\mathcal{M}(N)) < \mathcal{L}$ **then**
$\mathcal{L} \leftarrow \mathcal{L}(\mathcal{M}(N))$;
$\mathcal{M} \leftarrow \mathcal{M}(N)$;
end
end

Algorithm 3: Rank based enhanced backward detection.

Step3: The location change points within the squared residuals (i.e., R_t^2) are estimated by REBD. The estimated location change points with respect to squared residuals are assigned to be scale change points with respect to the original dataset.

Step4: The scale change points model is fitted with respect to the original sequence of random variables X_1, X_2, ..., X_T.

Step5: The residuals of the proposed model are obtained as follow:

$$R_t^* = X_t / \widehat{\sigma}_t.$$

Step6: REBD is conducted with respect to the R_t^* in order to update the estimated location change points model. The estimated locations change points with respect to $R_{t\,t=1}^{*T}$ are assigned to be the location change points of the original dataset.

Step7: Steps 2-7 are repeated until the location and scale change points models can not be updated.

4.4 Simulation Study

4.4.1 Assessing the Accuracy of REBD

In this section, REBD is conducted with respect to replications of the location change points scenarios which were introduced in details in Section 3.4. Recall that, all of the synthesized datasets in this section are generated from $M1$ to $M5$ change points models. Moreover, these datasets are constructed under the assumption of normality. For each of the change points models, 1000 replications are generated. Then REBD procedure is performed with respect to each replication and estimated location change points are collected and plotted in Figure 4.1. Ideally, the estimated location change points in Figure 4.1 should cluster around the true mean change points.

Based on the histogram in Figure 4.1, REBD underperformed in comparison to the likelihood based EBD. The relative inferiority of REBD in comparison to the parametric approach is expected, since the underlying assumptions of EBD are satisfied in this particular simulation study. In the location change points models $M1$, $M2$, $M3$ and $M4$, the parametric enhanced backward detection performs slightly better. Unlike the results in the aforementioned change points models, the relative underperformance of the rank based enhanced backward detection is more visible under the $M5$ change points model. The rank based enhanced backward detection grossly underestimated the number of the change points in $M5$ scenario. The poor accuracy of rank based enhanced backward detection with respect to $M5$ location change points model can be attributed to the choice of the Schwartz information criteria (4.8). Thus, we recommend different variation of the rank based penalized loss function which is inspired by Schwarz information criterion (Schwarz et al. (1978)). Our proposed rank based penalized loss function with respect to the candidate set $\mathcal{M}(N) = \{\tau_1, \tau_2, \ldots, \tau_N\}$ and the sequence of random variables $X_1, X_2, \ldots, X_T$

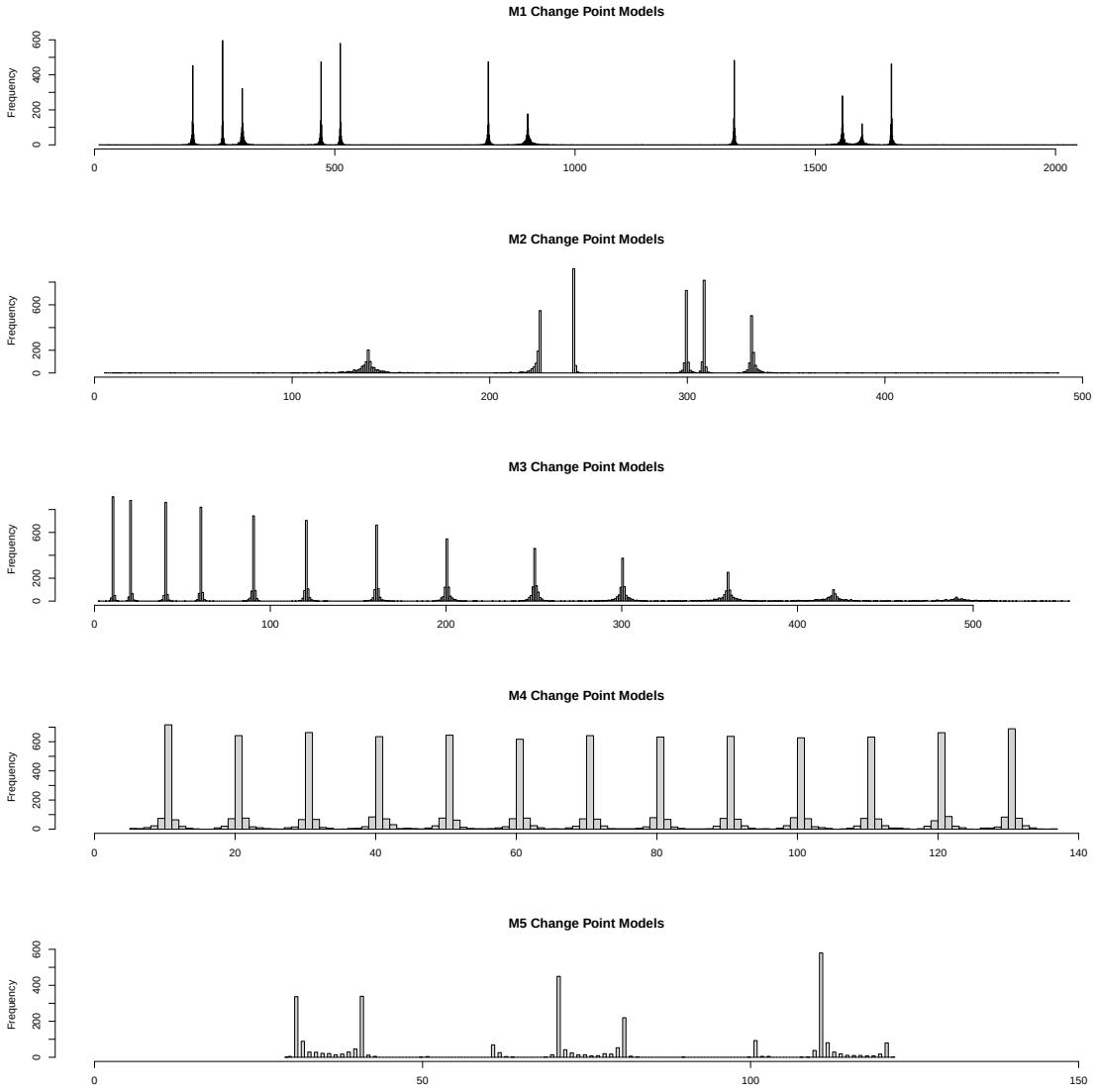

Figure 4.1: The spike in the histogram corresponds to true locations of the change points.

is calculated:

$$\mathcal{L}_2(\mathcal{M}(N)) = \frac{T}{2}\log\left(\frac{\sum_{j=1}^{N+1}\sum_{t=\tau_{j-1}+1}^{\tau_j}\left(r_t - \overline{r}_j\right)^2}{T}\right) + N\log(T), \quad \text{where}$$
$$\overline{r}_j = \frac{1}{\tau_j - \tau_{j-1}}\sum_{t=\tau_{j-1}}^{\tau_j} r_t, \quad \text{and} \tag{4.9}$$

and r_t is the rank of the t-th observations (i.e., $r_t = \sum_{s=1}^{T} 1(X_s < X_t) + 0.5\sum_{s=1}^{T} 1(X_t = X_s) + 0.5$). After replacing the Shwarz information criterion (4.8) with the penalized loss function $\mathcal{L}_2$ (4.9), the performance of rank based enhanced backward detection is boosted according to Figure 4.2. More precisely, the rank based enhanced backward detection has become more sensitive to change in location parameter.

According to the results from Figures 4.1 and 4.2, the performance of the rank based enhanced backward detection has improved after replacing the rank based BIC (4.8) with the penalized loss function (4.9). To further compare the performance of both approaches, we design a simulation study under more general scenarios, so that, the number of observations and change points, locations parameters between the change points and locations of change points are determined randomly. In this study, ten thousands sequence of random variables are generated. Then both rank based enhanced backward detection via rank Schwarz Information criterion (4.8) (REBD1) and the ranked based enhanced backward detection via loss function (4.9) (REBD2) are performed with respect to each generated dataset. Each sequence of random variables are generated as follow:

- The number of observations is randomly selected from the set {100, 200, 500, 1000, 2000} and denoted by T.
- The number of change points N is drawn from a Poisson distribution with a rate which is randomly selected from the set {4, 8}.
- The locations of the change points are selected randomly from the indexes of the random variables, i.e., $\{1, 2, \ldots, T-1\}$.
- Jump size after each change point is randomly drawn from a Gaussian distribution with mean zero and variance σ_J^2 which is randomly selected from the set {1, 2, 10}.
- Finally the Gaussian white noise with the unite variance is added to the constructed mean function.

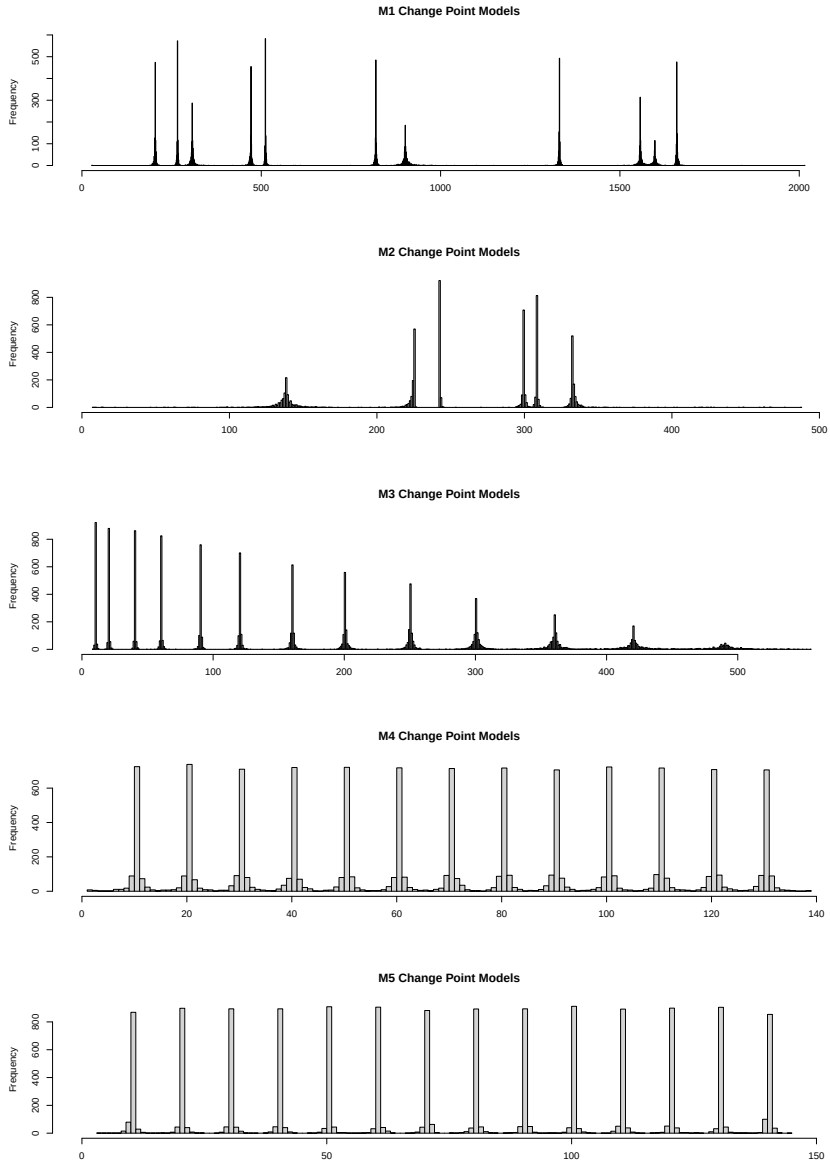

Figure 4.2: Performance of the rank based enhanced backward detection has improved after replacing the Bayesian Information Criteria (4.8) with the loss function (4.9).

Note that, this simulation study was designed by Fryzlewicz to determine the value of threshold for wild binary segmentation (Fryzlewicz et al. (2014)). The results of this simulation study is recorded in Table 4.1. Table 4.1 indicates the vast superiority of REBD2 over REBD1. Thus, the rank based loss function (4.9) is applied within the rank based enhanced backward detection from now on.

	REBD1	REBD2
MSE	0.97	0.40

Table 4.1: REBD1 is referred to the rank based enhanced backward detection with the rank based BIC (4.8). REBD2 is referred to the rank based enhanced backward detection with the loss function (4.9).

4.4.2 Performance of REBD on non-Gaussian data

Due to optimally of the test statistic used in the parametric EBD, the parametric enhanced backward detection is supposed to perform better than REBD, when the underlying distributions of the random variables are Gaussian. Here, we investigate the accuracy of the enhanced backward detection in non-parametric setting. More precisely, we inspect the performance of the enhanced backward detection when underlying distribution function is not Gaussian. Then the accuracy of the estimated location change points models by parametric and rank based enhanced backward detection methods are compared with each other. In the following simulation study, the additive normally distributed randomized noise in the change points models which were introduced in Section 3.4 is replaced with non Gaussian error terms. The aforementioned change points models are represented by $M1$, $M2$, $M3$, $M4$ and $M5$ in Table 4.2. Moreover, the rank and likelihood based enhanced backward detection methods are launched with respect to the sequence of random variables with no change points. We refer to this particular change points model as null model and it is denoted by $M0$. Even though, the simulated datasets are not normally distributed, the parametric enhanced backward detection is constructed based on the assumptions of the normal mean change points model to study its robustness to non-normalized datasets. Both variations of the enhanced backward detection are conducted with respect to the generated data and mean square errors of the estimated location change points models are calculated

$$\frac{1}{T}\sum_{t=1}^{T}(\mu_t - \widehat{\mu}_t)^2,$$

where μ_t and $\widehat{\mu}_t$ are true and estimated mean of the random variable X_t.The average of mean square errors are recorded in Table 4.2. Based on the results in Table 4.2, the enhanced backward detection is outperformed by the rank based enhanced backward detection when the underlying distributions are not Gaussian. Moreover, REBD is competitive with enhanced backward detection in normal mean change points models.

MSE								
Enhanced Backward Detection	Change Point Model	Standard Normal	Scaled t_{10}	Scaled t_5	Laplace	Skewed Normal	Skewed t_{10}	Skewed t_5
EBD	M0	0.00615	0.0303	0.146	0.00543	0.0126	0.0296	14.7
REBD		0.00373	0.00442	0.00593	0.000304	0.00383	0.00345	2.09
EBD	M1	2.44	4.27	15.5	5.26	2.66	5.268	0.0167
REBD		2.54	3.16	3.98	2.16	2.43	2.318	0.00653
EBD	M2	0.00361	0.00636	0.0182	0.00792	0.00407	0.00752	3.79
REBD		0.00587	0.00714	0.00893	0.00597	0.00609	0.00567	1.22
EBD	M3	1.57	2.30	4.88	2.33	1.64	2.18	0.0756
REBD		1.45	1.80	2.28	1.25	1.39	1.31	0.0563
EBD	M4	0.0572	0.105	0.182	0.0648	0.0596	0.0655	0.0353
REBD		0.0535	0.0922	0.145	0.0556	0.0567	0.0580	0.0291
EBD	M5	0.0219	0.0326	0.0564	0.0297	0.0221	0.0269	0.164
REBD		0.0193	0.0300	0.0522	0.0256	0.0201	0.0230	0.00372

Table 4.2: The rank based enhanced backward detection (REBD) is much more accurate than the enhanced backward detection (EBD), if the random variables are generated from non-Gaussian distribution.

4.4.3 Performance of REBD under no-change point model

One concern in the change points detection problem is false discovery of change points. In this section, we study the performance of the rank based enhanced backward detection in the models with zero change points. In our experiments, we consider datasets of size $T = 100, 200, 500, 1000,$ and 2000 generated under the assumption of no-change point from a Gaussian distribution. Note that, since these models contain no change points, they are defined based on the number of observations. For each of these change points model, 10000 replications are generated. Finally, the estimated number of change points by EBD and REBD are recorded and the results are summarized in Table 4.3.

The Number of Detected Change Points by rank based Enhanced Backward Detection ($\widehat{N}$)	T=100	T=200	T=500	T=1000	T=2000
$\widehat{N} = 0$	10000	10000	10000	10000	10000

Table 4.3: The rank based enhanced backward detection is not susceptible to false discovery of change points.

According to Tables 3.3 and 4.3, REBD is more robust to false discovery of change points in comparison to the parametric enhanced backward detection. Although the recorded results in Table 4.3 seems to be desirable at first look, one may worry about the potential insensitivity of REBD to detecting location change points. In other words, the magnitude of a change in a location parameter must be more significant when the rank based enhanced backward detection is applied in comparison to the occasions in which the parametric enhanced backward detection is the method of choice.

4.5 Data Analysis

4.5.1 Housing Price Indexes

In this section, the rank based enhanced backward detection is applied with respect to the housing price index datasets. We compare the estimated change points by rank based enhanced backward detection with our findings in Section 3.5.1. Furthermore, the estimated mean and variance change points by the rank based enhanced backward detection are aligned to the events that may have influenced the rise or fall of the housing price.

Although the changes in location and scale parameters are visible in some segments of the HPI of Hackney, the rank enhanced backward detection has not recovered any change points according to Figure 4.3. This observation implies the rank based enhanced backward detection method is more conservative than the enhanced backward detection. Recall that, enhanced backward detection recovered the increase in variance during the financial crisis.

With regard to HPI of the borough of Newham, the rank based enhanced backward detection has not recovered any scale change points. However, the rank based enhanced

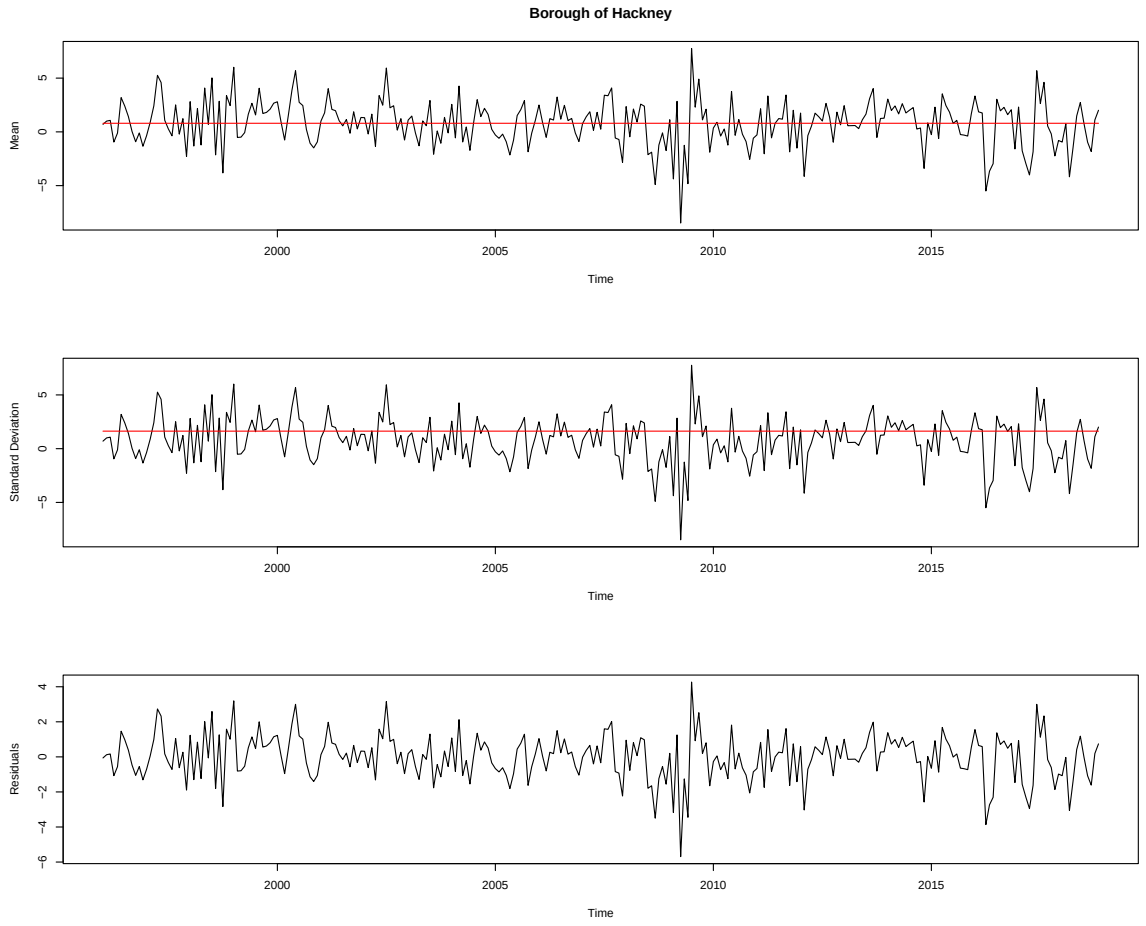

Figure 4.3: The top two plots indicates the variance change points. However, after eliminating the apparent outliers, the adjusted HPI of Hackney is stationary.

backward detection has recovered multiple location change points. One notable observation is that, the rank based enhanced backward detection has not recovered any change points during the financial crisis. This potential failure of the rank based enhanced backward detection can be due to the fact that scale change points which are associated with the financial rescission are too close to each other. In other words, the rank based enhanced backward detection has misidentified the significant increase in the scale parameter of random variables during the financial rescission as a cluster of outliers. This failure of the rank based enhanced backward detection is visible on the time series plot of the residuals of the fitted change points model in Figure 4.4.

Finally, the rank based enhanced backward detection is conducted with respect to HPI of the borough of Tower Hamlets. Unlike likelihood based enhanced backward detection, no change in location or scale parameters is detected by rank based enhanced backward detection. Similar to the HPI of the borough of Hackney and Newham, the rank based enhanced backward detection fails to recover the locations of change points due to narrowness of the segments which are constructed by change points. This shortcoming of the rank based enhanced backward detection can be attributed to the global ranking of the random variables.

4.5.2 Copy Number Variation

According to Wang et al. (2020), "the gene copy number (CN) is the number of copies of a particular gene in the genotype of an individual, which can be measured by microarray experiments. In healthy cells, there are two copies of DNA. In tumor cells, parts of chromosomes may be deleted or amplified several times and it will lead to the copy number of such regions being different from 2". The multiple change point detection methods can be applied to a copy number variation dataset in order to determine the region in which a copy number changes. Most often, the random variables are assumed to be generated from the normal mean change points model. Chen & Wang proposed a variation of the circular binary segmentation (Chen and Wang (2008)) which is sensitive to both change in mean and variance of the random variables. That is, the problem of interest is to identify the points in which both mean and variance have changed simultaneously. However, the copy number variation is mainly concerned with the change in the location parameter of the random variables and identifying the points that scale parameter changes is not the problem of interest. Thus, the rank based enhanced backward detection seems to be an attractive approach. In this section, the datasets GSE11976 and GSE29172 are uploaded from the package `acnr`. The estimated location change points model is represented with the red line in Figure 4.6.

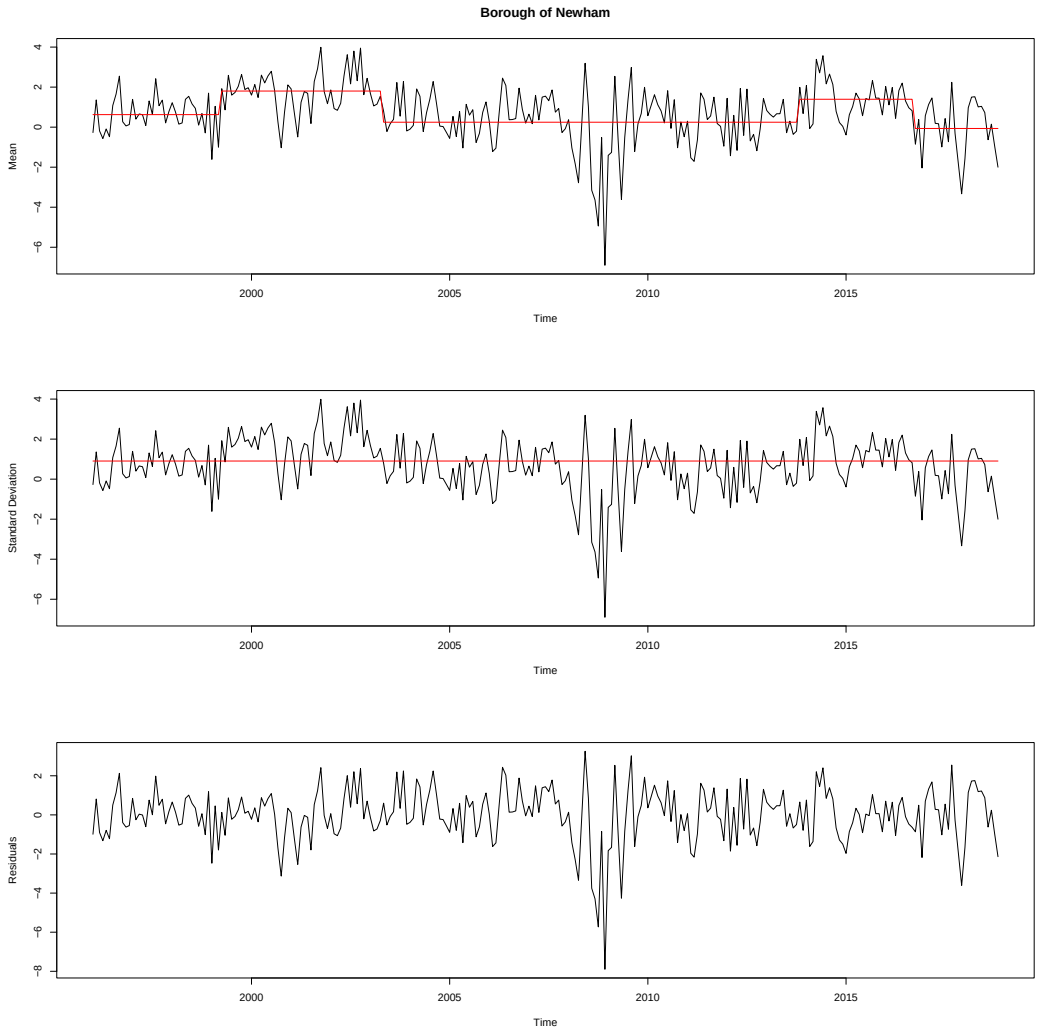

Figure 4.4: Rank based enhanced backward detection has failed to to recover change in variance.

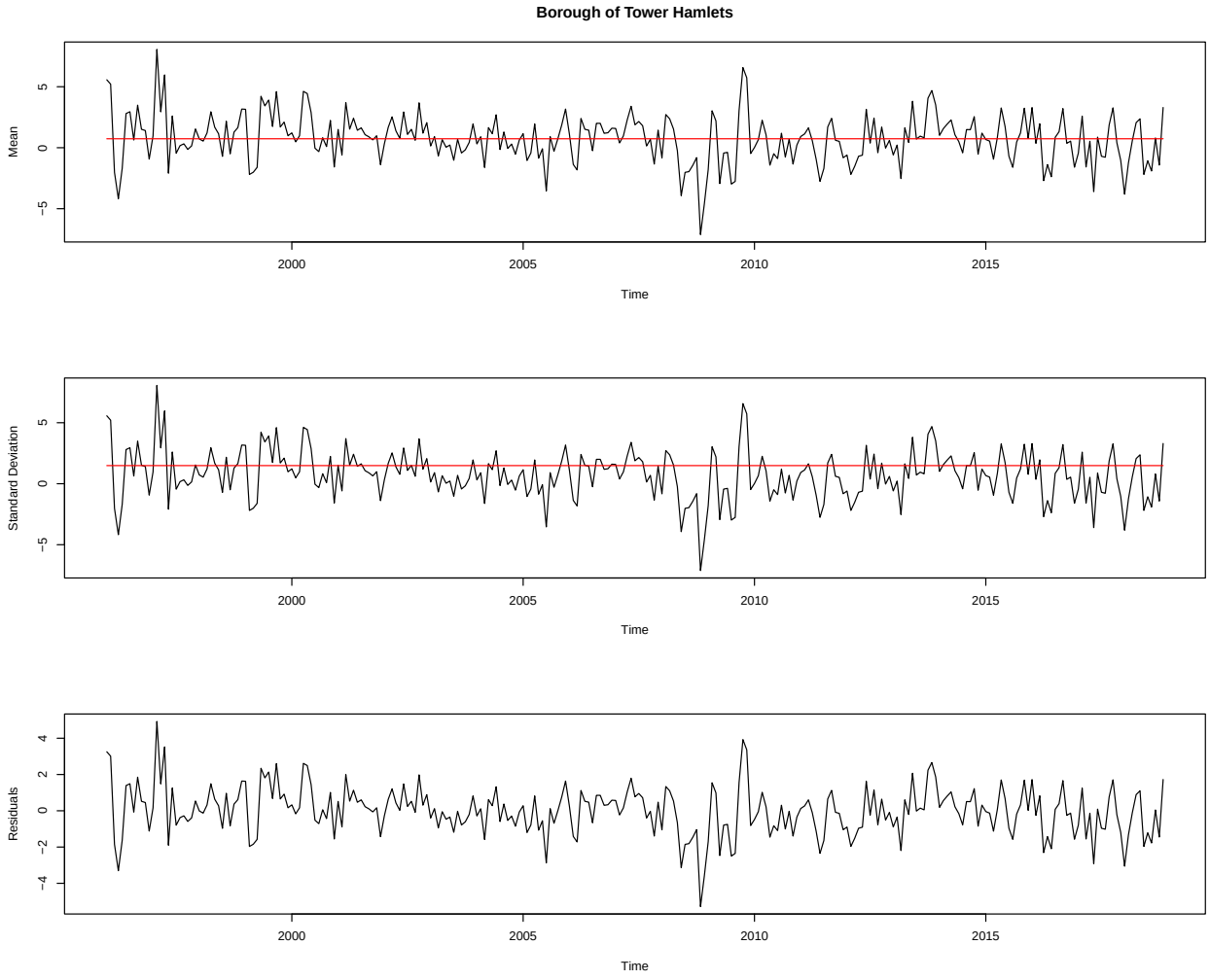

Figure 4.5: After elimination of the apparent outliers, the adjusted data looks weakly stationary.

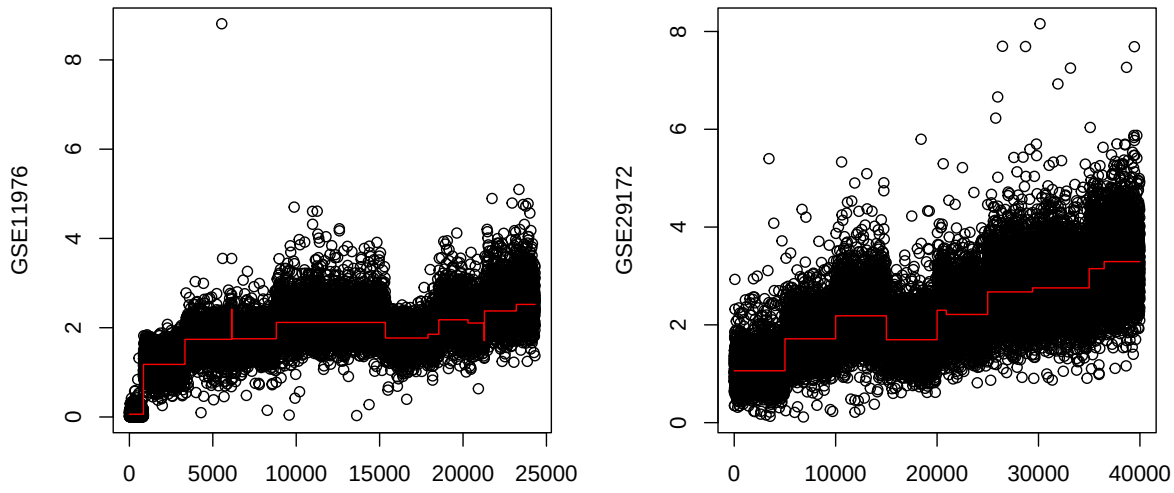

Figure 4.6: The fitted mean of CNV by REBD is represented by red line.

4.5.3 Polls

In this section, we have investigated the raw data from the national polls during the 2016 election term. The goal of the change point analysis on this dataset is comparing the popularity of the candidates of the mainstream parties throughout the various periods according to the national polls. These candidates are Hillary Clinton and Donald Trump. To assess the relative popularity of the two main presidential candidates, Trump's share vote is subtracted from Clinton's vote in each poll. The resulting quantity is denoted by H-T. Next, the mean of the votes difference H-T is estimated by the sample mean based on the recovered mean change points via rank based enhanced backward detection. The positive estimated mean implies the overall popularity of Hilary Clinton over Donald Trump and negative estimated mean suggests otherwise. In this section, we only use the surveys from Ipsos polling organization. We obtained this data set from `dslabs` R package.

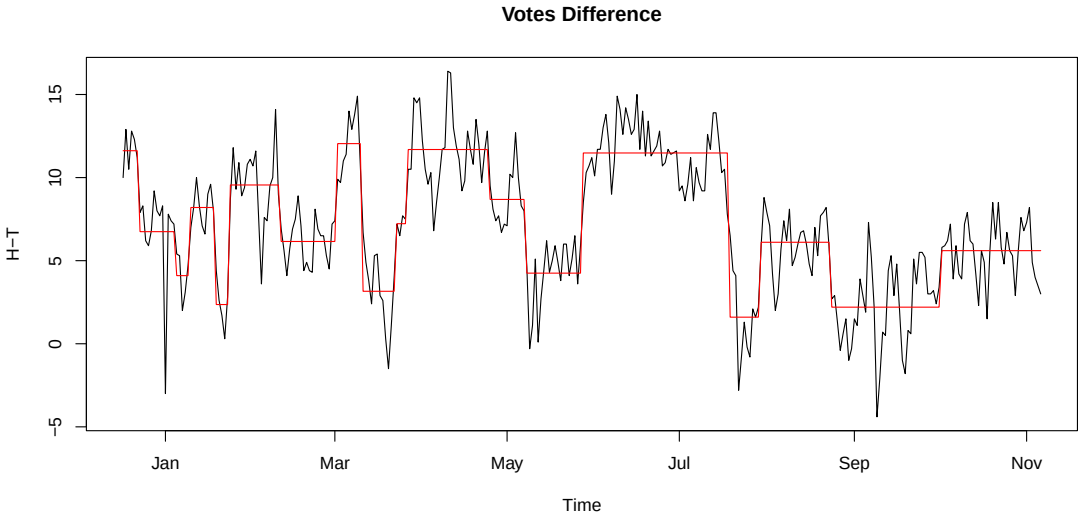

Figure 4.7: The solid red lines indicates the estimated mean,

Based on Figure 4.7, Hillary Clinton consistently had a larger share of the votes according to the national poll. Moreover, the mean of the last recovered segment is 5.6 %. That is, according to the national polls, Hillary Clinton's share of total vote would have exceeded Trump's vote by 5.6% in the election day. However, Hillary Clinton had only 2.8% more votes than Trump in 2016 US presidential election. Thus, the national polls had overestimated the popularity of Hillary Clinton.

Similar to HPI datasets, after fitting the estimated local change points models and obtaining the residuals, the scale change points are obtained. For this purpose, the rank based enhanced backward detection is launched with respect to the squared value of the residuals. However, no scale change points was detected by the rank based enhanced backward detection. According to Figure 4.8, the residual of the estimated mean change points has a constant variance.

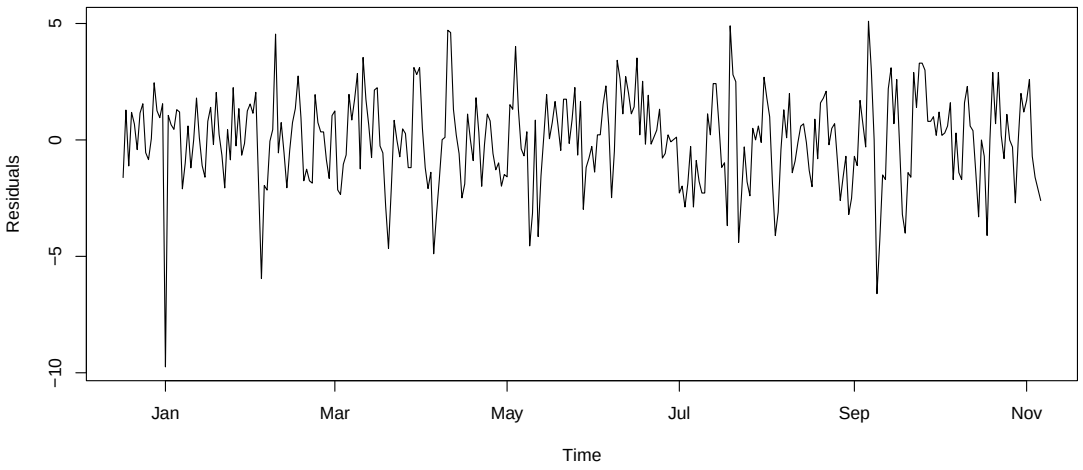

Figure 4.8: No change in variance is visible with respect to the residuals of the estimated mean change points model.

4.6 Final Remarks on REBD

As we discussed in Section 3.2, the sweeping process terminates, if the underlying rank based single change point test statistic in REBD is derived from some sort of a loss function as shown in equation (4.4). It is import to note that, a sequence of random variables are ranked globally prior to launch of the rank based enhanced backward detection. Otherwise, if the random variables are ranked locally, then the sweeping process and consequently REBD will not converge. Unlike REBD, the local search rank based methods are most often ranked locally. For instance, the observations in each of the randomly drawn intervals are ranked separately in context of the rank based wild binary segmentation (Ross (2021)). As a result of global ranking, the rank based enhanced backward detection is less likely to recover the locations of closely located change points. For example, the rank based enhanced backward detection did not perform accurately in the HPI datasets. Note that, ranking of the random variables locally can be time consuming. Since the sequence of random variables is ranked only once at the beginning of the rank based enhanced backward detection, this approach is more computationally efficient in comparison to other rank based local search methods. Therefore, despite of the loss of accuracy due to global ranking, the rank based enhanced backward detection remains competitive to locally ranked based

methods.

As demonstrated in Section 4.5, the rank based enhanced backward detection is much easier to use in comparison to the parametric enhanced backward detection. No assumption on the distribution function is required in the rank based enhanced backward detection. As demonstrated in Table 4.2, REBD is proven to be superior to EBD, when the underlying distribution of the random variables is non-Gaussian and competitive in the normal mean change point scenarios. Thus, when the parametric distribution of random variables cannot be specified accurately, the use of rank based enhanced backward detection is preferred.

Chapter 5

Narrowest Over Threshold via Interval Selection with Shorten Exhaustive Search

Local search methods estimate the positions of change points by performing a test within the intervals of random variables. If a change point test statistic is conducted within an interval with multiple change points, then the aforementioned change points may offset each other and become undetectable to the conducted test. Thus, one of the main concerns of local search methods is obtaining intervals with exactly one change point. The narrowest over threshold method attempts to obtain such intervals by constructing large number of randomized intervals. Since some of the intervals in which a predefined threshold is satisfied may contain multiple change points, the narrowest over threshold prioritizes shorter intervals over the longer ones. In other words, among the randomized intervals that are associated with substantially large test statistics, change points are estimated based on the narrower ones. As the number of change points or observations increases, larger number of randomized intervals is required. Note that, as a result of drawing insufficient number of randomized intervals, the number of change points may get underestimated. Since determining an appropriate number of randomized intervals is a challenging and yet crucial task, we propose a data adaptive interval generating mechanism which we name *interval selection via shorten exhaustive search* (IS.SES). IS.SES algorithm generates intervals based on the data configuration of the random variables. Thus, the number of intervals is not required to be determined by practitioners. Later on in this chapter, we propose substituting the randomized intervals with IS.SES generated ones in the narrowest over threshold. We call the resulting change point detection method *narrowest over threshold*

with interval selection via shorten exhaustive search (NOT-IS.SES).

IS.SES method relies on a single change point test statistic for recovering the locations of change points candidates. In the context of IS.SES algorithm, a single change point test must be derived from an additive loss function. Such a loss function is calculated as follow:

$$\mathcal{L}(0,\, \tau_1,\, \tau_2, \ldots,\, \tau_N,\, T) = \sum_{j=0}^{N} \mathcal{C}(\tau_j + 1 : \tau_{j+1}), \quad \text{where} \quad \tau_0 = 0, \quad \tau_{N+1} = T, \tag{5.1}$$

and the set $\{\tau_1,\, \tau_2,\, \ldots,\, \tau_N\}$ is a change point candidate set. Moreover, the function $\mathcal{C}(a + 1, b)$ is referred to as the cost function between indexes a and b. One notable example of the loss function (5.1) is sum of square errors. Consider the sequence of random variables $X_1,\, X_2,\, \ldots,\, X_T$ with piece-wise constant mean and constant variance, then sum of square errors of the mean change points candidate set $\{\tau_1,\, \tau_2,\, \ldots,\, \tau_N\}$ is calculated as follow:

$$\mathcal{L}_{\text{SSE}}(0,\, \tau_1,\, \tau_2,\, \ldots,\, \tau_N,\, T) = \sum_{j=0}^{N} \mathcal{C}(\tau_j + 1 : \tau_{j+1}), \quad \text{where}$$

$$\mathcal{C}(\tau_j + 1 : \tau_{j+1}) = \sum_{t=\tau_j+1}^{\tau_{j+1}} (X_t - \overline{X}_j)^2, \quad \text{and} \quad \overline{X}_j = \frac{1}{\tau_{j+1} - \tau_j} \sum_{t=\tau_j+1}^{\tau_{j+1}} X_t\,.$$

By simple algebraic manipulations, the loss function (5.1) can be represented as follow:

$$\mathcal{L}(0,\, \tau_1,\, \tau_2,\, \ldots,\, \tau_N,\, T) = \mathcal{L}(0,\, \tau_1,\, \ldots,\, \tau_{\ell-1},\, \tau_\ell) + \mathcal{L}(\tau_\ell,\, \tau_{l+1},\, \ldots,\, \tau_N, T).$$

This property of the loss function is referred to as the *additive* property. For the additive loss function $\mathcal{L}$, the corresponding change point detection test statistic $\mathcal{U}(\cdot \,|\, a,\, b)$ with respect to the sub-sequence $X_{a+1},\, X_{a+2},\, \ldots, X_b$ is calculated as follow:

$$\mathcal{U}(t \mid\, a,\, b) = \mathcal{L}(a,\, b) - \mathcal{L}(a,\, t,\, b) = \mathcal{C}(a + 1 : b) - \Big[\mathcal{C}(a + 1 : t) + \mathcal{C}(t + 1 :\, b)\Big], \tag{5.2}$$

and a candidate change point estimate is

$$\widehat{\tau} = \underset{t \in \{a+d,\, \ldots,\, b-d\}}{\arg\max} \Big[\mathcal{U}(t \mid\, a,\, b)\Big],$$

where d is the number of parameters that may change throughout random variables. For instance, consider the normal mean change point model (2.3); since the variance remains constant and change in distribution occurs with respect to the mean, then d is set to be one.

Consider the change point candidate set $\{\alpha,\ \beta\}$, the relationship between a loss function (5.1), aka $\mathcal{L}(0,\ \alpha,\ \beta,\ T)$ and change point test statistics (5.2) ($\mathcal{U}(\alpha \mid\ 0,\ \beta)$ and $\mathcal{U}(\beta \mid\ \alpha, T)$) is as follow:

$$\begin{aligned}
\mathcal{L}(0,\ \alpha,\ \beta,\ T) &= \mathcal{L}(0,\ \alpha\,\beta) + \mathcal{L}(\beta,\ T) \\
&= \Big[\mathcal{C}(1:\ \alpha) + \mathcal{C}(\alpha+1:\ \beta)\Big] + \mathcal{C}(\beta+1:\ T) \\
&= \Big[\mathcal{C}(1:\ \beta) - \mathcal{U}(\alpha \mid\ 0,\ \beta)\Big] + \mathcal{C}(\beta+1:\ T) \\
&= \Big[\mathcal{C}(1:\beta) + \mathcal{C}(\beta+1:T)\Big] - \mathcal{U}(\alpha \mid\ 0,\ \beta) \\
&= \mathcal{L}(0,\ T) - \mathcal{U}(\alpha \mid\ 0,\ \beta) - \mathcal{U}(\beta \mid\ 0,\ T),
\end{aligned} \tag{5.3}$$

and equivalently

$$\mathcal{L}(0,\ \alpha,\ \beta,\ T) = \mathcal{L}(0,\ T) - \mathcal{U}(\beta \mid\ \alpha,\ T) - \mathcal{U}(\alpha \mid\ 0,\ T).$$

Therefore,

$$\mathcal{U}(\beta \mid\ \alpha,\ T) = \mathcal{U}(\alpha \mid\ 0,\ \beta) + \mathcal{U}(\beta \mid\ 0,\ T) - \mathcal{U}(\alpha \mid\ 0,\ T). \tag{5.4}$$

Equation (5.4) is the main motivation of the IS.SES algorithm. By running the IS.SES algorithm, large number of intervals are discarded prior to conducting the change point test. According to equation (5.4), the calculated change point test statistic in the eliminated intervals will be smaller than a predefined threshold. Furthermore, IS.SES algorithm is able to reduce the *search space* of the change point test statistic. That is, the point which maximizes the change point test statistic is selected among the smaller set of indexes. Thus, the computational complexity of IS.SES is reduced in comparison to an exhaustive search. In the next section, IS.SES algorithm is elaborated more precisely.

5.1 Interval Selection via Shorten Exhaustive Search

The goal of the IS.SES algorithm is to generate relatively narrow intervals in which a predefined threshold is satisfied. One of the main component of IS.SES is the *shorten exhaustive search* (SES) algorithm which discards some of the unnecessary intervals and reduces the search space of the test. The SES algorithm searches for a single interval in which a predefined threshold is satisfied. After obtaining each interval, the SES algorithm is relaunched on the remaining dataset to recover another change point candidate. This process continues, until the SES algorithm can not recover any interval or change point candidate. Note that, if the SES algorithm does not recover any interval in which a

predefined threshold is satisfied, then we can conclude such an interval does not exist. We will discuss this property of the SES algorithm in more detail.

Suppose the SES algorithm is launched with respect to the sequence of random variables $X_1, X_2, \ldots, X_T$, where the threshold is set to be λ. Consider the interval $(\alpha, \tau]$ with data points $X_{\alpha+1}, X_{\alpha+2}, \ldots, X_\tau$, where $\alpha \geq d$. To calculate the change point test statistic within the interval $(\alpha, \tau]$, the following condition is required to be satisfied:

$$\max_{i \in \{\alpha+d, \ldots, \tau\}} \Big[\mathcal{U}(\alpha \mid 0, i)\Big] + \max_{j \in \{\alpha, \ldots, \tau-d\}} \Big[\mathcal{U}(j \mid 0, \tau)\Big] - \mathcal{U}(\alpha \mid 0, \tau) \geq \lambda. \tag{5.5}$$

If the condition (5.5) is not met and $\alpha \geq d$, then the maximum of the change point test statistic (5.2) will not exceed a predefined threshold, since

$$\begin{aligned}\mathcal{U}(\beta \mid \alpha, \tau) &= \mathcal{U}(\alpha \mid 0, \beta) + \mathcal{U}(\beta \mid 0, \tau) - \mathcal{U}(\alpha \mid 0, \tau) \\ &\leq \max_{i \in \{\alpha+d, \ldots, \tau\}} \Big[\mathcal{U}(\alpha \mid 0, i)\Big] + \max_{j \in \{d, \ldots, \tau-d\}} \Big[\mathcal{U}(j \mid 0, \tau)\Big] - \mathcal{U}(\alpha \mid 0, \tau) < \lambda,\end{aligned}$$

for $\alpha + d \leq \beta \leq \tau - d$. Note that, if $\alpha < d$, then the change point test statistic will be calculated without examining condition (5.5).

Suppose condition (5.5) is satisfied with respect to the interval $(\alpha, \tau]$ and $\alpha \geq d$. The index such $\beta \in \{\alpha + d, \alpha + d + 1, \ldots, \tau - d\}$ will be in the search space of the test, if the following conditions are hold true:

$$\begin{aligned}&\max_{i \in \{\alpha+d, \ldots, \tau\}} \Big[\mathcal{U}(\alpha \mid 0, i)\Big] + \mathcal{U}(\beta \mid 0, \tau) - \mathcal{U}(\alpha \mid 0, \tau) \geq \lambda \quad \text{and} \\ &\max_{j \in \{d, \ldots, \beta-d\}} \Big[\mathcal{U}(j \mid 0, \beta)\Big] + \mathcal{U}(\beta \mid 0, \tau) - \mathcal{U}(\alpha \mid 0, \tau) \geq \lambda.\end{aligned} \tag{5.6}$$

If $\alpha \geq d$ and the condition (5.6) is not satisfied, then $\mathcal{U}(\beta \mid \alpha, \tau) < \lambda$ by equation (5.4) and hence elimination of the index β from the search space of the test. Assuming none of the indexes in the set $\{\alpha+d, \alpha+d+1, \ldots, \tau-d\}$ satisfies the condition (5.6) and $\alpha \geq d$, then no test will be conducted with respect to the interval $(\alpha, \tau]$. Note that, if $\alpha < d$, then a change point test is performed within the interval $(\alpha, \tau]$ and the search space of the test includes all possible indexes within the aforementioned interval (i.e. $\{\alpha + d, \alpha + d + 1, \ldots, \tau - d\}$).

Recall that the SES algorithm investigates all possible intervals within the sequence of random variables. . This search is conducted as follow:

- Set $t = 2\,d$.

- While $t \leq T$ and the desirable interval has not been recovered yet do:
  - Set $\eta = \alpha = t - d$.
  - Set $\mathcal{U}_{\max} = 0$.
  - while the desirable interval has not been recovered and $\alpha > 0$ do:
    * if $\alpha \geq d$:
      · Calculate the test statistic $\mathcal{U}(\alpha \mid 0, t)$ and subsequently update the following quantity:

$$\max_{i \in \{\alpha+d, \ldots, t\}} \Big[\mathcal{U}(\alpha \mid 0, i)\Big] = \max\Big(\mathcal{U}(\alpha \mid 0, t), \max_{i \in \{\alpha+d, \ldots, t-1\}} \Big[\mathcal{U}(\alpha \mid 0, i)\Big]\Big),$$

$$\max_{j \in \{\alpha, \ldots, t-d\}} \Big[\mathcal{U}(j \mid 0, t)\Big] = \max\Big(\mathcal{U}(\alpha \mid 0, t), \max_{j \in \{\alpha+1, \ldots, t-d\}} \Big[\mathcal{U}(j \mid 0, t)\Big]\Big).$$

      · If $\mathcal{U}(\alpha \mid 0, t) > \mathcal{U}_{\max}$, then

$$\mathcal{U}_{\max} = \mathcal{U}(\alpha \mid 0, t) \quad \text{and} \quad \eta = \alpha.$$

      · If the condition in (5.5) is satisfied with respect to the interval $(\alpha, t]$, then collect all of the indexes which meet the requirements (5.6) in the set $\mathcal{S}$.
      · If $\mathcal{S} \neq \emptyset$, then estimate the change point as follow:

$$\widehat{\tau} = \operatorname*{argmax}_{i \in \mathcal{S}} \Big[\mathcal{U}(i \mid \alpha, t)\Big].$$

      Otherwise, if $\mathcal{S} = \emptyset$, then the interval $(\alpha, t]$ can be ignored and the change point test is not conducted with respect to the aforementioned interval.

    On the other hand, if $\alpha < d$, estimate the change point as follow:

$$\widehat{\tau} = \operatorname*{argmax}_{i \in \{\alpha+d, \ldots, t-d\}} \Big[\mathcal{U}(i \mid \alpha, t)\Big]$$

    * If the interval $(\alpha, t]$ is selected and $\mathcal{U}(\tau \mid \alpha, t) \geq \lambda$, then stop the procedure and assign the row vector

$$[\tau, \alpha, t, \mathcal{U}(\tau \mid \alpha, t)]$$

    to be the output of SES. Otherwise, set $\alpha = \alpha - 1$.

– If SES has not stopped yet, then if $\mathcal{U}_{\max} \geq \lambda$, stop the procedure and set the output to be

$$[\eta,\, 0,\, t,\, \mathcal{U}_{\max}].$$

On the other hand, if $\mathcal{U}_{\max} < \lambda$, then set $t = t + 1$.

Note that, if $i - j < d$, $k - i < d$ or $k - i < 2d$, then the test statistic $\mathcal{U}(i \mid j,\, k) = 0$.

The SES algorithm stops, when either the value of t exceeds the number of observations or an interval in which a predefined threshold is satisfied, is obtained. Therefore, the SES algorithm returns either no outcome or a row vector. For instance, if the SES algorithm recovers an interval $(\alpha,\, t]$, where τ maximizes the change point test statistic within the aforementioned interval and $\mathcal{U}(\tau \mid \alpha,\, t) \geq \lambda$, then the output of the SES is

$$[\tau,\, \alpha,\, t,\, \mathcal{U}(\tau \mid \alpha,\, t)].$$

On the other hand, if the SES algorithm returns no output, then we conclude that the threshold is not satisfied within any of the intervals in the sequence of random variables. That is,

$$\mathcal{U}(i \mid j,\, k) < \lambda,$$

where i, j, and k are indexes of the random variables.

To facilitate the computation of the SES algorithm, a $T \times 4$ matrix $\mathcal{R}$ is constructed prior to launch of the procedure. The first column of the matrix $\mathcal{R}$ contains the indexes of the random variables. The remaining entries are filled up during the execution of the SES algorithm. For instance, after calculating the quantity $\mathcal{U}(\alpha \mid 0,\, t)$ in the SES algorithm, the entries of the matrix $\mathcal{R}$ are updated as follow:

$$\begin{aligned}
\mathcal{R}[\alpha,\, 2] &= \max\Big(\mathcal{R}[\alpha,\, 2],\, \mathcal{U}(\alpha \mid 0,\, t)\Big),\\
\mathcal{R}[t,\, 3] &= \max\Big(\mathcal{U}(\alpha \mid 0,\, t),\, \mathcal{R}[t,\, 3]\Big)\,,\\
\mathcal{R}[\alpha,\, 4] &= \mathcal{U}(\alpha \mid 0,\, t).
\end{aligned} \tag{5.7}$$

Recall that, the change point test statistic is conducted with respect to the interval $(\alpha,\, \tau]$, if the condition (5.5) is satisfied. The requirement (5.5) can be rewritten in terms of the matrix $\mathcal{R}$:

$$\mathcal{R}[\alpha,\, 2] + \mathcal{R}[t,\, 3] - \mathcal{R}[\,\alpha,\, 4] \geq \lambda\,. \tag{5.8}$$

Furthermore, the conditions in (5.6) can be written as follow:

$$\begin{aligned} \mathcal{R}[\alpha,\, 2]+\mathcal{R}[\beta,\, 4] - \mathcal{R}[\alpha,\, 4] \geq \lambda\,, \\ \mathcal{R}[\beta,\, 3]+\mathcal{R}[\beta,\, 4] - \mathcal{R}[\alpha,\, 4] \geq \lambda\,. \end{aligned} \tag{5.9}$$

The SES algorithm is a prominent component of IS.SES algorithm and recovers a single interval in which a predefined threshold is satisfied. To obtain the locations of the remaining change point candidates, the SES is called frequently within the IS.SES algorithm. More precisely, the IS.SES algorithm is performed with respect to the sequence of random variables $X_1,\, X_2,\, \ldots,\, X_T$ as follow:

- Set $\ell = 0$.
- Construct the matrix $\mathcal{M}_{0\times 4}$.
- Set $STOP = 0$;
- While $STOP = 0$ do:
  - Conduct the SES algorithm with respect to the sub-sequence of random variables $X_{\ell+1},\, X_{\ell+2}, \ldots,\, X_T$.
  - If no change point is recovered by SES algorithm, then stop the procedure ($STOP = 1$).
  - Otherwise, if the SES algorithm recovers a change point τ from the interval $(\alpha,\, \beta]$, then
    - record the row vector $[\tau,\, \alpha,\, \beta,\, \mathcal{U}(\tau \mid \alpha,\, \beta)]$ in the matrix $\mathcal{M}$.
    - set $\ell = \alpha + 1$.

The output of the IS.SES is a matrix $\mathcal{M}$. Each row of the matrix $\mathcal{M}$ contains lower and upper bounds of intervals in which a predefined threshold is satisfied, a change point candidate and the value of change point test statistic. Suppose IS.SES identifies that the calculated change point test statistic (5.2) in the interval $(\alpha,\, \tau]$ is larger λ and the index β is the change point candidate within the aforementioned interval, then the row vector

$$[\beta,\, \alpha,\, \tau,\, \mathcal{U}(\beta \mid \alpha,\, \tau)]$$

is recorded in the matrix $\mathcal{M}$. Note that, the index β maximizes the test statistic $\mathcal{U}(\cdot \mid \alpha,\, \tau)$ within the interval $(\alpha,\, \tau\,]$. Generally, we refer to the indexes that maximize the test statistic

(5.2) within the recovered intervals as *change point candidates*. Suppose the change point candidates β and η are extracted from the intervals $(\alpha, \tau]$ and $(\gamma, \delta]$, respectively. Then we say indexes β and η are *conflicted*, if $\beta \in (\gamma, \delta)$ or $\eta \in (\alpha, \tau)$.

Notice that IS.SES-generated intervals can be used as replacements to randomized intervals in a narrowest over threshold (Baranowski et al. (2019)). Similar to the traditional narrowest over threshold (Baranowski et al. (2019)), the change point candidate which is extracted from the narrowest IS.SES-generated interval is assigned to be a change point estimate. Next, all of the change points candidates which are conflicted with the estimated change point are discarded. This procedure continues until all recorded change point candidates in the matrix $\mathcal{M}$ are either assigned to be change point estimates or discarded due to their conflict with the estimated change points. We refer to this variation of the narrowest over threshold algorithm as the *narrowest over threshold with interval selection via shorten exhaustive search* (NOT-IS.SES).

5.2 Ultimate Test Statistic Maximizer

In the previous section, the *interval selection via shorten exhaustive search* (IS.SES) was introduced. Then IS.SES was implemented in the narrowest over threshold algorithm as sort of an interval generating mechanism. We call the resulting method *narrowest over threshold with interval selection via shorten exhaustive search* (NOT-IS.SES). As we discussed, a primary component of IS.SES which is responsible for recovering intervals in which a predefined threshold is satisfied, is the *shorten exhaustive search* (SES) algorithm. In this section, we propose some modifications to SES algorithm in order to find the largest maximized change point test statistic value within all possible intervals of the random variables. That is, the modified variation of the SES algorithm is applied with respect to the sequence of random variables $X_1, X_2, \ldots, X_T$ in order to obtain the quantity Λ^* such that:

$$\Lambda^* = \max_{(i,j,k)} \Big[\mathcal{U}(j \mid i, k) \Big] \quad \text{where} \quad |i-j| \geq d \quad , \quad |k-j| \geq d. \tag{5.10}$$

and the single change point test statistic $\mathcal{U}$ is derived from a loss function as demonstrated in (5.2). We refer to this variation of the SES algorithm as *Ultimate Test Statistic Maximizer* (UTSM).

Recall that, the SES algorithm stops as soon as an interval in which a predefined threshold λ is satisfied, is obtained. In the UTSM algorithm, a predefined threshold λ

is replaced with the largest calculated change point test statistic up to the current stage of the UTSM algorithm. This quantity is denoted by Λ^* in this thesis. The quantity Λ^* which can be regarded as a threshold in the context of the UTSM algorithm may not remain constant. Suppose, the single change point test statistic is calculated with respect to the interval $(\alpha, \tau]$ as follow:

$$\beta = \underset{i \in \mathcal{S}}{\operatorname{argmax}} \Big[\mathcal{U}(i \mid \alpha, \tau) \Big],$$

where $\mathcal{S}$ is defined as a search space of the test. If $\mathcal{U}(\beta \mid \alpha, \tau) > \Lambda^*$, then the quantity Λ^* is updated as follow:

$$\Lambda^* = \mathcal{U}(\beta \mid \alpha, \tau).$$

Thus, unlike the SES algorithm in Section (5.1), the UTSM algorithm stops only when all of the possible intervals are examined. Similar to the SES algorithm, the satisfaction of the conditions (5.5) and (5.6) are assessed with respect to each interval and then the single change point test statistic (5.2) is calculated within the intervals that satisfy the aforementioned conditions. UTSM is conducted as follow:

- Set $t = 2\,d$.
- Set $\Lambda^* = 0$.
- The output vector is assinged to be

$$\mathcal{O} = [0,\, 0,\, 0,\, 0].$$

- While $t \leq T$ do:
  - Set $\eta = \alpha = t - d$.
  - Set $\mathcal{U}_{\max} = 0$.
  - while $\alpha > 0$ do:
    - if $\alpha \geq d$:
      - Calculate the test statistic $\mathcal{U}(\alpha \mid 0, t)$ and update the entries of the matrix $\mathcal{R}$ as demonstrated in (5.7).
      - If $\mathcal{U}(\alpha \mid 0, t) > \mathcal{U}_{\max}$, then

$$\mathcal{U}_{\max} = \mathcal{U}(\alpha \mid 0, t) \quad \text{and} \quad \eta = \alpha.$$

· If condition in (5.8) is satisfied with respect to the interval $(\alpha,\, t]$, then collect all of the indexes which meet the requirements (5.9) in the set $\mathcal{S}$.

· If $\mathcal{S} \neq \emptyset$, then evaluate the change point test statistic $\mathcal{U}$ in $\mathcal{S}$ and find:

$$\tau = \underset{i \in \mathcal{S}}{\operatorname{argmax}} \Big[\mathcal{U}(i \mid \alpha,\, t)\Big]$$

On the other hand, if $\alpha < d$, calculate the change point test statistic as follow:

$$\tau = \underset{i \in \{\alpha+d,\, \ldots,\, t-d\}}{\operatorname{argmax}} \Big[\mathcal{U}(i \mid \alpha,\, t)\Big]$$

∗ If $\mathcal{U}(\tau \mid \alpha,\, t) > \Lambda^{*}$, then output vector and Λ^{*} are updated as follow:

$$\Lambda^{*} = \mathcal{U}(\tau \mid \alpha,\, t) \quad \text{and}$$
$$\mathcal{O} = [\tau,\, \alpha,\, t,\, \Lambda^{*}].$$

∗ Set $\alpha = \alpha - 1$.

– If $\mathcal{U}_{\max} \geq \Lambda^{*}$, the output vector $\mathcal{O}$ and Λ^{*} are updated as follow:

$$\Lambda^{*} = \mathcal{U}(\eta \mid 0,\, t) \quad \text{and}$$
$$\mathcal{O} = [\eta,\, 0,\, t,\, \Lambda^{*}].$$

– Set $t = t + 1$.

In addition to estimating the locations of change points, another concern of the local search methods is setting up a threshold for avoiding false discovery of change points. The threshold is supposed to be constructed based on the probability distribution of the random variables. This goal can be achieved by a bootstrap re-sampling method. The general scheme of the threshold selection by parametric bootstrap re-sampling is as follow:

- Draw B bootstrap replications;
- Perform the underlying local search method with respect to each replication and record the largest calculated change point test statistic;
- The $(1-\alpha) \times 100$-th percentile of the recorded change point test statistic is assigned to be the threshold;

The value of α is selected based on the sensitivity of the practitioner to overestimation of the number of change points. As the value of α decreases, the threshold size increases and subsequently the estimated number of change points decreases. The largest calculated change point test statistic which is obtained by a local search method may not be equal to the quantity Λ^* in (5.10). Here, we propose obtaining the quantity Λ^* in each replication of the random variables by UTSM algorithm. Then $(1-\alpha) \times 100$-th percentile of the obtained Λ^* by UTSM is assigned to be the threshold. Our hope is that, by implementing UTSM within the parametric bootstrap re-sampling, the quality of the threshold will be improved and the false discovery of change points is reduced.

5.3 Alternative Exhaustive Search

The IS.SES and UTSM algorithms perform similarly to conducting exhaustive search. However, some of the intervals in which a predefined threshold is not satisfied are eliminated prior to conducting a change point test. These intervals are identified by the help of equation (5.5). Moreover, the search space of the conducted tests in the aforementioned algorithms is reduced with the help of the conditions in equation(5.6). Here, we propose an alternative algorithm to exhaustive search. That is, similar to the exhaustive search, all of the possible intervals are examined; However, a single change point test statistic is often calculated differently. We refer to this approach as *alternative exhaustive search* (AES).

Similar to IS.SES and UTSM algorithms, the underlying assumption of AES is that a single change point test statistic is derived from some sort of a loss function as shown in equation (5.2). In the AES algorithm, the single change point test statistic is required to be calculated, if either lower or upper bounds of the interval of interest are either 0 or T, respectively. That is, the single change point test statistic is calculated with respect to the interval $(\alpha, \beta]$, if $\alpha = 0$ or $\beta = 0$. The remaining intervals are calculated by equation (5.4). More precisely, consider the interval $(\alpha, \beta]$, where $\alpha \neq 0$ and $\beta \neq T$, then the single change point test statistic with respect to the index τ and interval $(\alpha, \beta]$ is calculated as follow:

$$\mathcal{U}(\tau \mid \alpha, \beta) = \mathcal{U}(\alpha \mid 0, \tau) + \mathcal{U}(\tau \mid 0, \beta) - \mathcal{U}(\alpha \mid 0, \beta),$$

or equivalently,

$$\mathcal{U}(\tau \mid \alpha, \beta) = \mathcal{U}(\beta \mid \tau, T) + \mathcal{U}(\tau \mid \alpha, T) - \mathcal{U}(\beta \mid \alpha, T).$$

Prior to the launch of AES, the following quantities are calculated and recorded:

$$\begin{aligned} &\mathcal{U}(i \mid 0, k) \quad \text{for} \quad i = d,\, d+1,\, \ldots,\, k-d \quad \text{and} \quad k = 2d,\, 2d+1,\, \ldots,\, T \quad \text{and} \\ &\mathcal{U}(i \mid j, T) \quad \text{for} \quad i = j+d,\, j+d+1,\, \ldots,\, T-d \quad \text{and} \quad j = 1,\, 2,\, \ldots,\, T-2d, \end{aligned} \tag{5.11}$$

where d is the dimension of the underlying parameter vector. The remaining values of change point test statistic are built upon the previously calculated change point test statistic in Equation (5.11) as shown in Equation (5.4). In summary AES algorithm is conducted as follow:

S1: Calculate all of the change point test statistics in Equation (5.11).

S2: Initiate the matrix $\mathcal{M}_{0\times 4}$.

S3: For $\beta = 2d+1,\, 2d+2,\, \ldots, T$ and $\alpha = \beta - 2d,\, \beta - 2d + 1, \ldots,\, 1$ do:

- Defined the search space for the test with respect to the interval $(\alpha,\, \beta]$ as follow:

$$\mathcal{S} = \{\alpha + d,\, \alpha + d + 1,\, \ldots,\, \beta - d\}.$$

- If $\alpha \geq d$, maximize the change point test statistic with respect to the interval $(\alpha,\, \beta]$ as follow:

$$\underset{t\in\mathcal{S}}{\operatorname{argmax}}\Big[\mathcal{U}(t \mid \alpha,\, \beta)\Big] = \underset{t\in\mathcal{S}}{\operatorname{argmax}}\Big[\mathcal{U}(\alpha \mid 0,\, t) + \mathcal{U}(t \mid\, 0,\, \beta) - \mathcal{U}(\alpha \mid\, 0,\, \beta)\Big].$$

  If $\alpha < d$, maximize the change point test statistic with respect to the interval $(\alpha,\, \beta]$ as follow:

$$\underset{t\in\mathcal{S}}{\operatorname{argmax}}\Big[\mathcal{U}(t \mid \alpha,\, \beta)\Big] = \underset{t\in\mathcal{S}}{\operatorname{argmax}}\Big[\mathcal{U}(\beta \mid t,\, T) + \mathcal{U}(t \mid\, \alpha,\, T) - \mathcal{U}(\beta \mid\, \alpha,\, T)\Big].$$

- If $\max\limits_{t\in\mathcal{S}}\Big[\mathcal{U}(t \mid \alpha,\, \beta)\Big] \geq \lambda$, then add the row vector

$$\left[\underset{t\in\mathcal{S}}{\operatorname{argmax}}\Big[\mathcal{U}(t \mid \alpha,\, \beta)\Big],\, \alpha,\, \beta,\, \max_{t\in\mathcal{S}}\Big[\mathcal{U}(t \mid \alpha,\, \beta)\Big]\right],$$

  into the matrix $\mathcal{M}$.

By the end of this procedure, all of the intervals in which a predefined threshold is satisfied are collected in the matrix $\mathcal{M}$. These intervals can be used in local search methods such as wild binary segmentation and narrowest over threshold.

Occasionally, the underlying single change point test statistic within the local search method are computationally expensive. For instance, consider sequence of random variables with time series dependency which are generated from a normal mean change point model. Even though, the variance is assumed to be constant through out the sequence of random variables, estimating the variance is a challenging task due to the dependency of the random variables and disruptions in mean. As long as the underlying change point test statistic is extracted from a loss function as shown in Equation (5.4), the single change point change point test statistic is not required to be calculated by traditional means in most instances. For instance, consider the interval $(\alpha, \beta]$, where $\alpha \geq d$, then change point test statistic with respect to the index τ can be obtained as follow:

$$\mathcal{U}(\tau \mid \alpha, \beta) = \mathcal{U}(\alpha \mid 0, \tau) + \mathcal{U}(\tau \mid 0, \beta) - \mathcal{U}(\alpha \mid 0, \beta).$$

Therefore, a single change point test statistic is obtained with simpler calculation and without a loss of accuracy.

5.4 Computational Complexity

5.4.1 NOT-IS.SES and UTSM

Since the IS.SES algorithm is a data adaptive interval generating mechanism, the computational complexity of this procedure depends on the data configuration. That is, the computational complexity of IS.SES increases or decreases based on the number and locations of change points, the magnitude of the change after each change point and a predefined threshold. Thus, the computational complexity of NOT-IS.SES algorithm can not be assessed by the traditional means. To address this problem, the computational complexity of NOT-IS.SES is guessed by `GuessCompx` package Agenis-Nevers et al. (2021). To estimate the computational complexity of NOT-IS.SES, multiple datasets, each with one thousand observations are generated. Each sequence of random variables is generated as follow:

- The number of observations is set to be 1000.
- The number of change points is selected from Poisson distribution, where the rate is sampled from the set $\{4, 8\}$.

	$O(\sqrt{T})$	$O(T)$	$O(T\log(T))$	$O(T^2)$	$O(T^3)$
$\alpha = 0.1$	45	292	381	273	9
$\alpha = 0.05$	38	297	346	306	13
$\alpha = 0.01$	34	258	345	348	34

Table 5.1: The estimated computational complexities of NOT-IS.SES.

	$O(T)$	$O(T\log(T))$	$O(T^2)$	$O(T^3)$
UTSM Computational Complexities	2	95	892	11

Table 5.2: The estimated computational complexity of UTSM.

- The locations of the change points are drawn randomly and uniformly from the set $\{1, 2 \ldots, T-1\}$.
- The change in mean of the random variables after each change points is selected from Gaussian distribution with mean zero and variance which is selected randomly from the set $\{1, 3, 10\}$.
- The Gaussian white noise with unit variance is added to the constructed mean function.

The NOT-IS.SES algorithm is conducted with respect to each generated sequence of random variables. As mentioned earlier, the magnitude of a predefined threshold can effect the computational complexity of the NOT-IS.SES algorithm. Thus, for each trial, the NOT-IS.SES algorithm is conducted with respect to three distinct thresholds. These thresholds are constructed by the help of a parametric bootstrap re-sampling and UTSM, where $\alpha = 0.1, 0.05,$ and 0.01. The estimated computational complexities of NOT-IS.SES are recorded in Table 5.1.

To investigate the computational complexity of UTSM, a different simulation study is performed. Since UTSM is mainly applied for defining the threshold by parametric bootstrap re-sampling, one thousands data points are randomly drawn from the Gaussian white noise process with unit variance for each replication. Then the UTSM algorithm is launched with respect to the generated datasets. Based on Table (5.2), the computational complexity of UTSM procedure seems to be quadratic. In the next section, we compare the computational complexity of UTSM to the exhaustive search (ES).

5.4.2 Computational Complexity of the UTSM versus the ES

The ultimate goal of the UTSM algorithm is to identify the maximum value of a change point test statistic with respect to all possible intervals within the sequence of random variables without an exhaustive search. Thus, An alternative approach to the UTSM is to evaluate the statistic with respect to all possible intervals within the sequence of random variables, namely the exhaustive search. Then, the largest calculated change point test statistic is assigned to be the maximized change point test statistic Λ^* (5.10). In this section, we want to demonstrate the computational efficiency of UTSM algorithm in comparison to the exhaustive search. For this purpose, the following simulation study is conducted:

- Ten thousands replications of a sequence of Gaussian white noise with unite variance are generated.
- Each replication contains 1000 data points.
- Total number of available intervals is calculated as follow:

$$T.I = \frac{999 \times 1000}{2} = 499500.$$

- UTSM is launched with respect to each replication and the number of calculated change point test statistic by UTSM is divided by $T.I$.
- The calculated proportions of conducted tests by the UTSM algorithm over total number of possible distinct change point tests are plotted in the histogram in Figure 5.1.

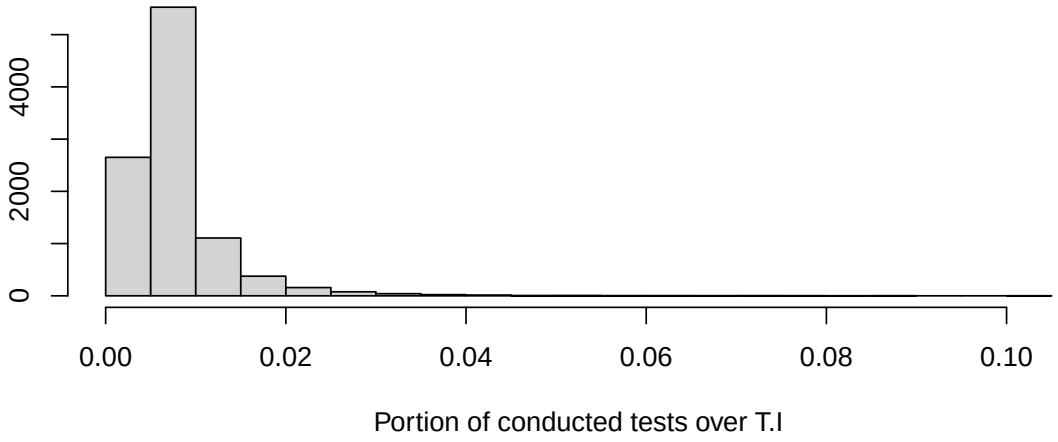

Figure 5.1: Overwhelming majority of the intervals are discarded by UTSM prior to conducting a change point test.

Based on the results in Figure 5.1, large number of available intervals are eliminated prior to conducting a change point test statistic in the UTSM algorithm. Therefore, although the UTSM finds the exact solution for the maximization problem (5.10), the computational complexity of the UTSM is reduced significantly in comparison to the exhaustive search method. Moreover, the search space of the conducted tests by UTSM algorithm is reduced significantly.

5.5 Simulation Study

In this section, we have designed multiple sets of simulation studies to access the performance of the NOT-IS.SES algorithm. We assume that the sequence of random variables are generated from the normal mean change points model. In this setting, the sequence of random variables are generated independently from normal distribution with unknown but fixed variance and either constant or piece-wise constant mean. Thus, the underlying change point test in the applied NOT-IS.SES procedure is generalized log-likelihood ratio

test Furthermore, the threshold is set to be $1.3\sqrt{(2\log(T))}$. Note that, this threshold is the recommended threshold for wild binary segmentation in normal mean change point setting (Fryzlewicz et al. (2014)).

5.5.1 Estimated Locations of Change Points by NOT-IS.SES

One of the main concerns of the multiple change point methods is estimating the locations of change points accurately. To assess the accuracy of the proposed change point method, the NOT-IS.SES algorithm is conducted with respect to the normal mean change point models in Section 3.4. Then the estimated mean change points are collected and plotted in the form of histogram in Figure 5.2. Ideally, the estimated change points are supposed to cluster around the true change points.

5.5.2 NOT-IS.SES and False Discovery of Change Points

Change point methods can suffer from false discovery of change points. To assess the susceptibility of NOT-IS.SES to false discovery of change points, the following simulation study is performed:

- Multiple sequence of random variables with $T = 100,\ 200,\ 500,\ 1000,\ 2000$ observations are generated from Gaussian white noise with unit variance.
- NOT-IS.SES is conducted with respect to each generated dataset and the number of the estimated change points are collected in Table 5.3.
- The threshold is set to $\lambda = 1.3\sqrt{2\ \log(T)}$.

As the number of observations increases, NOT-IS.SES is less likely to point out to the existence of mean change points in the white noise process.

5.5.3 NOT-IS.SES versus NOT

The narrowest over threshold (NOT) obtains change point candidates by conducting a test (usually generalized log-likelihood ratio test) with respect to randomized intervals. In this chapter, we propose replacing randomized intervals with IS.SES generated ones. Thus, comparing the accuracy of NOT-IS.SES and NOT through a simulation study seems

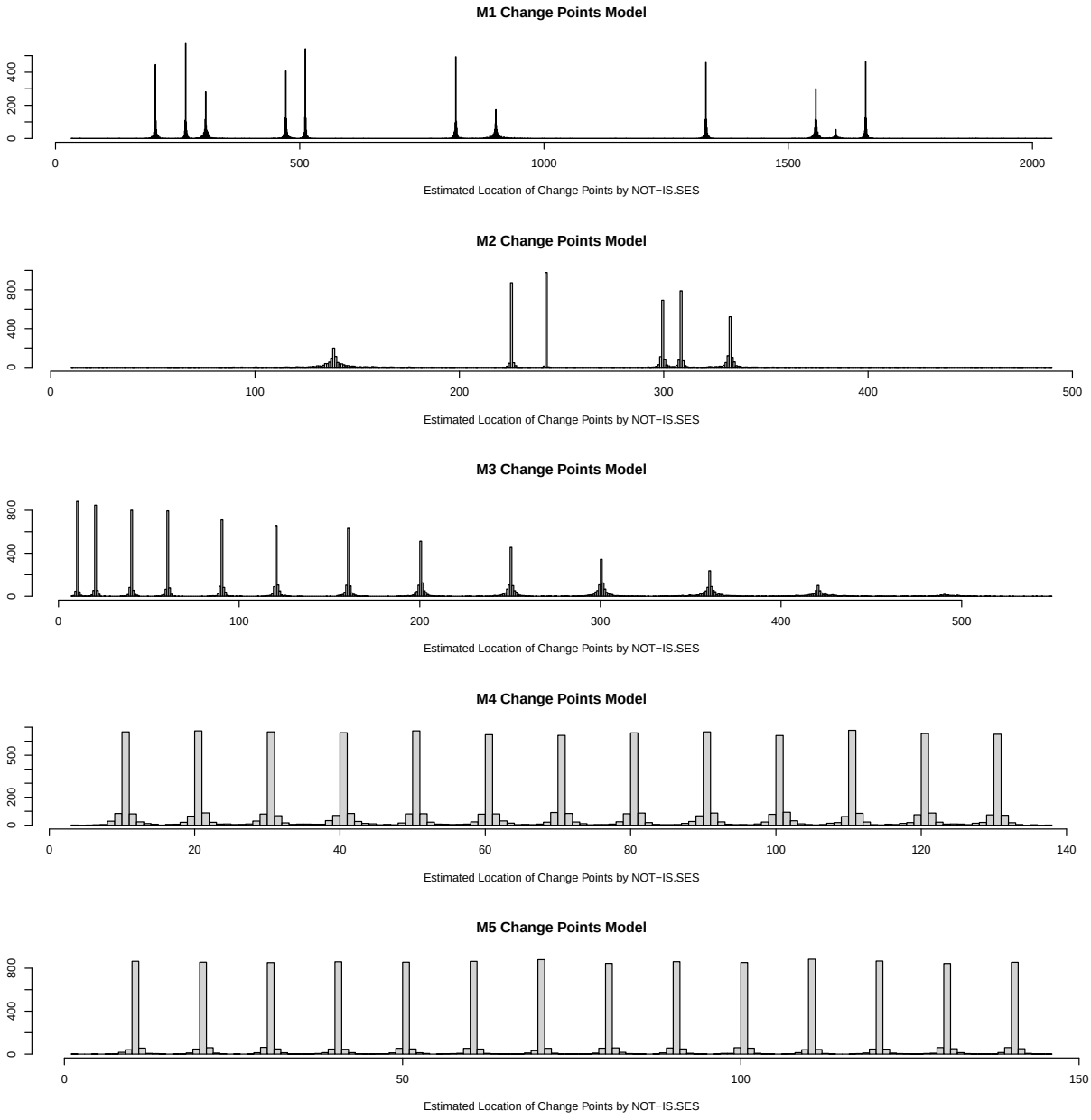

Figure 5.2: The estimated change points are concentrated about the locations of the true change points.

	T=100	T=200	T=500	T=1000	T=2000
$\widehat{N} = 0$	736	804	870	895	907
$\widehat{N} = 1$	147	107	89	77	65
$\widehat{N} = 2$	79	64	30	26	25
$\widehat{N} \geq 3$	38	25	11	2	3

Table 5.3: As the number of the observations increases, the risk of false discovery of change points reduces.

reasonable and worth the investigation. The narrowest over threshold and NOT-IS.SES are applied with respect to the generated sequence of random observations which are constructed based on the introduced mean change point models in Section 3.4. To equalize the circumstances for the narrowest over threshold and NOT-IS.SES, the threshold is set to $1.3\sqrt{2\log(T)}$ for both methods. Similar to some of the previous simulation studies, the accuracy of these methods are evaluated by the mean square error. The results of this simulation study is recorded in Table 5.4. Based on the recorded results in Table 5.4, NOT-IS.SES performs slightly better in term of the mean square error.

	MSE				
	M1	M2	M3	M4	M5
NOT	3.282	0.00663	1.827	0.0928	0.0890
NOT-IS.SES	3.014	0.00453	1.786	0.0822	0.0253

Table 5.4: IS.SES interval generating mechanism has boosted the performance of the narrowest over threshold procedure.

5.6 Data Analysis

In this section, the random variables are assumed to be independently distributed with either constant or piece-wise constant location and scale parameters. The aim is to locate the points within the sequence of independently distributed random variables after which the location and scale parameters have changed. In other words, consider the sequence of independently distributed random variables $X_1, X_2, \ldots, X_T$ with a location parameter μ_t and scale parameter σ_t^2, then the problem of interest is to find the indexes $\{\tau_1, \tau_2, \ldots, \tau_N\}$

and $\{\eta_1, \eta_2, \ldots, \eta_M\}$ such that:

$$\mu_1 = \cdots = \mu_{\tau_1} \neq \mu_{\tau_1+1} = \cdots = \mu_{\tau_2} \neq \mu_{\tau_2+1} = \cdots = \mu_{\tau_N} \neq \mu_{\tau_N+1} = \cdots = \mu_T \quad \text{and}$$
$$\sigma_1^2 = \cdots = \sigma_{\eta_1}^2 \neq \sigma_{\eta_1+1}^2 = \cdots = \sigma_{\eta_2}^2 \neq \sigma_{\eta_2+1}^2 = \cdots = \sigma_{\eta_M}^2 \neq \sigma_{\eta_M+1}^2 = \cdots = \sigma_T^2.$$

Since the underlying distribution function is not specified, the rank based approach seems to be a reasonable choice. To conduct a rank based change point test via the NOT-IS.SES procedure, the sequence of random variables $X_1, X_2, \ldots, X_T$ are ranked globally as follow:

$$r_t = \sum_{i=1}^{T} 1(X_i < X_t) + 0.5 \sum_{i=1}^{T} 1(X_i = X_t) + 0.5.$$

Through the application of NOT-IS.SES algorithm in this section, the locations of change points are estimated by a rank change point test statistic. Recall that, the rank based single change point test statistic $\mathcal{U}(t \,|\, a, b)$ is conducted with respect to the interval with data points $X_{\alpha+1}, X_{\alpha+2}, \ldots, X_\beta$ as follow:

$$\mathcal{U}(t \mid \alpha, \beta) = \frac{\left| \overline{r}_{(\alpha+1):t} - \overline{r}_{(t+1):\beta} \right|}{\sqrt{\frac{1}{t-\alpha} + \frac{1}{\beta-t}}}, \tag{5.12}$$

where, the ranking of the random variables is done with respect to entire sequence of random variables and not the interval in which the test is conducted. The rank CUSUM statistic (5.12) is mainly designed to recover the location change points (i.e, the points after which the location parameters has changed) within the sequence of random variables. Thus, the change point analysis is conducted in multiple stages here. The first stage is conducted to estimate the location change points. Next, the observations between the location change points are averaged and the residuals are obtained accordingly. Then, the calculated residuals are squared and subsequently the squared residuals are ranked globally. Finally, the rank based NOT-IS.SES is launched to recover the change in the underlying location parameter of the squared residuals. Note that, the change in the location parameter of the squared residuals is equivalent to change in the scaled parameters of the original dataset.

The threshold is set up by a parametric bootstrap re-sampling procedure. Since the rank based CUSUM statistic is distribution free, the replications of the random variables are generated from standard normal distribution. For each generated replication, the quantity Λ^* (5.10) is calculated by the UTSM algorithm. The quantity Λ^* which is associated with the b-th replication is denoted by Λ_b^*. The $(1-\alpha) \times 100$ percentile of the calculated Λ_b^*s is assigned to be the threshold aka λ. In this section, we set α to be 0.1.

5.6.1 Housing Price Index

In this section, we analysed the monthly percentage of the changes in the housing price index (HPI) of three boroughs of the city of London, UK (Hackney, Newham and Towerhamlets) from January 1996 to October 2018. The HPI provides an overall measure of completed house sale transactions on the monthly bases. For more information about the HPI dataset, we refer you to Fryzlewicz et al. (2018). To demonstrate the result, the estimated location and scale change point models are fitted with red lines in Figure 5.3. The location change point model is fitted with respect to the original data and the scale change point model is fitted with respect to the squared value of the residuals of the fitted locations parameters.

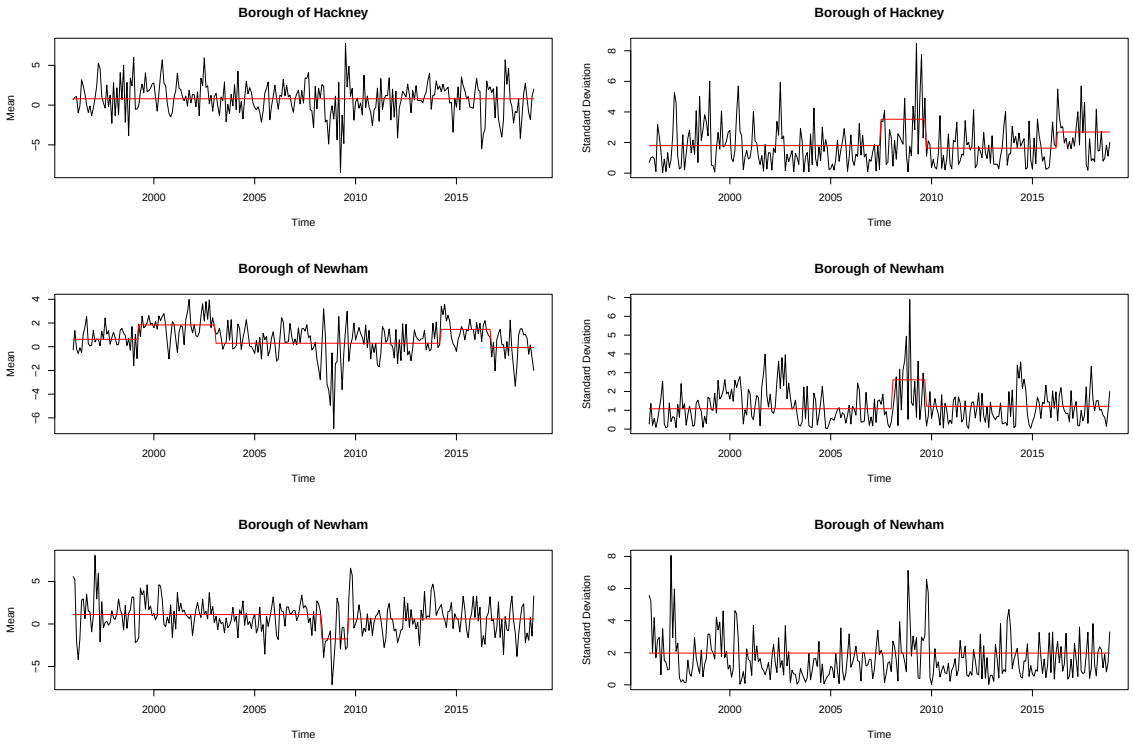

Figure 5.3: The red horizontal lines represent the fitted change point models. The rank based NOT-IS.SES seems to be less sensitive to detecting narrow segments.

One of the most recent evens that affect the world economy was 2007-2009 financial crisis. During this period, the scale parameter of the HPI increased in the Hackney and

Newham boroughs. However, no change in the location parameter has occurred in the aforementioned boroughs. On the other hand, no change in the scale parameter corresponding to the great rescission is recovered in the borough of Towerhamlets. The proposed model by NOT-IS.SES indicates to the negative changes in the housing price of this borough during the financial recession period. In comparison to the proposed models in Chapter 4 i.e., ranked based enhanced backward detection, NOT-IS.SES is proven to be more conservative. This quality of NOT-IS.SES can be attributed to the way that the threshold is set up.

5.6.2 2016 USA Presidential Election National Poll

In this section, we investigate the raw data from the US national polls during the 2016 election term. The goal of the change point analysis on this dataset is comparing the popularity of the candidates of the mainstream parties throughout various periods according to the national polls. These candidates are Hillary Clinton and Donald Trump. To assess the relative popularity of the two main presidential candidates, Trump's vote share is subtracted from that of Clinton in each poll. The resulting quantity is denoted by (H-T). Next, the the location parameter of the vote differences is estimated based on the estimated location change points by the rank based NOT-IS.SES. The positive estimated location parameter implies the overall popularity of Hilary Clinton over Donald Trump and negative estimated location parameter suggests otherwise. In this section, we only use the surveys from Ipsos polling organization. We obtained this data set from `dslabs` R package.

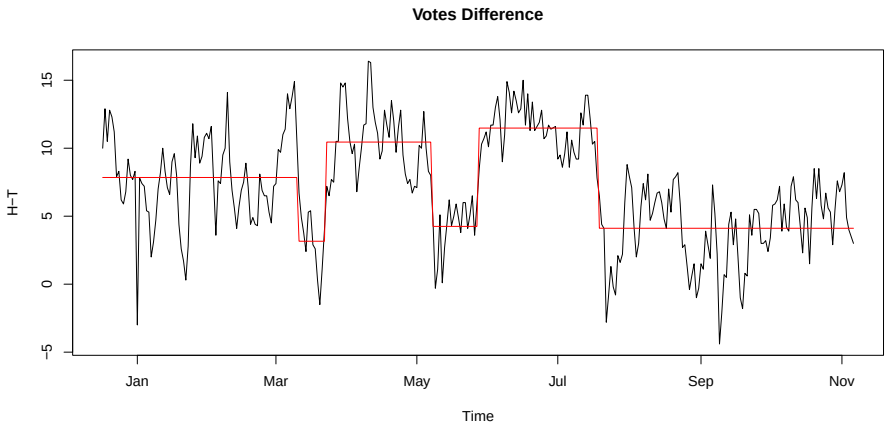

Figure 5.4: From August of 2016, the popularity of the both candidates remained the same.

As demonstrated in Figure 5.4, Hillary Clinton had consistently enjoyed a relative popularity over Trump according to the national polls. Furthermore, the estimated location parameter of H-T in the last segment is 4.11%. That is, on average, Clinton was 4.11% ahead of Trump according to the national polls in the last days of 2016 presidential campaign. This result is close to the popular votes in election day. Although the final prediction of the vote difference between Clinton and Trump is close to the final result, but NOT-IS.SES seems to miss out the last two mean change points in the months of September and October. In conclusion, the estimated location change point model by the rank based enhanced backward detection seems to be more accurate than the one that is extracted from rank based NOT-IS.SES .

5.7 Discussion

One of the main challenges of the IS.SES and UTSM algorithms is their computational complexities. One difficulty that we have faced, is recording the calculated test statistic values

$$\mathcal{U}(i \mid 0, j) \quad \text{for} \quad j = 2\, d \ldots, T \quad \text{and} \quad i = d, \ldots, j - d. \tag{5.13}$$

Since recording the aforementioned change point test statistic values requires large memory space in `R` statistical software, we decided to simply eliminate some of the intervals in which a predefined threshold is not satisfied (5.5), and reduce search space of the conducted tests (5.6). If a developer manages to find an efficient way to store test statistics value (5.13), SES and IS.SES will become much more computationally efficient. Although in this thesis, IS.SES algorithm is applied as an interval generating mechanism in narrowest over threshold, it can also be applied as a supplement to other local search methods. For instance, the IS.SES algorithm can be applied between the detected change points in order to recover the change points which were missed by the originally applied local search method.

The IS.SES interval generating mechanism can be applied in the narrowest significant pursuit (Fryzlewicz (2020*b*)) as well. The narrowest significant pursuit estimates the confidence regions within the sequence of random variables by drawing randomized intervals in two stages. Unlike the randomized intervals, IS.SES generated intervals contain no sub-intervals in which a predefined threshold is satisfied. For instance, suppose the interval $(\alpha, \tau]$ is recovered by IS.SES, then the predefined threshold will not be satisfied within any of the sub-intervals of the interval $(\alpha, \tau]$. Thus, the NOT-IS.SES algorithm can also

be applied for extracting confidence regions. That is, the intervals which are used for estimating the locations of change points can also be regarded as confidence regions.

One of the concern of the NOT-IS.SES algorithm is setting up a threshold based on the probability distribution of the underlying change point test statistic. Setting up a large threshold can result in underestimation of the number of change points. Moreover, in some cases eliminated change points may offset the remaining change points. The NOT-IS.SES algorithm is designed with the concern that the assigned threshold is overstated. Recall that, even if a threshold is satisfied within only one interval, the NOT-IS.SES is able to recover such an interval without conducting an exhaustive search. Thus, the NOT-IS.SES algorithm is ideal, when the goal is to recover the mean change points that are associated with a significant change.

Chapter 6

Hybrid Enhanced Backward Detection

In Chapter 3, enhanced backward detection via Bayesian information criteria (EBD) was introduced. In the EBD, the number of change points is estimated by Bayesian information criteria (BIC) (Schwarz et al. (1978)). In this chapter, we propose a new variation of EBD which estimates the number of change points by setting up a threshold based on the probability distribution of the underlying change point test statistic. We refer to this method as *threshold based enhanced backward detection* (λ-EBD). Similar to EBD, in λ-EBD, the locations of change points are estimated individually by conducting a single change point test within intervals of random variables. The λ-EBD method will be elaborated more precisely in Section 6.1.

Recall that, the local search methods estimate the locations of change points by conducting a test within the intervals of the random variables. Ideally, these intervals contain exactly one change point. However, some of the intervals which are selected by λ-EBD for estimating change points may contain multiple change points. Generally, to obtain change point estimates from desirable intervals, some of the local search methods such as narrowest over threshold (Baranowski et al. (2019)) attempt to estimate the locations of change points based on relatively short intervals. In this approach, all of the intervals in which a predefined threshold is satisfied are collected and change points estimates are extracted from the shortest intervals. As a result of prioritizing the narrowest intervals, the locations of change points are much more likely to be estimated based on the intervals with exactly one change point. Therefore, to obtain change point estimates form narrower intervals, IS.SES algorithm is launched with respect to the intervals from which change point estimates were extracted during the execuation of λ-EBD. Then the narrowest generated

sub-intervals by IS.SES replaces the intervals in which the IS.SES procedure performed. The shortening of the intervals which are used for locating the change point estimates will be discussed in Section 6.1

To recover change point estimates which λ-EBD may have failed to detect, IS.SES algorithm is launched with respect to the segments between the change point estimates from $\lambda - EBD$. The application of IS.SES algorithm alongside with λ-EBD will be discussed more precisely in Section 6.1. Since the resulting method employs λ-EBD and IS.SES in combination with each other, we refer to this approach as *hybrid enhanced backward detection* (HEBD). In Section (6.2), HEBD is performed with respect to a real world dataset. This dataset is the collection of all Ipsos national polls with regard to 2016 USA presidential election polls.

6.1 Hybrid Enhanced Backward Detection

Similar to EBD, λ-EBD is a member of bottom up detection methods. Recall that the bottom-up process eliminates the change point candidates sequentially. Therefore, in enhanced backward detection, the change point candidate with the smallest value of the change point test statistic is eliminated at each stage. However, in λ-EBD, the change point candidates are scored slightly differently. To track down which change point candidate to be eliminated next, the matrix $\mathcal{R}$ is constructed in the following manner. Suppose the current set of change point candidates is $\{\tau_1, \tau_2, \ldots, \tau_N\}$, then the matrix $\mathcal{R}$ is $(N+2)\times 4$ dimensional and its entries in each row are as follow:

$$
\begin{aligned}
&[0, -1, 1, \infty], \quad \text{for the first row} \\
&[T, T-1, T+1, \infty], \quad \text{for the last row, and} \\
&[\tau_j, \tau_{j-1}, \tau_{j+1}, \mathcal{U}(\tau_j \mid \alpha_{j-1}, \beta_{j-1})] \quad \text{for (j+1)-th row,}
\end{aligned}
$$

where α_j and β_j are the lower and upper bounds of the interval from which change point candidate τ_j is extracted. Note that, the interval $(\alpha_j, \beta_j]$ is a sub-interval of $(\tau_{j-1}, \tau_{j+1}]$ and index τ_j is the point which maximizes the change point test statistic $\mathcal{U}$ within the interval $(\alpha_j, \beta_j]$. Throughout this chapter, we refer to the elements in the fourth column of the matrix $\mathcal{R}$ as *scores*. For instance, the score of the change point candidate τ_j is $\mathcal{U}(\tau_j \mid \alpha_j, \beta_j)$. Similar to the methods in the previous chapters, the underlying single change point test statistic is derived from some sort of a loss function as demonstrated in equation (5.2).

At each stage of λ-EBD, the change point candidate with the smallest score is eliminated and then a variation of the sweeping process, which is modified to suit λ-EBD is launched.

We call this variation of the sweeping process *threshold based sweeping process* (TSP). To describe λ-EBD and TSP algorithms, the following example is constructed. Suppose the change point candidate τ_{j} is associated with the smallest score in the matrix $\mathcal{R}$. Therefore, the change point candidate τ_{j} and subsequently the $(j+1)$ row from the matrix $\mathcal{R}$ are eliminated. After eliminations of the change point candidate τ_{j}, TSP is launched. For this purpose, the first and last change point candidates which are required to be inspected by TSP are determined. We call the first and last change point candidates which are set to be examined, *count* (c) and *maximum count* (m.c), respectively. In this scenario, since the change point candidate τ_{j} is eliminated, the count and maximum count indexes are determined as follow:

$$c = \min(c-1, 1) \quad \text{and} \quad m.c = \min(c+2,\, N-1).$$

TSP algorithm with regard to the change point candidate τ_{c} is performed as follow:

- Examination of the change point candidate τ_{c} depends on its score and its corresponding interval in the matrix $\mathcal{R}$. These scenarios are as follow:

Case 1: If $\mathcal{R}[\,c+1, 4] \geq \lambda$ and $\mathcal{R}[c+1, 2] < \tau_{c-1}$ or $\mathcal{R}[c+1, 3] > \tau_{c+1}$, then

- The single change point test statistic is maximized with respect to the interval $(\tau_{c-1},\, \tau_{c+1}]$ as follow:

$$\tau = \underset{\tau_{c-1}<j<\tau_{c+1}}{\operatorname{argmax}}\ \mathcal{U}(j \mid \tau_{c-1},\, \tau_{c+1}).$$

- The entries in the $(c+1)$ row of the matrix $\mathcal{R}$ are updated as follow:

$$[\,\tau,\, \tau_{c-1},\, \tau_{c+1},\, \mathcal{U}(\tau \mid \tau_{c-1},\, \tau_{c+1})].$$

Case 2: If $\mathcal{R}[\,c+1,\, 4] < \lambda$, and $\mathcal{R}[c+1,\, 2] \neq \tau_{c-1}$ or $\mathcal{R}[c+1,\, 3] \neq \tau_{c+1}$ then

- The single change point test statistic is maximized with respect to the interval $(\tau_{c-1},\, \tau_{c+1}]$ as follow:

$$\tau = \underset{\tau_{c-1}<j<\tau_{c+1}}{\operatorname{argmax}}\ \mathcal{U}(j \mid \tau_{c-1},\, \tau_{c+1}).$$

- The entries in the $(c+1)$ row of the matrix $\mathcal{R}$ are updated as follow:

$$[\,\tau,\, \tau_{c-1},\, \tau_{c+1},\, \mathcal{U}(\tau \mid \tau_{c-1},\, \tau_{c+1})].$$

Case 3: If neither Case 1 or Case 2 occur, then TSP ignores the $(c+1)$-th row.

- If $\tau_c \neq \mathcal{R}[c,\, 1]$, then

$$c = \max(1, c-1) \quad \text{and} \quad \text{and} \quad m.c = \min(N-1,\, c+1).$$

  Otherwise, the count index is incremented with one i.e., $c = c + 1$.

- TSP is stopped as soon as $c > m.c$.

The elimination of change point candidates through λ-EBD procedure continues until all of the recorded scores in the matrix $\mathcal{R}$ exceeds a predefined threshold.

Suppose the change point estimates $\tau_1, \tau_2, \ldots, \tau_N$ are extracted by λ-EBD from the intervals $(\alpha_1, \beta_1]$, $(\alpha_2, \beta_2]$, $\ldots$, $(\alpha_N, \beta_N]$, respectively. Note that, the recorded intervals in the matrix $\mathcal{R}$ at the end of the run of λ-EBD contain exactly one change point estimate. In other words, the interval such as $(\alpha_j, \beta_j]$m which contains the change point estimate τ_j, is a sub-interval of $(\tau_{j-1}, \tau_{j+1}]$. However, the recorded intervals in the matrix $\mathcal{R}$ may contain more than one true change points. Thus, narrowing down these intervals is beneficial, since the narrower intervals are less likely to contain multiple change points. For this purpose, the IS.SES procedure is performed with respect to the intervals which are used by λ-EBD for estimating the locations of change points. For example, consider the change point estimate τ_j which is obtained by λ-EBD from the interval $(\alpha_j, \beta_j\,]$, then the IS.SES procedure is launched with respect to the aforementioned interval. The narrowest generated interval by IS.SES replaces the interval $(\alpha_j, \beta_j]$ and the point which maximizes the change point test statistic within the updated interval substitutes the change point estimates τ_j. Not only the IS.SES algorithm can narrow down the intervals in the matrix $\mathcal{R}$, but also, it can recover the change points which λ-EBD has failed to detect. Note that, the estimated change points must not *conflict* each other. Recall that, the change point candidates τ_1 and τ_2 which are extracted from the intervals $(\alpha_1, \beta_1]$ and $(\alpha_2, \beta_2]$, respectively, conflict with each others, when $\tau_1 \in (\alpha_2, \beta_2)$ or $\tau_2 \in (\alpha_1, \beta_1)$.

Suppose the indexes $\tau_1, \tau_2, \ldots, \tau_N$ are change point estimates within the sequence of random variables $X_1, X_2, \ldots, X_T$. Moreover, the change points estimates $\tau_1, \tau_2, \ldots, \tau_N$ and their corresponding intervals are recorded in the matrix $\mathcal{R}$. IS.SES generated intervals are added to the matrix $\mathcal{R}$ as follow:

Step 0: The matrix $\mathcal{R}^*$ is introduced and set equal to $\mathcal{R}$.

Step 1: IS.SES procedure is performed with respect to every recorded intervals in the matrix $\mathcal{R}$.

Step 2: IS.SES generated intervals along side with their corresponding change point candidates and the value of maximized change point test statistic are added as a row vector to the matrix $\mathcal{R}^*$. For instance, suppose the interval $(\alpha,\ \beta]$ is an IS.SES generated interval, where the underlying change point test statistic is maximized by the index τ, then the row vector

$$[\tau,\ \alpha,\ \beta,\ \mathcal{U}(\tau \mid\ \alpha,\ \beta)]$$

is recorded in the matrix $\mathcal{R}^*$.

Step 3: The row vectors in the matrix $\mathcal{R}^*$ are arranged with respect to the length of their corresponding intervals in the increasing order. That is, the row which contains an interval with the shorter length are located on the top rows of the matrix $\mathcal{R}^*$.

Step 4: The change point candidate in the matrix $\mathcal{R}^*$ which is extracted from the narrowest interval is assigned to be a change point estimate.

Step 5: The rows in the matrix $\mathcal{R}^*$ with a change point candidate which conflicts with estimated change points are eliminated.

Step 6: Steps 4 and 5 are repeated, until all of the change point candidates in the matrix $\mathcal{R}^*$ are either assigned to be change point estimates or eliminated due to their conflict with the estimated change points.

Step 7: Set the matrix $\mathcal{R}$ equal to the matrix $\mathcal{R}^*$.

As mentioned earlier, the change points which were not detected by λ-EBD can be identified by the help of the IS.SES procedure. For instance, suppose the indexes η and τ are assigned to be change points based on the previous analysis, then the remaining change points within the interval $(\eta,\ \tau]$ are obtained by IS.SES procedure as follow:

1: The IS.SES procedure is performed within the interval $(\eta,\ \tau]$.

2: The IS.SES generated intervals which conflict with change point estimates η and τ are eliminated.

3: The remaining IS.SES generated intervals are sorted in an increasing manner with respect to their length.

4: The change point estimate is extracted from the narrowest remaining IS.SES generated intervals.

5: All of the IS.SES generated change point candidates which conflict with the recently recovered change point estimate are eliminated.

6: Steps 4 and 5 are repeated until all of the IS.SES generated change point candidates are either assigned to be change point estimates or eliminated.

Similar to NOT-IS.SES, the intervals which are used for estimating the location of change points do not contain any sub-intervals in which a predefined threshold is satisfied other than themselves.

In this section, HEBD multiple change point detection method was introduced. In Section 6.2, the applicability of HEBD procedure is demonstrated with respect to a real world dataset.

6.2 Data Analysis

In this section, the change point analysis via HEBD is conducted with respect to the 2016 presidential election national polls from Ipsos. At first, vote difference between Hillary Clinton and Donald Trump in each poll is calculated. That is, Trump's share of the votes in each poll is subtracted from Clinton's. We refer to this dataset as H-T. Then the mean change point model with regard to H-T dataset is estimated. The segment of H-T with positive means implies that Hillary Clinton had captured the larger share of votes during that particular period. On the other hand, Negative estimated mean implies Donald Trump was ahead of Hillary Clinton according to the national election polls.

In this chapter, we assume the random variables in H-T dataset are independently and normally distributed with a unit variance. Therefore, the variance of the random variables is stabilized by box-cox transformation. Next, the transformed dataset is centered and scaled. To search the location of change points individually, the underlying change point test statistic is derived from the generalized log-likelihood ratio test under the assumption of normality and independency of the random variables. That is, the underlying single change point test statistic in HEBD is calculated with respect to the sub-sequence of random variables $X_{\alpha+1}, X_{\alpha+2}, \ldots, X_{\beta}$ as follow:

$$\tau = \underset{\alpha<t<\beta}{\operatorname{argmax}} \left[\frac{\left| \overline{X}_{(\alpha+1):t} - \overline{X}_{(t+1):\beta} \right|}{\sqrt{\frac{1}{t-\alpha} + \frac{1}{\beta-t}}} \right].$$

Similar to Chapter 5, the threshold is determined with the help of a parametric bootstrap re-sampling and the UTSM procedure.

According to the QQ-plot in Figure 6.1, the residuals which are obtained from the proposed mean change point model by HEBD is normally distributed. Furthermore, according to the time series plot in Figure 6.1, estimated residuals is weakly stationary. HEBD estimated mean change point model implies that Hillary Clinton was almost consistently ahead of Donald Trump during 2016 USA presidential election period, according to the Ipsos national polls. Note that, at the last segment, Hillary Clinton was ahead of Donald Trump by 5% . However according to the election results, Clinton had captured 3% more votes than Trump. In other words, the national polls may have overestimated Hillary Clinton's vote.

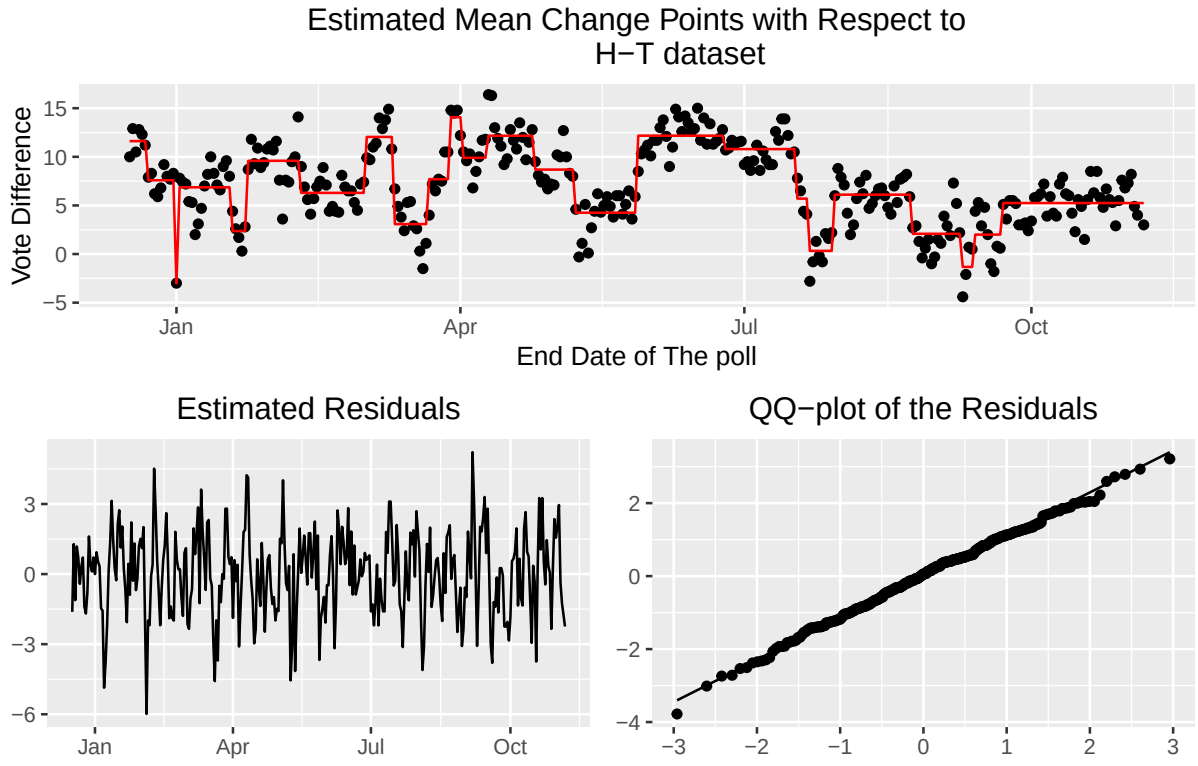

Figure 6.1: Time series and QQ plots indicates that our proposed change point model is accurate.

Chapter 7

Conclusion and Future Works

In this thesis, we introduce enhanced backward detection (EBD) and narrowest over threshold with interval selection via shorten exhaustive search (NOT-IS.SES). The fundamental requirement for these methods is that the single change point statistic must be extracted from a loss function. The generalized log-likelihood ratio test is one example of such a single change point test statistic. Enhanced backward detection is a member of the bottom of change point detection techniques, in which an estimated change point set is obtained by sequential elimination of change point candidates. On the other hand, the narrowest over threshold with interval selection via shorten exhaustive search is a member of top-down methods. The fundamental ingredient of NOT-IS.SES is the shorten exhaustive search (SES) algorithm, which is responsible for recovering an interval when a predefined threshold is satisfied. One significance of SES is that if SES recovers no interval, then we can conclude that there is no interval within the sequence of random variables in which a predefined threshold is satisfied.

Further, we introduce a modified version of the SES algorithm designed to calculate the largest change point test statistic within the sequence of random variables. That is, the single change point test statistic is maximized with respect to the possible location of a change point and the interval in which the test is calculated. We also propose a rank-based variation of our proposed methods. Using a rank-based single change point test statistic extracted from a loss function, EBD and NOT-IS.SES can also be applied to nonparametric problems.

EBD and NOT-IS.SES can be applied to a wide variety of problems, as long as the single change point test statistic is extracted from a loss functions as demonstrated in 3 and 5. However, in this thesis, we have only investigated the performance of these methods

concerning univariate datasets. Thus, applying EBD and NOT-IS.SES to multivariate and high-dimensional datasets is worth investigating. Moreover, the application of EBD and NOT-IS.SES concerning the dataset with time-series dependency can be an exciting topic for further research.

References

Agenis-Nevers, M., Bokde, N. D., Yaseen, Z. M. and Shende, M. K. (2021), 'An empirical estimation for time and memory algorithm complexities: newly developed r package', *Multimedia Tools and Applications* **80**(2), 2997–3015.

Anastasiou, A. and Fryzlewicz, P. (2019), 'Detecting multiple generalized change-points by isolating single ones', *arXiv preprint arXiv:1901.10852* .

Aue, A. and Horváth, L. (2013), 'Structural breaks in time series', *Journal of Time Series Analysis* **34**(1), 1–16.

Auger, I. E. and Lawrence, C. E. (1989), 'Algorithms for the optimal identification of segment neighborhoods', *Bulletin of mathematical biology* **51**(1), 39–54.

Bai, J. (1994), 'Least squares estimation of a shift in linear processes', *Journal of Time Series Analysis* **15**(5), 453–472.

Baranowski, R., Chen, Y. and Fryzlewicz, P. (2019), '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.

Barassi, M., Horváth, L. and Zhao, Y. (2020), 'Change-point detection in the conditional correlation structure of multivariate volatility models', *Journal of Business & Economic Statistics* **38**(2), 340–349.

Bleakley, K. and Vert, J.-P. (2011), 'The group fused lasso for multiple change-point detection', *arXiv preprint arXiv:1106.4199* .

Brodsky, B. and Darkhovsky, B. (1993), Applications of nonparametric change-point detection methods, *in* 'Nonparametric Methods in Change-Point Problems', Springer, pp. 169–182.

Brown, R. L., Durbin, J. and Evans, J. M. (1975), 'Techniques for testing the constancy of regression relationships over time', *Journal of the Royal Statistical Society: Series B (Methodological)* **37**(2), 149–163.

Chen, J. and Gupta, A. (2004), 'Statistical inference of covariance change points in gaussian model', *Statistics* **38**(1), 17–28.

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

Chen, J. and Gupta, A. K. (2011), *Parametric statistical change point analysis: with applications to genetics, medicine, and finance*, Springer Science & Business Media.

Chen, J. and Wang, Y.-P. (2008), 'A statistical change point model approach for the detection of dna copy number variations in array cgh data', *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **6**(4), 529–541.

Chenouri, S., Mozaffari, A. and Rice, G. (2020), 'Robust multivariate change point analysis based on data depth', *Canadian Journal of Statistics* **48**(3), 417–446.

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

Cho, H. and Fryzlewicz, P. (2020), 'Multiple change point detection under serial dependence: Wild energy maximisation and gappy schwarz criterion', *arXiv preprint arXiv:2011.13884* .

Cho, H. and Kirch, C. (2019), 'Localised pruning for data segmentation based on multiscale change point procedures', *arXiv preprint arXiv:1910.12486* .

Cobb, G. W. (1978), 'The problem of the nile: Conditional solution to a changepoint problem', *Biometrika* **65**(2), 243–251.

Darling, D. A., Erdös, P. et al. (1956), 'A limit theorem for the maximum of normalized sums of independent random variables', *Duke Math. J* **23**(1), 143–155.

Davis, R. A., Huang, D. and Yao, Y.-C. (1995), 'Testing for a change in the parameter values and order of an autoregressive model', *The Annals of Statistics* pp. 282–304.

Eichinger, B., Kirch, C. et al. (2018), 'A mosum procedure for the estimation of multiple random change points', *Bernoulli* **24**(1), 526–564.

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

Fryzlewicz, P. (2020*a*), 'Detecting possibly frequent change-points: Wild binary segmentation 2 and steepest-drop model selection', *Journal of the Korean Statistical Society* **49**(4), 1027–1070.

Fryzlewicz, P. (2020*b*), 'Narrowest significance pursuit: inference for multiple change-points in linear models', *arXiv preprint arXiv:2009.05431* .

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

Fryzlewicz, P. et al. (2018), 'Tail-greedy bottom-up data decompositions and fast multiple change-point detection', *Annals of Statistics* **46**(6B), 3390–3421.

Hampel, F. R. (1974), 'The influence curve and its role in robust estimation', *Journal of the american statistical association* **69**(346), 383–393.

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

Hawkins, D. M. (1977), 'Testing a sequence of observations for a shift in location', *Journal of the American Statistical Association* **72**(357), 180–186.

Haynes, K., Eckley, I. A. and Fearnhead, P. (2017), 'Computationally efficient changepoint detection for a range of penalties', *Journal of Computational and Graphical Statistics* **26**(1), 134–143.

Hinkley, D. V. (1970), 'Inference about the change-point in a sequence of random variables'.

Horváth, L. (1993), 'The maximum likelihood method for testing changes in the parameters of normal observations', *The Annals of statistics* pp. 671–680.

Horváth, L., Miller, C. and Rice, G. (2020), 'A new class of change point test statistics of rényi type', *Journal of Business & Economic Statistics* **38**(3), 570–579.

Horváth, L. and Rice, G. (2014), 'Extensions of some classical methods in change point analysis', *Test* **23**(2), 219–255.

Huang, T., Wu, B., Lizardi, P. and Zhao, H. (2005), 'Detection of dna copy number alterations using penalized least squares regression', *Bioinformatics* **21**(20), 3811–3817.

Hušková, M. and Slabỳ, A. (2001), 'Permutation tests for multiple changes', *Kybernetika* **37**(5), 605–622.

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

Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L. and Tsai, T. T. (2005), 'An algorithm for optimal partitioning of data on an interval', *IEEE Signal Processing Letters* **12**(2), 105–108.

Jen, T. and Gupta, A. K. (1987), 'On testing homogeneity of variances for gaussian models', *Journal of Statistical Computation and Simulation* **27**(2), 155–173.

Jiang, F., Zhao, Z. and Shao, X. (2020), 'Time series analysis of covid-19 infection curve: A change-point perspective', *Journal of econometrics* .

Jun Shin, S., Wu, Y. and Hao, N. (2020), 'A backward procedure for change-point detection with applications to copy number variation detection', *Canadian Journal of Statistics* **48**(3), 366–385.

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

Kim, H.-J. and Siegmund, D. (1989), 'The likelihood ratio test for a change-point in simple linear regression', *Biometrika* **76**(3), 409–423.

Kovács, S., Li, H., Bühlmann, P. and Munk, A. (2020), 'Seeded binary segmentation: A general methodology for fast and optimal change point detection', *arXiv preprint arXiv:2002.06633* .

Kruskal, W. H. and Wallis, W. A. (1952), 'Use of ranks in one-criterion variance analysis', *Journal of the American statistical Association* **47**(260), 583–621.

Lee, S., Ha, J., Na, O. and Na, S. (2003), 'The cusum test for parameter change in time series models', *Scandinavian Journal of Statistics* **30**(4), 781–796.

Lepage, Y. (1971), 'A combination of wilcoxon's and ansari-bradley's statistics', *Biometrika* **58**(1), 213–217.

Mann, H. B. and Whitney, D. R. (1947), 'On a test of whether one of two random variables is stochastically larger than the other', *The annals of mathematical statistics* pp. 50–60.

Messer, M., Kirchner, M., Schiemann, J., Roeper, J., Neininger, R., Schneider, G. et al. (2014), 'A multiple filter test for the detection of rate changes in renewal processes with varying variance', *The Annals of Applied Statistics* **8**(4), 2027–2067.

Mood, A. M. (1954), 'On the asymptotic efficiency of certain nonparametric two-sample tests', *The Annals of Mathematical Statistics* pp. 514–522.

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

Padilla, O. H. M., Yu, Y., Wang, D. and Rinaldo, A. (2019), 'Optimal nonparametric change point detection and localization', *arXiv preprint arXiv:1905.10019* .

Page, E. (1954), 'Continuous inspection schemes', *Biometrika* **41**(1/2), 100–115.

Page, E. (1955), 'A test for a change in a parameter occurring at an unknown point', *Biometrika* **42**(3/4), 523–527.

Rigaill, G. (2010), 'Pruned dynamic programming for optimal multiple change-point detection', *arXiv preprint arXiv:1004.0887* **17**.

Ross, G. J. (2021), 'Nonparametric detection of multiple location-scale change points via wild binary segmentation', *arXiv preprint arXiv:2107.01742* .

Schwarz, G. et al. (1978), 'Estimating the dimension of a model', *Annals of statistics* **6**(2), 461–464.

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

Shao, X. and Zhang, X. (2010), 'Testing for change points in time series', *Journal of the American Statistical Association* **105**(491), 1228–1240.

Silverman, B. W. (1985), 'Some aspects of the spline smoothing approach to non-parametric regression curve fitting', *Journal of the Royal Statistical Society: Series B (Methodological)* **47**(1), 1–21.

Srivastava, M. and Worsley, K. J. (1986), 'Likelihood ratio tests for a change in the multivariate normal mean', *Journal of the American Statistical Association* **81**(393), 199–204.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

Venkatraman, E. and Olshen, A. B. (2007), 'A faster circular binary segmentation algorithm for the analysis of array cgh data', *Bioinformatics* **23**(6), 657–663.

Wang, Y., Wang, Z. and Zi, X. (2020), 'Rank-based multiple change-point detection', *Communications in Statistics-Theory and Methods* **49**(14), 3438–3454.

Worsley, K. (1979), 'On the likelihood ratio test for a shift in location of normal populations', *Journal of the American Statistical Association* **74**(366a), 365–367.

Yao, Y.-C. (1988), 'Estimating the number of change-points via schwarz'criterion', *Statistics & Probability Letters* **6**(3), 181–189.

Yao, Y.-C. and Davis, R. A. (1986), 'The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates', *Sankhyā: The Indian Journal of Statistics, Series A* pp. 339–353.

Zhang, N. R. and Siegmund, D. O. (2007), 'A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data', *Biometrics* **63**(1), 22–32.

Zhou, J. and Liu, S. Y. (2009), 'Inference for mean change-point in infinite variance ar (p) process', *Statistics & Probability Letters* **79**(1), 6–15.