SEQUENTIAL CHANGE-POINT DETECTION IN LINEAR REGRESSION AND LINEAR QUANTILE REGRESSION MODELS UNDER HIGH DIMENSIONALITY

Suthakaran Ratnasingam

A Dissertation

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2020

Committee:

Wei Ning, Advisor

Andy Garcia,
Graduate Faculty Representative

Hanfeng Chen

Junfeng Shang

ABSTRACT

Wei Ning, Advisor

Sequential change point analysis aims to detect structural change as quickly as possible when the process state changes. A good sequential change point detection procedure is expected to minimize the detection delay time and the risk of raising false alarm. Existing sequential change point detection methods cannot be applicable for high-dimensional data because they are univariate in nature and thus present challenges.

In the first part of the dissertation, we develop a monitoring method to detect structural change in smoothly clipped absolute deviation (SCAD) penalized regression model for high-dimensional data after the historical sample with the sample size m. The unknown pre-change regression coefficients are replaced by the SCAD penalized estimator. The asymptotic properties of the proposed test statistics are derived. We conduct a simulation study to evaluate the performance of the propose method. The proposed method is applied to the gene expression in the mammalian eye data to detect changes sequentially.

In the second part of the dissertation, we develop a sequential change point detection method to monitor structural changes in SACD penalized quantile regression (SPQR) model for high-dimensional data. We derive the asymptotic distributions of the test statistic under the null and alternative hypotheses. Furthermore, to improve the performance of the SPQR method, we propose the Post-SCAD penalized quantile regression estimator (P-SPQR) for high-dimensional data. Simulations are conducted under different scenarios to study the finite sample properties of the SPQR and P-SPQR methods. A real data application is provided to demonstrate the effectiveness of the method.

In the third and fourth part of the dissertation, we investigate the change point problem for Skew-Normal distribution and three parameter Weibull distribution respectively. Besides detecting and obtaining the point estimate of a change location, we propose an estimation procedure based on the confidence distribution (CD) along with the modified information criterion (MIC) to

construct the confidence set for the change location. Simulations are conducted to evaluate the performance of the proposed method in terms of powers, coverage probabilities and average lengths of confidence sets. Real data applications are provided in each part to illustrate the performance of the proposed methods.

I would like to dedicate my dissertation to my beloved parents Theivanayagam Ratnasingam, Chandraleela Ratnasingam, brother Subakaran Ratnasingam, my wife Bhairavi Suthakaran and to our son Ygeshan Suthakaran for their unconditional love and support.

ACKNOWLEDGMENTS

I profusely thank my advisor Professor Wei Ning, for his constant support, guidance and inspiration. This dissertation would have not been completed without his persistent encouragement and constructive comments on its various stages. I could not have imagined having a better advisor and mentor for my PhD study.

I am extremely grateful to my dissertation committee members: Professors Junfeng Shang, Hanfeng Chen and Andy Garcia. I am very much appreciate for the feedback and comments to help improve this dissertation.

I would like to express my gratitude to Professor Craig Zirbel, for encouragement and advise he has provided throughout my graduate school years at BGSU. I would also like to thank the faculty and staff at the department of mathematics and statistics for their academic support and general help.

This section would be incomplete without acknowledging my colleagues and friends. A big thank you to my classmates who consistently offered me continuous support and encouragement through all of the ups and downs of graduate school. I am particularly grateful to Patrick Stewart for all the support that he has extended to me during my time at Bowling Green, and I hope we can work together in the future.

I wish to thank my lovely parents and brother for their lifelong support throughout my life, and I believe without their support, it is impossible for me to finish my college and graduate studies in my life.

Last not least, thank you to my beloved wife Bhairavi and our son Yugeshan for being so helpful and supportive throughout my graduate studies.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Table Page

CHAPTER 1 LITERATURE REVIEW

1.1 Introduction

The change point analysis plays an important role in finance, time series analysis, economy, quality control, genome research, signal processing, medical research and statistical calibration. The detection of change points attempts to identify points in time when the probability distribution of a stochastic process or time series changes. The study of the change-point problem was originally stated by Page (1954, 1955) who first proposed a procedure to detect only one change in a parameter. Offline change-point detection and sequential (online) change-point detection are two common types of change-point detection methods. We generally use log-likelihood or modified information criterion in the offline change point analysis while online detection method uses generalized likelihood ratio, cumulative sum method (CUSUM) or Shiryayev-Roberts (SR) procedure.

1.2 Change Point Analysis

A change point detection method has been widely used in various areas such as finance, climate monitoring, quality control, examination of gene expressions, etc. The change point problem is generally referred to as identifying the number changes and estimating respective locations. The change point problem can be formulated in several models including parametric, nonparametric, regression, time series, and Bayesian, see for example, Chernoff and Zacks (1964); Gupta and Chen (1996); Sen and Srivastava (1975a); Cosörgö and Horváth (1997); Chen and Gupta (1997).

Let $x_1, x_2, \ldots, x_n$ be a sequence of n independent random variables with probability distribution functions $F_1, F_2, \ldots, F_n$ respectively. The change point problem test the following hypothesis,

$$H_0 : \ F_1 = F_2 = \cdots = F_n \tag{1.2.1}$$

versus

$$H_1 : \ F_1 = \cdots = F_{k_1} \neq F_{k_1+1} = \cdots = F_{k_2} \neq F_{k_2+1} = \cdots = F_{k_q} \neq F_{k_q+1} = \cdots = F_n, \tag{1.2.2}$$

where $1 < k_1 < k_2 < \cdots < k_q < n$ and q is the number unknown change points and $k_1, k_2, \ldots, k_q$ are the corresponding change locations. In the case of multiple change points, Vostrikova (1981) introduced binary segmentation procedure which allows to identify the number of change points and estimate the locations simultaneously.

Sequential change-point detection has been extensively studied for univariate data. In sequential change-point detection method, the observations are coming one by one and an alarm must be raised as quickly as possible following any change only based on previous information, see for example, Page (1954); Shiryaev (1963); Lorden (1971). Suppose that $x_1, \ldots, x_k$ are independent and identically distributed to a pre-change distribution f_0 and $x_{k+1}, \ldots, x_n$ are independent and identically distributed to a post-change distribution f_1 where f_0 and f_1 are known. Let k be an unknown change point location and there exists a positive constant c, the stopping time of the Page CUSUM,Page (1954) procedure is defined as,

$$T_n = \inf\left\{n \geq 1 : \max_{1\leq k\leq n} \sum_{i=k}^{n} \log \frac{f_1(x_i)}{f_0(x_i)} \geq c\right\}. \tag{1.2.3}$$

The CUSUM method minimizes the detection delay. In the Page CUSUM method, however, Lai and Xing (2010) studied when pre- and post-change density functions were not known in advance and assumed they came from the same canonical exponential family. In this case, the log-likelihood ratio in Page CUSUM method (1.2.3) was replaced by the generalized likelihood ratio (GLR). Mei (2006) proposed a change point detection procedure for unknown pre-change parameter. Further, he investigated when both the pre-change distribution f_{θ_0} and post-change distribution f_{θ_1} involved with unknown parameters. The test statistics for the Mei (2006) procedure is given as follows.

$$T^*(a) = \inf\left\{n \geq 1 : \max_{1\leq k\leq n} \inf_{\theta\in\Theta} \sup_{\lambda\in\Lambda} \left\{\frac{1}{p(\theta)} \sum_{i=k}^{n} \log \frac{g_\lambda(x_i)}{f_\theta(x_i)}\right\} \geq a\right\}, \tag{1.2.4}$$

where a is a positive threshold value determine by controlling either the short or the long ARL at

a given level and $p(\cdot)$ is a positive continuous function on Θ. For all θ and λ,

$$p(\theta) = \inf_{\lambda \in \Lambda} \frac{I(\lambda, \theta)}{q(\lambda)} \text{ and } q(\lambda) = \inf_{\theta \in \Theta} \frac{I(\lambda, \theta)}{p(\theta)}, \tag{1.2.5}$$

where $I(\lambda, \theta) = E_\lambda \log\{g_\lambda(x)/f_\theta(x)\}$. The E_λ is the expectation associated with the post-change parameter λ.

1.3 Dissertation Structure

In Chapter 2 we develop a method for sequential change-point detection procedure for high-dimensional data in penalized regression model. Our proposed test statistics for the sequential change-point detection procedure based on SCAD penalized regression model with finite monitoring horizon for high-dimensional data. The asymptotic properties of the proposed test statistics are derived. The asymptotic critical values of the proposed method are obtained through simulations. Simulations are conducted to study the performance of the proposed method. Stopping time at various change point locations are obtained. The proposed method is applied to the gene expression in the mammalian eye data to detect the structural change sequentially.

In Chapter 3, we provide an overview of penalized quantile regression and summarize existing work related to our study. We propose a sequential change point detection procedure in SCAD penalized quantile regression for high-dimensional data. To improve the monitoring process, we provide a modified version of the test statistic based on the post-SCAD penalized quantile regression estimator. We conduct simulations to illustrates the testing procedure and to evaluate the performance of the proposed methods.

In Chapter 4, we discuss the confidence distributions for skew normal change-point model based on modified information criterion (MIC). Simulations are conducted to illustrate the advantages of the proposed method. We compare the existing method in terms of coverage probabilities and average lengths of the confidence sets, especially when the change occurs at the very beginning or in the very end. The proposed method is applied to two stock market data to illustrate the detection and the estimation procedures.

In Chapter 5, we study the change point problem for three-parameter Weibull distribution. We verify the asymptotic distribution of the MIC statistics and convergence of the change point estimator through simulations. Simulations are carried out to illustrate the performance of the MIC and SIC statistics. A rainfall data is used to illustrate the application of the proposed procedure.

The conclusions and some potential future research directions are provided in Chapter 6.

CHAPTER 2 SEQUENTIAL CHANGE POINT DETECTION PROCEDURE FOR HIGH-DIMENSIONAL DATA VIA SCAD PENALTY

2.1 Introduction

In the sequential change-point analysis the observations are received sequentially and we need to make decision whether to continue the process or not after every new observation. The decision has to be made solely based on previous information in real-time. There has been a rich literature in the sequential change-point detection analysis for univariate data, see for example, Page (1954); Shiryaev (1963); Roberts (1966); Lorden (1971). Horváth, Huškova, Kokoszka, and Steinebach (2004) proposed a sequential monitoring method to detect structural change based on weighted CUSUMs of residuals, in which the unknown in-control parameter has been replaced by its least-squares estimate from the training observations. Furthermore, their monitoring process continues until infinity when the null hypothesis is not rejected. In practice, for most real-world applications, we cannot continue to monitor the process until infinity if no change exists. Horváth, Kokoszka, and Steinebach (2007) further investigated the monitoring process for a linear model which stops even if no change is detected after a certain number of observations. Zhou, Wang, and Tang (2015) developed a method for sequential detection of structural changes in linear quantile regression models.

High-dimensional data analysis is a very popular research area in statistics. Generally, the number of explanatory variables (p) exceeds the number of observations (n) as we refer to high-dimensional data. In real-world scenario, we often deal with large number of explanatory variables. For example, genomic and healthcare data set have large number of explanatory variables for each observation. For large number of predictor variables, there is a chance that these variables can be correlated with other explanatory variables. Moreover, too many variables could lead to over-fitting in a regression model. Variable selection procedures such as forward selection, backward elimination, or stepwise method to select the best subset of predictor variables and to attain parsi-

mony and good fit. These methods, however, have certain limitations for collinear regressors and lead to inaccurate results for high-dimensional data.

The penalized regression techniques have been proven to be effective for high-dimensional data, see for example, Tibshirani (1996); Hastie, Tibshirani, and Friedman (2009). The idea behind the penalized regression method is to perform linear regression, while shrinking the coefficients towards zero. The benefit of this approach is that shrinking the coefficient estimates can significantly reduce their variances. There are two well-known techniques for shrinking the regression coefficients towards zero are the ridge regression and the least absolute shrinkage and selection operator (Lasso), see Tibshirani (1996). The major difference is that the ridge (ℓ_2 penalty) regression shrinks all the coefficients to a non-zero value whereas the Lasso (ℓ_1 penalty) shrinks some of the coefficients and sets other to be exactly equal to zero. Thus, Lasso method performs both variable selection and parameter estimation simultaneously. The Lasso-type estimator proposed by Knight and Fu (2000) where they minimized the residual sum of squares and a penalty proportions to the models parameter. Other penalized regression techniques, including smoothly clipped absolute deviation (SCAD) (Fan and Li (2001)), Elastice Net (Zou and Hastie (2005)) and adaptive Lasso (Zou (2006)) are few. The SCAD penalty performs both variable selection and estimation simultaneously and thus, it is computationally feasible for high-dimensional data. Fan and Li (2001) established the asymptotic properties of SCAD penalized likelihood. In addition, the SCAD penalty function satisfies the oracle property, see, Fan and Li (2001); Fan and Peng (2004).

A handful of literature is available on the Lasso estimation in change point analysis. Kim and Kim (2008) studied the asymptotic behavior of the least squares estimators in segmented multiple regression which has one or more change points. Harchaoui and Lévy-Leduc (2010) proposed an approach for estimation of the location of change-points in one-dimensional piecewise constant signals observed in white noise. They used a penalized least-squares criterion with a ℓ_1 penalty. Ciuperca (2014) studied the model selection procedure for adaptive Lasso with multiple change-points and investigated its asymptotic properties. To the best of our knowledge there is no previous study investigated the use of the SCAD penalty with regression model for sequential change point

analysis under high-dimensional scenario.

In this chapter, we propose test statistics for sequential change point detection procedure using SCAD penalized regression model with finite monitoring horizon for high-dimensional data. Consider the linear model,

$$Y_i = X_i^\top \beta + \mathcal{E}_i, \quad i = 1, 2, \ldots, \tag{2.1.1}$$

where X_i is the ith observation for the vector $(X_1, \ldots, X_p)$ of the explanatory variables, $\beta = (\beta_1, \ldots, \beta_p)$ denotes the true unknown coefficients and Y_i denotes the response variable Y. We assume that there exist historical observations of size m such that the matrix $m \times p$ of the variables $X = (X_1, \ldots, X_p) \in \mathbb{R}^{m \times p}$, the response variables $Y = (Y_1, \ldots, Y_m) \in \mathbb{R}^m$, and a vector of independent identically distributed errors $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_m)$. Let $\beta_0 = (\beta_{01}, \ldots, \beta_{0p})$ be the true unknown parameter vector.

2.1.1 LASSO

The least absolute shrinkage and selection operator (Lasso) method was introduced by Tibshirani (1996). This estimation method is defined as the minimization of the least squares penalized by the norm ℓ_1 of the vector β. Lasso can successfully shrink some coefficients to be exactly zero and give a sparse solution.

$$\hat{\beta}_m^{Lasso} = \underset{\beta \in \mathbb{R}^P}{\arg\min} \left\{ \sum_{i=1}^{m} \left(Y_i - X_i^\top \beta\right)^2 + \lambda_m \sum_{j=1}^{p} |\beta_j| \right\}, \tag{2.1.2}$$

where $\lambda_m (0 < \lambda_m < \infty)$ is tuning parameter and can be obtained via cross-validation, see Tibshirani (1996). Zhao and Yu (2006) noted if an irrelevant predictor is highly correlated with the predictors in the true model, Lasso may not be able to distinguish it from the true predictors with any amount of data and any amount of regularization. Further, Knight and Fu (2000) proved that the Lasso estimator is only $n^{1/2}$-consistent under some regularity conditions. Thus, it cannot be achieved simultaneous consistent variable selection and estimation. Therefore, the oracle property do not hold for the Lasso, see for example Fan and Li (2001); Zou (2006).

2.1.2 SCAD Penalized Regression

To improve the performance of the Lasso, Fan and Li (2001) introduced an oracle selection procedure referred to as the smoothly clipped absolute deviation (SCAD). The SCAD is intended to penalize small coefficients heavily and large coefficients lightly. The penalized least squares estimator, denoted by $\hat{\beta}_m^{SCAD}$ can be defined as,

$$\hat{\beta}_m^{SCAD} = \underset{\beta \in \mathbb{R}^P}{\arg\min} \left\{ \sum_{i=1}^{m} \left(Y_i - X_i^\top \beta\right)^2 + \sum_{j=1}^{p} p_{\lambda_m}(|\beta_j|) \right\}, \tag{2.1.3}$$

here $p_{\lambda_m}(\cdot)$ is the penalty function with tuning parameter λ_m. The SCAD penalty is symmetric and continuously differentiable on $(-\infty, 0) \cup (0, \infty)$. The first derivative of the SCAD penalty is given by,

$$P'_{\lambda_m}(\beta) = \lambda_m \{ I(\beta \leq \lambda_m) + \frac{(a\lambda_m - \beta)_+}{(a-1)\lambda_m} I(\beta > \lambda_m) \}. \tag{2.1.4}$$

Fan and Li (2001) discussed that the SCAD penalty function satisfies three requirements for variable selection coefficient estimation, including asymptotic unbiasedness, sparsity and continuity of the estimated parameters. The solution to the SCAD penalty can be given as, the using the

$$\hat{\beta}_j^{SCAD} = \begin{cases} \text{sign}(\hat{\beta}_j)(|\hat{\beta}_j| - \lambda_m)_+ & \text{if} \quad |\hat{\beta}_j| \leq \lambda_m, \\ \{(a-1)\hat{\beta}_j - \text{sign}(\hat{\beta}_j) a\lambda_m\}/(a-2) & \text{if} \quad 2\lambda_m < |\hat{\beta}_j| \leq a\lambda_m, \\ \hat{\beta}_j & \text{if} \quad |\hat{\beta}_j| > a\lambda_m, \end{cases}$$

where $\lambda_m (0 < \lambda_m < \infty)$ and a are two unknown parameters. Theoretically, the best pair (λ_m, a) can be obtained using two dimensional grids search, for example, cross validation method. Fan and Li (2001) suggested $a = 3.7$ is a good choice for various problems. In this research a was set to 3.7 and λ_m selected by the cross validation method. The tuning parameter λ_m controls the amount of shrinkage. The larger the value of λ_m, the greater the amount of shrinkage. Fan and Li (2001) discussed that SCAD penalty function satisfies three requirements for variable selection and coefficient estimation, including asymptotic unbiasedness, sparsity and continuity of the estimated

parameters. Without loss of generality, we assume that the first q regression coefficients are nonzeros and the remaining $(p-q)$ regression coefficients are 0s. Let $\mathcal{A} = \{\beta_{0,j} \neq 0 : j = 1, \ldots, p\}$ be the index set of the nonzero coefficients for the true parameter, where $\beta_{0,j}$ is the jth component of the parameter vector β_0. We denote the SCAD penalized regression estimate by $\hat{\beta}_m^{SCAD}$. Let $\mathcal{A}^* = \{\hat{\beta}_{m,j}^{SCAD} \neq 0 : j = 1, \ldots, p\}$ be the index set of the SCAD penalized regression estimator calculated using the historical sample size m, where $\hat{\beta}_{m,j}^{SCAD}$ is the jth element of the SCAD penalized regression estimator $\hat{\beta}_m^{SCAD}$. To obtain the limiting distribution, we make the following assumptions, also called the regularity conditions, which are needed to derive the asymptotics of the estimators:

A1. Let $C_m = \frac{1}{m}\sum_{i=1}^{m} X_i X_i^\top$. There exists a $p \times p$ positive definite matrix C and a constant $\rho > 2$ such that $|C_m - C| = O(m^{-\rho})$.

A2. The model errors $\mathcal{E}_1, \ldots, \mathcal{E}_m, \mathcal{E}_{m+1}, \ldots$ are independent and identically distributed random variables. $E(\mathcal{E}_i) = 0$, $Var(\mathcal{E}_i) = \sigma^2 < \infty$ and $E|\mathcal{E}_i|^v < \infty$ with some $v > 2$.

A3. If A1 holds, we have, $|C_m^{-1} - C^{-1}| = O(m^{-\rho})$ a.s.

A4. There exist constants $M_1, M_2 \in \mathbb{R}$ such that

$$M_1 < \lambda_{min}\left(\frac{1}{n}X^\top X\right) < \lambda_{max}\left(\frac{1}{n}X^\top X\right) < M_2,$$

where λ_{min} and λ_{max} are the smallest and largest eigenvlaues of C_m respectively.

Since SCAD penalized estimator satisfy the oracle property we have,

1. $\sqrt{n}\mathbb{A}_n\big(n^{-1}X^\top X\sigma^2\big)^{1/2}\big(\hat{\beta}_m^{SCAD} - \beta_0\big) \longrightarrow N(0, C_{11})$, where $\mathbb{A}_n$ is an arbitrary matrix such that $\mathbb{A}_n\mathbb{A}_n^\top \longrightarrow C_{11}$, and C_{11} is $q \times q$ non-negative symmetric matrix and contains the elements of the matrix C in the set $\mathcal{A}$.

2. Consistency in variable selection, $\lim_{m\longrightarrow\infty} P(\mathcal{A}^* = \mathcal{A}) = 1$.

2.2 Sequential Change-Point Problem

Let m be the size of the historical sample. We assume that there is no change in the historical sample. Similar to Chu, Stinchcombe, and White (1996), we use the historical sample to estimate the pre-change coefficients of the SCAD penalized regression model. After we select the significant explanatory variables, the future incoming observations $Y,\ X_1, \ldots, X_p$ are monitored sequentially following the historical sample size m. Let T_m be the monitoring horizon. The linear model after historical observations m is,

$$Y_i = X_i^\top \beta + \mathcal{E}_i, \quad i = m+1, m+2, \ldots \tag{2.2.1}$$

At each time point i, our goal is to test whether we have the same model as the one using the historical sample size m. Under the null hypothesis, if there is no change in the coefficients,

$$H_0 : \beta_i = \beta_0 \quad \text{for } i = m+1, m+2, \ldots$$

Under the alternative hypothesis, we consider at an unknown time point k the coefficients change from β_0 to β_1. There exists $k \geq 1$ such that,

$$H_1 : \begin{cases} \beta_i = \beta_0 \ ; & i = m+1, m+1, \ldots, m+k, \\ \beta_i = \beta_1 \ ; & i = m+k+1, \ldots m+T_m \quad \text{and } \beta_0 \neq \beta_1. \end{cases}$$

According to Horváth et al. (2004),

$$\Gamma(m,k) = \frac{1}{\hat{\sigma}_m} \left| \sum_{i=m+1}^{m+k} \hat{\mathcal{E}}_i \right|, \tag{2.2.2}$$

where $\hat{\mathcal{E}}_i = Y_i - X_i^\top \hat{\beta}_m^{SCAD}$ for $i = m+1, m+2, \ldots$ and $\hat{\sigma}_m^2$ is the error variance defined as,

$$\hat{\sigma}_m^2 = \frac{1}{(m-p^*)} \sum_{i=1}^{m} (Y_i - X_i^\top \hat{\beta}_m^{SCAD})^2, \tag{2.2.3}$$

where p^* is the number of non-zero coefficients of the SCAD penalized estimator. For a given constant $\gamma \in [0, 1/2)$, the $g(m, k, \gamma)$ is called the normalising function and defined as,

$$g(m, k, \gamma) = m^{1/2}\left(1 + \frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}, \tag{2.2.4}$$

where γ is called the control parameter. We propose the test statistic for the monitoring structural change,

$$\Omega = \sup_{1 \le k \le T_m} \frac{\Gamma(m, k)}{g(m, k, \gamma)}. \tag{2.2.5}$$

2.2.1 Open-end Procedure

Horváth et al. (2004) studied the monitoring process which may continue to infinity if no alarm is raised. This is referred to as open-end procedure. In the open-end procedure, the monitoring horizon $T_m = \infty$. The stopping time of the proposed test statistic when the monitoring process stop and reject the null hypothesis is defined as,

$$\Lambda(m) = \begin{cases} \inf\{k \ge 1; & \Gamma(m, k) \ge g(m, k, \gamma) c_\alpha(\gamma)\}, \\ \infty & \text{for all } k = 1, 2, 3, \ldots, \end{cases} \tag{2.2.6}$$

where $c_\alpha(\gamma)$ is the critical value which can be obtained through simulations at a given significance level $\alpha \in (0, 1)$. Under the null hypothesis,

$$\lim_{m \to \infty} P(\Lambda(m) < \infty) = \alpha, \tag{2.2.7}$$

and under the alternative hypothesis,

$$\lim_{m \to \infty} P(\Lambda(m) < \infty) = 1. \tag{2.2.8}$$

2.2.2 Closed-end Procedure

In the closed-end procedure, the monitoring process stops after a fixed number of observations even if no change is observed, see for example, Horváth et al. (2007); Zhou et al. (2015). Let $N > 0$. Suppose $T_m < \infty$ with $\lim_{m\to\infty} T_m/m = N$. Under closed-end procedure, the stopping time of the proposed test statistic when the monitoring process stop and reject the null hypothesis is defined as,

$$\Lambda^*(m) = \begin{cases} \inf\{k \geq 1; & \Gamma(m,k) \geq g(m,k,\gamma)c^*_\alpha(\gamma)\}, \\ T_m & \text{for all } k = 1, \ldots, T_m, \end{cases} \tag{2.2.9}$$

where $c^*_\alpha(\gamma)$ is the critical value which can be obtained through simulations at a given significance level $\alpha \in (0,1)$. Under the null hypothesis,

$$\lim_{m\to\infty} P(\Lambda^*(m) < \infty) = \alpha, \tag{2.2.10}$$

and under the alternative hypothesis,

$$\lim_{m\to\infty} P(\Lambda^*(m) < \infty) = 1. \tag{2.2.11}$$

The monitoring process stops immediately for large γ value. Thus, the large value of γ is preferred when the change in the regression coefficients happens shortly after m.

Theorem 2.2.12. *Under the assumptions, if the null hypothesis holds, for open-end procedure,*

$$\lim_{m\to\infty} P\big(\Omega \leq c_\alpha(\gamma)\big) = P\bigg(\sup_{1\leq t\leq 1} \frac{\|W(t)\|_\infty}{t^\gamma} \leq c_\alpha(\gamma)\bigg),$$

for closed-end procedure,

$$\lim_{m\to\infty} P\big(\Omega \leq c^*_\alpha(\gamma)\big) = P\bigg(\sup_{1\leq t\leq N/(N+1)} \frac{\|W(t)\|_\infty}{t^\gamma} \leq c^*_\alpha(\gamma)\bigg),$$

where $\{W(t), 0 \leq t < \infty\}$ *denotes a* p*-dimensional Weiner process.*

Using Theorem 2.2.12, under the null hypothesis, we can obtain the asymptotic distribution of the CUSUM test statistic for the open- and closed-end procedures. The asymptotic critical value $c_\alpha(\gamma)$ is obtained from,

$$P\left(\sup_{1\leq t\leq 1}\frac{\|W(t)\|_\infty}{t^\gamma}\geq c_\alpha(\gamma)\right)=\alpha,$$

and the asymptotic critical value for the closed-end procedure $c^*_\alpha(\gamma)$ can be obtained from,

$$P\left(\sup_{1\leq t\leq N/(N+1)}\frac{\|W(t)\|_\infty}{t^\gamma}\geq c^*_\alpha(\gamma)\right)=\alpha,$$

where $\alpha\in(0,1)$ and the tuning parameter $0\leq\gamma<1/2$. We obtain the asymptotic critical values through simulation. The Weiner process $\|W(t)\|_\infty/t^\gamma$ is approximated by $\|M^{-1/2}\sum_{i=1}^{tM}e_i\|_\infty/t^\gamma$ where $e_i\sim N(0,1)$ and M is a grid of 10000. In each iteration, the maximum value of the process for both open- and closed-end procedures obtained over $t\in(0,1)$ and $t\in(0,N/(N+1))$ respectively. The asymptotic critical values are given in Tables 2.1-2.2. The results are based on 50000 iterations.

Theorem 2.2.13. *Under the assumptions, if the alternative hypothesis holds, we have,*

$$\sup_{1\leq k\leq T_m}\frac{\Gamma(m,k)}{g(m,k,\gamma)}\longrightarrow\infty\quad\text{as}\quad m\longrightarrow\infty.$$

Proof. **Theorem 2.2.12**

The proof of Theorem 2.2.12 is based on the series of lemmas given in Horváth et al. (2004) and Horváth et al. (2007). In high-dimensional setting the Lemma 3.1 in Horváth et al. (2007) and Lemma 5.1 in Horváth et al. (2007) are still valid. We will show that the Lemma 5.2 in Horváth et al. (2004) is true for the SCAD penalized least square model. After that, we can adopt the Theorem 2.1 in Horváth et al. (2004) to prove the Theorem 2.2.12. □

Lemma 2.2.14. *[Lemma 5.2 in Horváth et al. (2004)] If the assumptions of Theorem 2.2.12 are*

satisfied, as $m \longrightarrow \infty$

$$\sup_{1\leq k\leq N} \frac{\left| \sum_{i=m+1}^{m+k} \hat{\mathcal{E}}_i - \left(\sum_{i=m+1}^{m+k} \mathcal{E}_i - \frac{k}{m} \sum_{i=1}^{m} \mathcal{E}_i \right) \right|}{g(m,k,\gamma)} = O_p(m^{-\rho}).$$

where $g(m,k,\gamma)$ *is defined in (3.2.11).*

Proof. **Lemma 2.2.14**

So,

$$\hat{\beta}_m^{SCAD} - \beta_0 = C_m^{-1} \frac{1}{m} \sum_{i=1}^{m} X_i \mathcal{E}_i$$

we have

$$\begin{aligned} \sum_{i=m+1}^{m+k} \hat{\mathcal{E}}_i &= \sum_{i=m+1}^{m+k} \left(\mathcal{E}_i - X_i^\top (\hat{\beta}_m^{SCAD} - \beta_0) \right) \\ &= \sum_{i=m+1}^{m+k} \mathcal{E}_i - \left(\sum_{i=m+1}^{m+k} X_i \right)^\top C_m^{-1} \frac{1}{m} \sum_{j=1}^{m} X_j \mathcal{E}_j. \end{aligned}$$

By the central limit theorem and **A1** we get,

$$\left| \sum_{i=1}^{m} X_i \mathcal{E}_i \right| = O_p(m^{1/2}). \tag{2.2.15}$$

Since the SCAD estimate $\hat{\beta}_{m,j}^{SCAD}$ satisfies the oracle property by consistency in variable selection, $\lim_{m\longrightarrow\infty} P(\mathcal{A}^* = \mathcal{A}) = 1$. Therefore, $\lim_{m\longrightarrow\infty} P(\mathcal{A}^* \cap \mathcal{A}^c = \emptyset) = 1$. Let $\mathbb{S} = \mathcal{A}^* \cap \mathcal{A}$. Putting together (A1), (A3) and (2.2.15), we get $(X_{\mathbb{S}}^\top X_{\mathbb{S}})^{-1} = \frac{1}{m} C_{\mathbb{S}}^{-1}(1 + o_p(1))$, where $C_{\mathbb{S}}$ contains the elements of the matrix C with the index in the set $\mathbb{S}$. Considering the Karush-Kuhn-Tucker (KKT) optimality conditions, for $j \in \mathcal{A} \cap \mathcal{A}^*$ we have,

$$\begin{aligned} 2X_j^\top (Y - X\hat{\beta}_m^{SCAD}) &= \lambda_m \text{sign}(\hat{\beta}_{m,j}^{SCAD}), \\ 2X_j^\top (\mathcal{E} - X(\hat{\beta}_m^{SCAD} - \beta_0)) &= \lambda_m \text{sign}(\hat{\beta}_{m,j}^{SCAD}), \\ X_j^\top (\mathcal{E} - X(\hat{\beta}_m^{SCAD} - \beta_0)) &= \frac{1}{2} \lambda_m \text{sign}(\hat{\beta}_{m,j}^{SCAD}). \end{aligned}$$

For any given $\epsilon > 0$ and , we have,

$$P\bigg(X(\hat{\beta}_m^{SCAD} - \beta_0) = X_{\mathbb{S}}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}}\bigg) > 1 - \epsilon, \tag{2.2.16}$$

$$P\bigg(X_{\mathbb{S}}^{\top}(\mathcal{E} - X_{\mathbb{S}}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}}) = \frac{1}{2}\lambda_m \text{sign}(\hat{\beta}_{m,j}^{SCAD})\bigg) > 1 - \epsilon. \tag{2.2.17}$$

By assumptions,

$$P\bigg((\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}} = (X_{\mathbb{S}}^{\top}X_{\mathbb{S}})^{-1}X_{\mathbb{S}}^{\top}\mathcal{E} - \frac{1}{2}\lambda_m(X_{\mathbb{S}}^{\top}X_{\mathbb{S}})^{-1}\text{sign}(\hat{\beta}_{m,j}^{SCAD})\bigg) > 1 - \epsilon. \tag{2.2.18}$$

For all $j \in \mathcal{A}$, we have,

$$X_j^{\top}\mathcal{E} = O_p(m^{1/2}) \quad \text{and} \quad (X_{\mathbb{S}}^{\top}X_{\mathbb{S}})^{-1}X_{\mathbb{S}}^{\top}\mathcal{E} = O_p(m^{-1/2}). \tag{2.2.19}$$

The asymptotic normality of the estimators implies,

$$(q/m)^{1/2}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}} = O_p(1). \tag{2.2.20}$$

Putting together (2.2.19), (2.2.18) and assumptions, we get

$$(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}} = (X_{\mathbb{S}}^{\top}X_{\mathbb{S}})^{-1}X_{\mathbb{S}}^{\top}\mathcal{E}(1 + o_p(1)). \tag{2.2.21}$$

Similarly, for all $k \geq 1$, we get

$$P\bigg(\sum_{i=m+1}^{m+k} X_i^{\top}(\hat{\beta}_m^{SCAD} - \beta_0) = \sum_{i=m+1}^{m+k} X_{i,\mathbb{S}}^{\top}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}}\bigg) > 1 - \epsilon, \tag{2.2.22}$$

Putting together (2.2.22), (2.2.18) and (2.2.19), we obtain the CUSUM of residuals

$$\begin{aligned}\sum_{i=m+1}^{m+k}\hat{\mathcal{E}}_i &= \sum_{i=m+1}^{m+k}\left(\mathcal{E}_i - X_{i,\mathbb{S}}^{\top}(\hat{\beta}_m^{SCAD}-\beta_0)_{\mathbb{S}}\right)\\ &= \sum_{i=m+1}^{m+k}\mathcal{E}_i - \sum_{i=m+1}^{m+k}X_{i,\mathbb{S}}^{\top}(\hat{\beta}_m^{SCAD}-\beta_0)_{\mathbb{S}} \qquad (2.2.23)\\ &= \sum_{i=m+1}^{m+k}\mathcal{E}_i - \left(\sum_{i=m+1}^{m+k}X_{i,\mathbb{S}}^{\top}\right)(X_{\mathbb{S}}^{\top}X_{\mathbb{S}})^{-1}X_{\mathbb{S}}^{\top}\mathcal{E}(1+o_p(1)).\end{aligned}$$

Now applying the Theorem 2.1 of Horváth et al. (2004) completes the proof. □

Proof. **Theorem 2.2.13**

Let $k' = m + k^*$. Consider,

$$\begin{aligned}\sum_{i=m+1}^{m+k^{'}}\hat{\mathcal{E}}_i &= \sum_{i=m+1}^{m+k^{'}}\left(\mathcal{E}_i - X_{i,\mathbb{S}}^{\top}(\hat{\beta}_m^{SCAD}-\beta_0)_{\mathbb{S}}\right)\\ &= \sum_{i=m+1}^{m+k^{'}}\mathcal{E}_i - \sum_{i=m+1}^{m+k^{*}}X_{i,\mathbb{S}}^{\top}(\hat{\beta}_m^{SCAD}-\beta_0)_{\mathbb{S}} - \sum_{i=m+k^{*}+1}^{m+k^{'}}X_{i,\mathbb{S}}^{\top}(\hat{\beta}_m^{SCAD}-\beta_0)_{\mathbb{S}}.\end{aligned} \qquad (2.2.24)$$

But Theorem 2.2.12 yields that,

$$\frac{\left|\sum_{i=m+1}^{m+k^{'}}\hat{\mathcal{E}}_i - \left(\sum_{i=m+1}^{m+k^{'}}X_{i,\mathbb{S}}\right)^{\top}(\hat{\beta}_m^{SCAD}-\beta_0)_{\mathbb{S}}\right|}{g(m,k^{'},\gamma)} = O_p(1).$$

Let C_{11} be the first column of C and let $C_{11}^{\top}(\beta_0 - \hat{\beta}_m^{SCAD})_{\mathbb{S}} \neq 0$. Using the assumptions we have,

$$\begin{aligned}\left(\sum_{i=m+k^{*}+1}^{m+k^{'}}X_{i,\mathbb{S}}\right)^{\top}(\beta_0-\hat{\beta}_m^{SCAD})_{\mathbb{S}} &= (k^{'}-k^{*})C_{11}^{\top}(\beta_0-\hat{\beta}_m^{SCAD})_{\mathbb{S}}\\ &\quad + O((m+k^{*})^{1-\rho}) + O((m+k^{'})^{1-\rho}).\end{aligned}$$

Thus, we get

$$\liminf_{m\longrightarrow\infty}\frac{\left|\left(\sum_{i=m+k^*+1}^{m+k'}X_{i,\mathbb{S}}\right)^{\top}(\hat{\beta}_m^{SCAD}-\beta_0)_{\mathbb{S}}\right|}{g(m,k',\gamma)}>0.$$

This completes the proof. □

Table 2.1 Asymptotic critical values for the open-end procedure when the dimensions $p=2$ and $\gamma\in\{0.00,0.15,0.25,0.35,0.45,0.49\}$, calculated on 50000 replications

γ/α	0.010	0.025	0.050	0.100	0.250
0.00	3.019468	2.734264	2.486683	2.234026	1.833559
0.15	3.078669	2.795639	2.554139	2.303043	1.911031
0.25	3.135282	2.864542	2.626865	2.372679	1.990685
0.35	3.247067	2.970504	2.742768	2.498331	2.128268
0.45	3.479984	3.229682	3.008465	2.784761	2.434316
0.49	3.764611	3.508515	3.303435	3.073143	2.720441

Table 2.2 Asymptotic critical values for the closed-end procedure when $N \in \{2, 4, 6, 9\}$, $p = 2$ and $\gamma \in \{0.00, 0.15, 0.25, 0.35, 0.45, 0.49\}$, calculated on 50000 replications

γ	N/α	0.010	0.025	0.050	0.100	0.250
0.00	2	2.447139	2.211787	2.020908	1.815498	1.487324
	4	2.686489	2.430562	2.222431	1.990774	1.629909
	6	2.785824	2.514810	2.297350	2.057684	1.688002
	9	2.847110	2.572124	2.351546	2.108416	1.730943
0.15	2	2.649895	2.404296	2.201695	1.987356	1.647810
	4	2.829618	2.569068	2.355556	2.121323	1.759586
	6	2.902434	2.632341	2.411011	2.170361	1.801032
	9	2.948248	2.676014	2.446939	2.209239	1.833419
0.25	2	2.816933	2.561952	2.359019	2.134979	1.790156
	4	2.954011	2.681545	2.469789	2.236936	1.874830
	6	3.004398	2.735205	2.510444	2.276272	1.907793
	9	3.045116	2.767461	2.539297	2.300939	1.930600
0.35	2	3.026753	2.766572	2.563014	2.338523	1.994179
	4	3.122079	2.854851	2.635548	2.404585	2.049275
	6	3.149867	2.883356	2.661403	2.427641	2.070099
	9	3.171244	2.904029	2.679934	2.445955	2.085801
0.45	2	3.400988	3.136741	2.935825	2.713851	2.375432
	4	3.433242	3.167601	2.965047	2.739027	2.400873
	6	3.442337	3.179246	2.974512	2.748497	2.408898
	9	3.446603	3.188064	2.981583	2.754953	2.414795
0.49	2	3.732576	3.463671	3.257303	3.037973	2.688347
	4	3.744958	3.474354	3.268213	3.050684	2.701117
	6	3.747245	3.478562	3.271668	3.054774	2.706148
	9	3.749885	3.479867	3.274127	3.057681	2.708363

2.3 Simulation Studies

In this section, we evaluate the performance of the sequential change point detection procedure for both open- and closed-end methods for the SACD penalised regression model. We consider that the number of explanatory variables $p = 10$. The following two settings are considered. The first setting is used to evaluate the Type I errors of the proposed method. In the first case, the true parameter vectors $\beta_0 \in \{-2,\ 0,\ 2,\ 0,\ 10,\ 1,\ 0,\ 0,\ 8,\ -5\}$ and X_i for all $i \in \{1, \ldots, 10\} \backslash \{3, 4, 5\}$ have standard normal distribution $N(0, 1)$ and $X_3 \sim N(2, 1),\ X_4 \sim N(4, 1)$ and $X_5 \sim N(5, 1)$. In the second setting, under the null hypothesis, the true parameter vectors

$\beta_0 \in \{0,\ 0,\ 2,\ 0,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0\}$ and under the alternative hypothesis, we consider the parameter vector $\beta_1 \in \{0,\ 0,\ 0,\ 3,\ 0,\ 0,\ 1,\ 0,\ 0,\ -1\}$. We consider the two different distributions of the explanatory variables $X_1,\ X_2, \ldots, X_{10}$. Under $H_0,\ X_i$ for all $i \in \{1, \ldots, 10\} \backslash \{3, 4, 5\}$ have standard normal distribution $N(0, 1)$ and $X_3 \sim N(2, 1),\ X_4 \sim N(4, 1)$ and $X_5 \sim N(5, 1)$. The second distribution the ith explanatory variable is $X_i + 0.8$ where $X_i \sim N(0, 1)$ for all $i \in \{1, \ldots, 10\}$. Moreover, for both settings, the model errors $\mathcal{E}_i$ are iid $N(0, 1)$.

Table 2.3 summarizes the Type I error for both open and closed end procedures. The various control parameter value γ and the different sizes of the historical observations m are considered. The $\gamma \in \{0, 0.25, 0.45\}$ and $m \in \{75, 100, 200\}$. The results are based on 2500 iterations. The Type I errors based on the closed-end procedure is always larger than the Type I errors obtained from the open-end procedure. For small γ value, the Type I errors of the open-end procedure is below the nominal level 0.05. When open-end procedure is considered, smaller N provides slightly deflated Type I errors. Thus, for small N, we suggest to use the tuning parameter γ close to 0.5. Type I errors are compared in Figure 2.1.

Table 2.3 Type I errors of both open- and closed-end procedures for SCAD penalized regression for various values of γ and the nominal significance level $\alpha = 0.05$

		Closed-end			Open-end		
m	N/γ	0	0.25	0.45	0	0.25	0.45
75	2	0.047	0.053	0.044	0.013	0.024	0.036
	4	0.038	0.037	0.037	0.020	0.028	0.035
	10	0.040	0.042	0.036	0.031	0.033	0.035
	20	0.050	0.052	0.046	0.046	0.047	0.045
	50	0.041	0.041	0.036	0.040	0.039	0.037
100	2	0.035	0.038	0.036	0.010	0.019	0.030
	4	0.034	0.036	0.031	0.017	0.026	0.030
	10	0.039	0.038	0.038	0.030	0.034	0.036
	20	0.033	0.032	0.025	0.028	0.028	0.025
	50	0.034	0.034	0.027	0.032	0.033	0.027
200	2	0.028	0.029	0.031	0.009	0.016	0.028
	4	0.028	0.031	0.030	0.015	0.020	0.024
	10	0.041	0.039	0.037	0.030	0.034	0.036
	20	0.026	0.026	0.024	0.020	0.024	0.023
	50	0.030	0.032	0.028	0.029	0.029	0.027

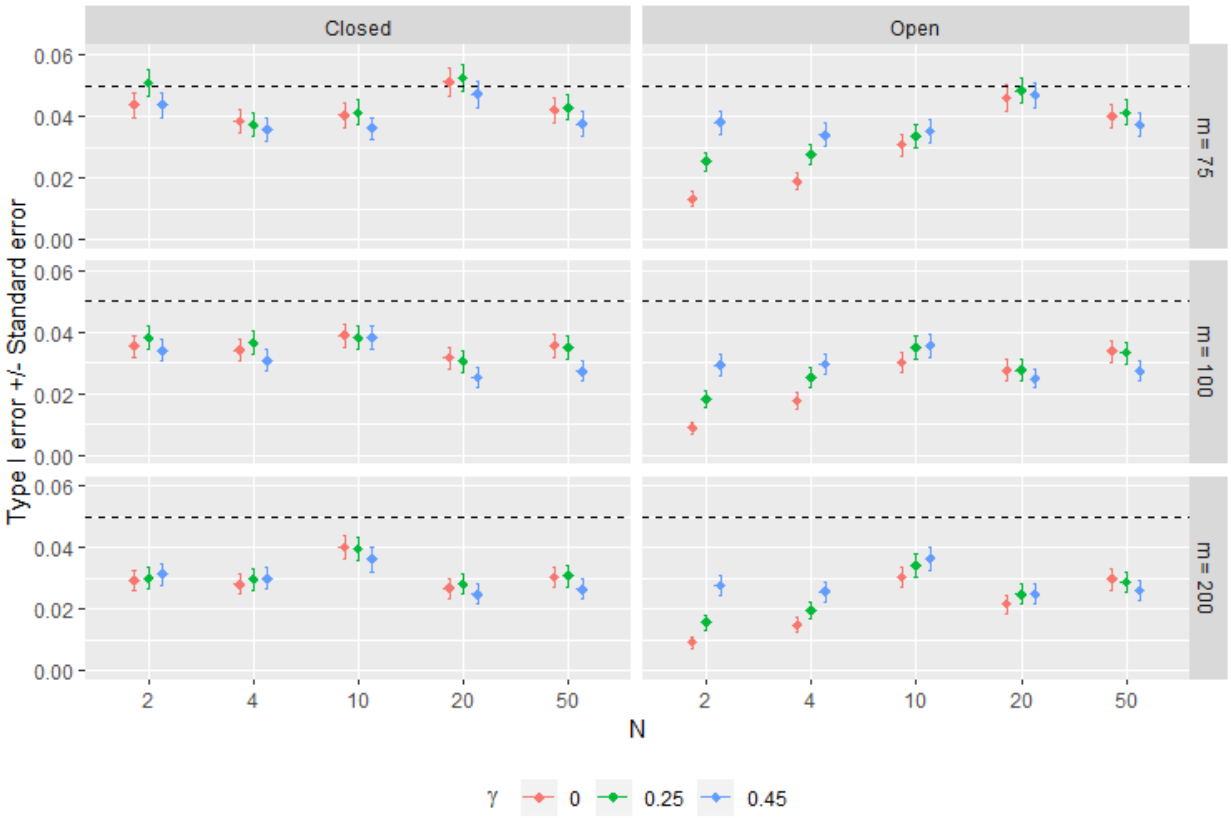

Figure 2.1 Type I error comparison for open- and closed-end procedures

We conducted the Monte Carlo simulation to investigate the performance of the proposed method. First we performed the power analysis for different control parameter $\gamma \in \{0, 0.25, 0.45\}$ while changing the size of a test α, considering $\alpha \in \{0.025, 0.05, 0.1\}$. Further, the simulations are carried out under various change point locations $k^* \in \{1, 25, 75\}$ with different historical sample size $m \in \{75, 100, 200\}$. We evaluate the power of the test based 1000 simulations and the results are summarized in Table 2.4. The results are sketched in Figure 2.2.

The summary statistics of the stopping time for both open and closed end procedures summarized in Tables 2.5 - 2.6. In all cases, the processes are monitored from $m + 1$ until time $9m$. In order to see the effect of the historical sample size, we performed the simulations for various m, such as $m \in \{100, 300, 600\}$. Further, we also change the true change point location, considering $k^* \in \{1, 25, 100\}$ and level $\alpha \in \{0.025, 0.05, 0.1\}$. Clearly, the selection of γ value influences the stopping time. As mentioned in Horváth et al. (2004), it is clear that the smaller value of γ takes

longer time to detect the structural change whereas larger γ value stops faster.

Table 2.4 Power comparison for closed-end procedure for change-points $k^* \in \{1, 25, 100\}$, $\alpha \in \{0.025, 0.05, 0.1\}$ and $\gamma \in \{0, 0.25, 0.45\}$

		$k^* = 1$			$k^* = 25$			$k^* = 75$		
γ	α/m	75	100	200	75	100	200	75	100	200
0.00	0.025	0.948	0.977	0.995	0.936	0.956	0.996	0.857	0.943	0.996
	0.050	0.993	0.999	0.999	0.999	0.997	1	0.988	0.995	1
	0.100	1	1	1	1	1	1	1	1	1
0.25	0.025	0.925	0.962	0.991	0.876	0.933	0.992	0.757	0.876	0.992
	0.050	0.969	0.988	0.999	0.973	0.989	0.999	0.952	0.985	0.999
	0.100	0.999	1	1	1	1	1	0.999	1	1
0.45	0.025	0.860	0.930	0.985	0.756	0.881	0.970	0.552	0.731	0.970
	0.050	0.937	0.965	0.995	0.918	0.960	0.997	0.846	0.931	0.997
	0.100	0.991	0.997	0.999	0.992	0.999	1	0.992	0.997	1

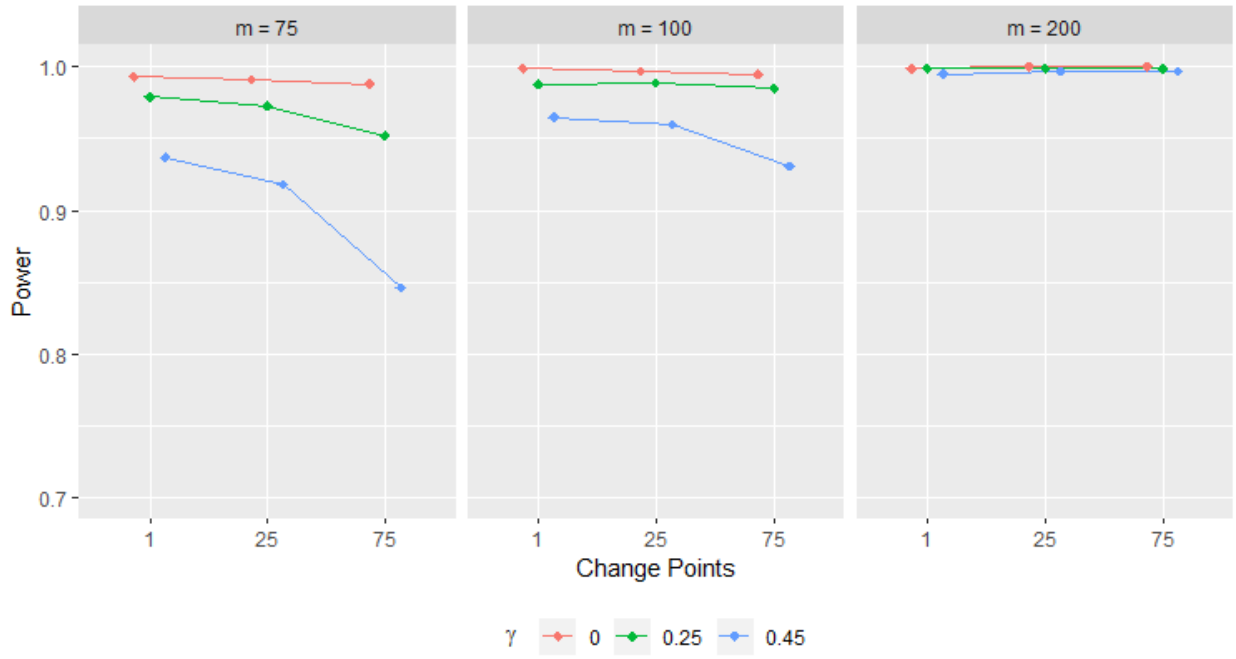

Figure 2.2 Power comparison for closed-end procedures with $\alpha = 0.05$

Table 2.5 Summary statistics for the detection time for the open-end procedure with $k^* \in \{1, 25, 100\}$, $\gamma \in \{0, 0.25, 0.45\}$ and $\alpha = 0.05$

		$m = 100$			$m = 300$			$m = 600$		
γ	Summary/k^*	1	25	100	1	25	100	1	25	100
0.00	min	4	27	72	6	31	102	10	33	109
	Q1	12	38	122	21	47	127	30	56	134
	Med	18	45	134	29	55	137	39	66	146
	Q3	26	57	141	40	67	152	48	79	161
	max	123	200	353	145	159	263	192	227	256
0.25	min	2	9	9	2	26	75	3	27	103
	Q1	5	34	119	7	36	118	9	39	121
	Med	8	40	130	12	42	128	15	45	130
	Q3	14	49	147	20	52	141	25	56	144
	max	102	197	353	95	149	244	153	187	239
0.45	min	1	1	1	2	3	3	2	13	13
	Q1	2	32	119	3	32	115	3	32	115
	Med	4	37	131	4	37	124	4	37	124
	Q3	7	46	148	8	44	137	8	44	135
	max	98	200	359	80	123	243	66	153	237

Table 2.6 Summary statistics for the detection time for the closed-end procedure with $k^* \in \{1, 25, 100\}$, $\gamma \in \{0, 0.25, 0.45\}$ and $\alpha = 0.05$

		$m = 100$			$m = 300$			$m = 600$		
γ	Summary/k^*	1	25	100	1	25	100	1	25	100
0.00	min	4	26	65	6	31	102	9	33	109
	Q1	11	37	120	20	45	125	28	54	132
	Med	16	44	132	27	53	135	37	64	144
	Q3	24	55	148	38	65	149	50	76	158
	max	122	197	335	144	157	262	192	226	255
0.25	min	2	8	8	2	26	74	3	27	102
	Q1	5	33	118	7	36	118	8	38	120
	Med	8	39	129	12	41	127	14	45	129
	Q3	14	48	145	20	51	140	24	55	142
	max	102	194	335	95	149	243	153	163	239
0.45	min	1	1	1	2	3	3	2	4	4
	Q1	2	32	119	3	32	115	3	32	115
	Med	4	37	131	4	37	124	4	36	124
	Q3	7	46	147	8	44	137	8	44	135
	max	54	200	359	80	123	243	66	151	236

We compare the estimated density of the stopping time at various change point locations, historical sample size, control parameter and size of a test. They are given in Figure 2.3. When α values change from 0.025 to 0.05, the estimated densities are roughly identical. Not surprisingly, the historical sample m has significant influence in the stopping time determination. When m changes from 100 to 300, we observe a high variability in the estimated densities. A small variation can be observed between the estimated densities for a fixed control value γ irrespective of the historical sample size.

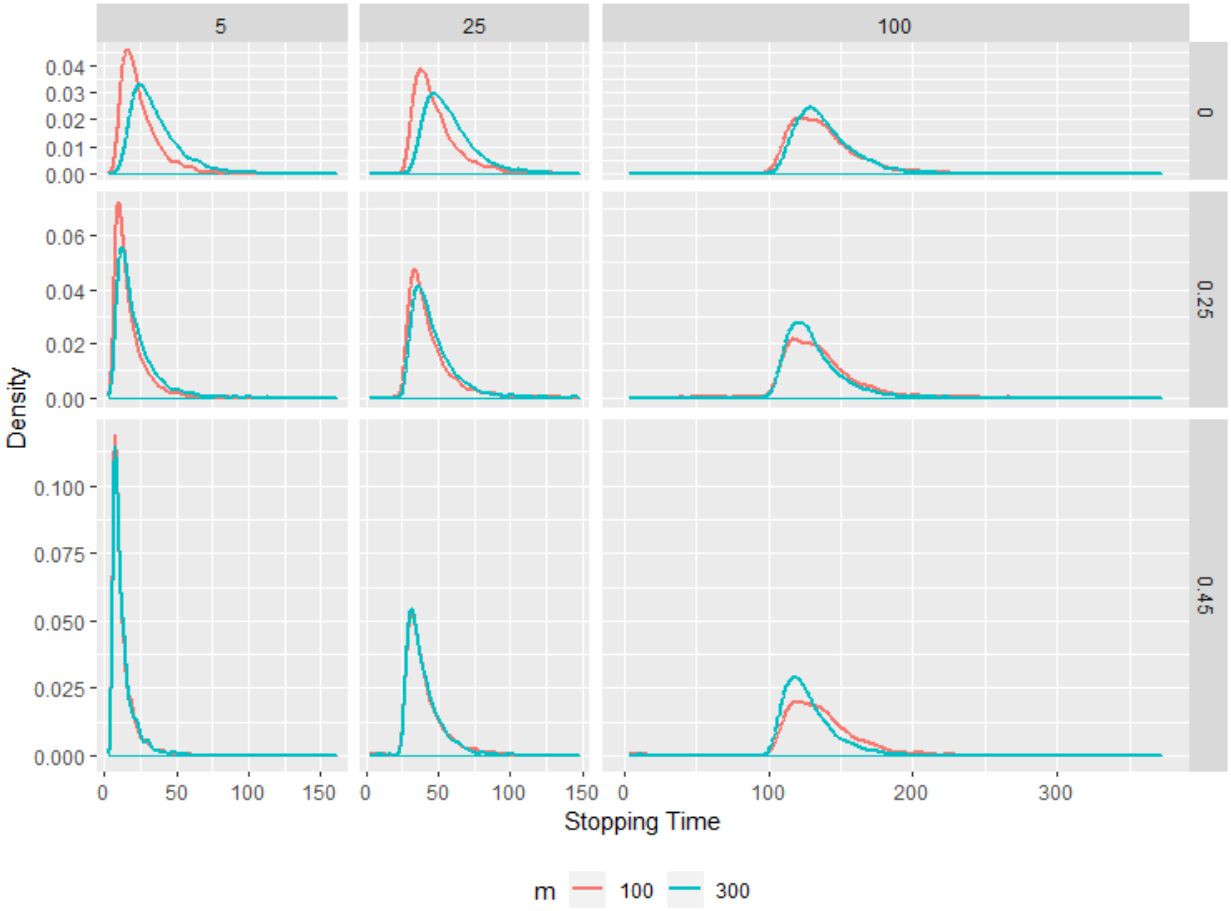

Figure 2.3 Estimated density of the stopping time for $k^* = \{5, 25, 100\}$ and $\gamma \in \{0,\ 0.25,\ 0.45\}$

2.3.1 Large p

To study the performance of the monitoring process in high-dimensional setting, we conduct a simulation study. We generate high-dimensional datasets with (p, m), considering

$(100, 75),\ (200, 100)$ and $(300, 200)$ are generated. We consider the following two settings. In the first settings, the non-zero components of the true parameters are $\beta_{0,1} = -5,\ \beta_{0,2} = 2,\ \beta_{0,3} = 5,\ \beta_{0,4} = 1,\ \beta_{0,5} = -3,\ \beta_{0,61} = -10$ and $\beta_{0,91} = 8$. The predictor variables X_i for all $i \in \{1, \ldots, p\}\backslash\{3, 4, 5\}$ have standard normal distribution $N(0, 1)$ and $X_3 \sim N(2, 1),\ X_4 \sim N(4, 1)$ and $X_5 \sim N(5, 1)$.

In the second setting, under the null hypothesis, the true parameter vectors $\beta_{0,1} = -1,\ \beta_{0,2} = 1,\ \beta_{0,3} = -1,\ \beta_{0,4} = 4,\ \beta_{0,5} = -2,\ \beta_{0,58} = -3$ and $\beta_{0,86} = 2$ and under the alternative hypothesis, we consider the parameter vector $\beta_{1,1} = 3,\ \beta_{1,2} = 2,\ \beta_{1,45} = -2$ and $\beta_{1,93} = 2$ and the two different distributions of the explanatory variables $X_1, X_2, \ldots, X_p$. Under H_0, X_i for all $i \in \{1, \ldots, p\}\backslash\{3, 4, 5\}$ have the normal distribution $N(0, 1)$ and $X_3 \sim N(2, 1),\ X_4 \sim N(4, 1)$ and $X_5 \sim N(5, 1)$. The second distribution the ith explanatory variable is $X_i + 0.8$ where $X_i \sim N(0, 1)$ for all $i \in \{1, \ldots, p\}$. Moreover, for both settings, the model errors $\mathcal{E}_i$ are iid $N(0, 1)$.

Table 2.7 summarizes the Type I error for both open and closed end procedures. The various control parameter value $\gamma \in \{0, 0.25, 0.45\}$ and the different size of the historical observations $m \in \{75, 100, 200\}$ are considered. The results are based on 2500 iterations. The Type I errors based on the closed-end procedure is always larger than the Type I errors based on the open-end procedure. In the open-end operation, it is vital to properly select the value of the control parameter. The Type I errors comparatively low in the open-end procedure for Small γ value. When open-end procedure is considered, smaller N provides slightly deflated Type I errors and it improves for large N. Thus, for small N, we suggest to use the tuning parameter γ close to 0.5. For large pair of (p, m), we have deflated Type I errors. Type I errors are sketched in Figure 2.4.

Table 2.7 Type I errors of both open- and closed-end procedures for SCAD penalized regression for various values of γ and the nominal significance level $\alpha = 0.05$

		Closed-end			Open-end		
(p, m)	N/γ	0	0.25	0.45	0	0.25	0.45
(100,75)	2	0.041	0.046	0.040	0.012	0.022	0.035
	4	0.043	0.045	0.038	0.025	0.034	0.034
	10	0.037	0.038	0.035	0.028	0.033	0.033
	20	0.044	0.043	0.034	0.036	0.039	0.033
	50	0.048	0.047	0.046	0.047	0.046	0.046
(200,100)	2	0.037	0.038	0.034	0.008	0.018	0.027
	4	0.031	0.032	0.027	0.016	0.022	0.024
	10	0.036	0.033	0.027	0.026	0.028	0.026
	20	0.035	0.032	0.032	0.027	0.029	0.032
	50	0.039	0.042	0.034	0.038	0.041	0.034
(300,200)	2	0.028	0.030	0.030	0.008	0.016	0.024
	4	0.024	0.027	0.022	0.012	0.016	0.020
	10	0.025	0.026	0.022	0.019	0.022	0.020
	20	0.022	0.019	0.026	0.017	0.018	0.025
	50	0.032	0.028	0.024	0.030	0.027	0.024

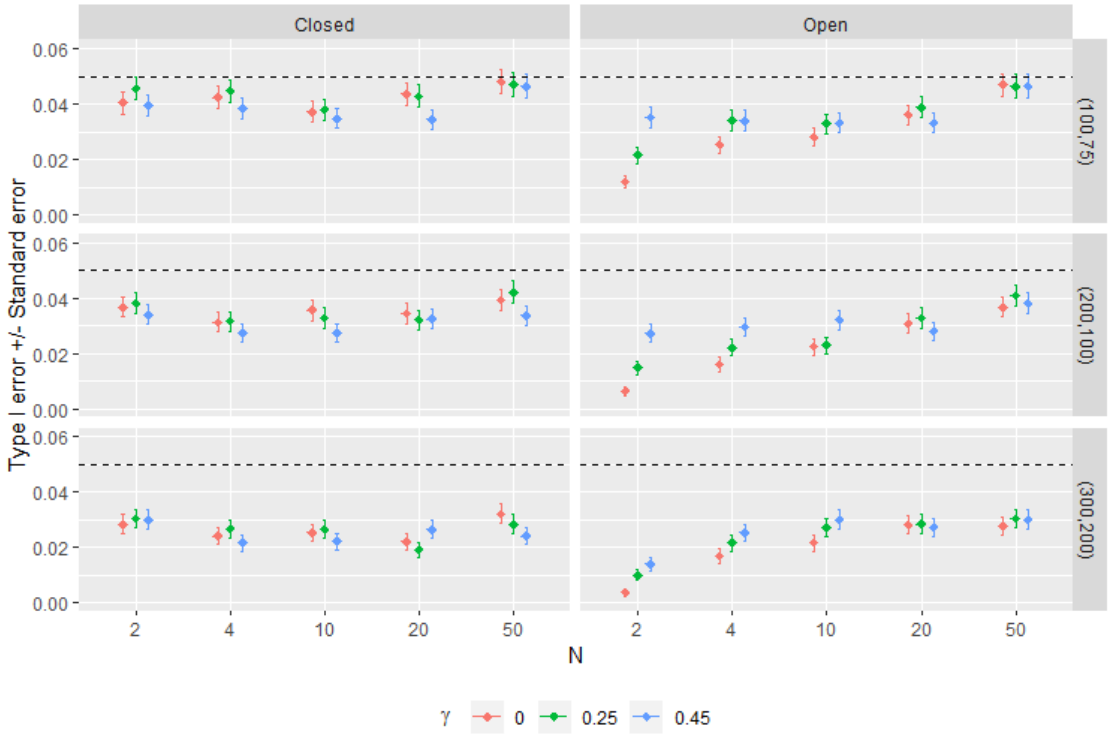

Figure 2.4 Type I error comparison for open- and closed-end procedures for various pairs of (p, m)

The power comparisons of the closed-end procedure is given in Table 2.8. We monitor the process until $9m$ observations. Figure 2.5 compares the power for the closed-end procedure. For large historical sample size m, the power is approximately equal to 1 regardless of α level. The five number summary of the detection time for both open- and closed-end procedure are given in Tables 2.9-2.10. The results are based on 2500 iterations. As we mentioned earlier, the larger γ value detect the change immediately when the change occurs shortly after the historical sample size m.

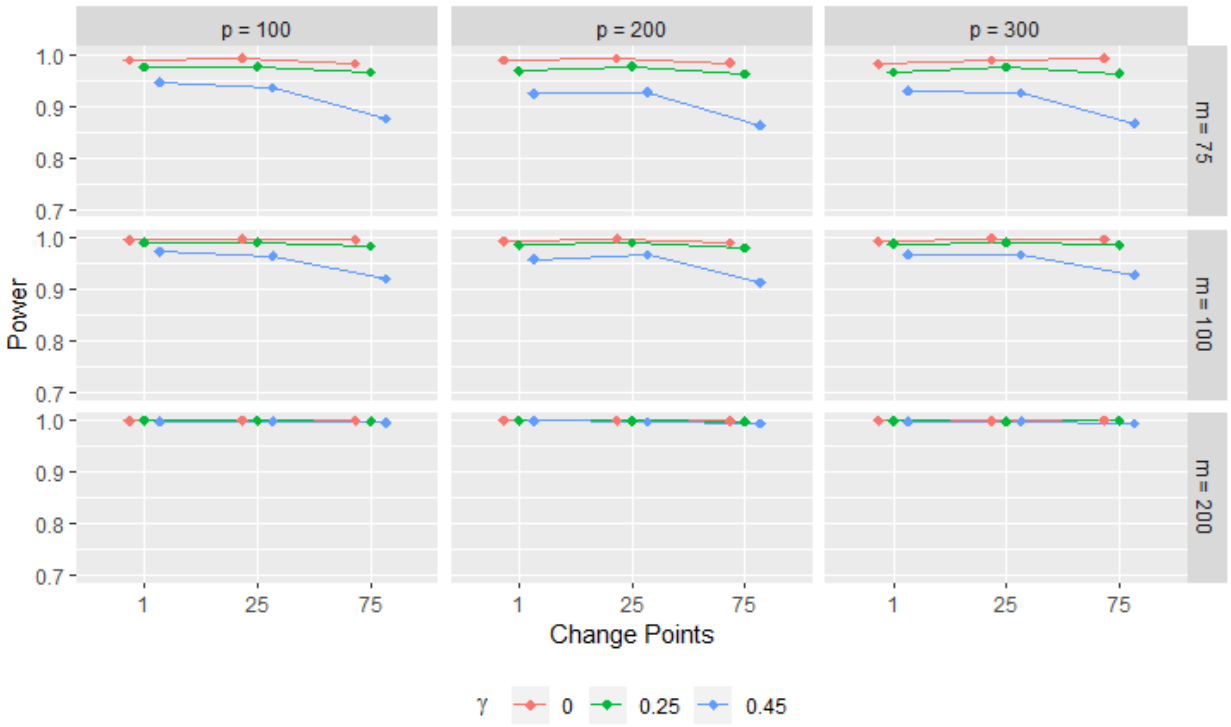

Figure 2.5 Power comparison for closed-end procedures with $\alpha = 0.05$

Figure 2.6 compares the estimated densities of the stopping time for various pair of (p, m) compare the estimated densities of the stopping time for various pair of (p, m) and the change-point locations. We observe variation in density plots, but decreases due to the large historical sample size.

Table 2.8 Power comparison for closed end procedure $k^* \in \{1, 25, 75\}$, $\alpha \in \{0.025, 0.05, 0.1\}$, $\gamma \in \{0, 0.25, 0.45\}$ and $p \in \{100, 200, 300\}$.

					$p = 100$					
		$k^* = 1$			$k^* = 25$			$k^* = 75$		
γ	α/m	75	100	200	75	100	200	75	100	200
0.00	0.025	0.962	0.980	0.999	0.935	0.970	0.999	0.875	0.938	0.996
	0.050	0.990	0.996	1	0.994	0.998	1	0.983	0.996	1
	0.100	1	1	1	1	1	1	1	1	1
0.25	0.025	0.941	0.969	0.999	0.886	0.946	0.997	0.779	0.875	0.993
	0.050	0.977	0.992	1	0.978	0.992	1	0.967	0.984	0.999
	0.100	0.998	0.998	1	0.999	1	1	1	1	1
0.45	0.025	0.874	0.936	0.999	0.768	0.879	0.995	0.599	0.727	0.976
	0.050	0.948	0.973	0.999	0.936	0.965	0.999	0.876	0.920	0.997
	0.100	0.989	0.996	1	0.995	0.998	1	0.990	0.997	1
					$p = 200$					
0.00	0.025	0.943	0.969	1	0.925	0.964	1	0.866	0.932	0.994
	0.050	0.990	0.993	1	0.993	0.998	1	0.984	0.990	1
	0.100	0.998	1	1	1	1	1	1	1	1
0.25	0.025	0.911	0.954	1	0.876	0.935	0.998	0.780	0.866	0.990
	0.050	0.969	0.986	1	0.978	0.991	1	0.963	0.981	0.999
	0.100	0.996	0.998	1	1	1	1	1	1	1
0.45	0.025	0.842	0.910	0.998	0.765	0.870	0.993	0.610	0.721	0.977
	0.050	0.925	0.958	1	0.928	0.967	0.998	0.864	0.913	0.994
	0.100	0.985	0.996	1	0.994	0.998	1	0.989	0.995	1
					$p = 300$					
0.00	0.025	0.946	0.976	1	0.921	0.966	0.999	0.887	0.944	0.999
	0.050	0.982	0.993	1	0.990	0.998	1	0.994	0.997	1
	0.100	1	1	1	1	1	1	1	1	1
0.25	0.025	0.915	0.962	0.999	0.886	0.932	0.999	0.785	0.875	0.992
	0.050	0.967	0.988	1	0.976	0.992	0.999	0.964	0.986	1
	0.100	0.998	1	1	1	1	1	1	1	1
0.45	0.025	0.859	0.928	0.996	0.793	0.869	0.996	0.624	0.746	0.968
	0.050	0.931	0.968	0.999	0.926	0.967	0.999	0.868	0.928	0.994
	0.100	0.985	0.994	1	0.994	0.998	1	0.988	0.999	1

Table 2.9 Summary statistics for the detection time for the open-end procedure at various change-point locations with $\gamma \in \{0, 0.25, 0.45\}$ and level $\alpha = 0.05$

	$(p, m) \rightarrow$	$(100, 75)$			$(200, 100)$			$(300, 200)$		
γ	Summary/k^*	1	25	75	1	50	100	1	100	200
0.00	min	2	26	44	2	33	33	3	101	151
	Q1	7	33	90	8	62	118	14	124	235
	Med	11	42	106	14	74	138	24	148	277
	Q3	20	58	139	26	99	182	45	201	369
	max	526	672	671	819	874	839	1335	1632	1658
0.25	min	2	10	10	2	15	15	2	70	70
	Q1	3	30	86	3	58	114	3	115	228
	Med	4	35	99	4	65	130	6	131	260
	Q3	7	45	127	8	88	167	10	165	340
	max	354	522	664	251	761	897	166	1433	1653
0.45	min	1	1	1	2	1	1	2	2	2
	Q1	2	29	86	2	56	113	2	112	227
	Med	2	33	99	2	63	130	3	127	260
	Q3	4	41	130	4	78	169	4	156	343
	max	72	525	665	37	685	893	32	1516	1789

Table 2.10 Summary statistics for the detection time for the closed-end procedure at various change-point locations with $\gamma \in \{0, 0.25, 0.45\}$ and level $\alpha = 0.05$

	$(p, m) \rightarrow$	$(100, 75)$			$(200, 100)$			$(300, 200)$		
γ	Summary/k^*	1	25	75	1	50	100	1	100	200
0.00	min	2	26	42	2	30	30	3	101	135
	Q1	6	32	88	7	61	116	12	121	231
	Med	10	40	102	12	71	134	22	142	268
	Q3	18	54	132	23	94	173	41	187	350
	max	525	608	670	874	870	890	1167	1481	1782
0.25	min	2	6	6	2	15	15	2	41	41
	Q1	3	29	85	3	57	113	3	114	226
	Med	4	34	97	4	64	128	5	129	255
	Q3	7	44	124	7	80	162	10	158	326
	max	353	521	671	250	874	876	166	1385	1784
0.45	min	1	1	1	1	1	1	2	2	2
	Q1	2	29	86	2	56	113	2	112	227
	Med	2	33	99	2	63	129	3	126	258
	Q3	4	41	128	4	78	167	4	155	341
	max	43	525	664	37	685	893	32	1472	1789

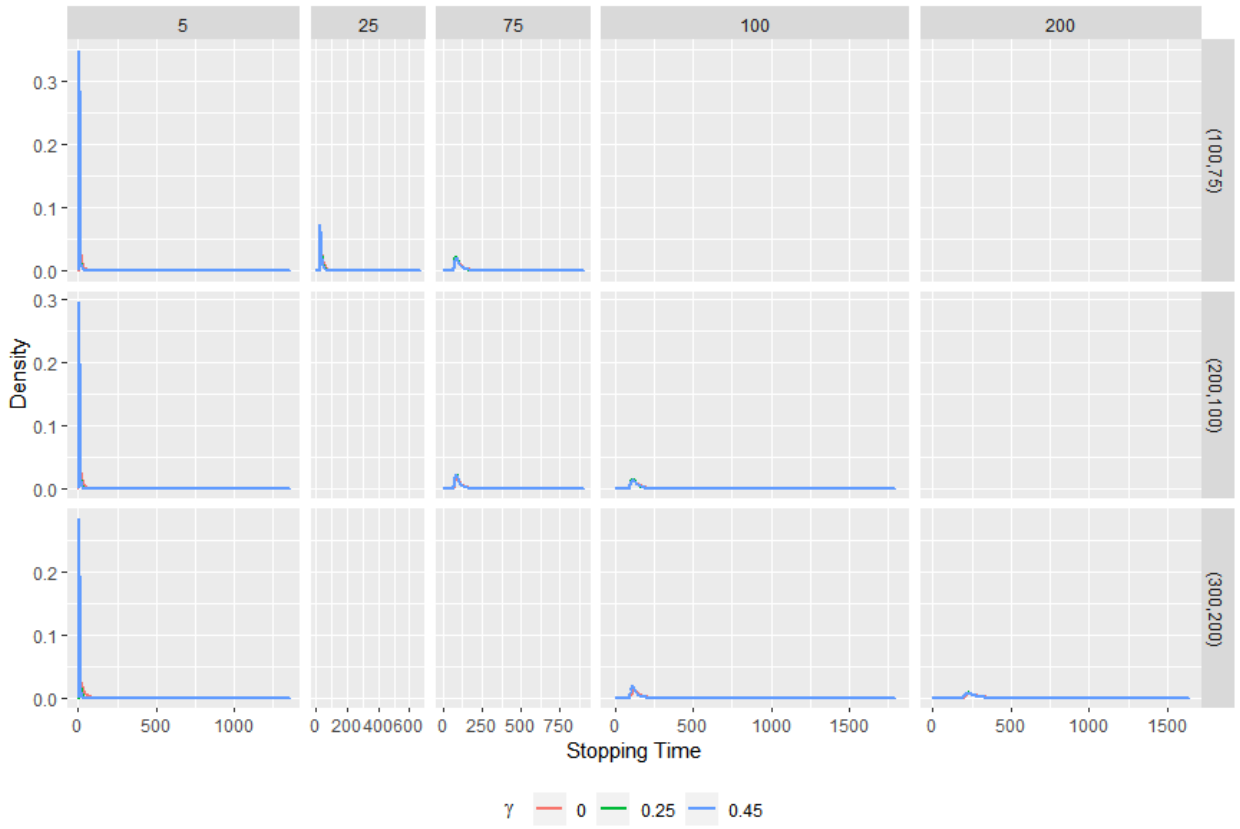

Figure 2.6 Estimated densities of the stopping time at various change points and various pair of (p, m)

2.4 Application

In this section, we apply the proposed method to a real data set. We consider the gene expression in the mammalian eye data. This dataset is originally described in Scheetz, Kim, Swiderski, Philp, Braun, Knudtson, Dorrance, DiBon, Huang, Casavant, Sheffield, and Stone (2006). The laboratory rats were examined to learn about gene expression and regulation in the mammalian eye. Inbred rat strains were crossed and tissue extracted from the eyes of 120 rats from the F2 generation. There are $n = 120$ observations and $p = 18,976$ explanatory variables. The outcome variable y is Gene expression measurement for Trim32 and the explanatory variables are the gene expression measurements for remaining genes. The outcome variable y is graphed in Figure 2.7.

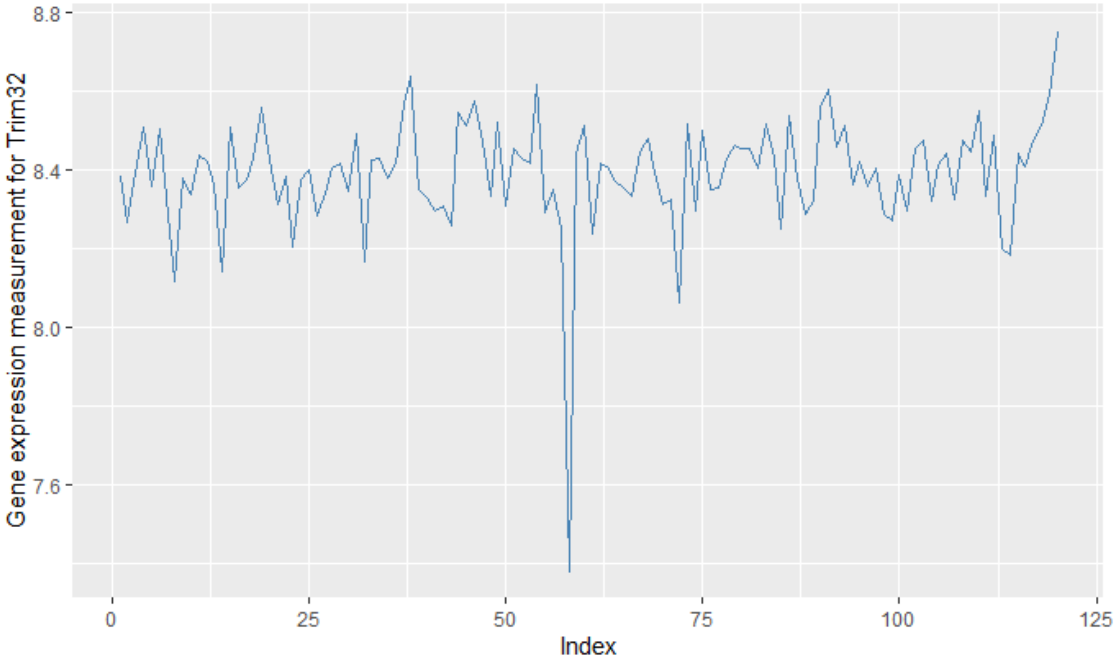

Figure 2.7 The outcome of the gene expression measurement for Trim32 in the mammalian eye data

According to the graph, there is no obvious jump in the data in the first 50 observations. Therefore, we consider the first 50 observations as the historical data. The proposed method is applied for the control parameter value $\gamma = 0.45$ with $\alpha = 0.05$. The first change point detected after the historical sample size $m = 50$ for a given control parameter value 0.45 is 8. That is, there is a change of gene expression of 58th rat comparing to the first 57 rats. We also consider the change point detection with log-likelihood method by assuming the normality of the data under the fixed sample size $n = 120$ situation. The result confirms the change occurring at 58th observation. With the binary segmentation method, there is only one change in the data. Comparing to the change point detection with the fixed sample size, the advantage of the sequential change point detection method is that only fewer samples are needed to make decision. In this application, our method only requires 50 observations and our monitoring process stops after 58 samples. The traditional log-likelihood method, however, needs entire observations $(n = 120)$ to estimate the change location. The change-point location corresponding to $\gamma = 0.45$ is graphed in Figure 2.8.

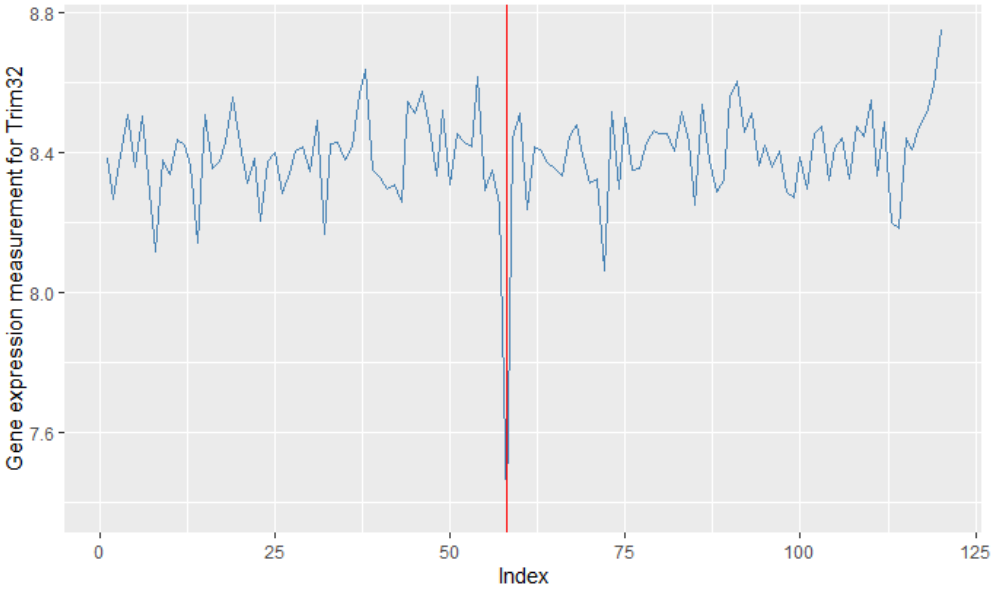

Figure 2.8 The first change point detection in the mammalian eye data

2.5 Conclusion

In this chapter, we proposed test statistics to monitor the structural change for high-dimensional data using SCAD penalized regression model for both open- and closed-end procedures. The asymptotic critical values for both monitoring process were provided. The Type I error probability of the open-end procedure is smaller than the closed-end procedure. The Type I error probability improves when N increases. However, the Type I error decreases when the historical sample size m increases. The power of the test is generally high and its only affected by the size of a test. We compute the stopping time for both open- and closed-end procedures. As similar to the Horváth et al. (2004) conclusion, if a change occurs immediately after historical sample size, we recommend a larger γ value close to 0.5 for high-dimensional data. In comparison to an open-end procedure, the monitoring process based on the closed-end procedure is usually quicker. We estimated the density of the stopping time for various cases. We apply the proposed detection procedure to analyze gene expression in the mammalian eye data to locate change point sequentially.

CHAPTER 3 SEQUENTIAL CHANGE POINT DETECTION FOR HIGH-DIMENSIONAL DATA USING PENALIZED QUANTILE REGRESSION

3.1 Introduction

In recent years, quantile regression has been widely used in many areas due to its appealing properties contrast to the traditional ordinary least square (OLS) regression model. The quantile regression method was introduced by Koenker and Bassett (1978) as an alternative to the least square regression. This is considered as an extension of least absolute deviation (LAD) regression or median regression. Unlike the least squares, quantile regression has been designed to model the changes in the conditional quantiles of the response variable with respect to the changes in the covariates. The OLS model examines the importance of the predictor X by modeling the conditional expectations of the response variable Y given X. The quantile regression produces much more information about the conditional response distribution and provides more robust analysis of data. Unlike the OLS, the quantile regression estimates are not sensitive to outliers. Therefore, the quantile regression can be used when the distribution of random errors is heavy-tailed, or when there are outliers in samples.

Change point detection for quantile regression has been extensively studied, see for example Bai (1996) proposed tests to detect changes in regression parameters as well as changes in variance. Additionally, it can be used to detect error heterogeneity in the data. Furno (2007) studied a likelihood ratio test based on quantile regressions. Qu (2008) proposed different test statistics for structural change occurring in a pre-specified quantile or across quantiles. Lagrange multiplier test for structural breaks in quantile regressions was proposed by Furno (2012). Wang and He (2007) proposed a test for detecting differences in certain quantiles of the intensity distributions. Aue, Cheung, Lee, and Zhong (2014) proposed a new methodology to simultaneously (or separately) detect break points, conduct variable selection, and estimate parameters in QR models. Zhang, Wang, and Zhu (2014) developed a new procedure for testing change points due to a covariate

threshold in regression quantiles. Their proposed test was based on the CUSUM of the subgradient of the quantile objective function and required fitting the model only under the null hypothesis. Although, very few studies examines the use of quantile regression in sequential change point detection. For instance, Zhou et al. (2015) developed a method for sequential detection of structural changes in linear quantile regression models. They established asymptotic properties of the test statistics. Ciuperca (2017) proposed the test statistic for sequential change point detection in a nonlinear quantile model.

In real-world scenarios, we often deal with large number of explanatory variables. For example, genomics, finance and healthcare data have large number of explanatory variables for each observation. In this chapter, we study the sequential change point method for quantile regression in high-dimensional covariates. Our approach enables us to analyze the conditional distribution of the response variable at different quantile level with large explanatory variables. To the best of our knowledge, there are no previous studies investigating the use of non-convex penalized quantile regression model in sequential change point analysis.

This chapter is organized as follows. In Section 3.2, the detection procedures based on SPQR and P-SPQR are proposed to detect changes sequentially under high dimensional scenarios. Corresponding asymptotic results are established. Simulations with various settings are conducted in Section 3.3 to investigate the performance of the proposed methods. The proposed P-SPQR method is applied to a breast cancer gene expression data to illustrate the detection and estimation process in Section 3.4. Some discussion is provided in Section 3.5.

3.2 Methodology

Let $\tau \in (0,1)$ be fixed and known quantile of interest. Suppose we have a random sample $\{Y_i, x_{i1}, \ldots, x_{ip}\}, i = 1, \ldots, m$ and a vector of independent identically distributed errors $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_m)$. Consider the model,

$$Y_i = \mathrm{x}_i^\top \beta + \mathcal{E}_i, \quad i = 1, \ldots, m, \tag{3.2.1}$$

where $\mathrm{x}_i = (x_{i1}, \ldots, x_{ip})^\top,\ i = 1, \ldots, m$ and $\beta = (\beta_{01}, \ldots, \beta_{0p})$ is the vector of unknown quantile regression parameters at the τth quantile level. Let $X = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_p)^\top$ be the $m \times p$ matrix of covariates, where $\mathrm{x}_1^\top, \ldots, \mathrm{x}_m^\top$ are the rows of X and $X = (X_1, \ldots, X_p)$ where $X_1, \ldots, X_p$ are the columns of X. The $\mathcal{E}_i$ is the error term satisfying $P(\mathcal{E}_i < 0|\mathrm{x}_i) = \tau$ for $i = 1, \ldots, m$. The model (3.2.1) can be expressed in a similar manner by specifying the τth conditional quantile as

$$Q_y(\tau|X_i) = \mathrm{x}_i^\top \beta. \tag{3.2.2}$$

The quantile coefficient β_τ can be estimated by,

$$\hat{\beta}_\tau = \underset{\beta \in \mathbb{R}^P}{\arg\min} \sum_{i=1}^{m} \rho_\tau(Y_i - \mathrm{x}_i^\top \beta), \tag{3.2.3}$$

where

$$\rho_\tau(u) = u(\tau - I(u < 0)), \tag{3.2.4}$$

is the quantile loss function introduced by Koenker and Bassett (1978). $I(\cdot)$ is the indicator function. When $\tau = 1/2$, it corresponds to the median regression. However, too many explanatory variables in the model may cause the problem of over-fitting. To remedy this issue, one can consider using the penalized quantile regression estimator in (3.2.3) as suggested in Koenker, Ng, and Portnoy (1994) and Koenker (2004).

3.2.1 SCAD Penalized Qunatile Regression (SPQR)

A number of penalty functions have been proposed in the literature, including Lasso ℓ_1 penalty by Tibshirani (1996), ℓ_2 penalty used in ridge regression by Hoerl and Kennard (1970). The Lasso has some rich properties which include shrinkage of the coefficients towards 0 for sufficiently large tuning parameter λ. However, the Lasso tends to produce biased estimates for large coefficients. The studies by Knight and Fu (2000), Fan and Li (2001) and Zou (2006) revealed that the variable selection in Lasso is consistent under certain conditions, but not in general. Thus, the Lasso does not possess the oracle property.

To overcome this issue, Fan and Li (2001) introduced a non-convex penalty called smoothly clipped absolute deviation (SCAD) and they suggested that it can be used for robust methods, such as median regression. The SCAD corresponds to a quadratic spline function with knots at λ and $a\lambda$. As mentioned in Fan and Li (2001), the SCAD penalty function satisfies three requirements for variable selection, including asymptotic unbiasednes, sparsity and continuity of the estimated parameters. It can estimate the zero coefficients as exactly zero with the probability approaching one. Regularized quantile regression with fixed p was studied by Zou and Yuan (2008), Wu and Liu (2009) and Kai, Li, and Zou (2011). Wu and Liu (2009) investigated the non-convex penalty for penalized quantile regression, including SCAD penalty for variable selection and showed that the SPQR satisfies the oracle property. Further, the oracle property of non-convex (SCAD and MCP) penalized linear quantile regression are established by Wang, Wu, and Li (2012) under high-dimensional settings. The SPQR model given as,

$$Q(\beta, \tau) = \sum_{i=1}^{m} \rho_\tau\big(Y_i - \mathrm{x}_i^\top \beta\big)^2 + \sum_{j=1}^{p} p_{\lambda_m}(|\beta_j|). \tag{3.2.5}$$

The SPQR solves the following minimization problem,

$$\hat{\beta}_m^\tau = \underset{\beta \in \mathbb{R}^P}{\arg\min} \left\{ \sum_{i=1}^{m} \rho_\tau\big(Y_i - \mathrm{x}_i^\top \beta\big)^2 + \sum_{j=1}^{p} p_{\lambda_m}(|\beta_j|) \right\}, \tag{3.2.6}$$

where $p_{\lambda_m}(\cdot)$ is the penalty function with tuning parameter $\lambda_m (\geq 0)$. The first derivative of the SCAD penalty function for some $a > 2$ and $\beta > 0$ is given as follows.

$$p'_{\lambda_m}(\beta) = \lambda_m \{ I(\beta \leq \lambda_m) + \frac{(a\lambda_m - \beta)_+}{(a-1)\lambda_m} I(\beta > \lambda_m) \}. \tag{3.2.7}$$

There are two unknown parameters λ_m and a. In practice, the best pair (λ_m, a) can be obtained by using two dimensional grids search, for example, cross validation method. Fan and Li (2001) suggested $a = 3.7$ is a good choice for various problems. In this research a is set to 3.7 to reduce the computational burden. The tuning parameter λ_m in the penalty function controls the

amount of shrinkage. The larger the value of λ_m, the greater the amount of shrinkage. Like all other penalized regression procedure, the performance of the penalized quantile regression depends on the selection of a tuning parameter λ_m. The tuning parameter selection methods are widely studied. Classical methods including Mallow's C_p (Mallows (1973)), Akaike information criterion (AIC; Akaike (1974)), Bayesian information criterion (BIC; Schwarz (1978)), cross-validation, and generalized cross-validation (Golub, Heath, and Wahba (1979)) have been used for model selection. In this research, the tuning parameter λ_m is selected using cross validation. Let $\beta_0 = (\beta_{0,1}, \beta_{0,2}, \ldots, \beta_{0,p})$ be the true parameter value and it is assumed to be sparse. Let $\mathbb{S}_0 = \{\beta_{0,j} \neq 0 : j = 1, \ldots, p\}$ be the index set of the nonzero coefficients for the true parameter, where $\beta_{0,j}$ is the jth component of the parameter vector β_0. Without loss of generality, we assume that the first q regression coefficients are nonzero and the remaining $(p - q)$ regression coefficients are 0. We denote the SCAD penalized quantile estimate by $\hat{\beta}_m^\tau$. Let $\mathbb{S}_* = \{\hat{\beta}_{m,j}^\tau \neq 0 : j = 1, \ldots, p\}$ be the index set of the SPQR estimator calculated using the historical sample size m, where $\hat{\beta}_{m,j}^\tau$ is the jth element of the SPQR estimator $\hat{\beta}_m^\tau$.

3.2.2 Asymptotic Properties

In this section, we establish the asymptotic properties of the proposed test statistic. We re-write $\mathrm{x}_i^\top = (\mathrm{z}_i^\top, \mathrm{w}_i^\top)$, where $\mathrm{z}_i = (x_{i1}, \ldots, x_{iq})^\top$ and $\mathrm{w}_i = (x_{iq+1}, \ldots, x_{ip})^\top$. We take into account the situation where the covariates are fixed. To prove the asymptotic properties, we impose following conditions.

C1. The model errors $\mathcal{E}_i$ are independent and identically distributed. Let $f(\cdot)$ and $F(\cdot)$ be the density function and the distribution function of $\mathcal{E}_i$ respectively and $f(\cdot)$ uniformly bounded away from zero.

C2. For all $n \in N$,

i. $\max_{1 \leq i \leq n} \|X_i\| = O_p(n^{1/4}/\sqrt{\log n})$,

ii. $\lim_{n \longrightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \|X_i\|^4 < \infty$ a.s, where $\| \cdot \|$ is the Euclidean norm.

C3. For all $n \in N$, there exists a positive definite matrix $\mathbb{C}$ such that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top = \mathbb{C}$.

C4. There exists a positive constant $L < \infty$ such that $\max_{1\le i\le n, 1\le j\le p} |x_{ij}| \le L$, for all $n \in N$.

C5. There exist positive constants $K_1 < K_2$ such that $K_1 \le \lambda_{\min}\big(\frac{1}{n}X_{\mathbb{S}_0}^\top X_{\mathbb{S}_0}\big) \le \lambda_{\max}\big(\frac{1}{n}X_{\mathbb{S}_0}^\top X_{\mathbb{S}_0}\big) \le K_2$ where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of $\frac{1}{n}X_{\mathbb{S}_0}^\top X_{\mathbb{S}_0}$ respectively. We also assumed that $\max_{1\le i\le n} \|\mathrm{z}_i\| = O_p(q^{1/2})$.

C6. There exists a constant $b > 0$ such that

$$\frac{1}{m^s}\left\|\sum_{i=m+k_m^*+1}^{m+k_m^*+m^s} \mathrm{x}_{i,\mathbb{S}_0}\left\{F(0) - F(\mathrm{x}_{i,\mathbb{S}_1}^\top \beta_{1,\mathbb{S}_1} - \mathrm{x}_{i,\mathbb{S}_0}^\top \beta_{0,\mathbb{S}_0})\right\}\right\| > b,$$

where $k_m^* = O(m^s)$, with the constant s and for open-end procedure $s > 1$ and for closed-end procedure $0 \le s \le 1$.

The (C1), (C2) and (C3) are used in literature on high-dimensional quantile regression models see for example Koenker (2005), Wang et al. (2012) and Zhou et al. (2015). The condition (C5) is on the true underlying model which is used in Wang et al. (2012). The condition (C6) is used in Ciuperca (2017).

3.2.3 Oracle Property

Suppose that the conditions are satisfied. If $\lambda_m \longrightarrow 0$ and $(q/m)^{1/2}\lambda_m \longrightarrow \infty$ as $m \longrightarrow \infty$, then the SPQR estimator $\hat{\beta}_m^\tau$ satisfy the oracle property,

i. Sparsity property for $\hat{\beta}_m^\tau$ happens in the historical data. Then,

$$P(\mathbb{S}_* = \mathbb{S}_0) = 1,$$

ii. Asymptotic normality: $(q/m)^{1/2}(\hat{\beta}_m^\tau - \beta_0)_{\mathbb{S}_0} \longrightarrow N\big(0, \tau(1-\tau)\mathbb{C}_{11}^{-1}/f(0)^2\big)$ in distribution as $m \longrightarrow \infty$ where $\mathbb{C}_{11}$ is top-left $q \times q$ matrix of $\mathbb{C}$.

Suppose that conditions hold, by adopting the Theorem 2.4 in Wang et al. (2012), we can conclude that the SCAD penalized quantile regression satisfies the oracle requirements for variable selection such as asymptotic unbiasedness, sparsity and continuity of the penalized estimator.

3.2.4 Sequential Change Point Problem

We study the change point problem with known pre-change coefficients but an unknown post-change parameters. Suppose there exists a historical sample size m such that $\beta_1 = \cdots = \beta_m = \beta_0$. This is called as noncontamination assumption and used in Zhou et al. (2015) and Ciuperca (2017). Now the pre-change parameters are obtained using the historical sample data. Let T_m be the monitoring horizon. After the historical sample size of m, we are interested in monitoring the process sequentially. The regression model after the historical observations m is,

$$Y_i = \mathrm{x}_i^\top \beta + \mathcal{E}_i, \quad i = m+1, m+2, \ldots \tag{3.2.8}$$

At each time point i, our goal is to test whether we have the same model as the first m observations. Under the null hypothesis, there is no change in the parameters,

$$H_0 : \beta_i = \beta_0\,; \quad \text{for } i = m+1, m+2, \ldots$$

Under the alternative hypothesis, we consider at an unknown time point k the parameters changing from β_0 to β_1. There exists $k \geq 1$ such that,

$$H_1 : \begin{cases} \beta_i = \beta_0\,; & i = m+1, m+1, \ldots, m+k, \\ \beta_i = \beta_1\,; & i = m+k+1, \ldots, m+T_m \quad \text{and } \beta_0 \neq \beta_1, \end{cases}$$

where $\beta_1 = (\beta_{1,1}, \beta_{1,2}, \ldots, \beta_{1,p})$ and it is unknown. Let $\mathbb{S}_1 = \{\beta_{1,j} \neq 0 : j = 1, \ldots, p\}$ be the index set of the nonzero coefficients under alternative hypothesis. Following Horváth et al. (2004) and Zhou et al. (2015), our monitoring process can be defined based on the following CUSUM

type process,

$$S(m,k) = J_{m,\mathbb{S}_*}^{-1/2} \sum_{i=m+1}^{m+k} \mathrm{x}_{i,\mathbb{S}_*} \psi_\tau(Y_i - \mathrm{x}_{i,\mathbb{S}_*}^\top \hat{\beta}_{m,\mathbb{S}_*}^\tau), \quad k = 1, \ldots, T_m, \tag{3.2.9}$$

where $J_{m,\mathbb{S}_*} = \tau(1-\tau) D_{m,\mathbb{S}_*}$ with $D_{m,\mathbb{S}_*} = \frac{1}{m} \sum_{i=1}^{m} \mathrm{x}_{i,\mathbb{S}_*} \mathrm{x}_{i,\mathbb{S}_*}^\top$ and $\psi_\tau(u) = \tau - I(u < 0)$. The proposed CUSUM based test statistic for the monitoring process of the SCAD penalized quantile regression is given as,

$$\Omega(m,k,\gamma) = \frac{\|S(m,k)\|_\infty}{g(m,k,\gamma)}, \tag{3.2.10}$$

where $g(m,k,\gamma)$ is called the normalising function and defined as,

$$g(m,k,\gamma) = m^{1/2} \left(1 + \frac{k}{m}\right) \left(\frac{k}{k+m}\right)^\gamma. \tag{3.2.11}$$

The γ is called the control parameter. The monitoring process stops immediately for large control parameter $\gamma \in [0, 1/2)$. Stopping time of the monitoring process based on the open-end procedure (see, Zhou et al. (2015)) is defined as,

$$\Lambda(k) = \begin{cases} \inf\{k \geq 1; & S(m,k)/g(m,k,\gamma) \geq c_\alpha(\gamma)\}, \\ \infty & \text{for all } k = 1, 2, \ldots, \end{cases} \tag{3.2.12}$$

where $c_\alpha(\gamma)$ is the critical value. Under the open-end procedure, the monitoring process may continue to infinity if no change is detected. Suppose $T_m < \infty$ with $\lim_{m\to\infty} T_m/m = N(> 0)$ with the possibility $N = \infty$. Under the open-end procedure, the monitoring boundary $T_m = \infty$. The critical value $c_\alpha(\gamma)$ is satisfying, under the null hypothesis,

$$\lim_{m\to\infty} P(\Lambda(k) < \infty) = \alpha, \tag{3.2.13}$$

and under the alternative hypothesis,

$$\lim_{m\to\infty} P(\Lambda(k) < \infty) = 1. \tag{3.2.14}$$

Suppose $T_m < \infty$, with $T_m/m \longrightarrow N$ $(N > 0)$ is defined as closed-end procedure and the stopping time is,

$$\Lambda^*(k) = \begin{cases} \inf\{k \geq 1; & S(m,k)/g(m,k,\gamma) \geq c^*_\alpha(\gamma)\}, \\ T_m & \text{for all } k = 1, \ldots, T_m. \end{cases} \tag{3.2.15}$$

where $c^*_\alpha(\gamma)$ is the critical value satisfying, under the null hypothesis,

$$\lim_{m\to\infty} P(\Lambda^*(k) < \infty) = \alpha, \tag{3.2.16}$$

and under the alternative hypothesis,

$$\lim_{m\to\infty} P(\Lambda^*(k) < \infty) = 1. \tag{3.2.17}$$

Theorem 3.2.18. *Under the assumptions (C1) to (C6) and for a given constant value of* $\gamma \in [0, 1/2)$*, if the null hypothesis holds, we have*

1. *for the open-end procedure,*

$$\lim_{m\to\infty} P\bigg(\sup_{1\leq k\leq T_m} \frac{S(m,k)}{g(m,k,\gamma)} \leq c^*_\alpha(\gamma)\bigg) = P\bigg(\sup_{1\leq t\leq 1} \frac{\|W(t)\|_\infty}{t^\gamma} \leq c^*_\alpha(\gamma)\bigg).$$

2. *for the closed-end procedure,*

$$\lim_{m\to\infty} P\bigg(\sup_{1\leq k\leq T_m} \frac{S(m,k)}{g(m,k,\gamma)} \leq c_\alpha(\gamma)\bigg) = P\bigg(\sup_{1\leq t\leq N/(N+1)} \frac{\|W(t)\|_\infty}{t^\gamma} \leq c_\alpha(\gamma)\bigg).$$

where $\{W(t), 0 \leq t < \infty\}$ *denotes a p-dimensional Weiner process,* $\alpha \in (0, 1)$ *and the control parameter* $0 \leq \gamma < 1/2$*.*

We obtain asymptotic critical values by simulation. The Weiner process $\|W(t)\|_\infty/t^\gamma$ is approximated by $\|M^{-1/2}\sum_{i=1}^{tM} e_i\|_\infty/t^\gamma$ where $e_i \sim N(0,1)$ and M is a grid of 10000. In each iteration, the maximum value of the process for both open- and closed-end procedures obtained over $t \in (0,1)$ and $t \in (0, N/(N+1))$ respectively.

Theorem 3.2.19. *Under the assumptions and for a given constant value of $\gamma \in [0, 1/2)$, if the alternative hypothesis holds, we have,*

$$\sup_{1\le k\le T_m} \frac{S(m,k)}{g(m,k,\gamma)} \longrightarrow \infty \quad \text{as} \quad m \longrightarrow \infty.$$

3.2.5 Post - SCAD Penalized Quantile Regression (P-SPQR)

In high-dimensional settings, the quantile regression model with SCAD penalized estimator is asymptotically unbiased. However, the direct use of this theorem will induce bias. To reduce the bias in the estimator, Belloni and Chernozhukov (2013) suggested so-called post-lasso estimator. They showed that the OLS post-lasso performs at least as well as the lasso under mild additional assumptions. Huang and Xie (2007) showed that under appropriate conditions, the SPQR is consistent for variable selection. As discuss earlier, the SPQR enjoys oracle property thus, the P-SPQR estimator becomes the oracle estimator as well. In order to improve the monitoring method, a modified test statistic based on the P-SPQR estimator is proposed. The variable selection procedure plays an important role in high-dimensional data set. In the first step, we select the important variables by regularizing quantile regression with a SCAD penalty function. Using the significant predictors, let $\hat{\beta}^*_{\tau,m}$ be the quantile coefficient estimator based on the historical data and can be obtained by minimizing,

$$\hat{\beta}^*_{\tau,m} = \arg\min_{\beta\in\mathbb{R}^{P'}} \sum_{i=1}^{m} \rho_\tau(Y_i - \mathrm{x}^\top_{i,\mathbb{S}^*}\beta), \tag{3.2.20}$$

where $\rho_\tau(\cdot)$ define in (3.2.4) and p' is the cardinality of the set $\mathbb{S}^*$. Following Horváth et al. (2004) and Zhou et al. (2015) , the subgradient-based CUSUM-type process,

$$S^*(m,k) = J_{m,\mathbb{S}^*}^{-1/2} \sum_{i=m+1}^{m+k} \mathrm{x}_{i,\mathbb{S}^*} \psi_\tau(Y_i - \mathrm{x}_{i,\mathbb{S}^*}^\top \hat{\beta}^*_{\tau,m}), \quad k = 1, \ldots, T_m, \tag{3.2.21}$$

where $J_{m,\mathbb{S}^*} = \tau(1-\tau) D_{m,\mathbb{S}^*}$ with $D_{m,\mathbb{S}^*} = \frac{1}{m} \sum_{i=1}^m \mathrm{x}_{i,\mathbb{S}^*} \mathrm{x}_{i,\mathbb{S}^*}^\top$ and $\psi_\tau(u) = \tau - I(u < 0)$. The modified CUSUM based test statistic for the monitoring process in P-SPQR model is given as,

$$\Omega^*(m,k,\gamma) = \frac{\|S^*(m,k)\|_\infty}{g(m,k,\gamma)}. \tag{3.2.22}$$

Stopping time for the open-end procedure,

$$\Lambda_{\text{modified}}(k) = \begin{cases} \inf\{k \geq 1; & S^*(m,k)/g(m,k,\gamma) \geq c_\alpha(\gamma), \}, \\ \infty & \text{for all } k = 1, 2, \ldots, \end{cases} \tag{3.2.23}$$

and for the closed-end procedure,

$$\Lambda^*_{\text{modified}}(k) = \begin{cases} \inf\{k \geq 1; & S^*(m,k)/g(m,k,\gamma) \geq c^*_\alpha(\gamma), \}, \\ T_m & \text{for all } k = 1, \ldots, T_m, \end{cases} \tag{3.2.24}$$

where $c_\alpha(\gamma)$ and $c^*_\alpha(\gamma)$ are asymptotic the critical values for open- and closed-end procedures respectively. For a given constant $\gamma \in [0, 1/2)$, the $g(m,k,\gamma)$ is called the normalising function defined in (3.2.11). Further, under the null hypothesis,

$$\lim_{m\to\infty} P(\Lambda_{\text{modified}}(k) < \infty) = \alpha \quad \text{and} \quad \lim_{m\to\infty} P(\Lambda^*_{\text{modified}}(k) < \infty) = \alpha, \tag{3.2.25}$$

and under the alternative hypothesis,

$$\lim_{m\to\infty} P(\Lambda_{\text{modified}}(k) < \infty) = 1 \quad \text{and} \quad \lim_{m\to\infty} P(\Lambda^*_{\text{modified}}(k) < \infty) = 1. \tag{3.2.26}$$

Under H_0, for the open- and closed-end procedure the test statistics given in (3.2.23) and (3.2.24) converges in distribution to $\sup_{1\leq t\leq 1}\frac{\|W(t)\|_\infty}{t^\gamma}$ and $\sup_{1\leq t\leq N/(N+1)}\frac{\|W(t)\|_\infty}{t^\gamma}$ respectively. Under H_1, the test statistics $\Omega^*(m,k,\gamma)$ converges in probability to ∞ as $m\longrightarrow\infty$.

3.2.6 Proofs of Theorems

Proof. Theorem 3.2.18:

Under the conditions, the convergence rate of the SCAD penalized quantile estimator is of order $(q/m)^{-1/2}$, see for example Wang et al. (2012). To establish the asymptotic properties of the test statistic defined in (3.2.10), we will consider for $u_m\in\mathbb{R}^{|\mathbb{S}_0|}$, $\|u_m\|\leq C$, the random vector,

$$\widetilde{R}_{m,k}(u_m)=\sum_{i=m+1}^{m+k}\boldsymbol{R_i}(u_m),$$

where

$$\begin{aligned}\boldsymbol{R_i}(u_m)=\mathrm{x}_{i,\mathbb{S}_0}&\Big\{\psi_\tau\big(Y_i-\mathrm{x}_{i,\mathbb{S}_*}^\top(\beta_{0,\mathbb{S}_*}+(q/m)^{-1/2}u_m)\big)-\psi_\tau(Y_i-\mathrm{x}_{i,\mathbb{S}_0}^\top\,\beta_{0,\mathbb{S}_0})\Big\}\\&+\Big\{\sum_{j=1}^{q}\big(p_{\lambda_m}(|\beta_{j0}+(q/m)^{-1/2}u_j|)-p_{\lambda_m}(|\beta_{j0}|)\big)\Big\}.\end{aligned}\tag{3.2.27}$$

For large m, Wu and Liu (2009) showed,

$$\Big\{\sum_{j=1}^{q}\big(p_{\lambda_m}(|\beta_{j0}+(q/m)^{-1/2}u_j|)-p_{\lambda_m}(|\beta_{j0}|)\big)\Big\}=0.$$

Thus, the $\boldsymbol{R_i}$ defined in (3.2.27) becomes,

$$\boldsymbol{R_i}(u_m)=\mathrm{x}_{i,\mathbb{S}_0}\Big\{\psi_\tau\big(Y_i-\mathrm{x}_{i,\mathbb{S}_*}^\top(\beta_{0,\mathbb{S}_*}+(q/m)^{-1/2}u_m)\big)-\psi_\tau(Y_i-\mathrm{x}_{i,\mathbb{S}_0}^\top\,\beta_{0,\mathbb{S}_0})\Big\}.\tag{3.2.28}$$

In particular, the vector u_m are deterministic and only the dimension of u_m is random, see Ciuperca (2017). Following Ciuperca (2017), for the proposed SPQR method, we consider the following situation. For fixed size $|\mathbb{S}_0|$, there exists u such that $\|u\|\leq C$. Following Ciuperca (2017), we

have,

$$\widetilde{R}_{m,k}(u) = \mathbb{E}\big(\widetilde{R}_{m,k}(u)\big) + O_P(r_{m,k}), \tag{3.2.29}$$

where $r_{m,k} = m^{-1/4}k^{1/2}(\log k)^{1/2}$. Then,

$$\widetilde{R}_{m,k}(u) = -\sum_{i=m+1}^{m+k} f(0)\mathrm{x}_{i,\mathbb{S}_0}\mathrm{x}_{i,\mathbb{S}_0}^{\top}(q/m)^{-1/2}u + O_P(r_{m,k}). \tag{3.2.30}$$

But we have,

$$\hat{\beta}^{\tau}_{m,\mathbb{S}_*} = \beta_{0,\mathbb{S}_0} + (q/m)^{1/2}J_{m,\mathbb{S}_0}\sum_{i=1}^{m}\mathrm{x}_{i,\mathbb{S}_0}\psi_{\tau}(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^{\top}\,\beta_{0,\mathbb{S}_0}) + o_p(m^{-1/2}).$$

Now considering $u = (q/m)^{1/2}(\hat{\beta}^{\tau}_{m,\mathbb{S}_*} - \beta_{0,\mathbb{S}_0})$ in (3.2.30), we get

$$\widetilde{R}_{m,k}\big((q/m)^{1/2}(\hat{\beta}^{\tau}_{m,\mathbb{S}_*} - \beta_{0,\mathbb{S}_0})\big) = -\sum_{i=m+1}^{m+k} f(0)\mathrm{x}_{i,\mathbb{S}_0}\mathrm{x}_{i,\mathbb{S}_0}^{\top}(\hat{\beta}^{\tau}_{m,\mathbb{S}_*} - \beta_{0,\mathbb{S}_0}) + O_P(r_{m,k}). \tag{3.2.31}$$

On the other hand,

$$\begin{aligned}
&\sum_{j=m+1}^{m+k}\mathrm{x}_{i,\mathbb{S}_0}\psi_{\tau}\big(Y_i - \mathrm{x}_{i,\mathbb{S}_*}^{\top}(\beta_{0,\mathbb{S}_*} + (q/m)^{-1/2}u)\big) - \sum_{j=m+1}^{m+k}\mathrm{x}_{i,\mathbb{S}_0}\psi_{\tau}(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^{\top}\beta_{0,\mathbb{S}_0})\\
&= \mathbb{E}\Big\{\sum_{j=m+1}^{m+k}\mathrm{x}_{i,\mathbb{S}_0}\psi_{\tau}\big(Y_i - \mathrm{x}_{i,\mathbb{S}_*}^{\top}(\beta_{0,\mathbb{S}_*} + (q/m)^{-1/2}u\big) - \sum_{j=m+1}^{m+k}\mathrm{x}_{i,\mathbb{S}_0}\psi_{\tau}(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^{\top}\beta_{0,\mathbb{S}_0})\Big\}\\
&\quad + O_p(r_{m,k})\text{ , note that }\mathrm{x}_{i,\mathbb{S}_0} \equiv \mathrm{z}_i\\
&= \sum_{j=m+1}^{m+k}\mathrm{x}_{i,\mathbb{S}_0}\Big\{\tau - P\big(y_i < \mathrm{x}_{i,\mathbb{S}_*}^{\top}(\beta_{0,\mathbb{S}_*} + (q/m)^{-1/2}u)\big) - \sum_{j=m+1}^{m+k}\mathrm{x}_{i,\mathbb{S}_0}\big(\tau - P(y_i < \mathrm{x}_{i,\mathbb{S}_0}^{\top}\beta_{0,\mathbb{S}_0})\big)\Big\}\\
&\quad + O_p(r_{m,k})\\
&= \sum_{j=m+1}^{m+k}\mathrm{x}_{i,\mathbb{S}_0}\Big\{F_y\big(\mathrm{x}_{i,\mathbb{S}_0}^{\top}\beta_{0,\mathbb{S}_0}\big) - F_y\big(\mathrm{x}_{i,\mathbb{S}_0}^{\top}(\beta_{0,\mathbb{S}_0} + (q/m)^{-1/2}u\big)\Big\} + O_p(r_{m,k})\\
&= -\sum_{j=m+1}^{m+k} f_y\big(a_i\big)\mathrm{x}_{i,\mathbb{S}_0}\mathrm{x}_{i,\mathbb{S}_0}^{\top}(q/m)^{-1/2}u + O_p(r_{m,k}),
\end{aligned}$$

where a_i is between $\big(\mathrm{x}_{i,\mathbb{S}_0}^\top \beta_{0,\mathbb{S}_0}\big)$ and $\big(\mathrm{x}_{i,\mathbb{S}_0}^\top (\beta_{0,\mathbb{S}_0} + (q/m)^{-1/2}u\big)$. By Wang et al. (2012), $(q/m)^{1/2}(\hat{\beta}^\tau_{m,\mathbb{S}_*} - \beta_{0,\mathbb{S}_0}) = O_p(1)$. Thus, by substituting $u = (q/m)^{1/2}(\hat{\beta}^\tau_{m,\mathbb{S}_*} - \beta_{0,\mathbb{S}_0})$ above and for large k using the conditions, we get,

$$\begin{aligned}
&\sum_{j=m+1}^{m+k} \mathrm{x}_{i,\mathbb{S}_0}\psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_*}^\top(\beta_{0,\mathbb{S}_*} + (q/m)^{-1/2}u)\big) - \sum_{j=m+1}^{m+k} \mathrm{x}_{i,\mathbb{S}_0}\psi_\tau(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^\top\beta_{0,\mathbb{S}_0}) \\
&\qquad = -\sum_{j=m+1}^{m+k} f_y(0)\mathrm{x}_{i,\mathbb{S}_0}\mathrm{x}_{i,\mathbb{S}_0}^\top(\hat{\beta}^\tau_{m,\mathbb{S}_*} - \beta_{0,\mathbb{S}_0}) + O_p(r_{m,k}) + O_p(km^{-1/2}) \qquad (3.2.32)\\
&\qquad = -\sum_{j=m+1}^{m+k} f_y(0)\mathrm{x}_{i,\mathbb{S}_0}\mathrm{x}_{i,\mathbb{S}_0}^\top(\hat{\beta}^\tau_{m,\mathbb{S}_*} - \beta_{0,\mathbb{S}_0}) + O_p(r_{m,k}) + O_p(km^{-1/2}).
\end{aligned}$$

Using (3.2.31) and (3.2.32), we obtain

$$\begin{aligned}
J_{m,\mathbb{S}_*}^{-1/2}\sum_{j=m+1}^{m+k} \mathrm{x}_{i,\mathbb{S}_*}\psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_*}^\top\hat{\beta}^\tau_{m,\mathbb{S}_*}\big) &= J_{m,\mathbb{S}_0}^{-1/2}\sum_{j=m+1}^{m+k} \mathrm{x}_{i,\mathbb{S}_0}\psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^\top\beta_{0,\mathbb{S}_0}\big) \\
&\quad - km^{-1/2}J_{m,\mathbb{S}_0}^{-1/2}\sum_{j=m+1}^{m+k} \mathrm{x}_{i,\mathbb{S}_0}\psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^\top\beta_{0,\mathbb{S}_0}\big) + O_p(r_{m,k}) \\
&\quad + O_p(km^{-1/2}).
\end{aligned} \qquad (3.2.33)$$

By K-M-T approximation (see, Komlós, Major, and Tusnády (1975, 1976)), for all m there exists two $p-$dimensional Wiener processes on $\{W_{1,m}, t \in [0,\infty)\}$ and $\{W_{2,m}, t \in [0,\infty)\}$ such that $m \longrightarrow \infty$,

$$\sup_{1\le k<\infty} k^{1/v}\left\| J_{m,\mathbb{S}_0}^{-1/2}\sum_{j=m+1}^{m+k} \mathrm{x}_{i,\mathbb{S}_0}\psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^\top\beta_{0,\mathbb{S}_0}\big) - W_{1,m}\left(\frac{k}{m}\right)\right\|_\infty = O_p(1),$$

and

$$\left\| J_{m,\mathbb{S}_0}^{-1/2}\sum_{j=1}^{m} \mathrm{x}_{i,\mathbb{S}_0}\psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^\top\beta_{0,\mathbb{S}_0}\big) - W_{2,m}\left(\frac{k}{m}\right)\right\|_\infty = O_p(m^{1/v}).$$

Next, following the similar steps of Horváth et al. (2004) of Theorem 2.1 and Ciuperca (2017) we complete the proof of Theorem 3.2.18. □

Proof. Theorem 3.2.19:

Let $k_m = k_m^* + m^s$. For the closed-end procedure, we consider case when $s = 1$. There exists a constant $C \in (0, \infty)$ such that,

$$\frac{\left\| J_{m,\mathbb{S}_0}^{-1/2} \sum_{i=m+1}^{m+k_m^*} \mathrm{x}_{i,\mathbb{S}_0} \psi_\tau(Y_i - \mathrm{x}_{i,\mathbb{S}_*}^\top \hat{\beta}_{m,\mathbb{S}_*}^\tau) \right\|_\infty}{g(m, k_m^*, \gamma)} \leq C, \tag{3.2.34}$$

with probability converging to 1, as $m \longrightarrow \infty$. Since the function $g(m, k, \gamma)$, is increasing in $k > 0$, we have

$$\frac{\left\| J_{m,\mathbb{S}_0}^{-1/2} \sum_{i=m+1}^{m+k_m^*} \mathrm{x}_{i,\mathbb{S}_0} \psi_\tau(Y_i - \mathrm{x}_{i,\mathbb{S}_*}^\top \hat{\beta}_{m,\mathbb{S}_*}^\tau) \right\|_\infty}{g(m, k_m, \gamma)} \leq C, \tag{3.2.35}$$

with probability converging to 1, as $m \longrightarrow \infty$. Consider the following random process, with $u \in \mathbb{R}^{|\mathbb{S}_*|}$, $\|u\| \leq C$,

$$\sum_{i=m+k_m^*+1}^{m+k_m} \boldsymbol{R_i}(u) = \sum_{i=m+k_m^*+1}^{m+k_m} \mathrm{x}_{i,\mathbb{S}_0} \Big\{ \psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_*}^\top (\beta_{0,\mathbb{S}_*} + (q/m)^{-1/2} u_m)\big) - \psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^\top \beta_{0,\mathbb{S}_0}\big) \Big\}.$$

Since $i = m + k_m^* + 1, \ldots, m + k_m$, the hypothesis H_1 is true, we have,

$$\sum_{i=m+k_m^*+1}^{m+k_m} \boldsymbol{R_i}(u) = \sum_{i=m+k_m^*+1}^{m+k_m} \mathrm{x}_{i,\mathbb{S}_0} \Big\{ \psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_*}^\top (\beta_{0,\mathbb{S}_*} + (q/m)^{-1/2} u_m)\big) - \psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^\top \beta_{1,\mathbb{S}_1}\big) \Big\}.$$

Since $\mathbb{E}\Big[\sum_{i=m+k_m^*+1}^{m+k_m} \mathrm{x}_{i,\mathbb{S}_0} \psi_\tau\big(Y_i - \mathrm{x}_{i,\mathbb{S}_0}^\top \beta_{0,\mathbb{S}_0}\big) \Big] = 0$ we have,

$$\mathbb{E}\Big[\sum_{i=m+k_m^*+1}^{m+k_m} \boldsymbol{R_i}(u) \Big] = \sum_{i=m+k_m^*+1}^{m+k_m} \mathrm{x}_{i,\mathbb{S}_0} \Big\{ F(0) - F(\mathrm{x}_{i,\mathbb{S}_1}^\top \beta_{1,\mathbb{S}_1} - \mathrm{x}_{i,\mathbb{S}_0}^\top (\beta_{0,\mathbb{S}_0} + (q/m)^{-1/2} u)) \Big\}.$$

Now by the condition (C6), there exists a constant $b > 0$, we get

$$\frac{1}{m} \left\| \sum_{i=m+k_m^*+1}^{m+k_m^*+m} \mathrm{x}_{i,\mathbb{S}_0} \Big\{ F(0) - F(\mathrm{x}_{i,\mathbb{S}_1}^\top \beta_{1,\mathbb{S}_1} - \mathrm{x}_{i,\mathbb{S}_0}^\top \beta_{0,\mathbb{S}_0}) \Big\} \right\| > b.$$

Since $f(\cdot)$ is bounded, and by conditions,

$$\begin{aligned}
&\left\| \sum_{i=m+k_m^*+1}^{m+k_m^*+m} \mathrm{x}_{i,\mathbb{S}_0} \left\{ F(0) - F(\mathrm{x}_{i,\mathbb{S}_1}^\top \beta_{1,\mathbb{S}_1} - \mathrm{x}_{i,\mathbb{S}_0}^\top (\beta_{0,\mathbb{S}_0} + (q/m)^{-1/2} u)) \right\} \right\| \\
&= \left\| \sum_{i=m+k_m^*+1}^{m+k_m^*+m} \mathrm{x}_{i,\mathbb{S}_0} \mathrm{x}_{i,\mathbb{S}_0}^\top (q/m)^{-1/2} f(b_i) \right\| \\
&\leq b(q/m)^{-1/2} \sum_{i=m+k_m^*+1}^{m+k_m^*+m} \max_{m+k_m^*+1 \leq i \leq m+k_m^*+m} \|X_i\| \\
&= o(k_m - k_m^*) = o(m),
\end{aligned}$$

where b_i is between 0 and $\left(\mathrm{x}_{i,\mathbb{S}_1}^\top \beta_{1,\mathbb{S}_1} - \mathrm{x}_{i,\mathbb{S}_0}^\top (\beta_{0,\mathbb{S}_0} + (q/m)^{-1/2} u)\right)$. Then following similar steps of Ciuperca (2017) complete the proof. Thus, we omit details. We can follow similar arguments to show that Theorem 3.2.19 still holds for $s = 1$. □

In order to proof the Theorems 3.2.18 & 3.2.19 for the P-SPQR we will follow the same strategy used in Ciuperca (2017). Thus, details are omitted to conserve space.

3.3 Simulation Study

In this section, we present the simulation study to demonstrate the good performance of our proposed method. The following two cases are considered in order to calculate the Type I error. In both cases, the number explanatory variables $p = 10$. All simulations are conducted using the R software.

- Case - I
  - In the first case, the true parameter vectors $\beta_0 \in \{1, 0, -1, 0, -15, 0, 0, 0, 0, 0\}$ and X_i for all $i \in \{1, \ldots, 10\}$ follows standard normal distribution $N(0, 1)$ and the model errors $\mathcal{E}_i \sim N(0, 1)$.
- Caese - II

- The true parameter vectors $\beta_0 \in \{1, 0, -1, 0, -15, 0, 0, -2, 0, 8\}$ and $X_i \sim$ Unif$(0, 2)$ for all $i \in \{1, \ldots, 10\}$. The model errors $\mathcal{E}_i \sim$ Cauchy$(0, 2)$.

Next we conduct the power analysis to illustrate the performance of the proposed test statistic. Under the null hypothesis, the true parameter vectors $\beta_0 \in \{-1, 0, 1, 8, 1, 0, 0, 0, -5, 0\}$ and under the alternative hypothesis, the parameter vector $\beta_1 \in \{0, -1, 0, 2, 0, 0, 1, 0, 0, -1\}$. We consider the two different distributions of the explanatory variables $X_1, X_2, \ldots, X_{10}$. Under H_0, X_i for all $i \in \{1, \ldots, 10\} \backslash \{3, 4, 5\}$ have standard normal distribution $N(0, 1)$ and $X_3 \sim N(2, 1), X_4 \sim N(4, 1)$ and $X_5 \sim N(5, 1)$. The second distribution for the ith explanatory variable is $X_i + 0.8$ for all $i \in \{1, \ldots, 10\}$. In both settings, the model errors $\mathcal{E}_i \sim N(0, 1)$. The second setting is used to compute the stopping time at different change point locations.

First, we evaluate the effect on the detection procedure of various control parameter values, considering $\gamma \in \{0, 0.25, 0.45\}$. Type I error for the open-and closed-end procedures for Case I at quantile level $\tau = 0.5$ are summarized in Table 3.1. The Table 3.1 results are summarized in Figure 3.1. Under open-end procedure, the P-SPQR method provides better Type I errors compare to the SPQR method. For the closed-end procedure, the SPQR method provides Type I errors close to nominal level. Under the closed-end procedure, the P-SPQR method provides slightly higher Type I errors for smaller N. We recommend larger N for small control parameter value γ.

In Case II, we consider the heavy tails of random errors. The results are summarized in Tables 3.2 - 3.3. For the open- and closed-end procedures and at $\tau \in \{0.5, 0.7\}$ with the historical sample size $m \in \{200, 400\}$ the SPQR method gives larger Type I error. Type I errors of the P-SPQR method close to the nominal level 0.05 for the closed-end procedure. Under open-end procedure the P-SPQR method produces deflated Type I errors for small control parameter value $\gamma = 0$. Thus, we suggest large N for smaller control parameter. Tables 3.2 - 3.3 are compared in Figure 3.2.

Table 3.1 Type I errors comparison for the open- and closed-end procedures of the SPQR and P-SPQR methods for various values of γ, the nominal significance level $\alpha = 0.05$ and $\tau = 0.5$ for Case I

			Closed-end			Open-end		
Method	m	N/γ	0	0.25	0.45	0	0.25	0.45
SPQR	75	2	0.044	0.050	0.046	0.008	0.020	0.040
		4	0.057	0.063	0.062	0.030	0.038	0.058
		6	0.047	0.051	0.060	0.030	0.039	0.056
		9	0.052	0.052	0.054	0.037	0.043	0.051
	100	2	0.052	0.053	0.050	0.010	0.026	0.040
		4	0.044	0.045	0.047	0.021	0.032	0.042
		6	0.038	0.038	0.038	0.022	0.027	0.037
		9	0.044	0.046	0.054	0.034	0.038	0.052
	150	2	0.040	0.038	0.046	0.008	0.021	0.038
		4	0.037	0.042	0.044	0.019	0.027	0.040
		6	0.043	0.044	0.041	0.028	0.035	0.040
		9	0.037	0.039	0.042	0.025	0.035	0.042
P-SPQR	75	2	0.058	0.060	0.052	0.012	0.030	0.043
		4	0.073	0.075	0.062	0.033	0.051	0.057
		6	0.070	0.065	0.056	0.043	0.054	0.052
		9	0.062	0.064	0.058	0.049	0.050	0.056
	100	2	0.064	0.065	0.050	0.016	0.034	0.045
		4	0.072	0.070	0.059	0.032	0.046	0.056
		6	0.067	0.064	0.054	0.040	0.048	0.052
		9	0.075	0.074	0.062	0.056	0.060	0.061
	150	2	0.058	0.062	0.058	0.015	0.029	0.049
		4	0.056	0.056	0.053	0.026	0.040	0.045
		6	0.068	0.068	0.048	0.042	0.051	0.045
		9	0.058	0.062	0.056	0.044	0.049	0.054

Table 3.2 Type I errors comparison for the open- and closed-end procedures of the SPQR and P-SPQR methods for various values of γ, $\alpha = 0.05$ and $\tau = 0.5$ for Case II

Method	m	N/γ	Closed-end			Open-end		
			0	0.25	0.45	0	0.25	0.45
SPQR	200	2	0.170	0.165	0.120	0.067	0.096	0.107
		4	0.229	0.213	0.163	0.160	0.174	0.154
		6	0.203	0.203	0.152	0.156	0.174	0.145
		9	0.243	0.226	0.165	0.210	0.203	0.163
	400	2	0.362	0.341	0.243	0.221	0.309	0.294
		4	0.396	0.377	0.289	0.397	0.432	0.374
		6	0.437	0.377	0.318	0.456	0.469	0.396
		9	0.442	0.412	0.331	0.505	0.512	0.427
P-SPQR	200	2	0.066	0.066	0.042	0.015	0.023	0.032
		4	0.070	0.071	0.074	0.031	0.047	0.069
		6	0.059	0.057	0.042	0.033	0.042	0.042
		9	0.064	0.056	0.049	0.041	0.049	0.047
	400	2	0.061	0.061	0.047	0.010	0.028	0.040
		4	0.065	0.063	0.048	0.027	0.040	0.042
		6	0.073	0.067	0.060	0.028	0.038	0.052
		9	0.077	0.069	0.050	0.042	0.044	0.048

Table 3.3 Type I errors comparison for the open- and closed-end procedures of the SPQR and P-SPQR methods for various values of γ, $\alpha = 0.05$ and $\tau = 0.7$ for Case II

Method	m	N/γ	Closed-end			Open-end		
			0	0.25	0.45	0	0.25	0.45
SPQR	200	2	0.386	0.358	0.295	0.191	0.264	0.269
		4	0.486	0.463	0.369	0.365	0.402	0.356
		6	0.502	0.482	0.403	0.421	0.442	0.394
		9	0.529	0.505	0.416	0.482	0.482	0.405
	400	2	0.721	0.705	0.587	0.493	0.590	0.560
		4	0.792	0.776	0.690	0.719	0.741	0.679
		6	0.824	0.805	0.716	0.763	0.710	0.396
		9	0.850	0.832	0.737	0.809	0.813	0.731
P-SPQR	200	2	0.053	0.052	0.051	0.014	0.025	0.038
		4	0.058	0.058	0.062	0.026	0.039	0.054
		6	0.048	0.048	0.062	0.029	0.040	0.059
		9	0.063	0.060	0.071	0.046	0.054	0.071
	400	2	0.058	0.062	0.068	0.018	0.032	0.055
		4	0.057	0.055	0.060	0.028	0.040	0.058
		6	0.042	0.044	0.054	0.027	0.031	0.051
		9	0.060	0.059	0.058	0.039	0.047	0.054

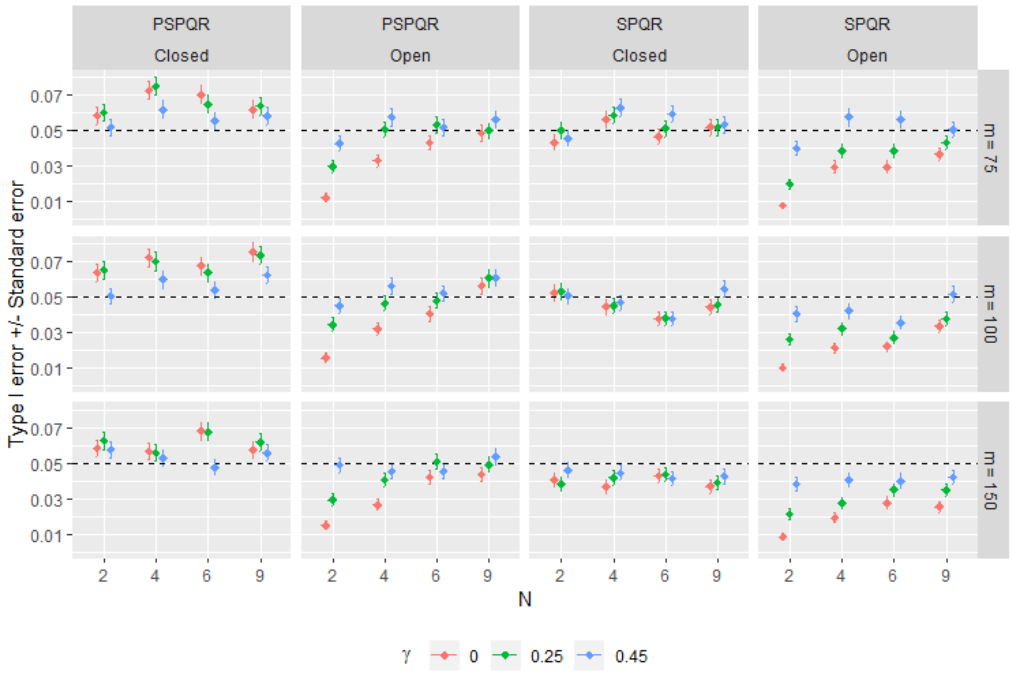

Figure 3.1 Case I: Type I error comparison for open- and closed-end procedures for P-SPQR and SPQR methods

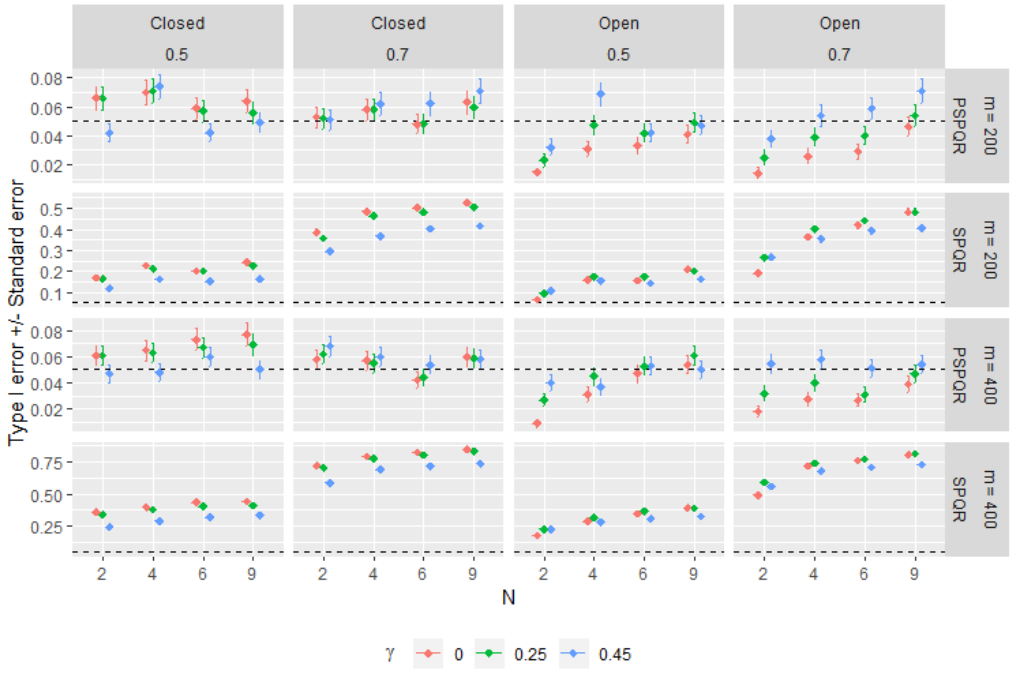

Figure 3.2 Case II: Type I error comparison for open- and closed-end procedures for P-SPQR and SPQR methods with $\tau \in \{0.5, 0.7\}$

Table 3.4 The power comparison for the closed-end procedure with $\tau = 0.5$, $p \in \{100, 300\}$, $k^* \in \{1, 25, 100\}$, and $\gamma \in \{0, 0.25, 0.45\}$

		$k^* = 1$				$k^* = 25$				$k^* = 100$			
	$m \rightarrow$	100		300		100		300		100		300	
γ	α	SPQR	P-SPQR	SPQR	P-SPQR	SPQR	P-SPQR	SPQR	P-SPQR	SPQR	P-SPQR	SPQR	P-SPQR
0.00	0.025	0.122	0.728	0.323	0.987	0.108	0.696	0.309	0.986	0.118	0.576	0.286	0.986
	0.050	0.193	0.812	0.439	0.990	0.175	0.780	0.425	0.990	0.209	0.689	0.398	0.988
	0.100	0.294	0.883	0.584	0.994	0.277	0.864	0.572	0.994	0.348	0.786	0.544	0.994
0.25	0.025	0.117	0.699	0.301	0.985	0.105	0.654	0.289	0.984	0.151	0.532	0.255	0.979
	0.050	0.183	0.788	0.412	0.999	0.171	0.749	0.396	0.988	0.240	0.639	0.366	0.987
	0.100	0.389	0.911	0.550	0.994	0.395	0.888	0.536	0.994	0.493	0.826	0.514	0.992
0.45	0.025	0.088	0.591	0.211	0.974	0.082	0.526	0.192	0.969	0.138	0.412	0.163	0.962
	0.050	0.136	0.678	0.304	0.982	0.139	0.612	0.278	0.979	0.198	0.489	0.242	0.975
	0.100	0.208	0.763	0.415	0.989	0.223	0.714	0.401	0.987	0.285	0.597	0.369	0.987

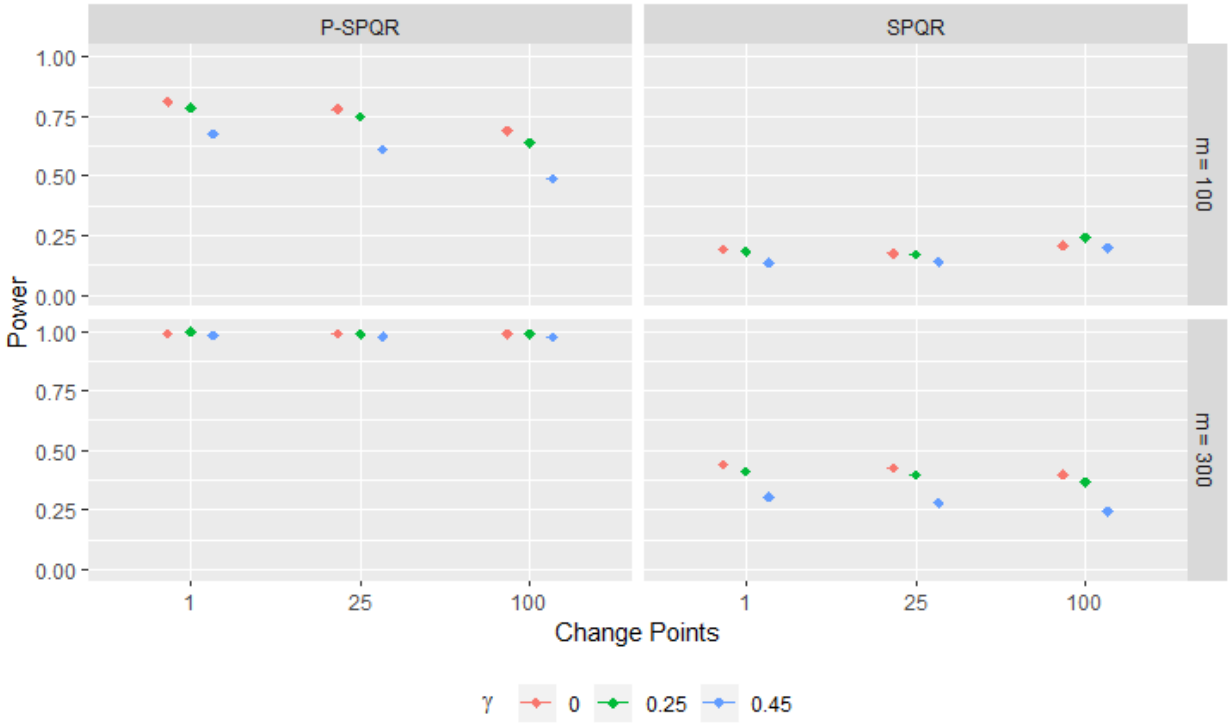

Figure 3.3 Power comparison for P-SPQR and SPQR methods under the closed-end procedure

We conduct the power simulation for both SPQR and P-SPQR methods. The results are summarised in Table 3.4. Figure 3.3 compares power for both procedures. We observe, in all cases, the P-SPQR method gives larger power compare to the SPQR method. We compute a five number summary of the stopping time for the open- and closed-end procedures. The results are summarised in Tables 3.5 - 3.6.

Table 3.5 Summary statistics of the detection time for the open-end procedure with $\tau = 0.5$, $\alpha = 0.05$ and $\gamma \in \{0, 0.25, 0.45\}$

		SPQR						P-SPQR					
	m		100			300			100			300	
γ	k^*	1	25	50	1	75	100	1	25	50	1	75	100
0.00	min	85	90	66	184	396	381	42	61	74	97	176	176
	Q1	226	321	390	653	578	629	138	179	235	206	320	359
	Med	337	370	468	854	1057	1132	238	292	363	282	423	465
	Q3	447	499	590	954	1135	1096	402	459	525	388	568	626
	max	879	877	895	2143	2488	2013	889	898	900	2343	2496	2340
0.25	min	23	11	11	106	64	64	20	17	17	42	65	65
	Q1	117	253	166	476	409	357	93	144	191	131	262	314
	Med	234	338	390	698	714	607	166	233	317	203	349	402
	Q3	348	485	522	927	1129	1137	345	413	503	302	488	577
	max	895	886	778	2493	2260	1830	888	899	892	2216	2263	2365
0.45	min	1	1	1	1	1	1	1	1	1	1	1	1
	Q1	18	10	7	19	24	22	62	100	156	79	238	290
	Med	60	17	15	82	48	39	127	192	274	144	344	402
	Q3	227	58	34	283	345	235	282	405	500	251	500	609
	max	872	875	798	2046	1345	1389	885	892	886	2235	2362	2367

Table 3.6 Summary statistics of the detection time for the closed-end procedure with $\tau = 0.5$, $\alpha = 0.05$ and $\gamma \in \{0, 0.25, 0.45\}$

		SPQR						P-SPQR					
	m		100			300			100			300	
γ	k^*	1	25	50	1	75	100	1	25	50	1	75	100
0.00	min	75	56	61	155	357	355	40	56	70	92	165	172
	Q1	201	282	354	612	623	617	123	166	214	184	305	338
	Med	297	389	466	825	972	912	205	266	327	250	389	434
	Q3	397	540	665	1504	1580	1430	373	451	498	366	518	584
	max	806	841	898	2451	2209	2446	873	894	883	2454	2570	2492
0.25	min	23	11	11	102	63	63	14	15	15	41	64	64
	Q1	116	195	111	397	397	247	89	137	106	124	247	296
	Med	193	286	392	544	806	422	158	223	300	189	335	386
	Q3	306	415	502	807	1115	911	309	388	479	281	455	542
	max	877	876	886	1784	2147	1996	889	895	899	2495	2471	2248
0.45	min	1	1	1	1	1	1	1	1	1	1	1	1
	Q1	21	9	9	24	22	22	58	97	149	78	228	288
	Med	55	17	16	88	49	44	119	186	269	141	335	397
	Q3	235	48	34	403	320	284	278	389	481	247	486	603
	max	870	872	717	2040	2263	2270	875	898	898	2227	2306	2362

3.3.1 Large p

In high-dimensional settings, we only consider P-SPQR method. We conduct simulation to study the finite sample properties of the proposed P-SPQR method. A high-dimensional data set with (p, m), considering $(100, 75), (200, 100)$ and $(300, 200)$ were generated. We consider the following two settings. In the first setting, the non-zero components of the true parameters are $\beta_{0,j} = 0$ for $j = \{1, \ldots, p\} \backslash \{1, 3, 5, 41, 77\}$ where $p \in \{100, 200, 300\}$. $\beta_{0,1} = 1$, $\beta_{0,3} = 15$, $\beta_{0,5} = -2$, $\beta_{0,41} = -2$, and $\beta_{0,77} = -8$. The explanatory variables $X_i \sim \text{Unif}(0, 1)$ and model errors $\mathcal{E}_i \sim N(0, 1)$ for $i = \{1, \ldots, p\}$.

In the second setting, under H_0, the regression coefficients are $\beta_{0,1} = 1,\ \beta_{0,3} = 15,\ \beta_{0,5} = -2,\ \beta_{0,6} = -13,\ \beta_{0,41} = -2,\ \beta_{0,77} = -8$ and $\beta_{0,j} = 0$ for all $j \in \{1, \ldots, p\} \backslash \{1, 3, 5, 8, 41, 77\}$ with $p \in \{100, 200, 300\}$. Under the alternative hypothesis the regression coefficients are $\beta_{1,2} = -1,\ \beta_{1,4} = 2,\ \beta_{1,7} = 1,\ \beta_{1,10} = -10.\ \beta_{1,51} = -8,\ \beta_{1,83} = -5$ and $\beta_{1,j} = 0$ for all $j \in \{1, \ldots, p\} \backslash \{2, 4, 7, 10, 51, 83\}$ with $p \in \{100, 200, 300\}$. We consider two different distributions of the explanatory variables. Under the null hypothesis the explanatory variables $X_i \sim \text{Unif}(0, 1)$. Under the alternative hypothesis, the ith explanatory variable is transformed to $X_i + 5$ and model errors $\mathcal{E}_i \sim N(0, 1)$ for $i \in \{1, \ldots, p\}$.

The first setting is used for the calculations of Type I error. Table 3.7 summarizes the Type I error for the open- and closed-end procedures. The various control parameter value, considering $\gamma \in \{0, 0.25, 0.45\}$ and the different size of the historical observations $m \in \{75, 100, 200\}$ are considered. The results are based on 1000 iterations. Type I errors based on the closed-end procedure is always higher than Type I errors computed from the open-end procedure. For small γ value, Type I errors of the open-end procedure is comparatively low. When open-end procedure is considered, smaller N provides slightly deflated Type I errors. Thus, for smaller N, we suggest to use the larger control parameter γ close to 0.5.

Table 3.7 Type I errors for P-SPQR method under the open- and closed-end procedures for various values of γ, the nominal significance level $\alpha = 0.05$ and $\tau = 0.5$

		Closed-end			Open-end		
(p, m)	N/γ	0	0.25	0.45	0	0.25	0.45
(100,75)	2	0.064	0.062	0.056	0.018	0.042	0.048
	4	0.062	0.060	0.052	0.032	0.040	0.050
	6	0.056	0.060	0.050	0.034	0.040	0.050
	9	0.046	0.048	0.054	0.032	0.046	0.052
(200,100)	2	0.070	0.076	0.068	0.020	0.048	0.062
	4	0.070	0.072	0.056	0.042	0.050	0.052
	6	0.060	0.062	0.060	0.038	0.056	0.058
	9	0.066	0.066	0.080	0.056	0.060	0.076
(300,200)	2	0.072	0.074	0.060	0.016	0.034	0.052
	4	0.082	0.072	0.074	0.036	0.050	0.066
	6	0.094	0.094	0.094	0.074	0.084	0.088
	9	0.088	0.090	0.072	0.052	0.074	0.070

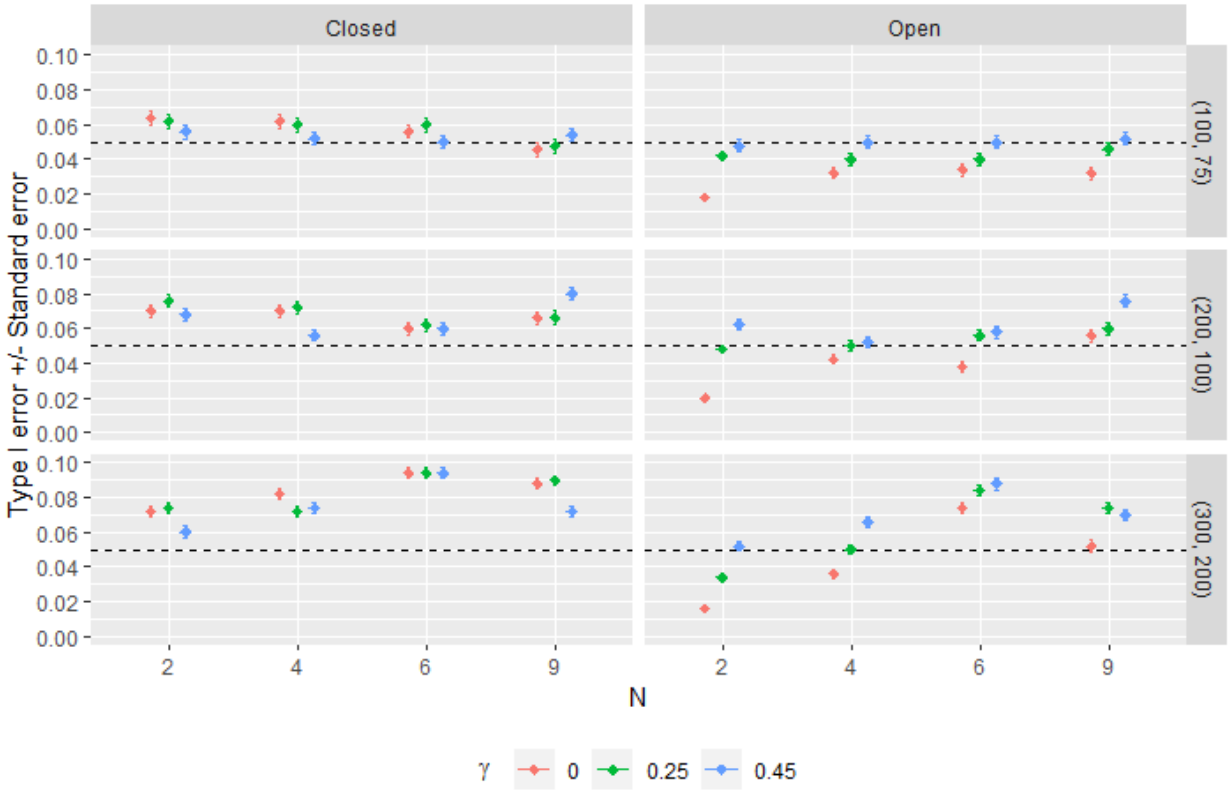

Figure 3.4 Type I error comparison for open- and closed-end procedures

We carry out the power analysis of the proposed P-SPQR procedure for high-dimensional data.

The results are summarized in Table 3.8. Figure 3.5 compares power for the P-SPQR method.

Table 3.8 Power comparison for P-SPQR method under the closed-end procedure for different pairs of (m, p), $\alpha \in \{0.025, 0.05, 0.1\}$, $\gamma \in \{0, 0.25, 0.45\}$, and various change-point locations

	(m,p)	$(75,100)$			$(100,200)$			$(200,300)$		
γ	α/k^*	1	25	75	1	50	100	1	100	200
0.00	0.025	0.827	0.791	0.670	0.779	0.727	0.639	0.771	0.707	0.644
	0.050	0.897	0.876	0.796	0.861	0.817	0.852	0.988	0.815	0.771
	0.100	0.954	0.946	0.894	0.932	0.903	0.865	0.921	0.898	0.870
0.25	0.025	0.800	0.737	0.600	0.755	0.684	0.589	0.742	0.663	0.590
	0.050	0.891	0.850	0.737	0.846	0.788	0.714	0.835	0.782	0.729
	0.100	0.941	0.923	0.852	0.907	0.875	0.825	0.910	0.874	0.842
0.45	0.025	0.676	0.565	0.385	0.640	0.512	0.417	0.631	0.495	0.410
	0.050	0.777	0.688	0.531	0.738	0.626	0.529	0.732	0.626	0.543
	0.100	0.870	0.808	0.678	0.834	0.758	0.654	0.824	0.749	0.682

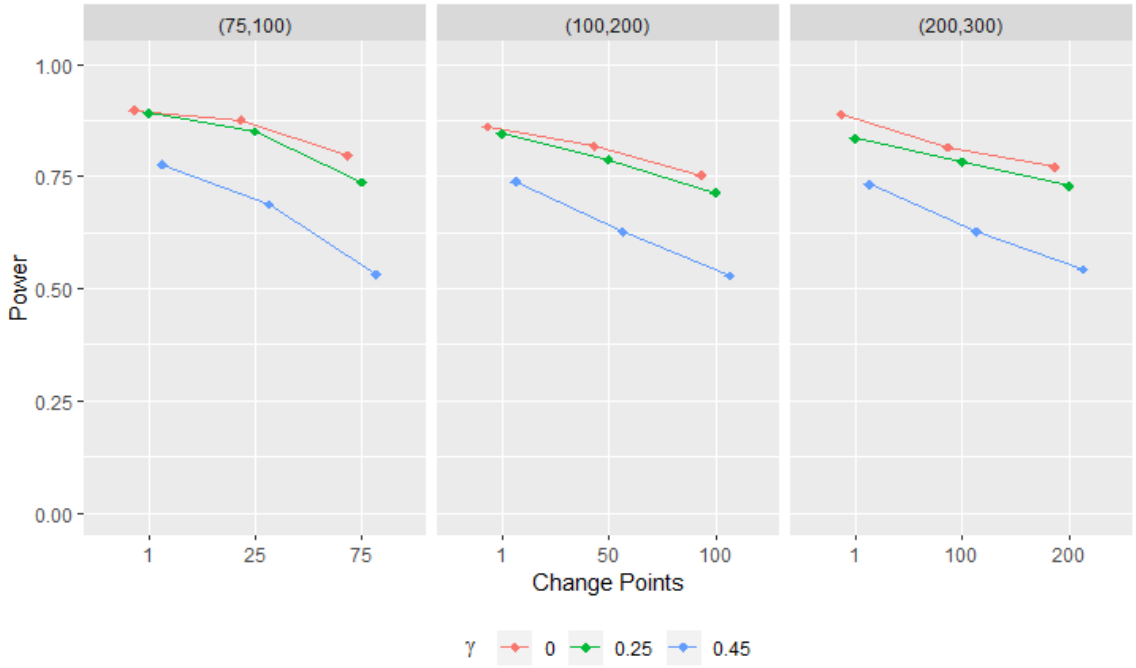

Figure 3.5 Power comparison for P-SPQR method under the closed-end procedure

We found the five number summary of the stopping time for the P-SPQR method. The results are summarized in Tables 3.9-3.10.

Table 3.9 Detection time for the open-end procedure at various change-point locations and different pairs of (m, p), $\gamma \in \{0, 0.25, 0.45\}$ and $\alpha = 0.05$

	(m,p)	$(75,100)$			$(100,200)$			$(200,300)$		
γ	Summary/k^*	1	25	75	1	50	100	1	100	200
0.00	min	21	38	47	32	58	58	59	112	103
	Q1	94	140	228	118	202	281	232	392	520
	Med	141	197	315	194	295	405	368	554	740
	Q3	224	295	431	312	429	559	653	877	1040
	max	667	674	674	898	896	899	1795	1760	1794
0.25	min	10	24	24	14	16	16	24	53	53
	Q1	61	118	206	77	157	250	139	318	438
	Med	102	173	301	135	254	383	248	486	670
	Q3	181	265	420	248	387	534	496	779	995
	max	671	663	675	887	890	895	1708	1789	1798
0.45	min	4	6	6	4	4	4	4	4	4
	Q1	42	107	207	49	142	213	71	204	387
	Med	85	172	328	103	255	375	172	464	680
	Q3	178	290	452	225	424	568	370	779	1013
	max	659	675	672	889	899	888	1798	1797	1798

Table 3.10 Detection time for the closed-end procedure at various change-point locations and different pairs of (m, p), $\gamma \in \{0, 0.25, 0.45\}$ and $\alpha = 0.05$

	(m,p)	$(75,100)$			$(100,200)$			$(200,300)$		
γ	Summary/k^*	1	25	75	1	50	100	1	100	200
0.00	min	21	37	46	28	54	56	57	90	90
	Q1	82	127	209	106	182	263	208	356	465
	Med	122	178	292	168	269	378	319	508	681
	Q3	189	264	399	274	390	520	551	803	966
	max	666	673	673	896	898	898	1754	1777	1790
0.25	min	10	23	23	14	15	15	24	50	50
	Q1	56	106	194	72	148	235	128	290	420
	Med	94	157	287	124	239	357	232	456	634
	Q3	164	249	398	230	365	505	451	746	950
	max	664	674	675	899	898	900	1788	1756	1784
0.45	min	4	6	6	4	3	3	4	4	4
	Q1	40	105	196	46	141	212	69	257	375
	Med	82	169	320	103	251	379	165	455	665
	Q3	170	281	439	220	420	572	365	765	1013
	max	650	675	667	896	900	893	1787	1777	1800

The estimated density of the stopping time are graphed in Figure 3.6. A similar pattern is evident in all three cases.

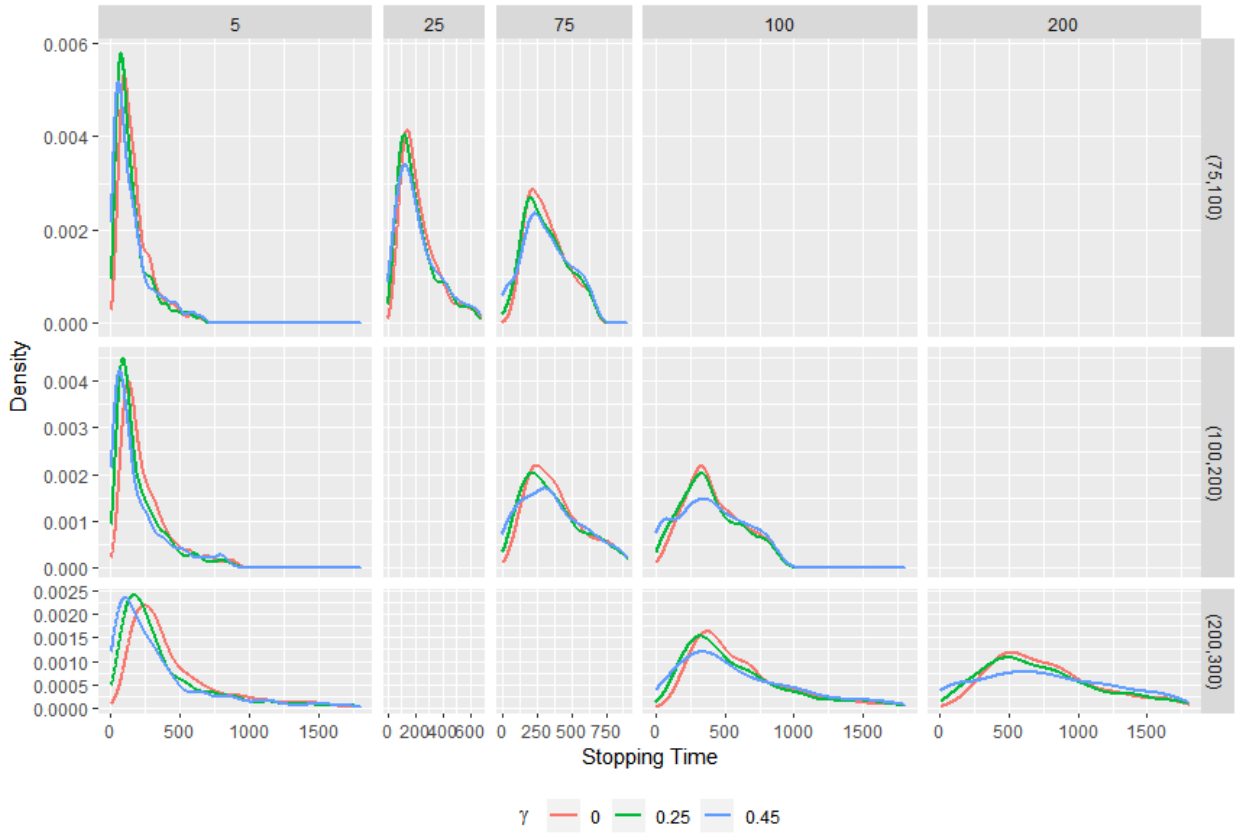

Figure 3.6 Estimated density of the stopping time for $\alpha = 0.05,\ \gamma \in \{0,\ 0.25,\ 0.45\}$ and various pair of (m, p)

3.4 Application

In this section, we apply our proposed method to a breast cancer gene expression data from The Cancer Genome Atlas (TCGA) project. The data contains expression measurements of $17,814$ genes from 536 patients. All expression measurements are recorded on the log scale. The response variable y is the gene expression measurement for BRCA1 and the explanatory variables are gene expression measurements for remaining genes. The response variable y is graphed below.

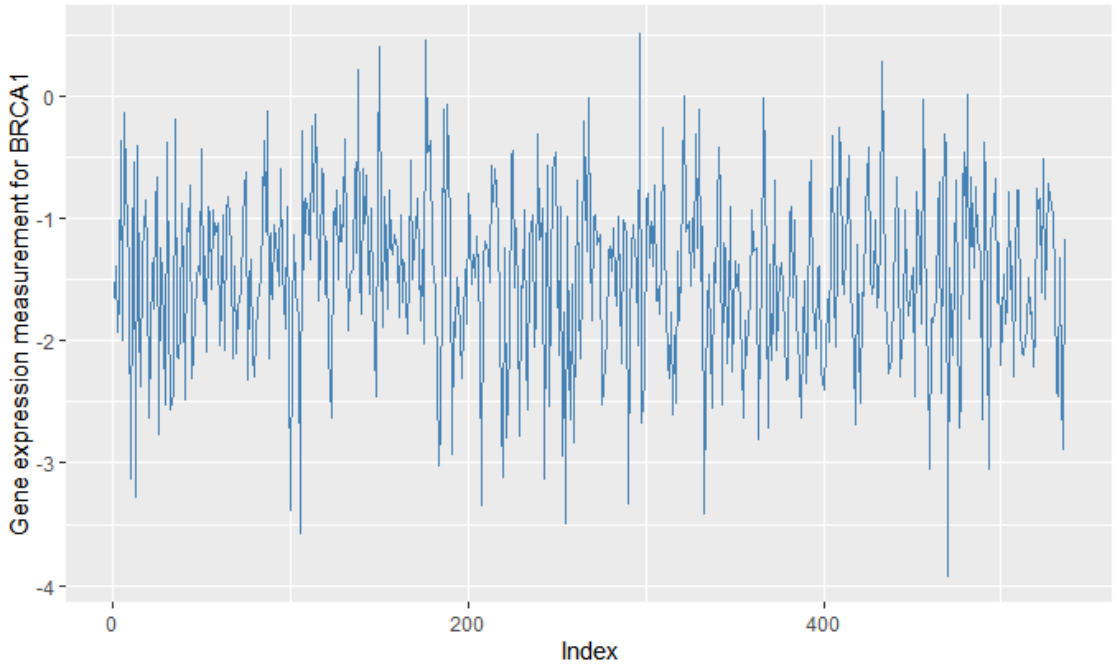

Figure 3.7 The gene expression measurement for BRCA1 of the breast cancer gene expression data

We can observe that there is no changes in the first 50 observations. Therefore, the first $m = 50$ observations are considered to be the historical sample size. The significant explanatory variables are chosen using by the post-SCAD penalized quantile regression at different quantile level of interest, considering $\tau \in \{0.5, 0.75\}$. Our proposed P-SPQR method is applied to monitor the future incoming observations sequentially. When quantile $\tau = 0.5$ our method detects nineteen change points. The corresponding multiple change points are $\{80, 86, 97, 124, 155, 161, 197, 203, 240, 254, 275, 332, 341, 368, 405, 441, 486, 509, 523\}$ and this graphed in Figure 3.8. However, we detect only six change points at qunatile level $\tau = 0.75$, they are $\{69, 108, 183, 206, 373, 468\}$ and they are plotted in Figure 3.9.

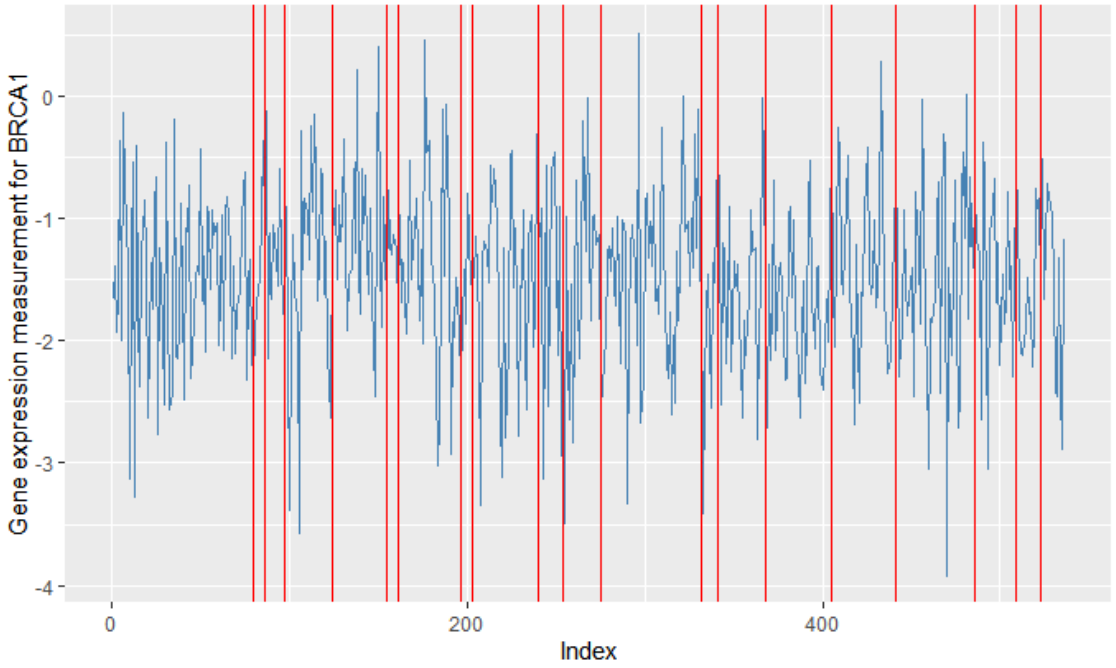

Figure 3.8 Change points at quantile level $\tau = 0.5$

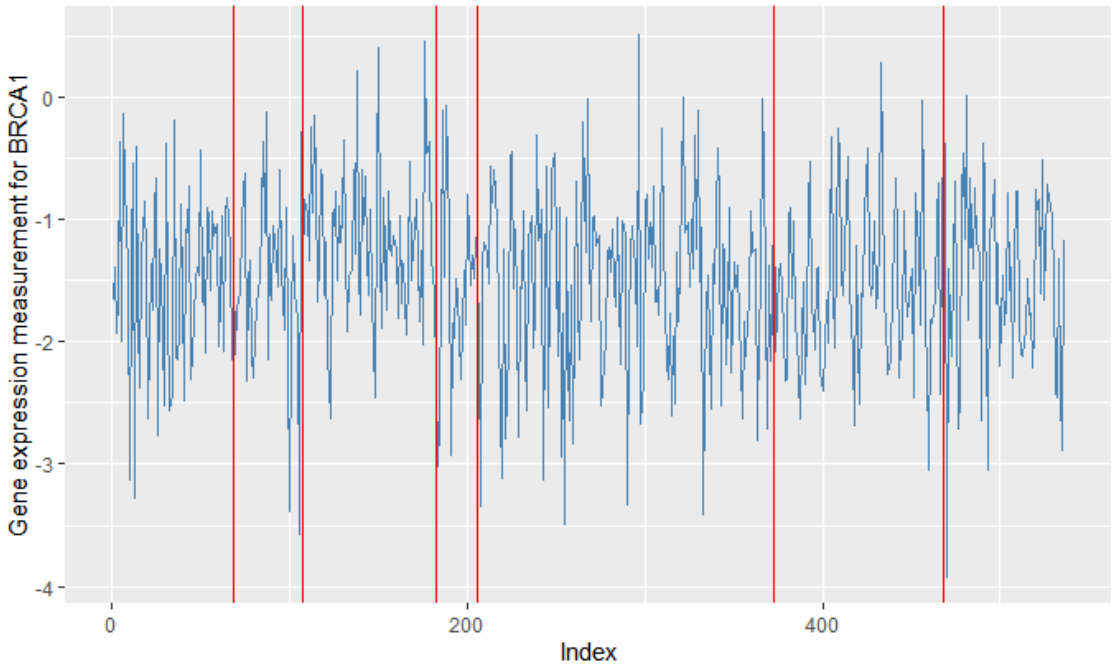

Figure 3.9 Change points at quantile level $\tau = 0.75$

3.5 Conclusion

In this chapter, we propose SPQR procedure for the sequential change point detection for high-dimensional data. The SPQR performs variable selection and estimation simultaneously. Moreover, to improve the SPQR based monitoring method, we develop the P-SPQR in sequential change point detection for high-dimensional data. Simulations are conducted to illustrate the performance of both methods. In Case II, the error distribution has heavy tails, the SPQR procedure gives much larger Type I errors. The P-SPQR method, however, significantly improves the Type I errors heavy tails of random errors. Thus,P-SPQR method is most suitable for heavy tails distributions. Further, the P-SPQR method shows higher power than the SPQR method. As expected, in both methods the power tends to increase when the historical sample size increases from 100 to 300. We have only considered the P-SPQR method for high-dimensional data. Our proposed P-SPQR method is applied to a breast cancer gene expression data to locate multiple change points sequentially.

CHAPTER 4 CONFIDENCE DISTRIBUTION FOR SKEW NORMAL CHANGE POINT MODEL BASED ON MODIFIED INFORMATION CRITERION

4.1 Introduction

The concept of a confidence distribution (CD) has its roots in Fisher's fiducial distribution. A CD is similar to a point estimator or an interval estimator but it uses a sample-dependent distribution function on the parameter space to estimate the parameter of interest. It also can provide confidence intervals of all nominal levels for a parameter of interest through confidence curves. Xie and Singh (2013) gave a detailed review of recent developments in confidence distributions. The first time of the terminology "confidence distribution" was used in a formal publication was dated back to Cox (1958). In his paper, Cox suggested that a confidence distribution "can either be defined directly, or can be introduced in terms of the set of all confidence intervals at different levels of probability". The formal modern definition for CD can be found in Schweder and Hjort (2002) and Singh, Xie, and Strawderman (2005); Singh and Strawderman (2007). The concept of CD is broad and it has wide range of applications including bootstrap distributions, p-value functions, normalized likelihood functions, and Bayesian posteriors, among others. For example, Schweder and Hjort (2002) used the CD to obtain the reduced likelihood function. Singh et al. (2005) proposed a method for combining information from independent studies through confidence distributions. Singh and Strawderman (2007) provided a formal definition of a CD and an asymptotic confidence distribution (aCD). Singh and Xie (2012) proposed a function called a CD posterior which uses the information from observed data with its corresponding prior information. Xie and Singh (2013) proposed a CD approach which incorporates the expert opinions while analyzing clinical data with binary outcomes. Shen, Liu, and Xie (2018) proposed a predictive distribution function with CDs. Cunen, Hermansen, and Hjort (2018) proposed a parametric method by using confidence distributions for detecting change points and obtained confidence curves for change locations.

The detection of change points is a process which attempts to identify points in time when the probability distribution of a stochastic process or time series changes. Change-point analysis plays an important role in financial time series analysis, economy, quality control, genome research, signal processing, medical research, statistical calibration and so on. The study of the change point problem was dated back to Page (1954, 1955) who first proposed a procedure to detect only one change in a parameter and has been extensively studied since then. For instance, Chernoff and Zacks (1964), Gardner (1969), Hawkins (1977) and Hawkins (1992) studied the testing and estimation of a change in the mean of a normal model. Hsu (1977) and Inclán (1993) studied change point problem for the variance in a normal model. Readers are referred to Cosörgö and Horváth (1997), Chen and Gupta (1999) and Chen and Gupta (2012) for more details of parametric and nonparametric methods on different types of change-point problems. The use of the information criteria in the view of the model selection as an alternative for change point detection has been extensive studied such as Hirotsu, Kuriki, and Hayter (1992), Chen and Gupta (1997), Chen, Gupta, and Pan (2006), Ngunkeng and Ning (2014), Hasan, Ning, and Gupta (2014), Cai, Said, and Ning (2016) and Said, Ning, and Tian (2019), to name a few.

As mentioned earlier, Cunen et al. (2018) suggested a change-point detection procedure based on a CD incorporating the likelihood function by profiling over the other parameters and obtained the confidence curve for the change location, consequently, the confidence sets at any confidence levels. However, as Chen et al. (2006) pointed out, the estimation of the change location through the regular likelihood function does not consider the contribution of the change location to the model complexity as a parameter. Clearly, when the change location k is located in the middle of the data, all the parameters are effective. However, the parameter space of the model becomes redundant when the location of change k is near 1 or n. To tackle this issue, Chen et al. (2006) proposed the modified information criterion (MIC) by considering the complexity of the penalty term which could relate to the change point locations. This method assigns larger penalty when the change point is near the beginning or the end of the data. In this chapter, we propose a CD-based procedure for a skew normal change-point model incorporating modified information criterion

(MIC). Furthermore, the presence of noise in the data can influence the intrinsic nature of data and cause changes. Therefore, we also verify the statistical significance of the detected change point through MIC-based test statistic.

This chapter is organized as follows. In Section 4.2, we go over the ideas of MIC and construction of confidence curves through CDs briefly, and then introduce the procedure based on CDs associated with MIC for a skew normal change-point model. In Section 4.3, simulations are conducted to investigate the performance of the proposed method and compare with some other existing method in terms of coverage probabilities and average lengths of confidence sets. The proposed method is applied to two stock market data to illustrate the detection and the estimation procedures in Section 4.4. Some discussion is provided in Section 4.5.

4.2 Methodology

4.2.1 Modified Information Criterion (MIC)

Let $x_1, \ldots, x_n$ be a random sample drawn from the density function $f(x; \Theta)$. The Schwarz information criterion (SIC) proposed by Schwarz (1978) is given as follows.

$$\text{SIC} = -2l_n(\hat{\Theta}) + \dim(\hat{\Theta}) \log(n), \tag{4.2.1}$$

where $l_n(\cdot)$ is the log-likelihood function of the random sample, $\hat{\Theta}$ is the maximum likelihood estimate (MLE) of the parameter Θ and $\dim(\hat{\Theta}_k)$ is the dimension of the parameter space. We denote Θ_L, Θ_R to be the pre-change and post-change parameters respectively and $\hat{\Theta}_L, \hat{\Theta}_R$ to be the MLEs of the pre-change and post-change parameters. In general, the change point problem can be treated as the model selection problem by selecting a better model between the null hypothesis of no change and the alternative hypothesis of at least one change existing. Therefore, the SIC in the context of having at least one change can be written as

$$\text{SIC}(k) = -2l_n(\hat{\Theta}_L(k), \hat{\Theta}_R(k), k) + \Big[2\dim(\hat{\Theta}_L(k)) + 1\Big] \log(n), \tag{4.2.2}$$

where $1 \leq k < n$ as well as (4.2.1) defines the SIC under the null hypothesis of no change which we denote it as SIC(n). However, as Chen et al. (2006) pointed out, (4.2.2) does not consider the change location to be a parameter which may cause the redundancy of the parameter space when the change occurs near the beginning or the end of data. Therefore, the modified information criterion (MIC) proposed by Chen et al. (2006) is given as follows. Under the null hypothesis of no change, the MIC is defined as,

$$\text{MIC}(n) = -2l_n(\hat{\Theta}) + \dim(\hat{\Theta}) \log(n), \tag{4.2.3}$$

where $\hat{\Theta}$ maximizes $l_n(\Theta)$. Therefore, under H_0, both SIC(n) and MIC(n) are same. Under the alternative hypothesis, the MIC is defined as,

$$\text{MIC}(k) = -2l_n(\hat{\Theta}_L(k), \hat{\Theta}_R(k), k) + \left[2\dim(\hat{\Theta}_L(k)) + \left(\frac{2k}{n} - 1\right)^2\right] \log(n), \tag{4.2.4}$$

where $1 \leq k < n$. The difference between (4.2.2) and (4.2.4) is that (4.2.4) considers the contribution of the change location k to the model as a parameter. If MIC$(n) > \min_{1 \leq k < n}$ MIC(k), then we select the model with a change point and the estimate of the change point is given by

$$\text{MIC}(\hat{k}) = \min_{1 \leq k < n} \text{MIC}(k). \tag{4.2.5}$$

Moreover, for the purpose of verifying the statistical significance of the detected change point, the associated MIC-based test statistic is defined as,

$$S_n = \text{MIC}(n) - \min_{1 \leq k < n} \text{MIC}(k) + \dim(\Theta) \log(n), \tag{4.2.6}$$

where MIC(n) and MIC(k) are defined in (4.2.3) and (4.2.5). Chen et al. (2006) showed, under Wald conditions and the regularity conditions, as $n \longrightarrow \infty$,

$$S_n \longrightarrow \chi^2_d, \tag{4.2.7}$$

in distribution under H_0, where d is the dimension of Θ.

4.2.2 Profile Log-likelihood and Deviance Function

Suppose observations $x_1, \ldots, x_k$ coming from the population with the density function $f(x, \Theta_L)$ and $x_{k+1}, \ldots, x_n$ coming from the population with the density function $f(x, \Theta_R)$. The log-likelihood function is,

$$l(k, \Theta_L, \Theta_R) = \sum_{i \leq k} \log(f(x, \Theta_L)) + \sum_{i \geq k+1} \log(f(x, \Theta_R)). \tag{4.2.8}$$

The profile log-likelihood function can be obtained by maximizing the log-likelihood function (4.2.8) over Θ_L and Θ_R for a given k. It can be defined as,

$$l_{prof}(k) = \max_{\Theta_L, \Theta_R} \{l(k, \Theta_L, \Theta_R)\} = l(k, \hat{\Theta}_L, \hat{\Theta}_R), \tag{4.2.9}$$

where $\hat{\Theta}_L$ and $\hat{\Theta}_R$ are MLEs of Θ_L and Θ_R for a given k respectively. Then the estimated change location $\hat{k}$ is given by $l_{prof}(\hat{k}) = \max_k(l_{prof}(k))$. After $\hat{k}$ is obtained, the deviance function is given by

$$D(k, \boldsymbol{x}) = 2\{l_{prof}(\hat{k}) - l_{prof}(k)\}, \tag{4.2.10}$$

where $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$. To construct a confidence curve for k based on the deviance function, we consider the estimated distribution of $D(k, \boldsymbol{x})$ at position k as follows.

$$\Psi_k(x) = P_{k, \hat{\Theta}_L, \hat{\Theta}_R}\{D(k, \boldsymbol{x}) < x\}, \tag{4.2.11}$$

where $x \in \mathbb{R}$. In the case of continuous parameters, Wilks theorem states that $\Psi_k(x)$ is approximately the distribution function of a χ^2_1. However, Wilks theorem does not hold for a discrete parameter k. Therefore, we compute Ψ_k through the simulations. The confidence curve can be

constructed as,

$$cc(k, \boldsymbol{x}_{obs}) = \Psi_k(D(k, \boldsymbol{x}_{obs})) = P_{k,\hat{\Theta}_L,\hat{\Theta}_R}\{D(k, \boldsymbol{x}) < D(k, \boldsymbol{x}_{obs})\}. \tag{4.2.12}$$

The probability that $cc(k, \boldsymbol{x}_{obs}) < \alpha$, under the true value of k, is often approximated well with α. Then, the confidence sets for k can be visualized using the plot $cc(k, \boldsymbol{x}_{obs})$. The $cc(k, \boldsymbol{x}_{obs})$ is the acceptance probability for k, or one minus the p-value for testing that value of k by using the deviance-based test which rejects the null hypothesis for high values of $D(k, \boldsymbol{x})$. We compute Ψ_k and hence $cc(k, \boldsymbol{x}_{obs})$ by simulations as follows.

$$cc(k, \boldsymbol{x}_{obs}) = \frac{1}{B}\sum_{j=1}^{B} I\{D(k, \boldsymbol{x}_j^*) < D(k, \boldsymbol{x}_{obs})\}, \tag{4.2.13}$$

for large number of B of simulated copies of data set $\boldsymbol{x}^*$. For each possible value of k, we simulate data $\boldsymbol{x}_j^*, j = 1, \cdots, B$ from $f(x, \Theta_L)$ and $f(x, \Theta_R)$ to the left and right side of k respectively. See Cunen et al. (2018) for more details. In our proposed procedure to construct the confidence curves, it is different from the Cunen et al. (2018) approach here to estimate k. Instead of estimating the change location k by maximizing the profile-likelihood function over all possible values of k, we estimates k using (4.2.5) by considering the impact of change locations. The MIC-based statistics S_n in (4.2.6) can be used to confirm a significant change statistically to avoid the fluctuations caused by noise.

4.3 Changes in All Three Parameters in a Skew Normal Distribution

The skew normal distribution (SN) was introduced by Azzalini (1985) which allows to regulate skewness in the data set. The probability distribution function of a skew normal random variable X is given by

$$f_X(x) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R} \tag{4.3.1}$$

and the cumulative distribution function (CDF) of the SN distribution is,

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) - 2T\left(\frac{x-\mu}{\sigma}, \lambda\right), \quad x \in \mathbb{R} \tag{4.3.2}$$

where T is Owen's function, ϕ and Φ are the probability distribution function and cumulative distribution function of the standard normal distribution. $\mu \in \mathbb{R}$ is the location parameter, $\sigma \in \mathbb{R}^+$ is the scale parameter and $\lambda \in \mathbb{R}$ is the shape parameter. We denote $X \sim SN(\mu, \sigma, \lambda)$. When $\lambda = 0$, the $SN(\mu, \sigma, \lambda)$ reduces to the normal $N(\mu, \sigma)$. Several basic properties of the skew normal distribution were studied by Azzalini (1985). Readers are referred to Azzalini and Capitanio (2014) for more details of skew normal distribution family and recent developments on this direction. The multivariate case of the skew normal distribution was investigated by Azzalini and Dalla Valle (1996). Although many methods have been proposed for making statistical inference for the skew normal distribution family, only a few of literature available on the change point detection. Arellano-Valle and Loschi (2013) proposed a Bayesian approach for detecting changes in parameters in a skew normal model. Ngunkeng and Ning (2014) proposed a skew normal change-point model based on the Schwarz information criterion (SIC). Their method was improved by Said et al. (2019) where they considered modified information criterion to detect changes in a skew normal model. A change-point problem for a skew normal distribution can be stated as follows.

$$x_i \sim \begin{cases} SN(\mu_L, \sigma_L, \lambda_L) & i = 1, \ldots, k, \\ SN(\mu_R, \sigma_R, \lambda_R) & i = (k+1), \ldots, n, \end{cases} \tag{4.3.3}$$

where $k \in \{1, \ldots, (n-1)\}$ is the unknown change point location and needed to be estimated. We are testing the following hypotheses,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu,$$

$$\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma,$$

$$\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda,$$

versus

$$H_1 : \underbrace{\mu_1 = \cdots = \mu_k}_{\mu_L} \neq \underbrace{\mu_{k+1} = \cdots = \mu_n}_{\mu_R},$$

$$\underbrace{\sigma_1 = \cdots = \sigma_k}_{\sigma_L} \neq \underbrace{\sigma_{k+1} = \cdots = \sigma_n}_{\sigma_R},$$

$$\underbrace{\lambda_1 = \cdots = \lambda_k}_{\lambda_L} \neq \underbrace{\lambda_{k+1} = \cdots = \lambda_n}_{\lambda_R},$$

where $1 \leq k < n$ is the unknown change point location with $(\mu_L, \sigma_L, \lambda_L)$ and $(\mu_R, \sigma_R, \lambda_R)$ being the parameters of left and right side of the change point location k. We assume there is at most one change in data since multiple-change detection can always be decomposed to multistages with at most a single change at each stage using the binary segmentation method proposed by Vostrikova (1981). The log-likelihood function under the null hypothesis defined as follows.

$$l_{H_0}(\mu, \sigma, \lambda) = n \log\left(\frac{2}{\sigma}\right) - \frac{1}{2}\sum_{i=1}^{n} \log\left(\phi\left(\frac{x_i - \mu}{\sigma}\right)^2\right) + \sum_{i=1}^{n} \log\left(\Phi\left(\lambda \frac{x_i - \mu}{\sigma}\right)\right). \quad (4.3.4)$$

To obtain the MLEs of μ, σ and λ, we take the first derivative of the log-likelihood function (4.3.4) with respect to the parameters μ, σ, λ and set the equations equal to zero. Under the alternative hypothesis,

$$\begin{aligned} l_{H_1}(k) = &\left\{ k \log\left(\frac{2}{\sigma_L}\right) - \frac{1}{2}\sum_{i=1}^{k} \log\left(\phi\left(\frac{x_i - \mu_L}{\sigma_L}\right)^2\right) + \sum_{i=1}^{k} \log\left(\Phi\left(\lambda_L \frac{x_i - \mu_L}{\sigma_L}\right)\right)\right\} \\ &+ \left\{ (n-k) \log\left(\frac{2}{\sigma_R}\right) - \frac{1}{2}\sum_{i=k+1}^{n} \log\left(\phi\left(\frac{x_i - \mu_R}{\sigma_R}\right)^2\right)\right. \\ &\left. + \sum_{i=k+1}^{n} \log\left(\Phi\left(\lambda_R \frac{x_i - \mu_R}{\sigma_R}\right)\right)\right\}. \end{aligned} \quad (4.3.5)$$

To obtain the MLEs of $\mu_L, \mu_R, \sigma_R, \sigma_L, \lambda_R$ andλ_L, we take the first derivative of the log-likelihood function (4.3.5) with respect to the parameters and set the equations equal to zero. Denote $\hat{\mu}_L, \hat{\mu}_R, \hat{\sigma}_R, \hat{\sigma}_L, \hat{\lambda}_R, \hat{\lambda}_L$ to be the MLEs of the $\mu_L, \mu_R, \sigma_R, \sigma_L, \lambda_R$ and λ_L respectively. The profiled log-likelihood function can be derived from (4.2.9). The modified information criteria can be obtained from (4.2.3). The estimate of the change point $\hat{k}$ is the value which minimizes the

equation (4.2.5). Further, the deviance function and the confidence curve can be obtained from (4.2.10) and (4.2.11) respectively.

4.4 Simulation Study

In this section, simulations will be conducted to investigate the performance of the proposed change-point detection method based on a CD, and compare with the one proposed by Cunen et al. (2018). To make fair comparisons, we perform two methods under the same settings and study their coverage probabilities and the average sizes of the confidence sets of the change points. A confidence set of a change point is defined by $\{k : cc(k, x) \leq \alpha\}$. Correspondingly, the size of a confidence set is defined by the number of k belonging to the confidence set for a given nominal level α.

Through all the simulations, we only consider a single change scenario since the multiple-change case can always be dealt with the binary segmentation method. The data before the change point k is always generated from $SN(0, 1, 1)$, and the data after the change point is generated from $SN(\mu, \sigma, \lambda)$ where $\mu = \{1, 1.5, 2\}$, $\sigma = \{2, 2.5, 3\}$ and $\lambda = \{3, 3.5, 4\}$. We consider two sample sizes $n = 50$ and $n = 100$. For the first sample size, we set up the changes occurring at $k = \{10, 20, 25\}$. For the second sample size $n = 100$, we set up the changes occurring at $k = \{10, 20, 40, 50\}$. The choices of k are approximately corresponding to the scenarios that a chance occurs at the very beginning, in the middle, and in the very end of data. We do not consider the changes occurring in the second half of the data, for example, $k = \{30, 40\}$ and $k = \{60, 80, 90\}$, since the performances will be similar due to the symmetric property. In our simulations, we consider three different approaches to obtain the estimated change location $\hat{k}$ to calculate the deviance function $D(k, x)$ given in (4.2.10). The first approach is simply based on the modified information criterion (MIC) given in (4.2.5). The second approach is based on the test statistic S_n which verifies statistically significance of the estimated change location $\hat{k}$ to avoid the impact of noise in the data. The third approach log-like method is by Cunen et al. (2018) which obtained $\hat{k}$ by maximizing the profile likelihood function $l_{prof}(k)$ given in (4.2.9) for all possible values of k without considering the impact of the location of the change.

1000 simulations are conducted for each scenario.

Table 4.1 lists simulation results of coverage probabilities for $n = 50$ and various confidence levels $0.50, 0.90, 0.95$ and 0.99. From the results, we observe that the MIC and S_n provide comparable coverage probabilities to the log-like method by Cunen et al. (2018), and in general, the MIC performs slightly better than the log-like method. As the increases of the differences among parameters, three methods perform similarly and the coverage probabilities get closer to the confidence levels we set up. Table 4.2 provides the average of sizes of confidence sets. MIC and S_n methods have similar average sizes. However, both of them provide smaller average sizes of confidence sets than the ones by log-like method by Cunen et al. (2018), especially when the change occurs at the beginning (equivalently, in the end) of the data. For example, for $SN(1, 2, 3)$ case with $k = 10$ and $\alpha = 0.99$, the average size of confidence sets by MIC method is 24.86 and is 24.68 by S_n method, comparing to 30.06 by log-like method. For $k = 25$, the average sizes of confidences are 15.38 for both MIC and S_n but 19.81 for log-like method. We also observe that, as the increases of differences among parameters, all three methods obtain similar average sizes of confidence sets. Figures 4.1 & 4.2 sketch the graphs of coverage probability and average size of confidence sets comparisons between S_n and the log-like, the likelihood method used in Cunen et al. (2018). Since MIC and S_n perform very similarly, therefore, we only graph the curves for S_n and log-like methods.

Table 4.1 The comparisons of coverage probabilities, $n = 50$

		$k = 10$			$k = 20$			$k = 25$		
	α	MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n
SN(1, 2, 3)	0.50	0.40	0.37	0.35	0.40	0.37	0.32	0.41	0.39	0.35
	0.90	0.73	0.72	0.70	0.68	0.65	0.63	0.69	0.66	0.62
	0.95	0.79	0.78	0.77	0.76	0.71	0.69	0.74	0.73	0.70
	0.99	0.91	0.89	0.89	0.89	0.86	0.83	0.86	0.86	0.82
SN(1.5, 2.5, 3.5)	0.50	0.43	0.41	0.40	0.44	0.43	0.40	0.46	0.44	0.38
	0.90	0.76	0.74	0.73	0.78	0.75	0.71	0.80	0.80	0.74
	0.95	0.83	0.81	0.79	0.87	0.85	0.81	0.88	0.87	0.83
	0.99	0.94	0.93	0.91	0.96	0.94	0.92	0.96	0.95	0.93
SN(2, 3, 4)	0.50	0.49	0.47	0.46	0.49	0.45	0.49	0.50	0.49	0.45
	0.90	0.82	0.81	0.78	0.90	0.80	0.82	0.90	0.89	0.81
	0.95	0.89	0.88	0.85	0.93	0.85	0.88	0.95	0.94	0.89
	0.99	0.96	0.96	0.95	0.99	0.93	0.96	0.99	0.99	0.96

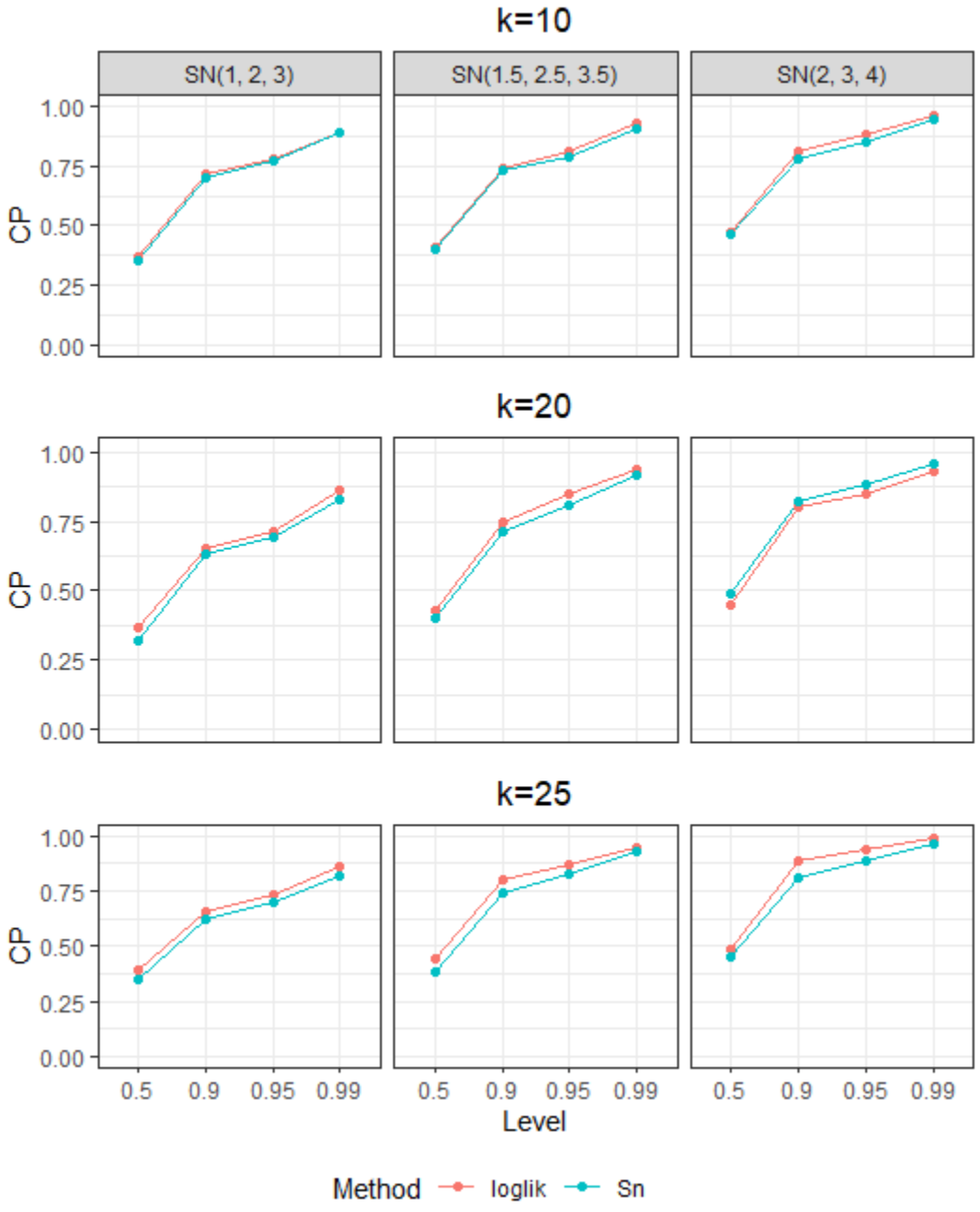

Figure 4.1 The coverage probability comparison when all parameters change at various change point positions, $k \in \{10, 20, 25\}$ and the sample size $n = 50$

Table 4.2 The comparisons of average sizes of confidence sets, $n = 50$

		$k = 10$			$k = 20$			$k = 25$		
	α	MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n
SN(1, 2, 3)	0.50	7.15	7.81	7.04	5.75	6.62	5.72	5.84	6.68	5.86
	0.90	13.78	17.04	13.48	9.71	11.82	9.65	9.32	11.54	9.34
	0.95	17.10	21.23	16.81	11.56	14.28	11.51	10.93	13.84	10.97
	0.99	24.86	30.06	24.68	16.33	20.38	16.25	15.38	19.81	15.38
SN(1.5, 2.5, 3.5)	0.50	4.91	5.61	4.87	4.79	5.13	4.77	5.02	5.17	5.04
	0.90	8.64	10.10	8.56	6.98	7.79	6.95	7.02	7.69	7.04
	0.95	10.59	12.53	10.47	7.99	9.16	7.96	7.98	8.89	7.99
	0.99	15.79	18.93	15.60	10.80	12.84	10.77	10.46	12.04	10.48
SN(2, 3, 4)	0.50	4.29	5.88	4.29	4.58	5.12	4.56	5.03	5.19	5.01
	0.90	6.97	8.58	6.96	6.22	7.21	6.21	6.61	7.17	6.61
	0.95	8.28	10.46	8.26	6.97	8.16	6.95	7.30	8.06	7.30
	0.99	12.14	15.56	12.12	8.93	10.79	8.92	9.11	10.44	9.11

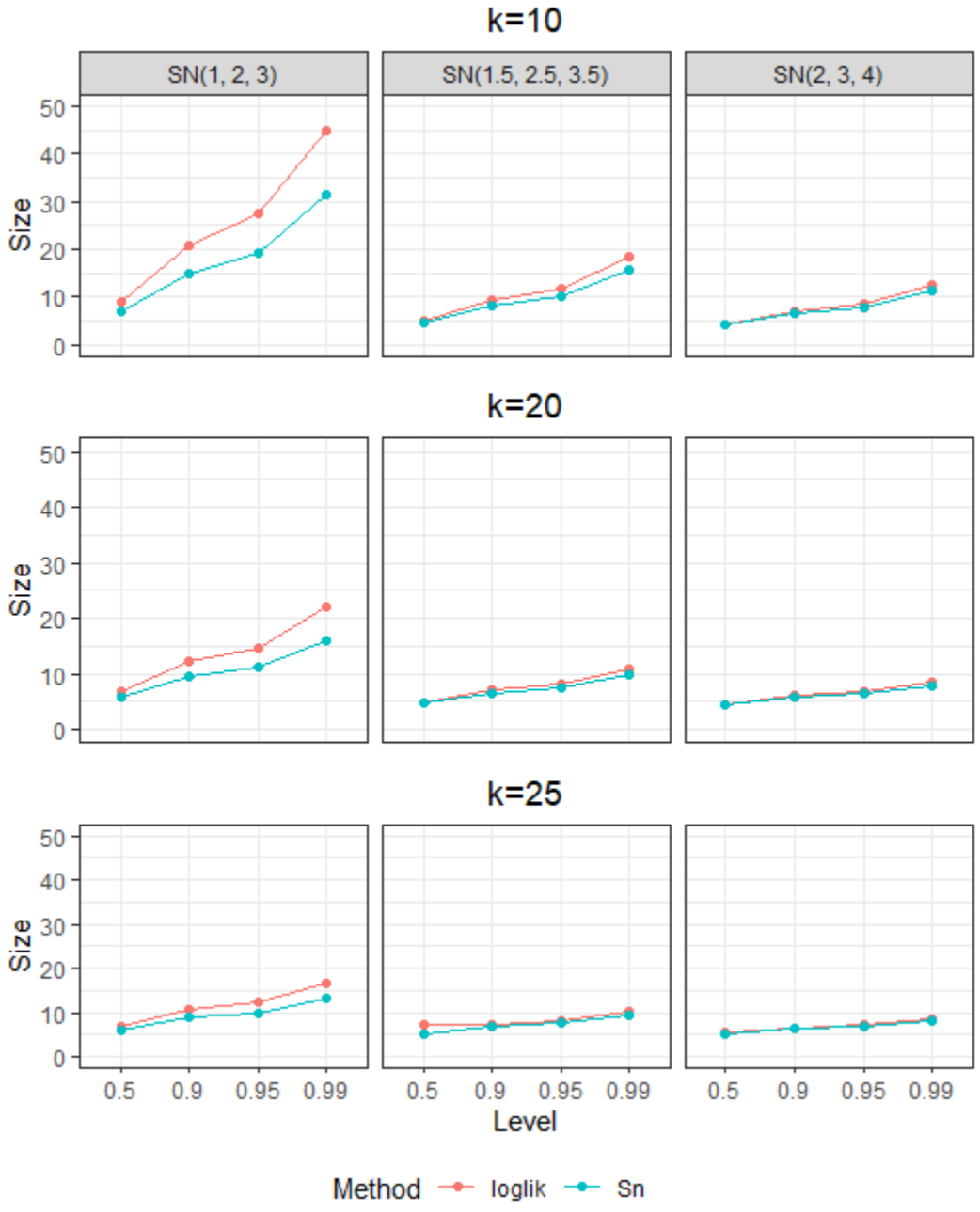

Figure 4.2 The confidence set comparison when change in all three parameters at various change point positions, $k \in \{10, 20, 25\}$ and the sample size $n = 50$

Tables 4.3 & 4.4 list the simulation results for the coverage probabilities and average sizes of confidence sets for $n = 100$ by three approaches with various confidence levels and change points. Same pattern is observed as the one observed from Tables 4.1 & 4.2. Figures 4.3 & 4.4 sketch the graphs for coverage probabilities and average sizes respectively.

Table 4.3 The comparisons of coverage probability, $n = 100$

	$k = 10$				$k = 20$			$k = 40$			$k = 50$		
	α	MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n
SN(1, 2, 3)	0.50	0.42	0.39	0.43	0.43	0.40	0.42	0.45	0.42	0.44	0.45	0.43	0.44
	0.90	0.68	0.66	0.68	0.69	0.68	0.69	0.70	0.67	0.69	0.70	0.68	0.70
	0.95	0.73	0.72	0.73	0.75	0.74	0.75	0.78	0.75	0.77	0.79	0.76	0.78
	0.99	0.83	0.82	0.85	0.86	0.84	0.86	0.87	0.86	0.86	0.89	0.85	0.88
SN(1.5, 2.5, 3.5)	0.50	0.45	0.44	0.45	0.46	0.44	0.46	0.48	0.47	0.48	0.49	0.47	0.48
	0.90	0.72	0.70	0.72	0.73	0.71	0.72	0.82	0.78	0.82	0.88	0.85	0.88
	0.95	0.77	0.76	0.75	0.77	0.76	0.77	0.88	0.85	0.88	0.91	0.88	0.91
	0.99	0.86	0.85	0.85	0.89	0.82	0.89	0.95	0.94	0.95	0.97	0.96	0.97
SN(2, 3, 4)	0.50	0.49	0.47	0.48	0.49	0.47	0.48	0.49	0.46	0.48	0.52	0.50	0.50
	0.90	0.82	0.80	0.82	0.83	0.85	0.83	0.85	0.83	0.85	0.88	0.86	0.87
	0.95	0.88	0.86	0.88	0.88	0.86	0.86	0.93	0.85	0.81	0.93	0.92	0.92
	0.99	0.96	0.95	0.96	0.96	0.96	0.96	0.97	0.94	0.96	0.98	0.96	0.97

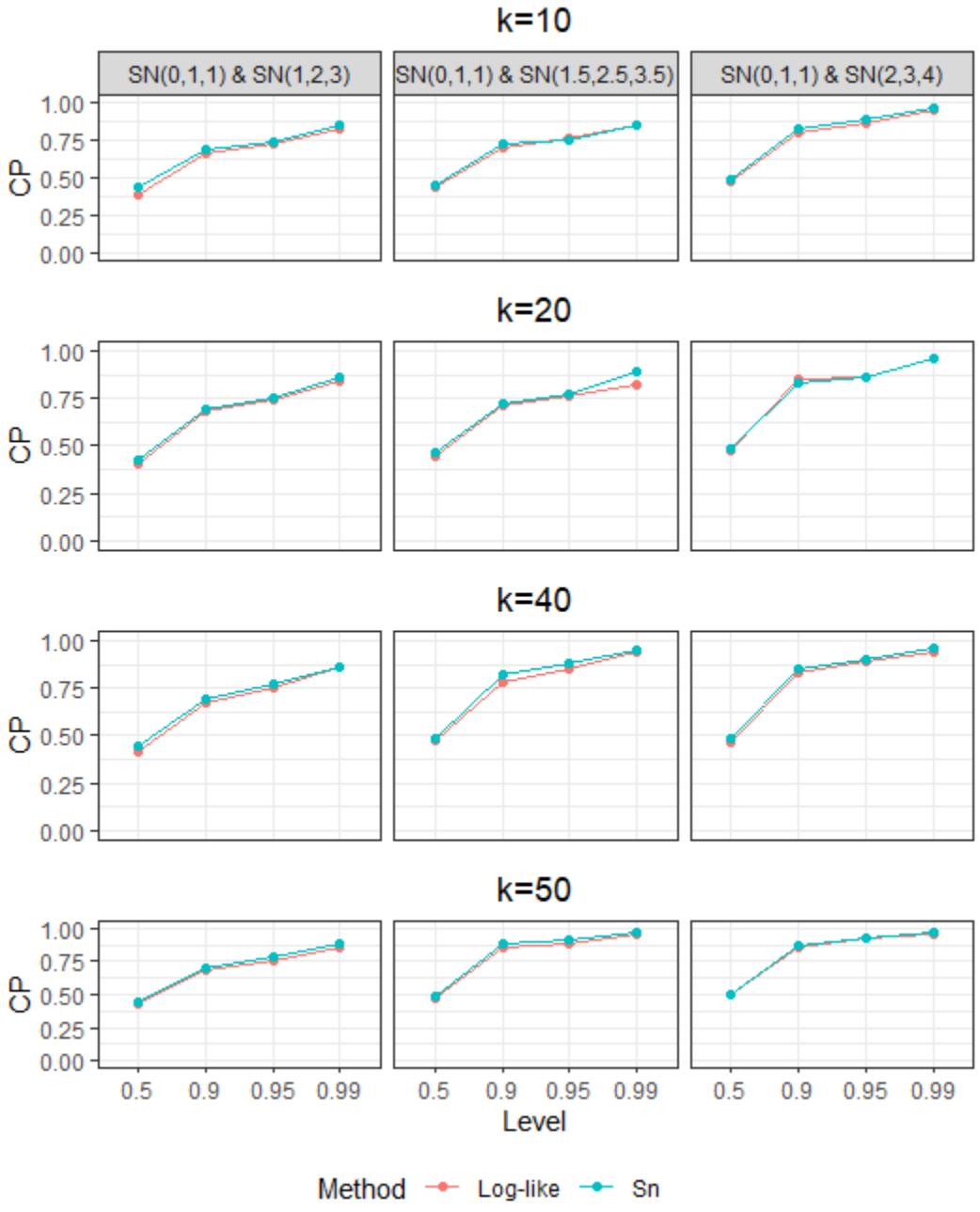

Figure 4.3 The coverage probability comparison when all parameters change at various change point positions, $k \in \{10, 20, 40, 50\}$ and the sample size $n = 100$

Table 4.4 The comparisons of average sizes of confidence sets, $n = 100$

		$k = 10$			$k = 20$			$k = 40$			$k = 50$		
	α	MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n
SN(1, 2, 3)	0.50	7.02	8.86	6.94	5.95	6.91	5.90	5.61	6.55	5.58	5.90	6.89	5.90
	0.90	14.82	20.88	14.71	9.69	12.22	9.64	8.48	10.56	8.45	8.74	10.74	8.74
	0.95	19.31	27.63	19.16	11.43	14.71	11.37	9.70	12.36	9.68	9.96	12.43	9.96
	0.99	31.71	45.11	31.34	15.98	21.92	15.88	12.71	16.75	12.64	12.99	16.78	12.99
SN(1.5, 2.5, 3.5)	0.50	4.85	5.17	4.75	4.70	4.87	4.69	4.77	4.92	4.75	5.38	7.29	5.20
	0.90	8.35	9.28	8.21	6.61	7.02	6.60	6.43	6.75	6.40	6.82	7.24	6.82
	0.95	10.40	11.60	10.22	7.50	8.08	7.48	7.13	7.64	7.10	7.50	8.02	7.50
	0.99	15.98	18.48	15.73	9.94	10.84	9.92	8.89	9.58	8.87	9.24	10.01	9.24
SN(2, 3, 4)	0.50	4.14	4.38	4.11	4.31	4.41	4.30	4.72	4.67	4.71	5.21	5.46	5.16
	0.90	6.57	7.12	6.53	5.71	5.97	5.70	5.88	5.95	5.88	6.38	6.45	6.38
	0.95	7.88	8.62	7.85	9.07	6.70	6.37	6.39	6.54	6.38	6.90	7.03	6.90
	0.99	11.47	12.68	11.42	8.02	8.42	8.00	7.63	7.93	7.62	8.20	8.47	8.20

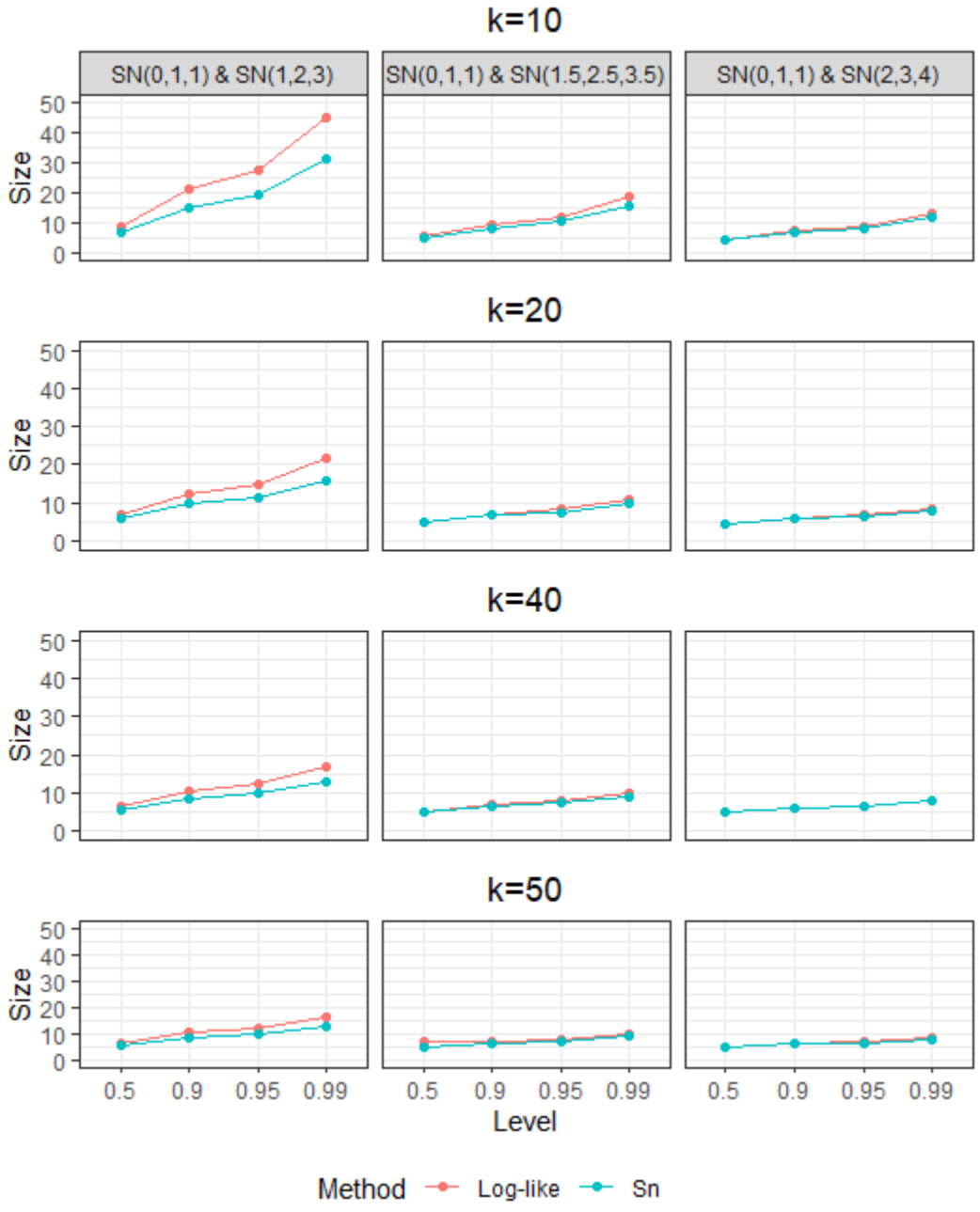

Figure 4.4 The confidence sets comparison when all three parameters changes at various change point locations, $k \in \{10, 20, 40, 50\}$ and the sample size $n = 100$

4.5 Application

In this section we consider the stock returns for Brazilian and Chilean markets to apply the proposed method. The stock returns for both countries were recorded weekly from October 31, 1995 to October 31, 2000. These data sets were used in Arellano-Valle and Loschi (2013) and

Ngunkeng and Ning (2014). Instead of using the data directly, we use the stock return ratio as recommended in Ngunkeng and Ning (2014). In both analyses, we also assume changes occurring in all three parameters simultaneously. The binary segmentation method proposed by Vostrikova (1981) is used to detect possible multiple changes in the data. The stock return ratio is obtained by the following transformation,

$$R_t = \frac{P_{t+1} - P_t}{P_t},\ t = 1, 2, \cdots, n-1.$$

The independence of the transformed data can be verified by the Portmanteau test. See Hsu (1979) and Ngunkeng and Ning (2014) for details.

4.5.1 Brazilian Market Return Ratio Data

The Brazilian market return ratio data is graphed in Figure 4.5.

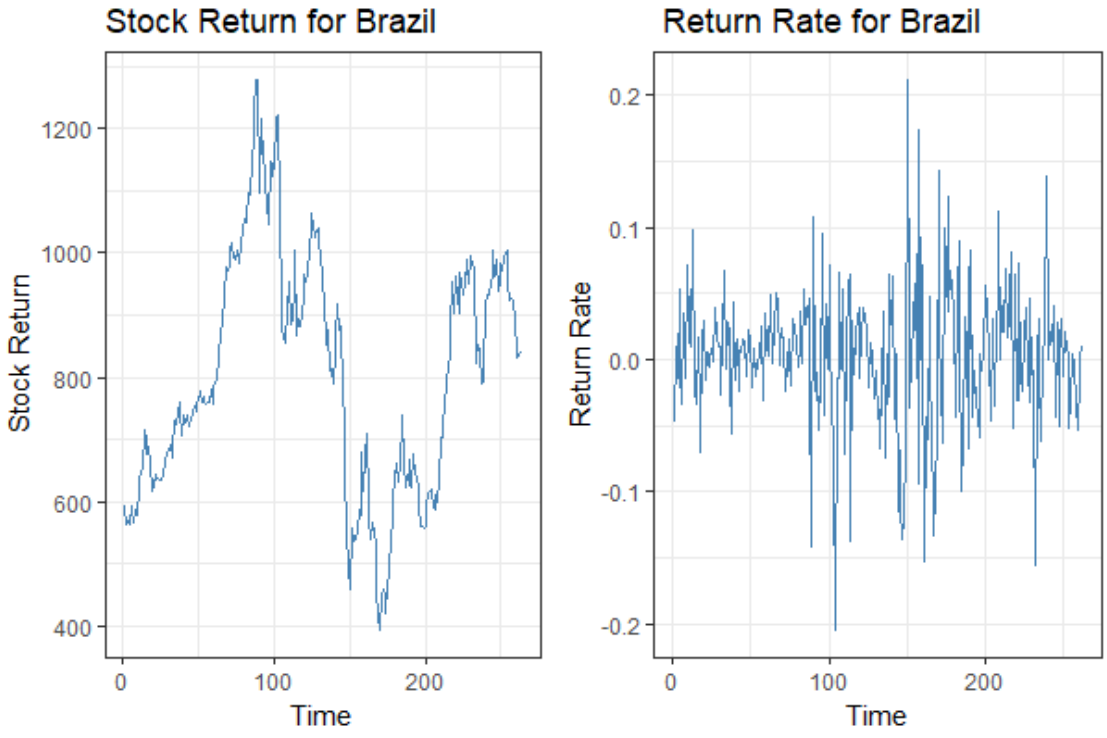

Figure 4.5 The weekly stock return data for Brazil

We apply the proposed approach along with the binary segmentation procedure to detect

multiple change points and construct corresponding confidence sets. $\text{MIC}(n) = \text{MIC}(262) = -781.8335 > \min_{1\leq k<n} \text{MIC}(k) = \text{MIC}(87) = -824.4513$ provides the estimation of the change location to be $\hat{k} = 87$. To confirm that it is a statistical significant change instead of being caused by noises, we calculate the test statistic $S_n = 59.3228$ associated with the critical value being $\chi^2_{0.95,3} = 7.815$ and the p-value being 8.202×10^{-13}. Therefore, we confirm that the change occurring at 87^{th} location which corresponds to the change 88^{th} location in the raw data is statistically significant. The maximum likelihood estimates (MLEs) of the parameters before the change point are $(\hat{\mu}_L, \hat{\sigma}_L, \hat{\lambda}_L) = (-0.0079, 0.0328, 0.8732)$, and the MLEs of the parameters are $(\hat{\mu}_R, \hat{\sigma}_R, \hat{\lambda}_R) = (-0.0006, 0.0611, 0.0022)$ after the change. Furthermore, the 95% confidence set for the change point is $\{71, 78, \ldots, 90\}$.

We then divide the data sets into two subsequences that are below k (≤ 87) and above k (> 87) and repeat the above detection process to detect changes in these two subsequences. Such an iterative process stops till no further change points being found. We detect all possible change points being $\{19, 88, 144, 170, 240\}$. The confidence curves for all change point estimates are shown in Figures 4.6 - 4.8. 95% confidence sets are marked by red dashed lines. Comparing to the ones obtained by Ngunkeng and Ning (2014), We detect an additional change at 170^{th}. These change points are shown in Figure 4.9. Moreover, as suggested by one of the reviewers, we apply the normal model by Cunen et al. (2018) to the data set to detect changes and obtain the change point set {19, 88, 144, 192, 240}. It has the change point 192 different from our method and the one by Ngunkeng and Ning (2014). However, the violation of normality of the data was verified by Ngunkeng and Ning (2014). Therefore, the model based on the skew normal distribution is more appropriate than the one based on the normal distribution by Cunen et al. (2018). Consequently, the results obtained based on the skew normal model are more convincible.

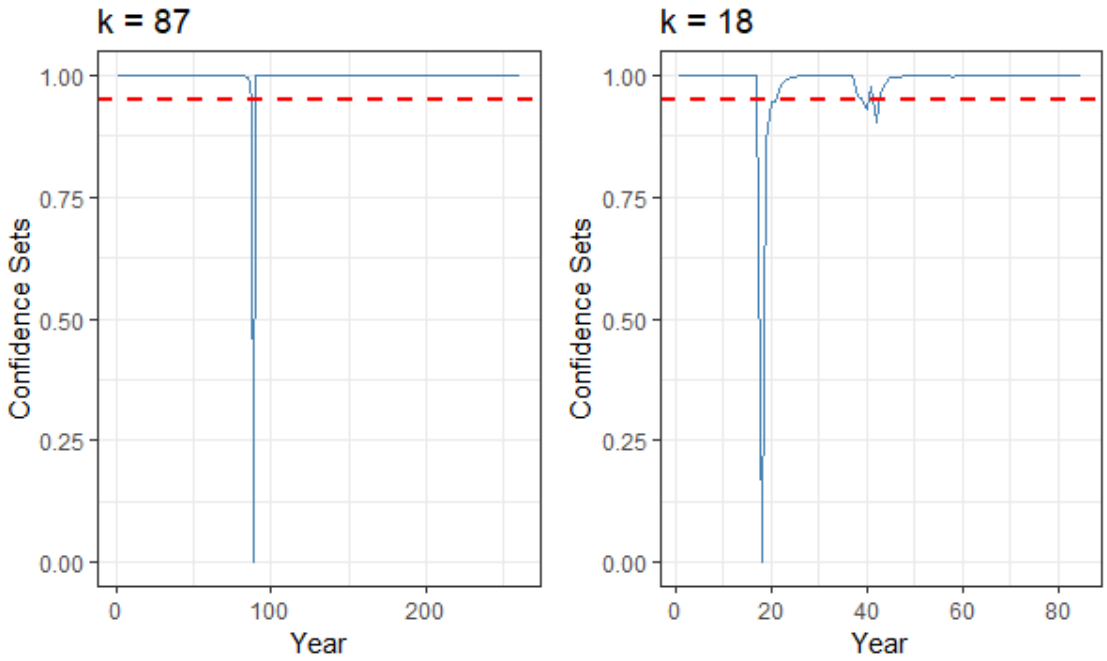

Figure 4.6 Left: Confidence curve for change point at $\hat{k} = 87$, Right: Confidence curve for the fist subset below $(k \leq 87)$, the $\hat{k} = 18$

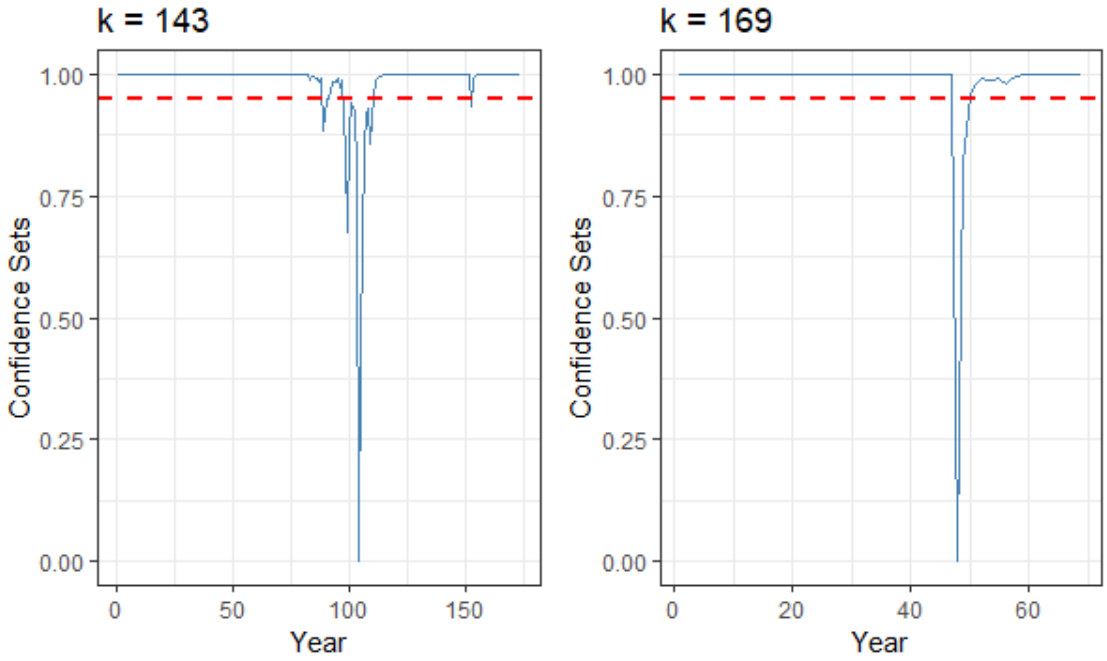

Figure 4.7 Left: Confidence curve for the second subset $(k \geq 88)$ after change point at $\hat{k} = 143$, Right: Confidence curve for the subset $(k \geq 144)$, the $\hat{k} = 169$

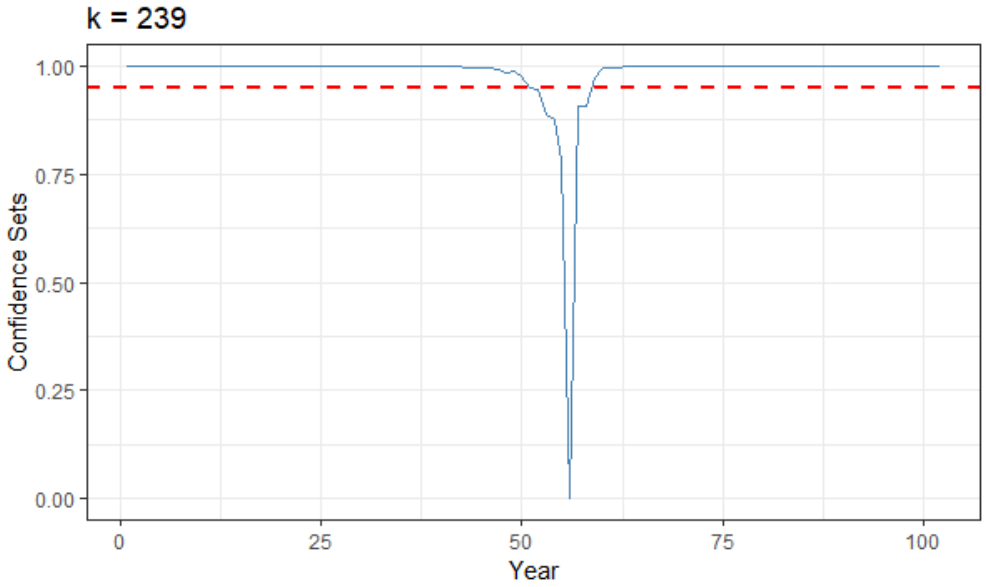

Figure 4.8 Confidence curve for the subset $(k \geq 170)$, the $\hat{k} = 239$

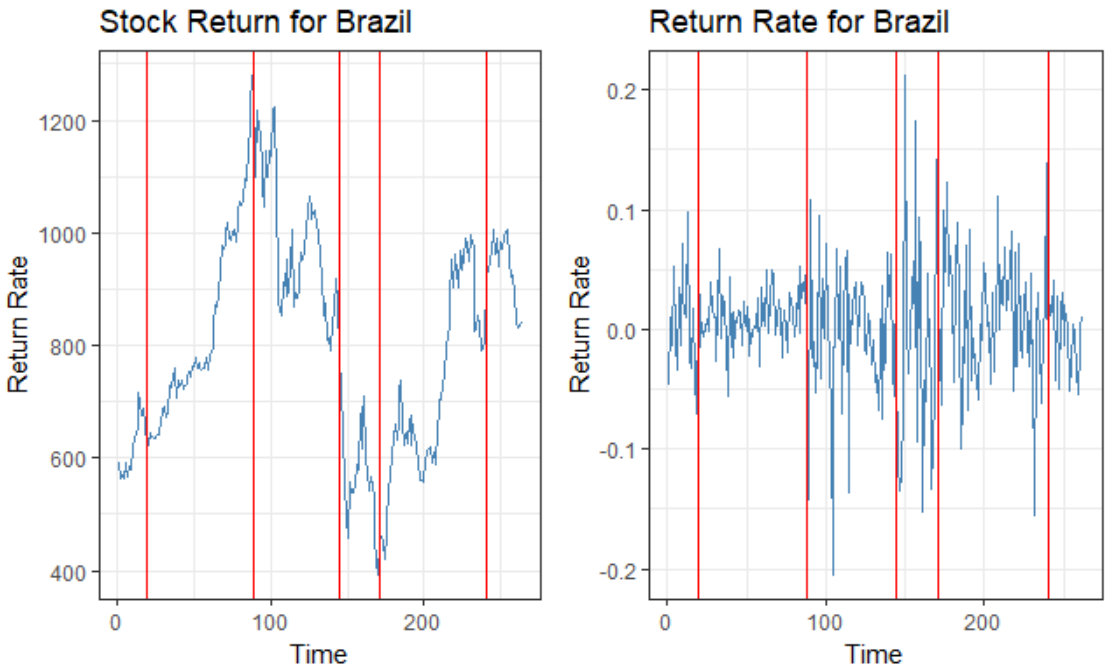

Figure 4.9 The weekly stock return data for Brazil with change point estimates

4.5.2 Chilean Market Return Ratio Data

The Chilean market return data and return ratio data are graphed in Figure 4.10.

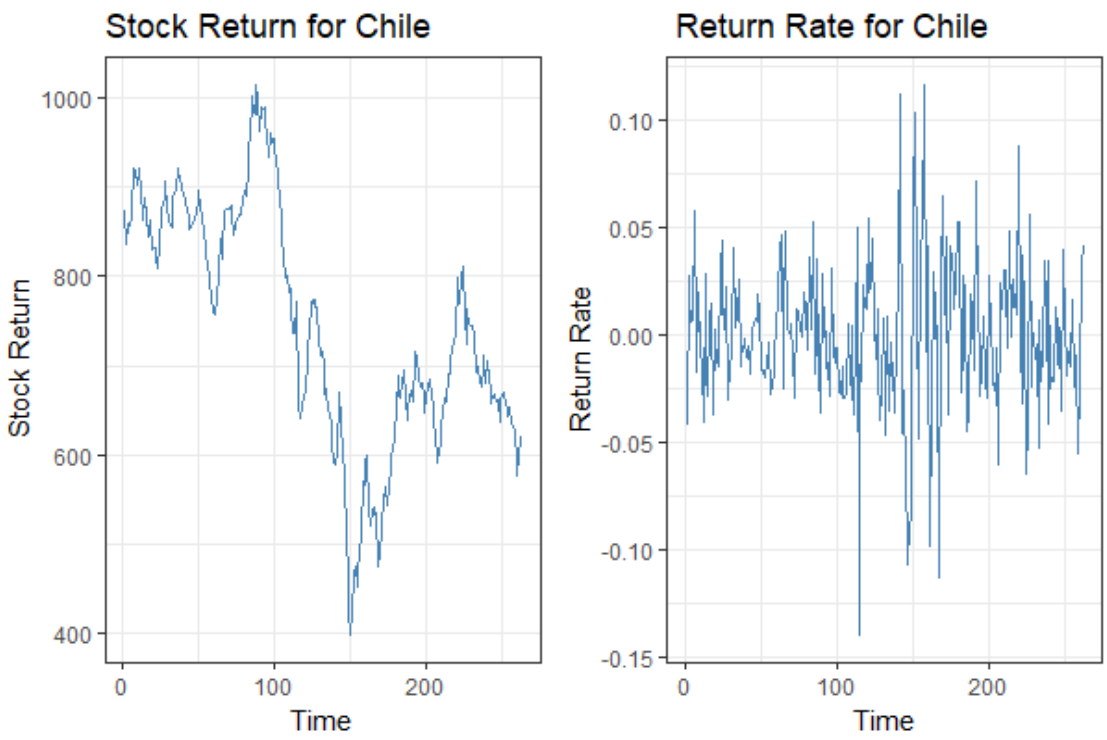

Figure 4.10 The weekly stock return data for Chile

The MIC(n) = MIC(262) = -1015.536 > $\min_{1\leq k<n}$ MIC(k) = MIC(87) = -1051.643. The estimated change location is $\hat{k} = 112$. The test statistic $S_n = 52.8011$ with the critical value $\chi^2_{0.95,3} = 7.815$ and the p-value 2.0214×10^{-11} confirms that it is a statistically significant change. The corresponding change point in the data is 113. The MLEs of the parameters are $(\hat{\mu}_L, \hat{\sigma}_L, \hat{\lambda}_L) = (-0.0253, 0.0323, 2.5638)$ and $(\hat{\mu}_R, \hat{\sigma}_R, \hat{\lambda}_R) = (-0.0004, 0.0402, 0.0018)$ before and after the change respectively. With the binary segmentation method, we found all four change points in the Chilen stock data. They are $\{61, 113, 170, 181\}$. Figures 4.11 and 4.12 show the confidence curves for all change point estimates and 95% confidence sets. These change points are graphed in Chilean Market data in Figure 4.13. The findings from our proposed approach are matched with the ones obtained by Ngunkeng and Ning (2014). Same as the Brazilian data, we also apply the normal model by Cunen et al. (2018) to detect changes. We obtain the change point set $\{7, 113, 170, 259\}$ which has only the change 170 matching our result and the one by Ngunkeng and Ning (2014). The difference is also due to the violation of normality of the data.

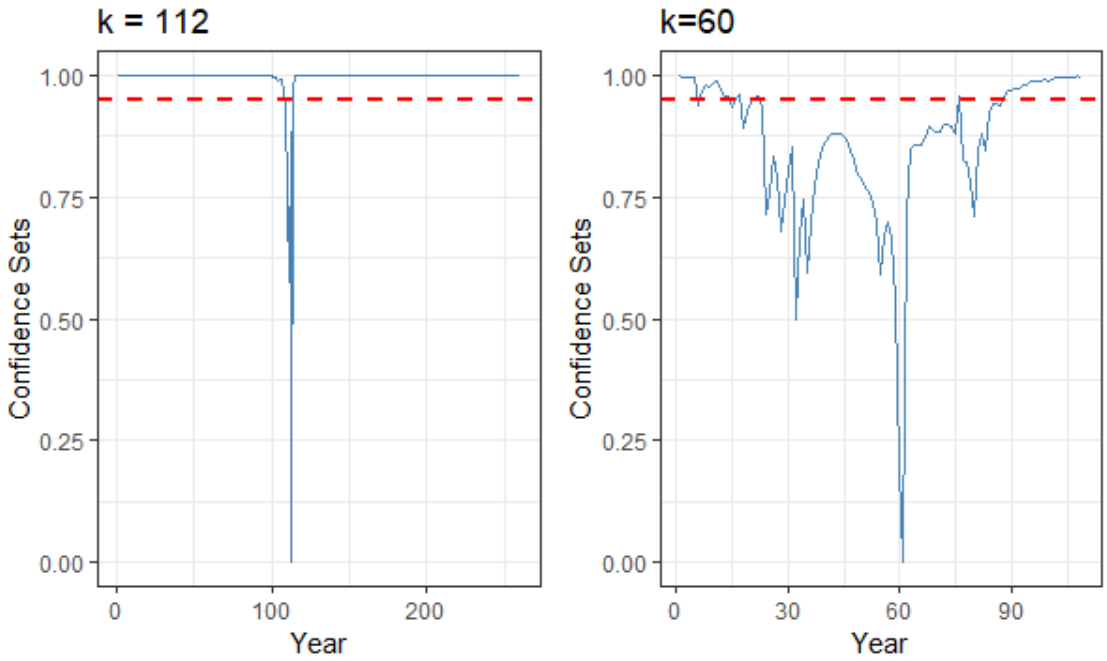

Figure 4.11 Left: Confidence curve for change point at $\hat{k} = 112$, Right: Confidence curve for the fist subset below $(k \leq 112)$, the change point estimate $\hat{k} = 60$

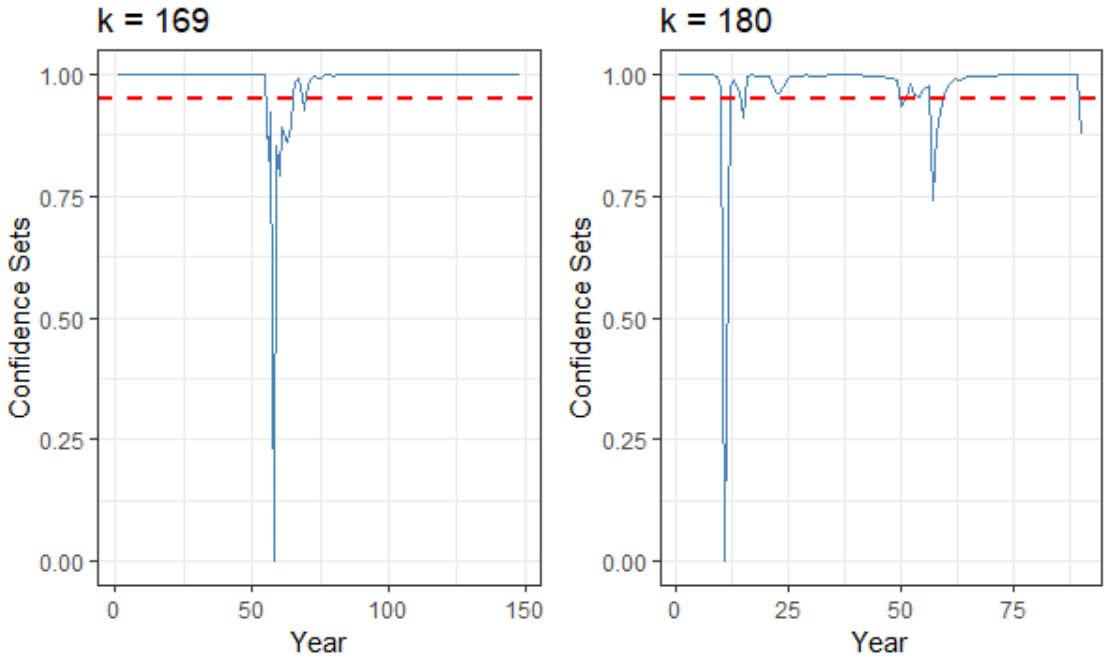

Figure 4.12 Left: Confidence curve for second subset $(k \geq 113)$, the change point estimate $\hat{k} = 169$, Right: Confidence curve for the subsequence after the change point $(k \geq 170)$, the change point estimate $\hat{k} = 180$

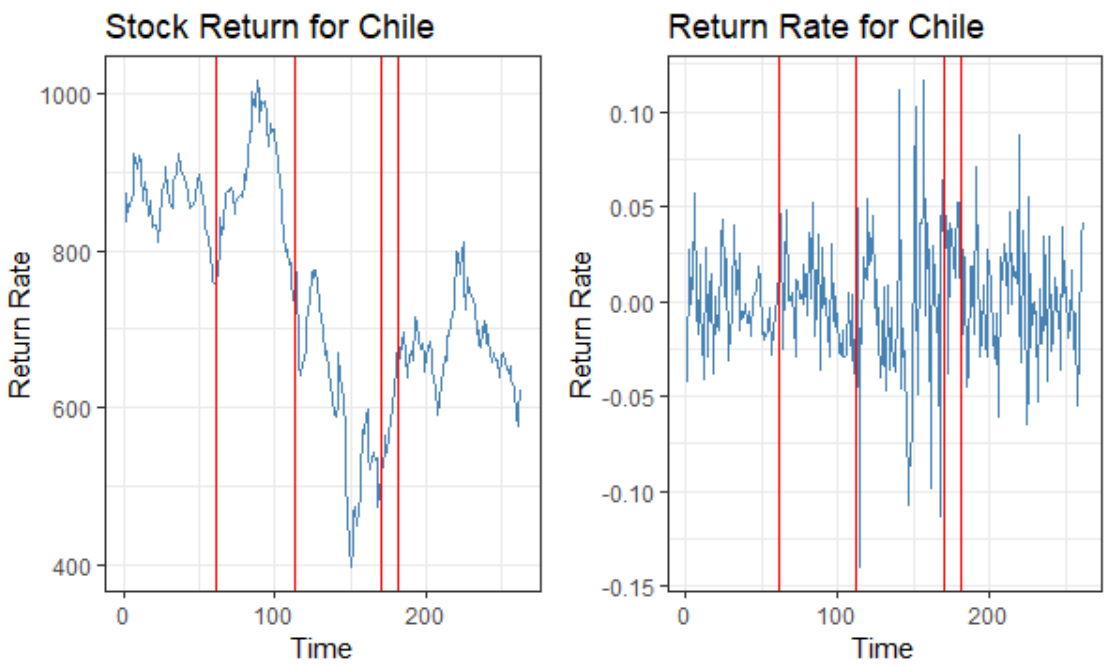

Figure 4.13 The weekly stock return data for Chile with change point estimates

4.6 Conclusion

In this chapter, we propose a CD-based procedure incorporating the modified information criterion for a skew normal change-point model. Different from other existing methods, the proposed method can provide confidence sets for the change point for a given nominal level instead of giving the point estimate only. Moreover, the proposed method considers the impact of the location of the change in terms of the model complexity. Consequently, it provides better coverage probability and comparatively smaller average sizes of confidence sets. Simulations are conducted under different scenarios which indicate the advantages of the proposed method. Two stock market data are given to illustrate the detecting procedure by the proposed method.

CHAPTER 5 CHANGE POINT DETECTION IN THREE PARAMETER WEIBULL DISTRIBUTION BASED ON THE MODIFIED INFORMATION CRITERION

5.1 Introduction

In the statistical analysis, the change point analysis plays an important role in identifying points in time when the probability distribution of stochastic processes or time series changes. When a change point exists, it is not advisable to perform a statistical analysis without taking into account of the existence of that change point because it could lead to misleading results. The change point analysis attempts to identify the number of change point(s) and the corresponding location(s). Change-point analysis has been extensively explored since Page (1954, 1955). Sen and Srivastava (1975a,b) derived the exact and asymptotic distribution of their test statistic for testing a single change in the mean of a sequence of normal random variables. Worsley (1979) studied the power of likelihood ratio and cumulative sum tests for a change in a binomial probability model. Srivastava and Worsley (1986) studied the multiple changes in the multivariate normal mean and approximated the null distribution of the likelihood ratio test statistic based on an improved Bonferroni inequality. Chen and Gupta (1995) studied the likelihood procedure for testing the change point hypothesis under multivariate Gaussian model. Asymptotic results of the test based on the likelihood ratio test can be found in Cosörgö and Horváth (1997). Gurevich and Vexler (2005) investigated the change-point problem in logistic regression. Wu (2008) provided a simultaneous change-point analysis and variable selection in a regression problem. Ning and Gupta (2009) studied the change-point problem for the generalized lambda distribution. Ramanayake and Gupta (2010) considered the problem of detecting a change-point in an exponential distribution with repeated values. Chen and Gupta (2012) discussed change-point problems for various parametric models with different approaches. Arellano-Valle and Loschi (2013) presented a Bayesian approach to study the change-point problem of the skew normal distribution. Ngunkeng and Ning (2014) studied the different change-point scenarios for the skew normal distribution from the viewpoint of the model selection.

Recently, Alghamdi, Ning, and Gupta (2018) proposed the change-point detection procedure for the Rayleigh Lomax distribution using Schwartz and modified information criterion.

Among all the distributions, the Weibull distribution is one of the most widely used lifetime distributions in reliability engineering. Due to its versatility, it can take on the characteristics of other types of distributions based on the value of the shape parameter. The three parameter Weibull distribution is defined as follows.

$$f_X(x) = \begin{cases} \dfrac{\alpha}{\beta}\left(\dfrac{x-\theta}{\beta}\right)^{\alpha-1} \exp\left(-\left(\dfrac{x-\theta}{\beta}\right)^{\alpha}\right) & ; x \geq \theta, \\ 0 & ; x < \theta, \end{cases} \tag{5.1.1}$$

and the cumulative distribution function is,

$$F_X(x) = 1 - \exp\left(-\left(\frac{x-\theta}{\beta}\right)^{\alpha}\right), \tag{5.1.2}$$

where θ is the location parameter, $\beta > 0$ is the scale parameter, and $\alpha > 0$ is the shape parameter. In short hand, we denote $W(\theta, \alpha, \beta)$. When $\theta = 0$, it reduces to two parameter Weibull distribution.

Jandhyala and Evaggelopoulos (1999) proposed a change-point methodology for identifying changes in the scale and shape parameters of a two-parameter Weibull distribution. The asymptotic results of the likelihood ratio test (LRT) statistic for detecting unknown changes in the parameters were derived as well as maximum likelihood estimate (MLE) of the unknown change point were obtained. They applied such a Weibull change-point model to model daily minimal temperatures measured in Uppsala. Jurušková (2007) investigated the asymptotic behavior of a log-likelihood ratio statistic for testing a change in a three-parameter Weibull distribution. In this chapter, we propose a detection procedure based on the modified information criterion (MIC) to detect simultaneous changes in parameters of a three-parameter Weibull distribution. Moreover, instead of only providing point estimates of change locations, we propose a method based on the confidence distribution (CD) to construct the confidence sets for change locations at a given significance level.

This chapter is organized as follows. In Section 5.2, the detecting procedure based on the MIC

is proposed. The asymptotic results of the test statistic associated with the detecting procedure is shown to be a standard χ^2 distribution. The procedure of constructing the confidence set of a change through the confidence distribution is provided in Section 5.3. Simulations to investigate the performance of the proposed procedures and compare with other existing methods in terms of powers, coverage probabilities and lengths of confidence sets are given in Section 5.4. A real data example is given in Section 5.5 to indicate an application of the proposed method. Some discussion is provided in Section 5.6.

5.2 Methodology

5.2.1 Modified Information Criterion

In general, change point detection can be treated in the sense of model selection. That is, testing the null hypothesis of no change versus the alternative hypothesis of at least one change is equivalent to choose the better model between the model under the null hypothesis and the model under the alternative hypothesis. Therefore, various model selection criteria can be adapted to the change-point detection procedure. The Schwarz information criterion (SIC) (Schwarz (1978)) is one of the most popular criteria for model selection. Zhang and Siegmund (2007) noted that the conventional SIC could detect change points more effectively when changes take place in the middle of the data. However, as Chen et al. (2006) pointed out, the conventional SIC method did not consider the complexity of the model which may cause the redundancy of the parameter space, especially a change occurring near the beginning or the end of data. To overcome this issue, Chen et al. (2006) proposed the modified information criterion by adjusting the penalty term in SIC so that it reflects the contributions of change-point locations to model complexity. This approach assigns lager penalty when the change point is close to 1 or n.

Let $x_1, \ldots, x_n$ be a random sample drawn from the density function $f(x; \Theta)$. The Schwarz information criterion (SIC) proposed by Schwarz (1978) is given as follows.

$$\text{SIC} = -2l_n(\hat{\Theta}) + \dim(\hat{\Theta}) \log(n), \tag{5.2.1}$$

where $l_n(\cdot)$ is the log-likelihood function of the random sample, $\hat{\Theta}$ is the maximum likelihood estimate (MLE) of the parameter Θ and $\dim(\hat{\Theta}_k)$ is the dimension of the parameter space. We denote Θ_L and Θ_R to be the pre-change and post-change parameters respectively and $\hat{\Theta}_L$ and $\hat{\Theta}_R$ to be the MLEs of the pre-change and post-change parameters. In general, the change point problem can be treated as the model selection problem by selecting a better model between the null hypothesis of no change and the alternative hypothesis of at least one change existing. Therefore, the SIC in the context of having at least one change can be written as

$$\text{SIC}(k) = -2l_n(\hat{\Theta}_L(k), \hat{\Theta}_R(k), k) + \left\{2\dim(\hat{\Theta}_L(k)) + 1\right\}\log(n), \tag{5.2.2}$$

where $1 \leq k < n$ as well as (5.2.1) defines the SIC under the null hypothesis of no change which we denote it as $\text{SIC}(n)$. However, as Chen et al. (2006) pointed out (5.2.3) does not consider the change location to be a parameter which may cause the redundancy of the parameter space when the change occurs near the beginning or the end of data. Therefore, the modified information criterion (MIC) proposed by Chen et al. (2006) is given as follows. Under the null hypothesis of no change, the MIC is defined as,

$$\text{MIC}(n) = -2l_n(\hat{\Theta}) + \dim(\hat{\Theta})\log(n), \tag{5.2.3}$$

where $\hat{\Theta}$ maximizes $l_n(\Theta)$. Therefore, under H_0, both $\text{SIC}(n)$ and $\text{MIC}(n)$ are same. Under the alternative hypothesis, the MIC is defined as,

$$\text{MIC}(k) = -2l_n(\hat{\Theta}_L(k), \hat{\Theta}_R(k), k) + \left\{2\dim(\hat{\Theta}_L(k)) + \left(\frac{2k}{n} - 1\right)^2\right\}\log(n), \tag{5.2.4}$$

where $1 \leq k < n$. The difference between (5.2.2) and (5.2.4) is that (5.2.4) considers the contribution of the change location k to the model as a parameter. If $\text{MIC}(n) > \min_{1\leq k<n} \text{MIC}(k)$, then

we select the model with a change point and the estimate of the change point is given by

$$\text{MIC}(\hat{k}) = \min_{1 \le k < n} \text{MIC}(k). \tag{5.2.5}$$

Moreover, for the purpose of verifying the statistical significance of the detected change point, the associated MIC-based test statistic is defined as,

$$S_n = \text{MIC}(n) - \min_{1 \le k < n} \text{MIC}(k) + \dim(\Theta) \log(n), \tag{5.2.6}$$

where $\text{MIC}(n)$ and $\text{MIC}(k)$ are defined in (5.2.3) and (5.2.4). Chen et al. (2006) showed, under Wald conditions and the regularity conditions, as $n \longrightarrow \infty$,

$$S_n \longrightarrow \chi^2_d, \tag{5.2.7}$$

in distribution under H_0, where d is the dimension of Θ. For the purpose of comparison later, we also provide the test statistic associated with the conventional SIC procedure as follows.

$$T_n = \text{SIC}(n) - \min_{1 \le k < n} \text{SIC}(k) + \dim(\Theta) \log n. \tag{5.2.8}$$

Chen and Gupta (1997) and Cosörgö and Horváth (1997) pointed out the asymptotic distribution of the related statistic for the SIC is found to have type I extreme value distribution.

5.2.2 MIC-based Detection Procedure for Three-Parameter Weibull Distribution

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables belong to a three-parameter Weibull distribution. The change point problem for a three-parameter Weibull distribution is defined as follows.

$$X_i \sim \begin{cases} W(\theta_L, \alpha_L, \beta_L) & i = 1, \ldots, k, \\ W(\theta_R, \alpha_R, \beta_R) & i = (k+1), \ldots, n, \end{cases} \tag{5.2.9}$$

where the pdf and cdf of three-parameter Weibull distribution are given in (5.1.1) and (5.1.2). We are testing the following hypotheses.

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_n = \theta,$$
$$\alpha_1 = \alpha_2 = \cdots = \alpha_n = \alpha,$$
$$\beta_1 = \beta_2 = \cdots = \beta_n = \beta,$$

versus

$$H_1 : \underbrace{\theta_1 = \cdots = \theta_k}_{\theta_L} \neq \underbrace{\theta_{k+1} = \cdots = \theta_n}_{\theta_R},$$
$$\underbrace{\alpha_1 = \cdots = \alpha_k}_{\alpha_L} \neq \underbrace{\alpha_{k+1} = \cdots = \alpha_n}_{\alpha_R},$$
$$\underbrace{\beta_1 = \cdots = \beta_k}_{\beta_L} \neq \underbrace{\beta_{k+1} = \cdots = \beta_n}_{\beta_R},$$

where (α, θ, β), $(\alpha_L, \theta_L, \beta_L)$ and $(\alpha_R, \theta_R, \beta_R)$ are unknown parameters and need to be estimated. k is the unknown change location and needs to be estimated as well. Under the null hypothesis, the log-likelihood function is given as,

$$l_n(\theta, \alpha, \beta) = n\log(\alpha) - n\log(\beta) + (\alpha - 1)\sum_{i=1}^{n} \log\left(\frac{x_i - \theta}{\beta}\right) - \sum_{i=1}^{n}\left(\frac{x_i - \theta}{\beta}\right)^{\alpha}. \quad (5.2.10)$$

The maximum likelihood estimators (MLEs) of θ, α and β can be obtained by setting these partial derivatives equal to zero.

$$\frac{\partial l_n(\theta, \alpha, \beta)}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log\left(\frac{x_i - \theta}{\beta}\right) - \sum_{i=1}^{n}\left(\frac{x_i - \theta}{\beta}\right)^{\alpha} \log\left(\frac{x_i - \theta}{\beta}\right),$$
$$\frac{\partial l_n(\theta, \alpha, \beta)}{\partial \beta} = \frac{-n\alpha}{\beta} + \frac{\alpha}{\beta}\sum_{i=1}^{n}\left(\frac{x_i - \theta}{\beta}\right)^{\alpha+1},$$
$$\frac{\partial l_n(\theta, \alpha, \beta)}{\partial \theta} = (\theta - 1)\sum_{i=1}^{n}\left(\frac{1}{x_i - \theta}\right) + \frac{\alpha}{\beta}\sum_{i=1}^{n}\left(\frac{x_i - \theta}{\beta}\right)^{\alpha-1}.$$

The MIC(n) is defined as,

$$\text{MIC}(n) = -2l_n(\hat{\theta}, \hat{\alpha}, \hat{\beta}) + 3\log(n), \tag{5.2.11}$$

where $\hat{\theta}, \hat{\alpha}$ and $\hat{\beta}$ are the MLEs of θ, α and β respectively. Similarly, under the alternative hypothesis, the log-likelihood function is,

$$\begin{aligned} l_{H_1} = l(k, \theta_L, \alpha_L, \beta_L, \theta_R, \alpha_R, \beta_R) &= \sum_{i=1}^{k} \log(f(x_i, \theta_L, \alpha_L, \beta_L) + \sum_{i=k+1}^{n} \log(f(x_i, \theta_R, \alpha_R, \beta_R) \\ &= \left\{ k\log(\alpha_L) - k\log(\beta_L) + (\alpha_L - 1)\sum_{i=1}^{k} \log\left(\frac{x_i - \theta_L}{\beta_L}\right) \right. \\ &\left. - \sum_{i=1}^{k} \left(\frac{x_i - \theta_L}{\beta_L}\right)^{\alpha_L} \right\} + \left\{ (n-k)\log(\alpha_R) - (n-k)\log(\beta_R) \right. \\ &\left. + (\alpha_R - 1)\sum_{i=k+1}^{n} \log\left(\frac{x_i - \theta_R}{\beta_R}\right) - \sum_{i=k+1}^{n} \left(\frac{x_i - \theta_R}{\beta_R}\right)^{\alpha_R} \right\}. \end{aligned}$$

The MLEs of the pre-change parameters θ_L, α_L and β_L can be obtained by solving the following equations.

$$\begin{aligned} \frac{\partial l_{H_1}}{\partial \alpha_L} &= \frac{k}{\alpha_L} + \sum_{i=1}^{k} \log\left(\frac{x_i - \theta_L}{\beta_L}\right) - \sum_{i=1}^{k} \left(\frac{x_i - \theta_L}{\beta_L}\right)^{\alpha_L} \log\left(\frac{x_i - \theta_L}{\beta_L}\right), \\ \frac{\partial l_{H_1}}{\partial \beta_L} &= \frac{-k\alpha_L}{\beta_L} + \frac{\alpha_L}{\beta_L} \sum_{i=1}^{k} \left(\frac{x_i - \theta_L}{\beta_L}\right)^{\alpha_L + 1}, \\ \frac{\partial l_{H_1}}{\partial \theta_L} &= (\theta_L - 1)\sum_{i=1}^{k} \left(\frac{1}{x_i - \theta_L}\right) + \frac{\alpha_L}{\beta_L} \sum_{i=1}^{k} \left(\frac{x_i - \theta_L}{\beta_L}\right)^{\alpha_L - 1}. \end{aligned}$$

and the MLEs of the post-change parameters θ_R, α_R and β_R are the solutions of the following equations.

$$\frac{\partial l_{H_1}}{\partial \alpha_R} = \frac{(n-k)}{\alpha_R} + \sum_{i=k+1}^{n} \log\left(\frac{x_i - \theta_R}{\beta_R}\right) - \sum_{i=k+1}^{n} \left(\frac{x_i - \theta_R}{\beta_R}\right)^{\alpha_R} \log\left(\frac{x_i - \theta_R}{\beta_R}\right),$$
$$\frac{\partial l_{H_1}}{\partial \beta_R} = \frac{-(n-k)\alpha_R}{\beta_R} + \frac{\alpha_R}{\beta_R} \sum_{i=k+1}^{n} \left(\frac{x_i - \theta_R}{\beta_R}\right)^{\alpha_R+1},$$
$$\frac{\partial l_{H_1}}{\partial \theta_R} = (\theta_R - 1) \sum_{i=k+1}^{n} \left(\frac{1}{x_i - \theta_R}\right) + \frac{\alpha_R}{\beta_R} \sum_{i=k+1}^{n} \left(\frac{x_i - \theta_R}{\beta_R}\right)^{\alpha_R-1}.$$

Now MIC(k) is given by,

$$\text{MIC}(k) = -2l(k, \hat{\theta}_L, \hat{\alpha}_L, \hat{\beta}_L, \hat{\theta}_R, \hat{\alpha}_R, \hat{\beta}_R) + \left\{6 + \left(\frac{2k}{n} - 1\right)^2\right\} \log n, \tag{5.2.12}$$

where $(\hat{\theta}_L, \hat{\alpha}_L, \hat{\beta}_L)$ and $(\hat{\theta}_R, \hat{\alpha}_R, \hat{\beta}_R)$ are MLEs of the parameters before and after the change respectively. If there is a change, the change point k is estimated by

$$\text{MIC}(\hat{k}) = \min_{1 \le k < n} \text{MIC}(k).$$

Furthermore, S_n can be obtained from (5.2.6). Under the Wald conditions and the regularity conditions provided by Chen et al. (2006), we have the following theorems.

Theorem 5.2.13. *As $n \to \infty$,*

$$S_n \longrightarrow \chi^2_3\,,$$

in distribution under the null hypothesis.

Theorem 5.2.14. *As $n \to \infty$, the change point satisfies $0 < k/n < 1$. Then the change point estimator $\hat{k}$ satisfies*

$$\hat{k} - k = O_p(1)\,.$$

The proofs of both Theorems are similar to ones in Chen et al. (2006).

5.3 Confidence Curve for Three Parameter Weibull Distribution

In this section, we provide steps to construct a confidence curve for the change point in a three-parameter Weibull distribution based on MIC. Most existing literature on change point problem focused on providing the point estimate of the change location. Recently, Cunen et al. (2018) proposed the confidence curve along with the confidence sets for the change point estimate through the confidence distribution (CD). The concept of a CD has its roots in Fisher's fiducial distribution. A CD is similar to a point estimator or an interval estimator but it uses a sample-dependent distribution function on the parameter space to estimate the parameter of interest. It also can provide confidence intervals of all nominal levels for a parameter of interest through confidence curves. More details and recent developments are referred to Xie and Singh (2013).

Cunen et al. (2018) used traditional log-likelihood function to obtain the point estimate of the change location, then constructed the confidence curves and confidence sets at given nominal levels through CD. However, as we mentioned in Section 5.2, this method does not consider the complexity of the model. Therefore, it is not effective due to the possible redundancy of the parameter space, especially when the change point is near 1 or n. Thus, we modify their approach by estimating the change location $\hat{k}$ using (5.2.5). The confidence curve for a three-parameter Weibull distribution can be obtained as follows.

Step 1: The profile log-likelihood function can be obtained by maximizing the log-likelihood function (5.2.12) over the parameters for each candidate value of k where $1 \leq k < n-1$.

$$\begin{aligned} l_{prof}(k) &= \max\Big(l(k, \theta_L, \alpha_L, \beta_L, \theta_R, \alpha_R, \beta_R)\Big) \\ &= l(k, \hat{\theta}_L, \hat{\alpha}_L, \hat{\beta}_L, \hat{\theta}_R, \hat{\alpha}_R, \hat{\beta}_R)\,. \end{aligned}$$

Step 2: The deviance is given by,

$$D(k, \boldsymbol{x}) = 2\{l_{prof}(\hat{k}) - l_{prof}(k)\}\,, \tag{5.3.1}$$

where $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ and $x_1, x_2, \cdots, x_k$ is a sample coming from the distribution $W(\theta_L, \alpha_L, \beta_L)$ and $x_{k+1}, \cdots, x_n$ coming from $W(\theta_R, \alpha_R, \beta_R)$. $l_{prof}(\hat{k}) = \max_{1 \leq k < n}(l_{prof}(k))$ and $\hat{k}$ is obtained by using (5.2.5) which is different from Cunen et al. (2018).

Step 3: To construct a confidence curve for k based on the deviance function, we consider the estimated distribution of $D(k, \boldsymbol{x})$ at position k as follows.

$$\Psi_k(x) = P_{k,\hat{\Theta}_L,\hat{\Theta}_R}\{D(k, \boldsymbol{x}) < x\}, \tag{5.3.2}$$

where $x \in \mathbb{R}$. In the case of continuous parameters, Wilks theorem states that $\Psi_k(x)$ is approximately the distribution function of a χ^2_1. However, Wilks theorem does not hold for a discrete parameter k. Therefore, we compute Ψ_k through the simulations. The confidence curve can be constructed as,

$$cc(k, \boldsymbol{x}_{obs}) = \Psi_k(D(k, \boldsymbol{x}_{obs})) = P_{k,\hat{\Theta}_L,\hat{\Theta}_R}\{D(k, \boldsymbol{x}) < D(k, \boldsymbol{x}_{obs})\}. \tag{5.3.3}$$

The probability that $cc(k, \boldsymbol{x}_{obs}) < \alpha$, under the true value of k, is often approximated well with α. Then, the confidence sets for k can be visualized using the plot $cc(k, \boldsymbol{x}_{obs})$. The $cc(k, \boldsymbol{x}_{obs})$ is the acceptance probability for k, or one minus the p-value for testing that value of k by using the deviance-based test which rejects the null hypothesis for high values of $D(k, \boldsymbol{x})$. We compute Ψ_k and hence $cc(k, \boldsymbol{x}_{obs})$ by simulations as follows.

$$cc(k, \boldsymbol{x}_{obs}) = \frac{1}{B}\sum_{j=1}^{B} I\{D(k, \boldsymbol{x}^*_j) < D(k, \boldsymbol{x}_{obs})\}, \tag{5.3.4}$$

for large number of B of simulated copies of data set $\boldsymbol{x}^*$. For each possible value of k, we simulate data $\boldsymbol{x}^*_j, j = 1, \cdots, B$ from $f(x, \Theta_L)$ and $f(x, \Theta_R)$ to the left and right side of k respectively. See Cunen et al. (2018) for more details. In our proposed procedure to construct the confidence curves, it is different from the Cunen et al. (2018) approach

here to estimate k. Instead of estimating the change location k by maximizing the profile-likelihood function over all possible values of k, we estimates k using (5.2.5) by considering the impact of the locations of changes. The MIC-based statistics S_n in (5.2.6) can be used to confirm a significant change statistically to avoid the fluctuations caused by noise.

5.4 Simulation Study

In this section, we conduct simulations at various values of the change point location k with different sample sizes $n = 50, 100, 150$. The pre-change distribution is always set to be $W(1, 1, 2)$ and the post-change distribution after change point k is $W(1.25, 1.25, 2.25)$, $W(1.5, 1.5, 2.5)$, and $W(1.75, 1.75, 2.75)$. Since S_n defined in (5.2.6) and T_n defined in (5.2.8) have different null rejection rates, therefore, we cannot simply compare the power of S_n and T_n directly. Thus, we consider T_n^* which denotes the power of T_n after the null rejection rate of T_n is equal to the corresponding S_n by adjusting its critical values (increase or decrease).

First, we verify the null asymptotic distribution of S_n which is stated in Theorem 5.2.13 numerically. For different sample sizes $n \in \{100, 200, 400\}$ we obtain the χ^2_3 quantile-quantile (Q-Q) plot for S_n values in Figures 5.1 - 5.3. From the plots, we observe that the null asymptotic distribution of S_n can be approximated to χ^2_3 when sample size increases. This confirms the result given in Theorem 5.2.13.

Second, we investigate the convergence of the change point estimator $\hat{k}$ in terms of the empirical probability distribution of $|\hat{k} - k| \leq \delta$ at various sample sizes and different values of the parameters where k is the true value of change location and δ is set to be the difference between the estimated and the true value of k. The results are summarized in Tables 5.1-5.4. Simulations results indicate that the change point estimates achieve the better convergence rate as the sample size increases.

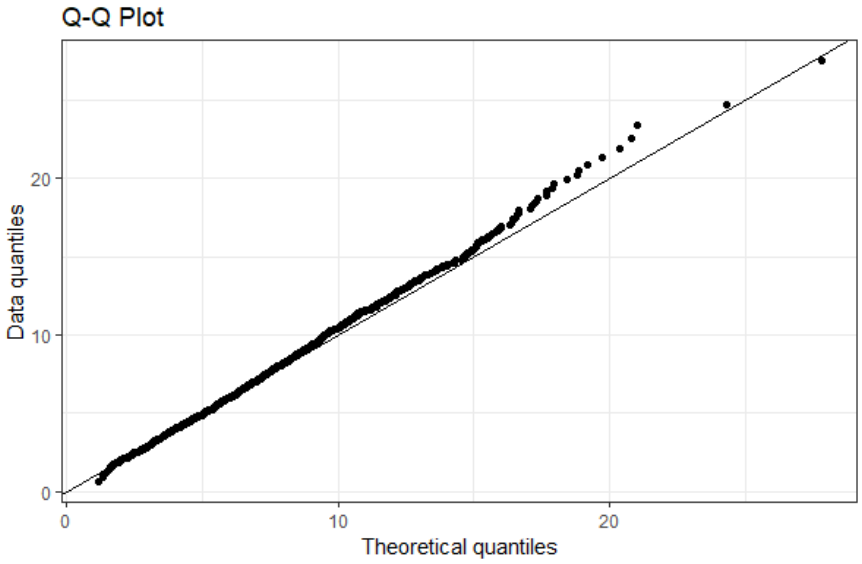

Figure 5.1 The Chi-square Q-Q plot of S_n as $n = 100$

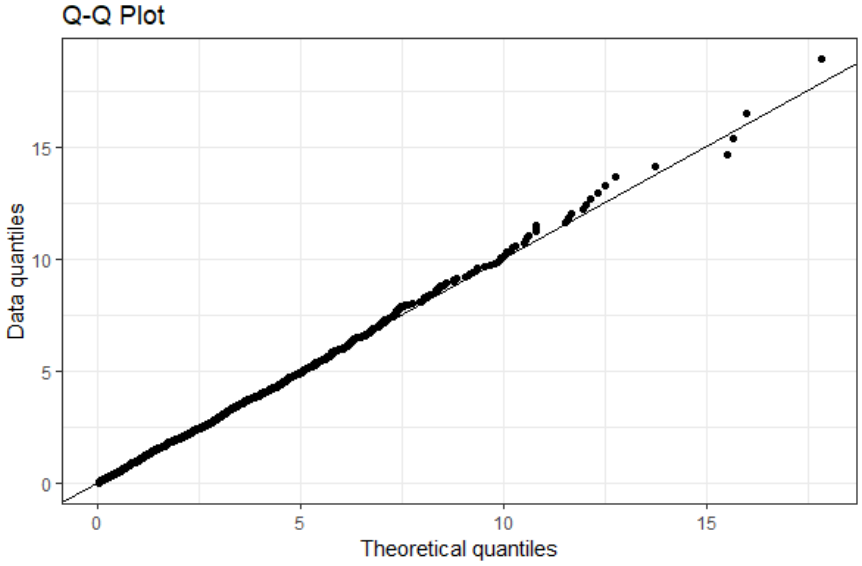

Figure 5.2 The Chi-square Q-Q plot of S_n as $n = 200$

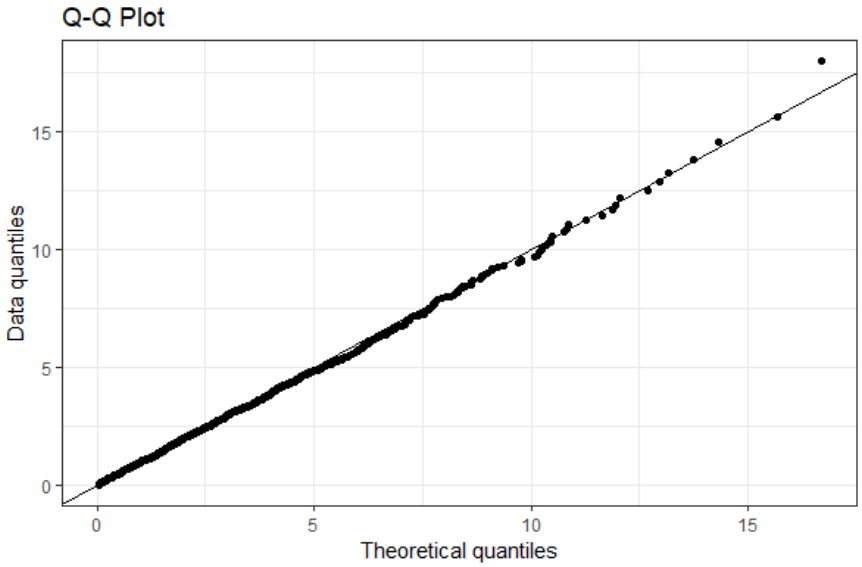

Figure 5.3 The Chi-square Q-Q plot of S_n as $n = 400$

Table 5.1 Probability distribution of $|\hat{k} - k| \leq \delta$ and the sample size is $n = 50$ at various change point locations k

	$n = 50$			
	k	$P(\|\hat{k} - k\| \leq 1)$	$P(\|\hat{k} - k\| \leq 2)$	$P(\|\hat{k} - k\| \leq 3)$
W(1.25, 1.25, 2.25)	10	0.192	0.281	0.354
	15	0.186	0.252	0.324
	25	0.191	0.278	0.331
W(1.5, 1.5, 2.5)	10	0.390	0.539	0.621
	15	0.422	0.542	0.617
	25	0.430	0.540	0.621
W(1.75, 1.75, 2.75)	10	0.593	0.703	0.768
	15	0.599	0.722	0.789
	25	0.623	0.727	0.788

The power comparisons between S_n and T_n^* with different values of parameters, samples sizes and change locations are reported in Tables 5.4-5.6. The true change point k considering $\{10, 15, 25\}, \{20, 40, 505\}, \{25, 50, 75\}$ for sample size $50, 100, 150$ respectively. When sample size $n = 50$, the change location is beyond the midpoint of the data, for example, $k \in \{40, 35\}$ are symmetric with $k \in \{10, 15\}$. In our simulation, we only consider the true change-point locations below or equal to the midpoint of the data set. From the tables, it can be clearly seen, in both pro-

Table 5.2 Probability distribution of $|\hat{k}-k| \leq \delta$ and the sample size is $n = 100$ at various change point locations k

	$n = 100$			
	k	$P(\|\hat{k}-k\| \leq 1)$	$P(\|\hat{k}-k\| \leq 2)$	$P(\|\hat{k}-k\| \leq 3)$
W(1.25, 1.25, 2.25)	20	0.234	0.351	0.418
	40	0.244	0.344	0.425
	50	0.234	0.330	0.402
W(1.5, 1.5, 2.5)	20	0.506	0.634	0.714
	40	0.508	0.641	0.722
	50	0.494	0.629	0.705
W(1.75, 1.75, 2.75)	20	0.666	0.773	0.834
	40	0.650	0.753	0.804
	50	0.647	0.753	0.796

Table 5.3 Probability distribution of $|\hat{k}-k| \leq \delta$ and the sample size is $n = 150$ at various change point locations k

	$n = 150$			
	k	$P(\|\hat{k}-k\| \leq 1)$	$P(\|\hat{k}-k\| \leq 2)$	$P(\|\hat{k}-k\| \leq 3)$
W(1.25, 1.25, 2.25)	25	0.277	0.385	0.468
	50	0.286	0.401	0.477
	75	0.300	0.398	0.469
W(1.5, 1.5, 2.5)	25	0.548	0.680	0.767
	50	0.507	0.640	0.701
	75	0.507	0.627	0.697
W(1.75, 1.75, 2.75)	25	0.689	0.796	0.846
	50	0.633	0.733	0.783
	75	0.621	0.708	0.752

cedures, the power increases as the difference between the parameters increase. Further, the power tends to increase when sample size increases. Unlike T_n^*, the S_n method considers the change point location. As a results, the power based on the S_n method is higher than the T_n^* method when the change location is further away from the middle of the data. The results of these two methods are graphed in Figure 5.4.

Table 5.4 Power comparison for sample size $n = 50$ at different change point locations $k \in \{10, 15, 25\}$

	k	W(1.25, 1.25, 2.25)	W(1.5, 1.5, 2.5)	W(1.75, 1.75, 2.75)
T_n^*	10	0.452	0.531	0.691
	15	0.499	0.686	0.801
	25	0.462	0.748	0.907
S_n	10	0.490	0.568	0.716
	15	0.542	0.724	0.821
	25	0.523	0.786	0.927

Table 5.5 Power comparison for sample size $n = 100$ at different change point locations $k \in \{20, 40, 50\}$

	k	W(1.25, 1.25, 2.25)	W(1.5, 1.5, 2.5)	W(1.75, 1.75, 2.75)
T_n^*	20	0.657	0.736	0.964
	40	0.752	0.927	0.998
	50	0.714	0.932	0.999
S_n	20	0.710	0.768	0.970
	40	0.806	0.939	0.998
	50	0.767	0.953	0.999

Table 5.6 Power comparison for sample size $n = 150$ as at different change point locations $k \in \{25, 50, 75\}$

	k	W(1.25, 1.25, 2.25)	W(1.5, 1.5, 2.5)	W(1.75, 1.75, 2.75)
T_n^*	25	0.765	0.796	0.992
	50	0.895	0.970	1.000
	75	0.878	0.990	1.000
S_n	25	0.811	0.829	0.994
	50	0.916	0.979	1.000
	75	0.902	0.991	1.000

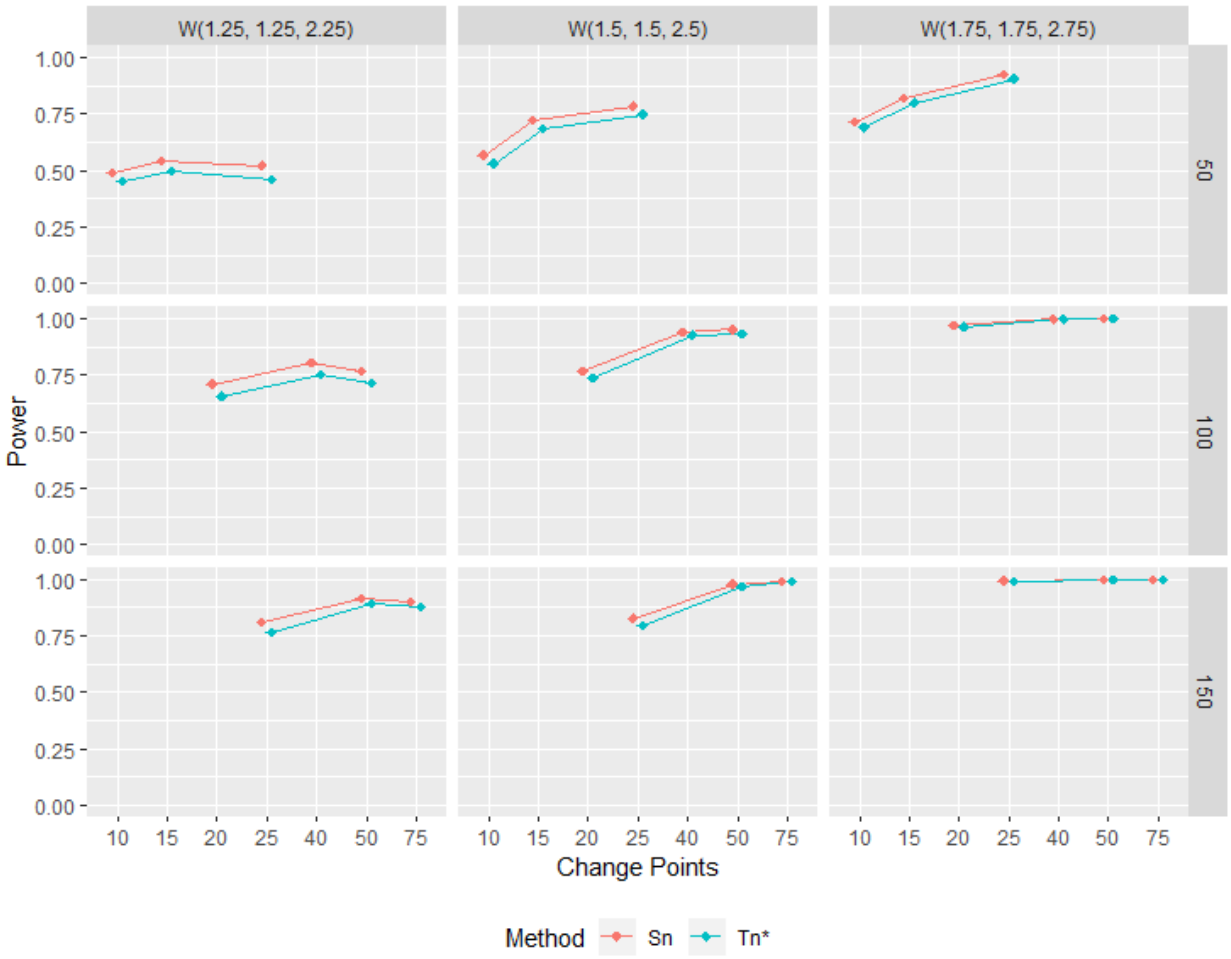

Figure 5.4 The power comparison of two methods with various change point locations and different sample size $n \in \{50, 100, 150\}$

We also conduct simulations to compare the proposed method for the confidence curve based on MIC in Section 5.3 and the method proposed by Cunen et al. (2018) in terms of coverage probabilities and average confidence sets of change points at various scenarios. A confidence set of a change point can be defined as $\{k : cc(k, x) \leq \alpha\}$. The size of a confidence set is determined by the number of k belonging to the confidence set with a given nominal level. For sample size $n = 50$, we generate the data set with true change point locations $k \in \{10, 15, 25\}$ and $k \in \{20, 40, 50\}$ are the true change point locations for sample size $n = 100$. The results are summarized in Tables 5.7 - 5.10. The MIC based method provides better coverage probabilities when the change point is located in the beginning or the ending of the data set. Not surprisingly,

both methods provide approximately same coverage probabilities when the change occurs in the middle of the data set. Further, the change point location k is close to 1 or n, MIC based approach provides thinner confidence sets compare to the traditional log-likelihood based approach used by Cunen et al. (2018). Similarly, if the change occurs in the center, the average size of the confidence sets are roughly equal for both methods. Thus, the comparisons confirms our method outperforms Cunen et al. (2018) method in terms better coverage probability and thinner confidence sets.

Table 5.7 The coverage probability comparison for all parameters change at various change point locations, $k \in \{10, 15, 25\}$ and the sample size $n = 50$

		$k = 10$		$k = 15$		$k = 25$	
	α	MIC	loglik	MIC	loglik	MIC	loglik
W(1.25, 1.25, 2.25)	0.50	0.61	0.61	0.68	0.67	0.66	0.67
	0.90	0.80	0.78	0.82	0.80	0.81	0.80
	0.95	0.83	0.81	0.86	0.84	0.85	0.84
	0.99	0.89	0.87	0.90	0.88	0.91	0.90
W(1.5, 1.5, 2.5)	0.50	0.69	0.68	0.73	0.72	0.72	0.72
	0.90	0.83	0.82	0.85	0.84	0.88	0.88
	0.95	0.88	0.87	0.89	0.88	0.91	0.91
	0.99	0.92	0.91	0.94	0.93	0.96	0.95
W(1.75, 1.75, 2.75)	0.50	0.70	0.70	0.72	0.72	0.75	0.76
	0.90	0.82	0.82	0.86	0.86	0.88	0.89
	0.95	0.87	0.87	0.90	0.90	0.91	0.92
	0.99	0.92	0.92	0.96	0.95	0.96	0.96

Table 5.8 The confidence sets comparison for all parameters change at various change point locations, $k \in \{10, 15, 25\}$ and the sample size $n = 50$

		$k = 10$		$k = 15$		$k = 25$	
	α	MIC	loglik	MIC	loglik	MIC	loglik
W(1.25, 1.25, 2.25)	0.50	17.39	17.66	18.13	18.48	18.41	18.96
	0.90	25.26	25.28	25.57	25.75	25.86	26.22
	0.95	28.25	28.18	28.36	28.40	28.67	28.93
	0.99	33.51	33.26	33.84	33.64	34.22	34.11
W(1.5, 1.5, 2.5)	0.50	13.37	13.96	13.35	13.94	13.69	14.02
	0.90	18.51	19.03	18.36	19.01	18.72	18.99
	0.95	20.82	21.33	20.57	21.22	21.03	21.25
	0.99	26.12	26.48	25.79	26.36	26.21	26.30
W(1.75, 1.75, 2.75)	0.50	10.53	10.75	10.26	10.47	10.48	10.66
	0.90	14.18	14.42	13.41	13.63	13.83	14.00
	0.95	15.92	16.19	14.92	15.15	15.38	15.54
	0.99	20.35	20.61	18.87	19.08	19.21	19.30

Table 5.9 The coverage probability comparison for all parameters change at various change point locations, $k \in \{20, 40, 50\}$ and the sample size $n = 100$

		$k = 20$		$k = 40$		$k = 50$	
	α	MIC	loglik	MIC	loglik	MIC	loglik
W(1.25, 1.25, 2.25)	0.50	0.73	0.72	0.76	0.75	0.75	0.76
	0.90	0.88	0.87	0.89	0.88	0.90	0.90
	0.95	0.91	0.90	0.93	0.92	0.93	0.90
	0.99	0.96	0.94	0.97	0.96	0.96	0.96
W(1.5, 1.5, 2.5)	0.50	0.75	0.73	0.76	0.74	0.78	0.78
	0.90	0.90	0.89	0.91	0.90	0.91	0.91
	0.95	0.94	0.93	0.95	0.95	0.95	0.94
	0.99	0.97	0.96	0.98	0.98	0.98	0.98
W(1.75, 1.75, 2.75)	0.50	0.79	0.78	0.80	0.80	0.78	0.78
	0.90	0.91	0.91	0.90	0.90	0.90	0.90
	0.95	0.94	0.94	0.93	0.93	0.95	0.95
	0.99	0.97	0.97	0.97	0.97	0.98	0.98

Table 5.10 The confidence sets comparison for all parameters change at various change point locations, $k \in \{20, 40, 50\}$ and the sample size $n = 100$

		$k = 20$		$k = 40$		$k = 50$	
	α	MIC	loglik	MIC	loglik	MIC	loglik
W(1.25, 1.25, 2.25)	0.50	27.64	30.09	26.84	29.05	27.08	28.61
	0.90	38.45	40.78	36.57	38.85	37.34	38.92
	0.95	43.09	45.24	41.01	43.22	41.71	43.27
	0.99	54.08	55.74	50.91	52.79	51.89	53.26
W(1.5, 1.5, 2.5)	0.50	18.66	19.33	17.61	18.03	18.08	18.19
	0.90	24.34	25.10	22.24	22.68	23.03	23.10
	0.95	26.95	27.73	24.42	24.85	25.20	25.25
	0.99	33.27	34.06	29.71	30.15	30.71	30.693
W(1.75, 1.75, 2.75)	0.50	12.56	12.66	13.38	13.47	14.27	14.26
	0.90	15.84	15.97	16.03	16.15	16.98	16.98
	0.95	17.43	17.57	17.28	17.40	18.26	18.26
	0.99	21.32	21.48	20.50	20.63	21.34	21.33

5.5 Application

In this section, we used annual maximum rainfall data at one rain gauge in Fort Collins, Colorado from 1900 through 1999. The data is available under extRemes package in R software, see Gilleland and Katz (2016). The data consists 100 observations and it is graphed in Figure 5.5. The Figure 5.6 shows the autocorrelation function (ACF) for the data.

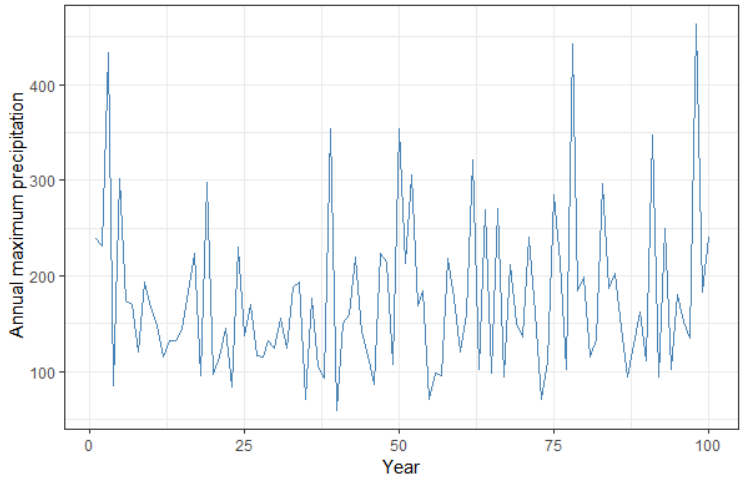

Figure 5.5 The annual maximum rainfall data at one rain gauge in Fort Collins, Colorado

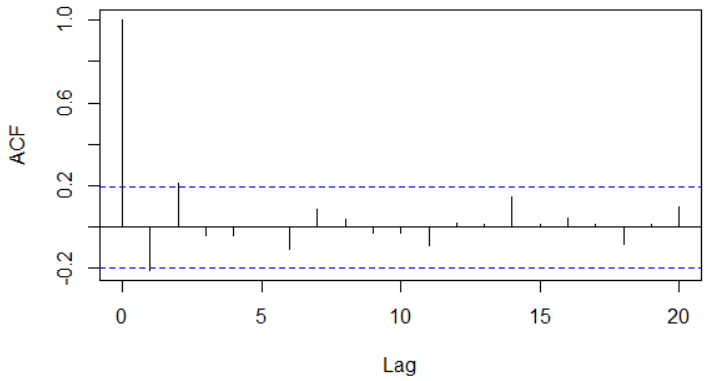

Figure 5.6 The auto-correlation plot for annual maximum rainfall data at one rain gauge in Fort Collins, Colorado

To ensure the independence in the time series data, we use the Portmanteau test statistic to check independence and normality of the transform data set. The Portmanteau test statistic is

given as,

$$Q_k = n \sum_{i=1}^{k} r_i^2,$$

where r_i are the autocorrelation coefficient at lag i, and k is the lag up to which the autocorrlation coefficient function of the data. Under the null hypothesis $Q_k \sim \chi_k^2$. Using the Portmanteau test,

$$Q_{20} = 100 \sum_{i=1}^{20} r_i^2 = 16.025735 < \chi_{0.95}^2(20) = 31.41043.$$

Thus, we fail to reject the null hypothesis which leads to the independence of the data. Since $MIC(n) = MIC(100) = 1156.3874 > \min_{2\leq k\leq n-2} MIC(k) = MIC(2) = 1143.4111$, the estimated change location is $\hat{k} = 2$. The corresponding test statistic $S_n = 27.6097$ with the critical value $\chi_{0.05,3}^2 = 7.815$ and the p-value 4.3858×10^{-6}. It confirms the change in the data. With the conventional SIC, we have $\min_{2\leq k\leq n-2} SIC(k) = SIC(2)$. Therefore, two methods provide the same conclusion. For potential multiple changes in the data, the binary segmentation method by Vostrikova (1981) is applied. Such a method decomposes the detecting procedure into several steps with assuming at most one change at each step. This process is repeated until no more change point is detected. With the binary segmentation method, the multiple change locations in the data are $\{2, 6, 17, 39, 94\}$. The confidence sets for the change point estimates $\hat{k} \in \{2, 6, 17, 39, 94\}$ are $\{2\}$, $\{5, 6, 7, 8\}$, $\{17, 18, 39, 40\}$, $\{34, 35, 37, 38, 39, 47, 50\}$, $\{3, 4, 44, 47, 71, 72, 73, 74, 75, 94, 95\}$ respectively. Figures 5.7 - 5.9 show the confidence curves for the estimated change locations and the 95% confidence set is marked by the horizontal red dashed line. Figure 5.10 indicates all the estimated changes in the data.

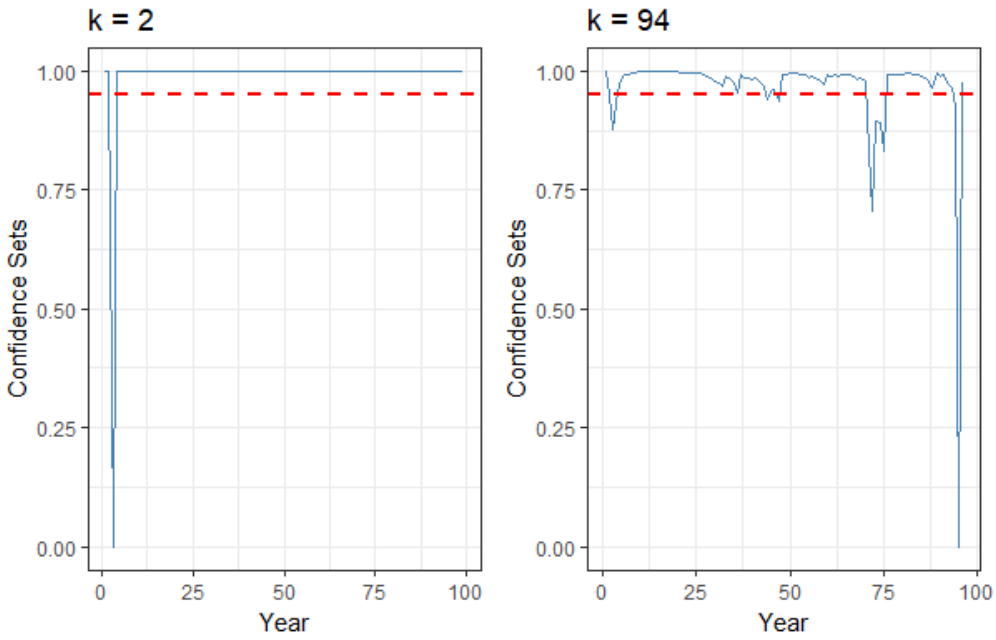

Figure 5.7 Left: Confidence curve for change point at $\hat{k} = 2$, Right: Confidence curve for the second subset above $(2 < k \leq 100)$, the $\hat{k} = 94$

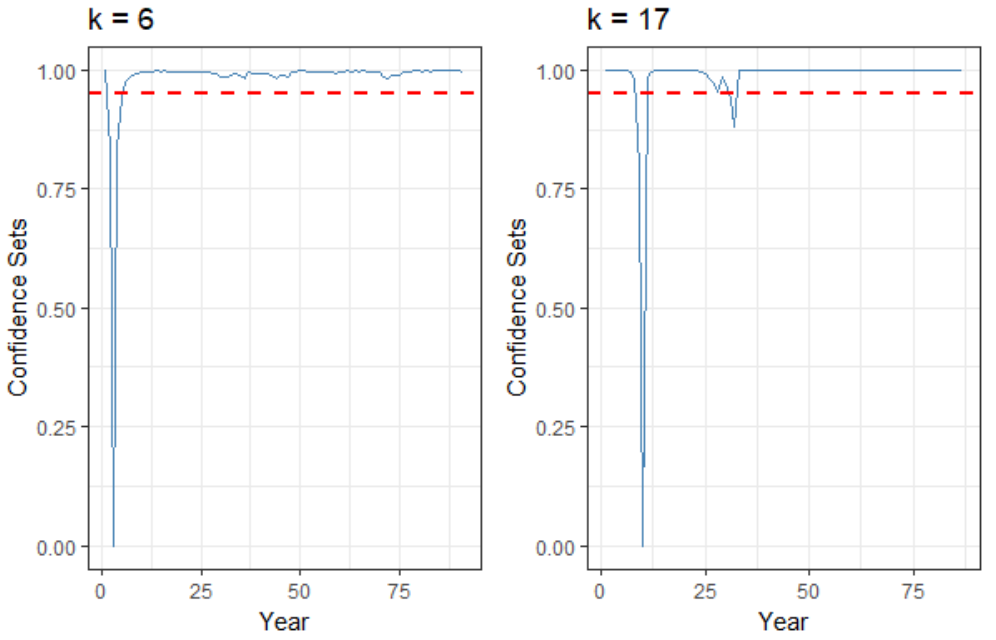

Figure 5.8 Left: Confidence curve for the third subset below $(2 < k \leq 94)$, the $\hat{k} = 6$, Right: Confidence curve for the fourth subset $(7 \leq k \leq 94)$, the $\hat{k} = 17$

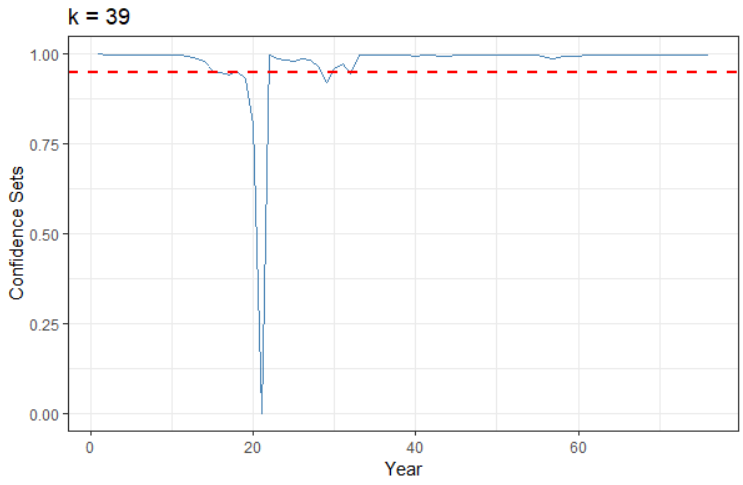

Figure 5.9 Left: Confidence curve for the fifth subset $(17 < k \leq 94)$, the $\hat{k} = 39$

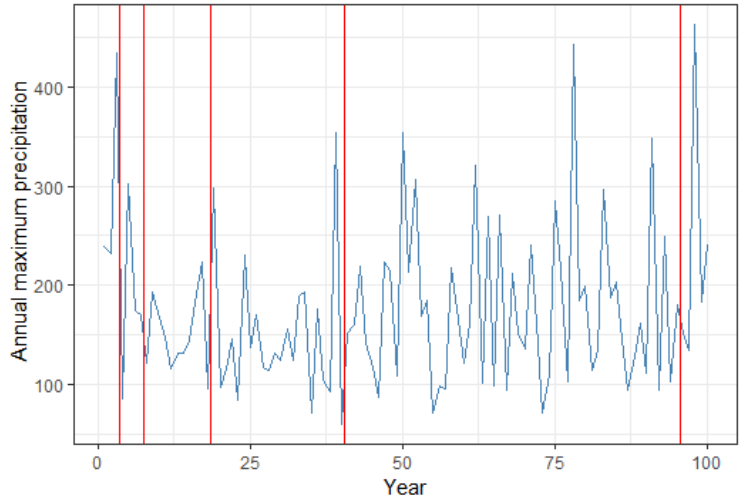

Figure 5.10 The annual maximum rainfall data at one rain gauge in Fort Collins, Colorado with change point locations

5.6 Conclusion

In this chapter, we propose a change point detection method for a three parameter Weibull distribution simultaneous based on the modified information criterion. The simultaneous changes in the parameters are considered. The asymptotic properties for the associated test statistic has been established. Moreover, we propose a modified approach to construct a confidence curve and to find a corresponding confidence sets for a change locations at a given significance level through confidence distributions. Simulations are conducted to compare the proposed methods

with the procedure based on the conventional SIC and the method provided in Cunen et al. (2018) in terms of powers, coverage probabilities and average lengths of confidence sets at various scenarios. Simulations indicate that our method is competitive to other existing methods and even better when the change happen at the beginning or in the end of the data. Along with binary segmentation method, the methods are applied to detect multiple change points and construct corresponding confidence sets for the annual maximum rainfall data.

CHAPTER 6 CONCLUSION AND FUTURE WORK

6.1 Conclusion

In this dissertation, we proposed sequential change-point detection procedure in linear regression and linear quantile regression models for high-dimensional data. Sequential change-point problem in linear regression under high-dimensional settings is studied in Chapter 2. We proposed test statistics for the sequential change-point detection procedure based on SCAD penalized regression model with finite monitoring horizon for high-dimensional data. The pre-change coefficients are replaced by the SCAD penalized estimator. In addition, we examined the advantages of both open- and closed-end procedures. The monitoring process for the open-end procedure may continue to infinity if no alarm is raised. Under the closed-end procedure the monitoring process stops after a certain number of observations being collected even if no change is detected. The asymptotic properties of the proposed test statistics were derived. The asymptotic critical values for both methods obtained through simulation. Based on the simulation study the open-end procedure produced slightly deated Type I error for smaller N. The larger control parameter value would be preferred when the change occurs shortly after the historical sample size.

In Chapter 3 we studied the monitoring process using penalized quantile regression for high-dimensional data sequentially. In this study, the SCAD penalized quantile regression (SPQR) was considered. For a specific quantile level α , we developed the CUSUM test statistic for the sequential monitoring process for both open- and closed-end procedures. In order to improve the monitoring method, a modified test statistic based on the post - SCAD penalized quantile regression (P-SPQR) estimator was proposed. The asymptotic distribution of the test statistic under the null and alternative hy-pothesis were derived. A simulation study was conducted to examine the performance of the proposed method.

In Chapter 4 and Chapter 5 we developed a change point detection procedure based on the confidence distribution combining with the modified information criterion to construct the confidence

set for the change location for Skew Normal and Weibull distribution respectively. Simulations results indicate the advantages of the proposed method comparing to the existing method in terms of coverage probabilities and average lengths of the confidence sets, especially when the change occurs at the very beginning or in the very end. Our proposed method applied to real datasets to illustrate the advantages.

6.2 Future Work

Recently, Li, Xu, Zhong, and Li (2019) proposed a method to detect change points in the mean of high-dimensional time series data. We would like to extend their method to monitor the structural change sequentially. In this research, we will establish the asymptotic properties of the test statistic and conduct simulation studies to evaluate the performance through the comparison with other existing methods.

We are interested to study the sequential change-point detection based on penalized high-dimensional empirical likelihood method. In such a scenario, the sample size is a random variable and the null hypothesis of sequential structural stability will be rejected as soon as a change is detected. Therefore, the objective is to detect such a change with a minimum number of false alarms. A nonparametric testing procedure based on penalized high-dimensional empirical likelihood method will be proposed and related asymptotic results will be studied.

Principal component analysis (PCA) is a fundamental dimension reduction tool in statistics. We will study the rich properties of PCA and to select the significant explanatory variables which influence the response variable Y. Corresponding asymptotic properties and simulations will be conducted. Comparisons to other existing methods will be studied to illustrate advantages.

BIBLIOGRAPHY

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC-19*, 716–723.

Alghamdi, A., W. Ning, and A. K. Gupta (2018). An information approach for the change point problem of the rayleigh lomax distribution. *International Journal of Intelligent Technologies and Applied Statistics 11(4)*, 233–254.

Arellano-Valle, R. B., C. L. M. and R. H. Loschi (2013). Change point detection in the skew-normal model parameters. *Communications in Statistics - Theory and Methods 42*, 603–618.

Aue, A., R. C. Y. Cheung, T. C. M. Lee, and M. Zhong (2014). Segmented model selection in quantile regression using the minimum description length principle. *Journal of the American Statistical Association 109*, 1241–1256.

Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scandinavian Journal of Statistics 12*, 171–178.

Azzalini, A. and A. Capitanio (2014). *The Skew-Normal and Related Families*. Cambridge University Press: New York.

Azzalini, A. and A. Dalla Valle (1996). The multivariate skew-normal distribution. *Biometrika 83*, 715–726.

Bai, J. (1996). Testing for parameter constancy in linear regressions: An empirical distribution function approach. *Econometrica 64(3)*, 597–622.

Belloni, A. and V. Chernozhukov (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli 19(2)*, 521–547.

Cai, X., K. K. Said, and W. Ning (2016). Change-point analysis with bathtub shape for the exponential distribution. *Journal of Applied Statistics 43*, 2740–2750.

Chen, J. and A. K. Gupta (1995). Likelihood procedure for testing change points hypothesis for multivariate gaussian model. *Random Operators and Stochastic Equations 3*, 235–244.

Chen, J. and A. K. Gupta (1997). Testing and locating variance change points with application to stock prices. *Journal of the American Statistical Association 92*, 739—-747.

Chen, J. and A. K. Gupta (1999). Change point analysis of a gaussian model. *Statistical Papers 40*, 323—-333.

Chen, J. and A. K. Gupta (2012). *Parametric Statistical Change Point Analysis with Applications to Genetics, Medicine, and Finance*. 2nd ed. Boston: Birkhauser.

Chen, J., A. K. Gupta, and J. Pan (2006). Information criterion and change point problem for regular models. *The Indian Journal of Statistics 68(2)*, 252–282.

Chernoff, H. and S. Zacks (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Annals of Mathematical Statistics 35*, 999—-1018.

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

Ciuperca, G. (2014). Model selection by lasso methods in a change-point model. *Statistical Papers 55*, 349–374.

Ciuperca, G. (2017). Real time change-point detection in a nonlinear quantile model. *Journal Sequential Analysis 36(1)*, 87–110.

Cosörgö, M. and L. Horváth (1997). Limit theorems in change-point analysis. *Wiley, New York*.

Cox, D. (1958). Some problems with statistical inference. *The Annals of Mathematical Statistics 29*, 357–372.

Cunen, C., G. Hermansen, and N. L. Hjort (2018). Confidence distributions for change-points and regime shifts. *Journal of Statistical Planning and Inference 195*, 14–34.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*, 1348—-1360.

Fan, J. and H. Peng (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics 32(3)*, 928–961.

Furno, M. (2007). Parameter instability in quantile regression. *Statistical Modelling 7*, 345–362.

Furno, M. (2012). Tests for structural break in quantile regressions. *AStA Advances in Statistical Analysis 96(4)*, 493–515.

Gardner, L. A. (1969). On detecting changes in the mean of normal variables. *Annals of Mathematical Statistics 40*, 116–126.

Gilleland, E. and R. W. Katz (2016). extremes 2.0: An extreme value analysis package in r. *Journal of Statistical Software 72(8)*, 1–39.

Golub, G. H., M. Heath, and G. Wahba (1979). Generalised cross validation as a method for choosing a good ridge parameter. *Technometrics 21(2)*, 215–223.

Gupta, A. K. and J. Chen (1996). Detecting changes of mean in multidimensional normal sequences with application to literature and geology. *Computational Statistics 11*, 211—221.

Gurevich, G. and A. Vexler (2005). Change point problems in the model of logistic regression. *Journal of Statistical Planning and Inference 131*, 313–331.

Harchaoui, Z. and C. Lévy-Leduc (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association 105(402)*, 1480–1493.

Hasan, A., W. Ning, and A. K. Gupta (2014). An information-based approach to the change-point problem of the noncentral skew t distribution with applications to stock market data. *Sequential Analysis: Design Methods and Applications 33*, 458–474.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning*. New York, NY, USA, (2001): Springer Series in Statistics Springer New York Inc.

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

Hawkins, D. M. (1992). Detecting shifts in functions of multivariate location and covariance parameters. *Journal of Statistical Planning and Inference 33*, 233–244.

Hirotsu, C., S. Kuriki, and A. J. Hayter (1992). Multiple comparison procedure based on themaximal component of the cumulative chi-squared statistic. *Biometrika 79*, 381–392.

Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12(1)*, 55–67.

Horváth, L., M. Huškova, P. Kokoszka, and J. Steinebach (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference 126*, 225–251.

Horváth, L., P. Kokoszka, and J. Steinebach (2007). On sequential detection of parameter changes in linear regression. *Statistics Probability Letters 77(9)*, 885–895.

Hsu, D. A. (1977). Tests for variance shifts at an unknown time point. *Applied Statistics 26, No.3*, 279—284.

Hsu, D. A. (1979). Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis. *Journal of the American Statistical Association 74*, 31—40.

Huang, J. and H. Xie (2007). Asymptotic oracle properties of scadpenalized least squares estimators. *Asymptotics: Practices, Processes and Inverse Problems (IMS Lecture Notes-Monograph Series) 55*, 149–166.

Inclán, C. (1993). Detection of multiple changes of variance using posterior odds. *Journal of Business Economic Statistics 11*, 189–300.

Jandhyala, V. K., F. S. B. and N. Evaggelopoulos (1999). Change point methods for weibull models with applications to detection of trends in extreme temperatures. *Environmetrics 10*, 547–564.

Jurušková, D. (2007). Maximum log-likelihood ratio test for a change in three param- eter weibull distribution. *Journal of Statistical Planning and Inference 137*, 1805–1815.

Kai, B., R. Li, and H. Zou (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of Statistics 39(1)*, 305–332.

Kim, J. and H. J. Kim (2008). Asymptotic results in segmented multiple regression. *Journal of Multivariate Analysis 99(9)*, 2016–2038.

Knight, K. and W. Fu (2000). Asymptotics for lasso-type estimators. *Annals of Statistics 28(5)*, 1356–1378.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis 91(1)*, 74–89.

Koenker, R. (2005). Quantile regression. *Cambridge University Press. The Edinburgh Building, Cambridge, UK*.

Koenker, R. and G. Bassett (1978). Regrssion quantiles. *Econometrica 46*, 33–50.

Koenker, R., P. Ng, and S. Portnoy (1994). Quantile smoothing splines. *Biometrika 81*, 673–680.

Komlós, J., P. Major, and G. Tusnády (1975). An approximation of partial sums of independent rv's, and the sample df. i. z. *Wahrsch. Verw. Gebiete 32*, 111–131.

Komlós, J., P. Major, and G. Tusnády (1976). An approximation of partial sums of independent rv's, and the sample df. ii. z. *Wahrsch. Verw. Gebiete 34*, 33–58.

Li, J., M. Xu, P.-S. Zhong, and L. Li (2019). Change point detection in the mean of high-dimensional time series data under dependence. *https://arxiv.org/pdf/1903.07006.pdf accessed 08/25/2019*.

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

Mallows, C. L. (1973). Some comments on c_P. *Technometrics 15 (4)*, 661–675.

Mei, Y. (2006). Sequential change-point detection when unknown parameters are present in the pre-change distribution. *The Annals of Statistics 34(1)*, 92–122.

Ngunkeng, G. and W. Ning (2014). Information approach for the change-point detection in the skew normal distribution and its applications. *Journal Sequential Analysis 33(4)*, 475–490.

Ning, W. and A. K. Gupta (2009). Change point analysis for generalized lambda distributions. *Communications in Statistics-Simulation and Computation 38*, 1789–1802.

Page, E. S. (1954). Continue inspection schemes. *Biometrika 41*, 100–135.

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

Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics 146(1)*, 170–184.

Ramanayake, A. and A. K. Gupta (2010). Testing for a change point in a sequence of exponential random variables with repeated values. *Journal of Statistical Computation and Simulation 80*, 191–199.

Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics 8*, 411–430.

Said, K. K., W. Ning, and Y. B. Tian (2019). Modified information criterion for testing changes in skew normal model. *Brazilian Journal of Probability and Statistics 33*, 280–300.

Scheetz, T. E., A.-Y. Kim, E. R. Swiderski, A. R. Philp, T. A. Braun, K. L. Knudtson, A. M. Dorrance, G. F. DiBon, J. Huang, T. L. Casavant, V. C. Sheffield, and E. M. Stone (2006). Reg-

ulation of gene expression in the mammalian eye and its relevance to eye disease. *Proceedings of the National Academy of Sciences 103*, 14429–14434.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*, 461–464.

Schweder, T. and N. Hjort (2002). Confidence and likelihood. *Scandinavian Journal of Statistics 29(2)*, 309—332.

Sen, A. K. and M. S. Srivastava (1975a). On tests for detecting change in mean. *Annals of Statistics 3*, 98—-108.

Sen, A. K. and M. S. Srivastava (1975b). Some one-sided tests on change in level. *Technometrics 17*, 61—64.

Shen, J., R. Y. Liu, and M. Xie (2018). Prediction with confidence-a general framework for predictive inference. *Journal of Statistical Planning and Inference 195*, 126–140.

Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl. 8(1)*, 22—-46.

Singh, K. and M. Xie (2012). Cd posterior-combining prior and data through confidence distributions. in contemporary developments in bayesian analysis and statistical decision theory:. *A Festchrift in Honor of Williams E. Strawderman*, (D.Fourdrinier et al. eds.). IMS Collection 8, 200–214. IMS, Beachwood, OH.

Singh, K., M. Xie, and W. E. Strawderman (2005). Combining information from independent sources through confidence distributions. *The Annals of Statistics 33(1)*, 159—-183.

Singh, K., X. M. and W. E. Strawderman (2007). Confidence distribution (cd): distribution estimator of a parameter. *Lecture Notes-Monograph Series Vol. 54, Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, 132—150.

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

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

Vostrikova, L. J. (1981). Detecting "disorder" in multidimensional random processes. *Soviet Mathematics: Doklady 24*, 55–59.

Wang, H. and X. He (2007). Detecting differential expressions in genechip microarray studies: a quantile approach. *Journal of the American Statistical Association 102*, 104–112.

Wang, L., Y. Wu, and R. Li (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association 107(497)*, 214–222.

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

Wu, Y. (2008). Simultaneous change point analysis and variable selection in a regression problem. *Journal of Multivariate Analysis 99*, 2154–2171.

Wu, Y. and Y. Liu (2009). Variable selection in quantile regression. *Statistics Sinica 19*, 801–817.

Xie, M. and K. Singh (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review 81(1)*, 2—-39.

Zhang, L., H. J. Wang, and Z. Zhu (2014). Testing for change points due to a covariate threshold in quantile regression. *Statistica Sinica 24*, 1859–1877.

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

Zhao, P. and B. Yu (2006). On model selection consistency of lasso. *Journal of Machine Learning Research 7*, 2541–2563.

Zhou, M., H. J. Wang, and Y. Tang (2015). Sequential change point detection in linear quantile regression models. *Statistics and Probability Letters 100*, 98–103.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101(476)*, 1418–1428.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society 67(2)*, 301–320.

Zou, H. and M. Yuan (2008). Composite quantile regression and the oracle model selection theory. *Annals of Statistics 36(3)*, 1108–1126.

APPENDIX A SELECTED R PROGRAMS

Chapter 2

For Small p : Type I error - Lasso - SCAD - Open-end

```
# return the g function:
Gstar=function(m,Tm, gamma){
  N=m*Tm
  k=1:N
  Gvalue=sqrt(m)*(1+k/m)*(k/(m+k))^(gamma)
  return(Gvalue)
}
p = 10 #Number of predictor variables
stat=function(y, X, m, Tm, gamma){
  y_train=y[1:m]
  x_train=X[(1:m),]
  y_test=y[-(1:m)]
  x_test=X[-(1:m),]
  lasso.scad = cv.ncvreg(x_train, y_train, penalty = "SCAD")
  bestlam = lasso.scad$lambda.min
  coef = coef(lasso.scad, s = "lambda.min")
  pstar = length(coef[which(coef != 0)])
  lasso.pred = predict(lasso.scad, X = x_test,
            type = "response", lambda = bestlam)
  lasso.pred1 = predict(lasso.scad, X = x_train,
            type = "response", lambda = bestlam)
  score1= cumsum(y_test - lasso.pred)
```

```
  Gvalue = Gstar(m, Tm, gamma)
  sigma1= sqrt((m-1)/(m-pstar))*sd(y_train - lasso.pred1)
  value = abs(score1)/sigma1/Gvalue
  result=list(stat = value)
  return(result)
}
p = 10
sim=function(nsim = 2500, m,Tm,gamma,alpha,cval1, k0){
  max.Gamma1=cpt1=NULL
  for (i in 1:nsim){
    N=m*Tm
    n=m+N
    idx=1:n
    beta0 = c(-2, 0, 2, 0, 10, 1, 0, 0, 8, -5)
    X = matrix(rnorm(n * p, 0,1), ncol = p)
    y=vector(mode="numeric",length=n)
    X[,3]=rnorm(n,2,1)
    X[,4]=rnorm(n,4,1)
    X[,5]=rnorm(n,5,1)
    ei_N=rnorm(n,0,1)    # the model error
    y= X%*%beta0 + ei_N
    Gstat1=stat(y,X,m,Tm,gamma)
    Gamma1=Gstat1$stat
    temp.cpt1 = apply(matrix(cval1,ncol=1), 1,
                      function(x) which(Gamma1>x)[1])
    cpt1 = rbind(cpt1, temp.cpt1)
    max.Gamma1 = c(max.Gamma1, max(Gamma1))
```

```
    #if(i%%500==0) print(i)
  }
  ##type I error (reject when cpt is not NA)
  TypeI1 = 1-apply(cpt1, 2, function(x) sum(is.na(x))/nsim)
  out = list(TypeI1=TypeI1, cpt1=cpt1, max.Gamma1=max.Gamma1)
  return(out)
}
# Example
d = 0
for(Tm in c(2, 4, 10, 20, 50)){
  #cval1=cvalue_closed(gamma,N,alpha)
  cval1 = 2.486683
  for(m in ms){
    sim1=sim(nsim, m,Tm,gamma, alpha,cval1, k0)
    cat("Type I :", c(sim1$TypeI1), "\n")
  }}
```

Detection Time: Open-end Procedure

```
Gstar=function(m,Tm, gamma){
  N=m*Tm
  k=1:N
  Gvalue=sqrt(m)*(1+k/m)*(k/(m+k))^(gamma)
  return(Gvalue)
}
### Calculate the test statistics
p = 10 #Number of predictor variables
stat=function(y, X, m, Tm, gamma){
  y_train=y[1:m]
```

```
  x_train=X[(1:m),]
  y_test=y[-(1:m)]
  x_test=X[-(1:m),]
  lasso.scad = cv.ncvreg(x_train, y_train, penalty = "SCAD")
  bestlam = lasso.scad$lambda.min
  coef = coef(lasso.scad, s = "lambda.min")
  pstar = length(coef[which(coef != 0)])
  lasso.pred = predict(lasso.scad, X = x_test,
                type = "response", lambda = bestlam)
  lasso.pred1 = predict(lasso.scad, X = x_train,
                type = "response", lambda = bestlam)
  score1= cumsum(y_test - lasso.pred)
  Gvalue = Gstar(m, Tm, gamma)
  sigma1= sqrt((m-1)/(m-pstar))*sd(y_train - lasso.pred1)
  value = abs(score1)/sigma1/Gvalue
  result=list(stat = value)
  return(result)
}
p  = 10 #Number of predictor variables
sim=function(nsim=2500,m,Tm,gamma, cvalue, k0){
  max.dstat= max.loc = cpt=NULL
  for (i in 1:nsim){
    N=m*Tm
    n=m+N
    idx=1:n
    beta1=c(0,0,2,0,0,1,0,0,1,0)
    beta2=c(0,0,0,3,0,0,1,0,0,-1)
```

```
    X = matrix(rnorm(n * p, 0,1), ncol = p)
    X2 = matrix(rnorm(n * p)+0.8, ncol = p)
    y = vector(mode = "numeric", length = n)
    X[, 3] = rnorm(n, 2, 1)
    X[, 4] = rnorm(n, 4, 1)
    X[, 5] = rnorm(n, 5, 1)
    ei_N = rnorm(n, 0, 1)  # the model errors
    y[1:(m + k0)] = X[1:(m + k0), ] %*% beta1 + ei_N[1:(m+k0)]
    y[(m + k0 + 1):n] = X2[(m + k0 + 1):n, ] %*% beta2 +
                        ei_N[(m + k0 + 1):n]
    Gstat=stat(y,X,m,Tm,gamma)
    dstat = Gstat$stat
    temp.cpt = apply(matrix(cvalue ,ncol=1), 1,
                     function(x) which(dstat>x)[1])
    cpt = rbind(cpt, temp.cpt)
    max.dstat = c(max.dstat, max(dstat))
    max.loc = c(max.loc, which.max(dstat))
    #if(i%%500==0) print(i)
  }
  out = list(max.dstat=max.dstat, max.loc = max.loc, cpt=cpt)
  return(out)
}
# Example
value = sim(nsim=2500, 100, 9, 0, 2.734264, 1)
allk = na.omit(value$cpt)
ceiling(summary(allk[,1]))
```

```
Gstar = function(m, Tm, gamma) {
    N = m * Tm
    k = 1:N
    Gvalue = sqrt(m) * (1 + k/m) * (k/(m + k))^(gamma)
    return(Gvalue)
}
p = 10  #Number of predictor variables
stat = function(y, X, m, Tm, gamma) {
    y_train = y[1:m]
    x_train = X[(1:m), ]
    y_test = y[-(1:m)]
    x_test = X[-(1:m), ]
    lasso.scad = cv.ncvreg(x_train, y_train, penalty = "SCAD")
    bestlam = lasso.scad$lambda.min
    coef = coef(lasso.scad, s = "lambda.min")
    pstar = length(coef[which(coef != 0)])
    lasso.pred = predict(lasso.scad, X = x_test,
                type = "response", lambda = bestlam)
    lasso.pred1 = predict(lasso.scad, X = x_train,
                type = "response", lambda = bestlam)
    score1 = cumsum(y_test - lasso.pred)
    Gvalue = Gstar(m, Tm, gamma)
    sigma1 = sqrt((m - 1)/(m - pstar)) * sd(y_train - lasso.pred1)
    value = abs(score1)/sigma1/Gvalue
    result = list(stat = value)
    return(result)
}
```

```
### Simulation
p = 10  #Number of predictor variables
sim = function(nsim = 2500, m, Tm, gamma, cvalue, k0) {
    max.dstat = max.loc = cpt = NULL
    for (i in 1:nsim) {
        N = m * Tm
        n = m + N
        idx = 1:n
        beta1 = c(0, 0, 2, 0, 0, 1, 0, 0, 1, 0)
        beta2 = c(0, 0, 0, 3, 0, 0, 1, 0, 0, -1)
        X = matrix(rnorm(n * p, 0, 1), ncol = p)
        X2 = matrix(rnorm(n * p) + 0.8, ncol = p)
        y = vector(mode = "numeric", length = n)
        X[, 3] = rnorm(n, 2, 1)
        X[, 4] = rnorm(n, 4, 1)
        X[, 5] = rnorm(n, 5, 1)
        ei_N = rnorm(n, 0, 1)  # the model errors
        y[1:(m + k0)] = X[1:(m + k0), ] %*% beta1 + ei_N[1:(m + k0)]
        y[(m + k0 + 1):n] = X2[(m + k0 + 1):n, ] %*% beta2 +
                            ei_N[(m + k0 + 1):n]
        Gstat = stat(y, X, m, Tm, gamma)
        dstat = Gstat$stat
        temp.cpt = apply(matrix(cvalue, ncol = 1), 1,
                 function(x) which(dstat > x)[1])
        cpt = rbind(cpt, temp.cpt)
        max.dstat = c(max.dstat, max(dstat))
        max.loc = c(max.loc, which.max(dstat))
```

```
        # if(i%%500==0) print(i)
    }
    out = list(max.dstat = max.dstat, max.loc = max.loc, cpt = cpt)
    return(out)
}
# Plot 1 m = 100 alpha = 0.025
## gamma = 0
value1 = sim(nsim = 2500, 100, 9, 0, 2.572124, 5)
allk1 = na.omit(value1$cpt)
ceiling(summary(allk1[, 1]))
## gamma = 0.25
value2 = sim(nsim = 2500, 100, 9, 0.25, 2.767461, 5)
allk2 = na.omit(value2$cpt)
ceiling(summary(allk2[, 1]))
## gamma = 0.45
value3 = sim(nsim = 2500, 100, 9, 0.45, 3.188064, 5)
allk3 = na.omit(value3$cpt)
ceiling(summary(allk3[, 1]))
mydata = data.frame(k = c(x1, x2, x3), gamma = c(rep(0, 2500),
    rep(0.25, 2500), rep(0.45, 2500)))
mydata$gamma = as.factor(mydata$gamma)
p1 = ggplot(mydata) + geom_density(aes(x = k, color = gamma),
    show.legend = FALSE) + stat_density(aes(x = k, colour = gamma),
    geom = "line", position = "identity", size = 0.8) +
    xlab("Stopping time") + ylab("Density") +
    geom_hline(yintercept = 0, colour = "white",
    size = 1) + labs(color = expression(gamma ~ "  "),
```

```
    title = expression(alpha ~ "= 0.025")) +
    theme(legend.position = "bottom")
# graph 2 m = 100 alpha = 0.05
## gamma = 0
value11 = sim(nsim = 2500, 100, 9, 0, 2.351546, 5)
allk11 = na.omit(value11$cpt)
ceiling(summary(allk11[, 1]))
value21 = sim(nsim = 2500, 100, 9, 0.25, 2.539297, 5)
allk21 = na.omit(value21$cpt)
ceiling(summary(allk21[, 1]))
## gamma = 0.45
value31 = sim(nsim = 2500, 100, 9, 0.45, 2.981583, 5)
allk31 = na.omit(value3$cpt)
ceiling(summary(allk31[, 1]))
mydata1 = data.frame(k = c(x11, x21, x31), gamma = c(rep(0, 2500),
    rep(0.25, 2500), rep(0.45, 2500)))
names(mydata1)
mydata1$gamma = as.factor(mydata1$gamma)
p2 = ggplot(mydata1) + geom_density(aes(x = k, color = gamma),
    show.legend = FALSE) + stat_density(aes(x = k, colour = gamma),
    geom = "line", position = "identity", size = 0.8) +
    xlab("Stopping time") + ylab("Density") +
    geom_hline(yintercept = 0, colour = "white",
    size = 1) + labs(color = expression(gamma ~ "  "),
    title = expression(alpha ~ "= 0.05")) +
    theme(legend.position = "bottom")
ggarrange(p1, p2, ncol = 2, nrow = 1,
```

```
    common.legend = TRUE, legend = "bottom")
```

Chapter 3

Small p: Type I error - SPQR - PSPQR - Open-end - tau = 0.5

```
rm(list=ls())
library(ncvreg)
library(glmnet)
library(quantreg)
library(rqPen)
# return the g function:
G = function(m, Tm, gamma) {
  N = m * Tm
  k = 1:N
  Gvalue = sqrt(m) * (1 + k/m) * (k/(m + k))^gamma
  return(Gvalue)
}
sqroot <- function(A) {
  e = eigen(A)
  v = e$vectors
  val = e$values
  sq = v %*% diag(sqrt(val)) %*% solve(v)
  return(t(sq))
}
p = 10 #Number of predictor variables
stat=function(y, X, m, Tm, gamma, tau){
  y_train=y[1:m]
  x_train=X[(1:m),]
  y_test=y[-(1:m)]
```

```
x_test=X[-(1:m),]
fit_lasso = rq.fit.scad(x_train,y_train,tau=0.5,alpha=3.7,
                        lambda=seq(1,100,length.out=10),
                        start="rq",beta=0.9995,eps = 1e-06)
coef1 = fit_lasso$coefficients
coef1[abs(coef1)<0.000001]=0
psi = tau - (y_test - x_test %*% coef1 < 0)
score1 = apply((x_test * as.vector(psi)), 2, cumsum)
Jvalue = t(x_train) %*% x_train/m * tau * (1 - tau)
Jsroot = sqroot(Jvalue)
Gvalue = G(m, Tm, gamma)
statvalue1 = t(solve(Jsroot) %*% t(score1))/Gvalue
statmax1 = pmax(abs(statvalue1[, 1]), abs(statvalue1[, 2]))
#Post QL
x_train = x_train[,c(coef1!=0)]
x_test = x_test[,c(coef1!=0)]
fit1 = rq(y_train ~ x_train - 1, tau)
coef1 = fit1$coef
psi = tau - (y_test - x_test %*% coef1 < 0)
score1 = apply((x_test * as.vector(psi)), 2, cumsum)
Jvalue = t(x_train) %*% x_train/m * tau * (1 - tau)
Jsroot = sqroot(Jvalue)
Gvalue = G(m, Tm, gamma)
statvalue1 = t(solve(Jsroot) %*% t(score1))/Gvalue
statmax2 = pmax(abs(statvalue1[, 1]), abs(statvalue1[, 2]))
result = list(stat1 = statmax1, stat2 = statmax2)
return(result)
```

```
}
p = 10
sim=function(nsim=nsim, m,Tm,gamma,alpha,tau, cval1, k0){
  max.Gamma1=cpt1=max.Gamma2=cpt2=NULL
  for (i in 1:nsim){
    N=m*Tm
    n=m+N
    idx=1:n
    beta0 = c(1, 0, -1, 0, -15, rep(0,5))
    X = matrix(rnorm(n * p, 0,1), ncol = p)
    y=vector(mode="numeric",length=n)
    #X[,3]=rnorm(n,2,1)
    #X[,4]=rnorm(n,4,1)
    #X[,5]=rnorm(n,5,1)
    ei_N=rnorm(n,0,1)    # the model errors
    y= X%*%beta0+ei_N
    Gstat1=stat(y,X,m,Tm,gamma,tau)
    Gamma1=Gstat1$stat1
    Gamma2=Gstat1$stat2
    temp.cpt1 = apply(matrix(cval1,ncol=1), 1,
                      function(x) which(Gamma1>x)[1]) # SPQR
    temp.cpt2 = apply(matrix(cval1,ncol=1), 1,
                      function(x) which(Gamma2>x)[1]) # PSPQR
    cpt1 = rbind(cpt1, temp.cpt1)
    cpt2 = rbind(cpt2, temp.cpt2)
    max.Gamma1 = c(max.Gamma1, max(Gamma1))
    max.Gamma2 = c(max.Gamma2, max(Gamma2))
```

```
    if(i%%500==0) print(i)
  }
  ##type I error (reject when cpt is not NA)
  TypeI1 = 1-apply(cpt1, 2, function(x) sum(is.na(x))/nsim) # SPQR
  TypeI2 = 1-apply(cpt2, 2, function(x) sum(is.na(x))/nsim) # PSPQR
  out = list(TypeI1=TypeI1, cpt1=cpt1, max.Gamma1=max.Gamma1,
             TypeI2=TypeI2, cpt2=cpt2, max.Gamma2=max.Gamma2)
  return(out)
}
#Example
alpha=0.05; gamma=0; k0=0; tau = 0.5
Ns = c(2, 4, 6, 9)
ms = c(75)
nsim = 2500
## Type I error
d = 0
for(Tm in c(2, 4, 6, 9)){
  #cval1=cvalue_closed(gamma,N, alpha)
  cval1 = 2.486683
  for(m in ms){
    sim1=sim(nsim, m,Tm,gamma, alpha,tau,cval1, k0)
    cat("Type I SPQR :", c(sim1$TypeI1), "\n")
    cat("Type I PSPQR :", c(sim1$TypeI2), "\n")
  }}
```

Type I error - SPQR - PSPQR - Closed-end - tau = 0.5

```
cvalue_closed = function(gamma, Tm, alpha) {
  rep = 50000
```

```
  num = 10000
  ind = floor(Tm/(Tm + 1) * num)
  t = (1:ind)/num
  stat = NULL
  for (i in 1:rep) {
    set.seed(seeds[i])
    e1 = rnorm(num, 0, 1)
    e2 = rnorm(num, 0, 1)

    W1 = 1/sqrt(num) * cumsum(e1)[1:ind]
    W2 = 1/sqrt(num) * cumsum(e2)[1:ind]

    absW = pmax(abs(W1), abs(W2))
    stat = c(stat, max(absW/t^gamma))
  }
  critical = quantile(stat, 1 - alpha)
  return(critical)
}
# return the g function:
G = function(m, Tm, gamma) {
  N = m * Tm
  k = 1:N
  Gvalue = sqrt(m) * (1 + k/m) * (k/(m + k))^gamma
  return(Gvalue)
}
sqroot <- function(A) {
  e = eigen(A)
```

```
  v = e$vectors
  val = e$values
  sq = v %*% diag(sqrt(val)) %*% solve(v)
  return(t(sq))
}
p = 10 #Number of predictor variables
stat=function(y, X, m, Tm, gamma, tau){
  y_train=y[1:m]
  x_train=X[(1:m),]
  y_test=y[-(1:m)]
  x_test=X[-(1:m),]
  fit_lasso = rq.fit.scad(x_train,y_train,tau=0.5,alpha=3.7,
                          lambda=seq(1,100,length.out=10),
                          start="rq",beta=0.9995,eps=1e-06)
  coef1 = fit_lasso$coefficients
  coef1[abs(coef1)<0.000001]=0
  psi = tau - (y_test - x_test %*% coef1 < 0)
  score1 = apply((x_test * as.vector(psi)), 2, cumsum)
  Jvalue = t(x_train) %*% x_train/m * tau * (1 - tau)
  Jsroot = sqroot(Jvalue)
  Gvalue = G(m, Tm, gamma)
  statvalue1 = t(solve(Jsroot) %*% t(score1))/Gvalue
  statmax1 = pmax(abs(statvalue1[, 1]), abs(statvalue1[, 2]))
  #Post QL
  x_train = x_train[,c(coef1!=0)]
  x_test = x_test[,c(coef1!=0)]
  fit1 = rq(y_train ~ x_train - 1, tau)
```

```
  coef1 = fit1$coef
  psi = tau - (y_test - x_test %*% coef1 < 0)
  score1 = apply((x_test * as.vector(psi)), 2, cumsum)
  Jvalue = t(x_train) %*% x_train/m * tau * (1 - tau)
  Jsroot = sqroot(Jvalue)
  Gvalue = G(m, Tm, gamma)
  statvalue1 = t(solve(Jsroot) %*% t(score1))/Gvalue
  statmax2 = pmax(abs(statvalue1[, 1]), abs(statvalue1[, 2]))
  result = list(stat1 = statmax1, stat2 = statmax2)
  return(result)
}
p = 10
sim=function(nsim=nsim, m,Tm,gamma,alpha,tau, cval1, k0){
  max.Gamma1=cpt1=max.Gamma2=cpt2=NULL
  for (i in 1:nsim){
    N=m*Tm
    n=m+N
    idx=1:n
    beta0 = c(1, 0, -1, 0, -15, rep(0,5))
    X = matrix(rnorm(n * p, 0,1), ncol = p)
    y=vector(mode="numeric",length=n)
    #X[,3]=rnorm(n,2,1)
    #X[,4]=rnorm(n,4,1)
    #X[,5]=rnorm(n,5,1)
    ei_N=rnorm(n,0,1)    # the model errors
    y= X%*%beta0+ei_N
    Gstat1=stat(y,X,m,Tm,gamma,tau)
```

```
    Gamma1=Gstat1$stat1
    Gamma2=Gstat1$stat2
    temp.cpt1 = apply(matrix(cval1,ncol=1), 1,
                  function(x) which(Gamma1>x)[1]) # SPQR
    temp.cpt2 = apply(matrix(cval1,ncol=1), 1,
                  function(x) which(Gamma2>x)[1]) # PSPQR
    cpt1 = rbind(cpt1, temp.cpt1)
    cpt2 = rbind(cpt2, temp.cpt2)
    max.Gamma1 = c(max.Gamma1, max(Gamma1))
    max.Gamma2 = c(max.Gamma2, max(Gamma2))
    if(i%%500==0) print(i)
  }
  ##type I error (reject when cpt is not NA)
  TypeI1 = 1-apply(cpt1, 2, function(x) sum(is.na(x))/nsim) # SPQR
  TypeI2 = 1-apply(cpt2, 2, function(x) sum(is.na(x))/nsim) # PSPQR
  out = list(TypeI1=TypeI1, cpt1=cpt1, max.Gamma1=max.Gamma1,
             TypeI2=TypeI2, cpt2=cpt2, max.Gamma2=max.Gamma2)
  return(out)

}
# Example
alpha=0.05; gamma=0; k0=0; tau = 0.5
ms = c(75,100,150)
nsim =2500

## Type I error
for(Tm in c(2)){
```

```
  cval1=cvalue_closed(gamma,Tm, alpha)
  for(m in ms){
    sim1=sim(nsim, m,Tm,gamma, alpha, tau, cval1, k0)
    cat("Type I SPQR :", c(sim1$TypeI1), "\n")
    cat("Type I PSPQR :", c(sim1$TypeI2), "\n")
  }}
```

Power - Closed-end - SPQR - PSPQR - tau - 0.5

```
G = function(m, Tm, gamma) {
  N = m * Tm
  k = 1:N
  Gvalue = sqrt(m) * (1 + k/m) * (k/(m + k))^gamma
  return(Gvalue)
}
sqroot <- function(A) {
  e = eigen(A)
  v = e$vectors
  val = e$values
  sq = v %*% diag(sqrt(val)) %*% solve(v)
  return(t(sq))
}
p = 10 #Number of predictor variables
stat=function(y, X, m, Tm, gamma, tau){
  y_train=y[1:m]
  x_train=X[(1:m),]
  y_test=y[-(1:m)]
  x_test=X[-(1:m),]
  fit_lasso = rq.fit.scad(x_train,y_train,tau = 0.5,alpha=3.7,
```

```
                                lambda=seq(1, 100, length.out = 10),
                                start="rq",beta=0.9995,eps = 1e-06)
  coef1 = fit_lasso$coefficients
  coef1[abs(coef1)<0.00001]=0
  psi = tau - (y_test - x_test %*% coef1 < 0)
  score1 = apply((x_test * as.vector(psi)), 2, cumsum)
  Jvalue = t(x_train) %*% x_train/m * tau * (1 - tau)
  Jsroot = sqroot(Jvalue)
  Gvalue = G(m, Tm, gamma)
  statvalue1 = t(solve(Jsroot) %*% t(score1))/Gvalue
  statmax1 = pmax(abs(statvalue1[, 1]), abs(statvalue1[, 2]))
  #Post QL
  x_train = x_train[,c(coef1!=0)]
  x_test = x_test[,c(coef1!=0)]
  fit1 = rq(y_train ~ x_train - 1, tau)
  coef1 = fit1$coef
  psi = tau - (y_test - x_test %*% coef1 < 0)
  score1 = apply((x_test * as.vector(psi)), 2, cumsum)
  Jvalue = t(x_train) %*% x_train/m * tau * (1 - tau)
  Jsroot = sqroot(Jvalue)
  Gvalue = G(m, Tm, gamma)
  statvalue1 = t(solve(Jsroot) %*% t(score1))/Gvalue
  statmax2 = pmax(abs(statvalue1[, 1]), abs(statvalue1[, 2]))
  result = list(stat1 = statmax1, stat2 = statmax2)
  return(result)
}
p  = 10 #Number of predictor variables
```

```
sim=function(nsim,m, Tm, gamma, tau, cvalue, k0){
  max.dstat1= max.loc1 = cpt1=max.dstat2= max.loc2 = cpt2=NULL
  for (i in 1:nsim){
    N=m*Tm
    n=m+N
    idx=1:n
    beta1=c(-1,0,1,8,1,0,0,0,-5,0)
    beta2=c(0,-1,0,2,0,0,1,0,0,-1)
    X = matrix(runif(n * p, 0,1), ncol = p)
    X2 = matrix(runif(n * p, 0,1)+0.8, ncol = p)
    y = vector(mode = "numeric", length = n)
    X[, 3] = rnorm(n, 2, 1)
    X[, 4] = rnorm(n, 4, 1)
    X[, 5] = rnorm(n, 5, 1)
    ei_N = rnorm(n, 0, 1)  # the model errors
    #ei_N = rcauchy(n, 0, 1)
    y[1:(m + k0)] = X[1:(m + k0), ] %*% beta1 + ei_N[1:(m+k0)]
    y[(m + k0 + 1):n] = X2[(m + k0 + 1):n, ] %*% beta2 +
                        ei_N[(m + k0 + 1):n]
    Gstat=stat(y,X,m,Tm,gamma, tau)
    dstat1 = Gstat$stat1
    dstat2 = Gstat$stat2
    temp.cpt1 = apply(matrix(cvalue ,ncol=1), 1,
                function(x) ifelse(max(dstat1)>x, 1, 0)) # SPQR
    temp.cpt2 = apply(matrix(cvalue ,ncol=1), 1,
                function(x) ifelse(max(dstat2)>x, 1, 0)) # PSPQR
    cpt1 = rbind(cpt1, temp.cpt1)
```

```
    cpt2 = rbind(cpt2, temp.cpt2)
    #if(i%%200==0) print(i)
  }
  #out = list(max.dstat=max.dstat, max.loc = max.loc, cpt=cpt)
  out = list(cpt1=cpt1, cpt2=cpt2)
  return(out)
}
# Example
value = sim(nsim=2500, 100, 9, 0, 0.5,   2.572124, 1)
allk1 = na.omit(value$cpt1)
mean(allk1) # SPQR
allk2 = na.omit(value$cpt2)
mean(allk2) # PSPQR
```

Dtection Time: Open-end - Tau = 0.5

```
G = function(m, Tm, gamma) {
  N = m * Tm
  k = 1:N
  Gvalue = sqrt(m) * (1 + k/m) * (k/(m + k))^gamma
  return(Gvalue)
}
sqroot <- function(A) {
  e = eigen(A)
  v = e$vectors
  val = e$values
  sq = v %*% diag(sqrt(val)) %*% solve(v)
  return(t(sq))
}
```

```
p = 10 #Number of predictor variables
stat=function(y, X, m, Tm, gamma, tau){
  y_train=y[1:m]
  x_train=X[(1:m),]
  y_test=y[-(1:m)]
  x_test=X[-(1:m),]
  fit_lasso = rq.fit.scad(x_train, y_train, tau = 0.5, alpha = 3.7,
                          lambda = seq(1, 100, length.out = 10),
                          start = "rq", beta = 0.9995, eps = 1e-06)
  coef1 = fit_lasso$coefficients
  coef1[abs(coef1)<0.00001]=0
  coef1
  psi = tau - (y_test - x_test %*% coef1 < 0)
  pstar = length(coef1[which(coef1 != 0)])
  score1 = apply((x_test * as.vector(psi)), 2, cumsum)
  Jvalue = t(x_train) %*% x_train/m * tau * (1 - tau)
  Jsroot = sqroot(Jvalue)
  Gvalue = G(m, Tm, gamma)
  statvalue1 = t(solve(Jsroot) %*% t(score1))/Gvalue
  statmax1 = pmax(abs(statvalue1[, 1]), abs(statvalue1[, 2]))
  # Post QR
  x_train = x_train[,c(coef1!=0)]
  x_test = x_test[,c(coef1!=0)]
  fit1 = rq(y_train ~ x_train - 1, tau)
  coef1 = fit1$coef
  psi = tau - (y_test - x_test %*% coef1 < 0)
  score1 = apply((x_test * as.vector(psi)), 2, cumsum)
```

```
  Jvalue = t(x_train) %*% x_train/m * tau * (1 - tau)
  Jsroot = sqroot(Jvalue)
  Gvalue = G(m, Tm, gamma)
  statvalue1 = t(solve(Jsroot) %*% t(score1))/Gvalue
  statmax2 = pmax(abs(statvalue1[, 1]), abs(statvalue1[, 2]))
  result = list(stat1 = statmax1, stat2 = statmax2)
  return(result)
}
p  = 10 #Number of predictor variables
sim=function(nsim,m, Tm, gamma, tau, cvalue, k0){
  max.dstat= max.loc = cpt=max.dstat1= max.loc1 = cpt1=NULL
  for (i in 1:nsim){
    N=m*Tm
    n=m+N
    idx=1:n
    beta1=c(-1,0,1,8,1,0,0,0,-5,0)
    beta2=c(0,-1,0,2,0,0,1,0,0,-1)
    X = matrix(rnorm(n * p, 0,1), ncol = p)
    X2 = matrix(rnorm(n * p, 0,1)+0.8, ncol = p)
    y = vector(mode = "numeric", length = n)
    X[, 3] = rnorm(n, 2, 1)
    X[, 4] = rnorm(n, 4, 1)
    X[, 5] = rnorm(n, 5, 1)
    ei_N = rnorm(n, 0, 1)  # the model errors
    y[1:(m + k0)] = X[1:(m + k0), ] %*% beta1 + ei_N[1:(m+k0)]
    y[(m + k0 + 1):n] = X2[(m + k0 + 1):n, ] %*% beta2
                  + ei_N[(m + k0 + 1):n]
```

```
    Gstat1=stat(y,X,m,Tm,gamma, tau)
    Gstat2=stat(y,X,m,Tm,gamma, tau)
    dstat1 = Gstat1$stat1
    dstat2 = Gstat2$stat2
    temp.cpt = apply(matrix(cvalue ,ncol=1), 1,
                function(x) which(dstat1>x)[1]) # SPQR
    temp.cpt1 = apply(matrix(cvalue ,ncol=1), 1,
                function(x) which(dstat2>x)[1]) # PSPQR
    cpt = rbind(cpt, temp.cpt)
    cpt1 = rbind(cpt1, temp.cpt1)
    #max.dstat = c(max.dstat, max(dstat))
    #max.loc = c(max.loc, which.max(dstat))
    #if(i%%500==0) print(i)
  }
  #out = list(max.dstat=max.dstat, max.loc = max.loc, cpt=cpt)
  out = list(cpt=cpt, cpt1=cpt1)
  return(out)
}
# Example
value = sim(nsim=1000, 100, 9, 0, 0.5, 2.486683, 1)
allk = na.omit(value$cpt) # SPQR
ceiling(summary(allk[,1]))
```

Chapter 4

```
rm(list=ls())
library(bbmle)
library(sn)
EPS = sqrt(.Machine$double.eps)
```

```
llik.skew <- function(xi, omega, alpha, x){
  sum(dsn(x, xi, omega, alpha, dp=NULL, log=TRUE))
}
thetahat.skew <- function(x){
  toptim<-function(theta)-llik.skew(theta[1],theta[2],theta[3],x)
  xi.guess = 0
  omega.guess = 1
  alpha.guess = 0
  res = nlminb(c(xi.guess, omega.guess, alpha.guess),
               toptim, lower=EPS)
  c(res$objective)
  #c(res)
}
cc_prob1 = list()
N = 1000 # Number of interations
M = 1
tau = list()
all.cover = list()
avg_func = function(nl,xil,omegal,alphal,nr,xir, omegar,
          alphar, true.tau){
  for(l in 1:M){
    for(k in 1:N){
      y = c(rsn(nl,xil,omegal,alphal),rsn(nr,xir,omegar,alphar))
      n = length(y)
      # Compute the log-likelihood
      loglike_prof = matrix(NA, nrow = (n-1),1)
      for (i in 3:(n - 3)) {
```

```
    loglike_prof[i,1]=(thetahat.skew(y[1:i])+thetahat.skew(y[(i+1):n]))
    }
  #log-likelihood  based approach
  #which.min(loglike_prof[-c(1:2,(n-3):(n-1)),1]) + 1
  tau[[k]] = which.min(loglike_prof) + 1
  #Compute Deviance
  deviance_obs = matrix(NA, nrow = (n-1), ncol = 1)
  for (i in 1:(n - 1)) {
    deviance_obs[,1] = -2*(rep(loglike_prof[tau[[k]],1],
            (n - 1)) - loglike_prof)
  }
  cc_prob1[[k]] = pchisq(deviance_obs[-c(1:2,(n-3):(n-1)),1],2)
}
tau = matrix(unlist(tau), nrow = N, ncol = 1)
#table(as.vector(tau))
cc_prob = matrix(unlist(cc_prob1), ncol = 1, byrow = TRUE)
cc_prob = matrix(cc_prob, ncol = N)
coverage = matrix(NA, nrow = 4, ncol = ncol(cc_prob))
for (i in 1:ncol(cc_prob)) {
  my_func <- function(x){
    ifelse(true.tau %in% x, 1, 0)
  }
  coverage[1, i] = my_func(which(cc_prob[, i] <= 0.5))
  coverage[2, i] = my_func(which(cc_prob[, i] <= 0.90))
  coverage[3, i] = my_func(which(cc_prob[, i] <= 0.95))
  coverage[4, i] = my_func(which(cc_prob[, i] <= 0.99))
}
```

```
    coverage_size = rowMeans(coverage)
  all.cover = matrix(data= c(0.5, 0.9, 0.95, 0.99, coverage_size ),
        nrow = 4, ncol = 2, dimnames = list(c(),
        c("Alpha","Coverage")))
  }
  return(all.cover)
}
# Example
mylist = avg_func(10, 0, 1, 1, 40, 0, 1, 2, 10)
apply(simplify2array(mylist), c(1,2), mean)
```

For S_n

```
cc_prob1 = list()
N = 1000  # Number of interations
M = 1
tau = list()
all.cover = list()
avg_func = function(nl, xil, omegal, alphal, nr, xir, omegar,
    alphar, true.tau) {
    for (l in 1:M) {
        for (k in 1:N) {
            y = c(rsn(nl, xil, omegal, alphal), rsn(nr, xir,
                omegar, alphar))
            n = length(y)
            # Compute the log-likelihood
            loglike_prof = matrix(NA, nrow = (n - 1), 1)
            for (i in 3:(n - 3)) {
                loglike_prof[i, 1] = (thetahat.skew(y[1:i]) +
```

```
        thetahat.skew(y[(i + 1):n]))
    }
    # MIC based approach
    mic = matrix(NA, nrow = (n - 1), 1)
    for (i in 1:(n - 1)) {
        mic[i, 1] = 2 * loglike_prof[i, 1] + (2 * 3 +
          ((2 * i/n) - 1)^2) * log(n)
    }
    tau[[k]] = which.min(mic) + 1
    # Compute Deviance
    deviance_obs = matrix(NA, nrow = (n - 1), ncol = 1)
    # for (i in 1:(n - 1)) {
    mic_n = 2 * thetahat.skew(y) + 3 * log(n)
    # mic_n
    S_n = mic_n - min(mic[-c(1:2, (n - 2):(n - 1)), 1]) +
        3 * log(n)
    # S_n
    chi_cric = qchisq(0.95, df = 3)
    if (S_n < chi_cric) {
        next
    }
    deviance_obs[, 1] = -2 * (rep(loglike_prof[tau[[k]],
        1], (n - 1)) - loglike_prof)
    # } }
    cc_prob1[[k]] = pchisq(deviance_obs[-c(1:2, (n -
        3):(n - 1)), 1], 2)
}
```

```
tau = matrix(unlist(tau), nrow = N, ncol = 1)
# table(as.vector(tau))
cc_prob = matrix(unlist(cc_prob1), ncol = 1, byrow = TRUE)
cc_prob = matrix(cc_prob, ncol = N)
# Performance for alpha values 0.5, 0.9, 0.95, 0.95
performance = matrix(NA, nrow = 4, ncol = ncol(cc_prob))
for (i in 1:ncol(cc_prob)) {
    performance[1, i] = length(which(cc_prob[, i] <= 0.5))
    performance[2, i] = length(which(cc_prob[, i] <= 0.9))
    performance[3, i] = length(which(cc_prob[, i] <= 0.95))
    performance[4, i] = length(which(cc_prob[, i] <= 0.99))
}
performance_size = rowMeans(performance)
coverage = matrix(NA, nrow = 4, ncol = ncol(cc_prob))
for (i in 1:ncol(cc_prob)) {
    my_func <- function(x) {
        ifelse(true.tau %in% x, 1, 0)
    }
    coverage[1, i] = my_func(which(cc_prob[, i] <= 0.5))
    coverage[2, i] = my_func(which(cc_prob[, i] <= 0.9))
    coverage[3, i] = my_func(which(cc_prob[, i] <= 0.95))
    coverage[4, i] = my_func(which(cc_prob[, i] <= 0.99))
}
coverage_size = rowMeans(coverage)
all.cover = matrix(data = c(0.5,0.9,0.95,0.99,coverage_size,
            performance_size), nrow = 4, ncol = 3,
            dimnames=list(c(),c("Alpha","Coverage","Size")))
```

```
    }
    return(all.cover)
}
```

Chapter 5

```
library(bbmle)
library(FAdist)
weib3_start <- function(x) {
    mu <- mean(log(x))
    sigma2 <- var(log(x))
    logshape <- log(1.2/sqrt(sigma2))
    logscale <- mu + (0.572/logshape)
    logthres <- log(0.5* min(x))
    list(logshape = logshape, logsc = logscale, logthres = logthres)
}
dweib3 <- function(x, shape, scale, thres, log = TRUE) {
    dweibull3(x, shape, scale, thres, log = log)
}
# sink('sink-examp.txt')
eps <- 0.01
# MIC = 0 tau.hat = 0
myfunc = function(nl, shapel, scalel, thresl, nr, shaper, scaler,
    thresr, true.tau) {
    n = nl + nr
    # Generate the data for left side of the change point k
    datl <- data.frame(x = rweibull3(nl, shapel, scalel, thresl))
    tmin <- log(1e-08 * min(datl$x))
    mleft <- mle2(x~dweib3(exp(logshape),exp(logsc),exp(logthres)),
```

```
    data = datl, lower = c(shape = 0, scale = 0, thres = -Inf),
    upper = c(logshape = Inf, logsc = Inf, logthres = tmin -
        eps), start = weib3_start(datl$x), method = "L-BFGS-B")
# Generate the data for right side of the change point k
datr <- data.frame(x = rweibull3(nr, shaper, scaler, thresr))
tmin <- log(1e-08 * min(datr$x))
mright <- mle2(x~dweib3(exp(logshape),exp(logsc),exp(logthres)),
    data = datr, lower = c(shape = 0, scale = 0, thres = -Inf),
    upper = c(logshape = Inf, logsc = Inf, logthres = tmin -
        eps), start = weib3_start(datr$x), method = "L-BFGS-B")
dat <- data.frame(x = c(datl$x, datr$x))
tmin <- log(1e-08 * min(dat$x))
mall <- mle2(x~dweib3(exp(logshape),exp(logsc),exp(logthres)),
    data = dat, lower = c(shape = 0, scale = 0, thres = -Inf),
    upper = c(logshape = Inf, logsc = Inf, logthres = tmin -
        eps), start = weib3_start(dat$x), method = "L-BFGS-B")
for (i in 2:(n - 2)) {
    datl = dat$x[1:i]
    datr = dat$x[(i + 1):n]
    MIC_n = 2 * (mall@min) + 3 * log(n)
    SIC = 2 * (mleft@min + mright@min) + 6 * log(n)
    SIC_k = 2 * (mleft@min + mright@min) + 7 * log(n)
    MIC_k = 2 * (mleft@min) + 2 * (mright@min) + (6 + ((2 *
        i)/n - 1)^2) * log(n)
    # S_n = (2*(mall@min) + 3*log(n)) - (2*(mleft@min +
    # mright@min) + (2*3 + (2*(true.tau)/n-1)^2 )*log(n)) +
    # 3*log(n)
```

```
    }
    S_nF = MIC_n - min(MIC_k) + 3 * log(n)
    T_nF = MIC_n - min(SIC_k) + 3 * log(n)
    # tau.hat = which(MIC == min(MIC)) print(list(min(MIC),MIC))
    # print(which.min(MIC)) cric = qchisq(.95, df=3) print(S_nF)
    # print(S_nF)
    p1 = ifelse(MIC_n > min(SIC) , 1, 0)
    p2 = ifelse(T_nF > 8.202864, 1, 0)
    p3 = ifelse(S_nF > qchisq(0.95,3), 1, 0)
    # p3 = ifelse(min(S_nF) > 10.2, 1, 0) k = ifelse(MIC_n >
    # min(MIC), 1, 0); k = ifelse(S_nF > 13.566464, 1, 0);
    #9.946725
    k = c(p1, p2, p3)
    k
}
d = 1
k = p1 = p2 = p3 = 0
count = 0
mypower = function(nl, shapel, scalel, thresl, nr, shaper, scaler,
    thresr, true.tau) {
    for (j in 1:1000) {
        count = count + myfunc(nl, shapel, scalel, thresl, nr,
            shaper, scaler, thresr, true.tau)
        d = d + 1
    }
    return(list(Power = count/1000))
}
```

```
# Example
mypower(25, 1, 1, 2, 125, 1.75, 1.75, 2.75, 25)

# Coverage Probability
library(FAdist)
library(bbmle)
rm(list=ls())
EPS = sqrt(.Machine$double.eps) # "epsilon" for very small numbers
llik.weibull <- function(shape, scale, thres, x)
{
  #sum(dweibull(x - thres, shape, scale, log=T))
  sum(dweibull3(x,shape, scale, thres,log=TRUE))
}
thetahat.weibull <- function(x)
{
  if(any(x <= 0)) stop("x values must be positive")
  toptim <- function(theta) -llik.weibull(theta[1],
                theta[2], theta[3], x)
  mu = mean(log(x))
  sigma2 = var(log(x))
  shape.guess = 1.2 / sqrt(sigma2)
  scale.guess = exp(mu + (0.572 / shape.guess))
  #thres.guess = 1
  thres.guess = 0.5 * min(x)
  res = nlminb(c(shape.guess, scale.guess, thres.guess), toptim,
                          lower=EPS)
  c(res$objective)
}
```

```
cc_prob1 = list()
N = 1000 # Number of interations
M = 1
tau = list()
all.cover = list()
p = 6
avg_func = function(nl,shapel,scalel,thresl,nr,shaper,
                    scaler, thresr, true.tau){
  for(l in 1:M){
    for(k in 1:N){
      y = c(rweibull3(nl,shapel,scalel,thresl) ,
                      rweibull3(nr,shaper, scaler, thresr))
      #set.seed(20)
      n = length(y)
      # Compute the log-likelihood
      loglike_prof = matrix(NA, nrow = (n-1),1)
      for (i in 3:(n - 3)) {
        loglike_prof[i,1] = (thetahat.weibull(y[1:i]) +
thetahat.weibull(y[(i+1):n]))
        }

      #which.min(loglike_prof[-c(1:2,(n-3):(n-1)),1]) + 2
      #tau[[k]] = which.min(loglike_prof[,1])
      tau[[k]] = which.min(loglike_prof[-c(1:2,(n-3):(n-1)),1])+1
      #MIC based approach
      #mic = matrix(NA, nrow = (n-1),1)
      #for (i in 1:(n - 1)) {
```

```
    #mic[i,1] = 2*loglike_prof[i,1] + (2*p + (2*i/n-1)^2)*log(n)
  #}
  #which.min(mic[-c(1:2,(n-3):(n-1)),1])+2
  #tau[[k]] = which.min(mic[,1])
  #tau[[k]] = which.min(mic[-c(1:2,(n-3):(n-1)),1]) + 1
  #Compute Deviance
  deviance_obs = matrix(NA, nrow = (n-1), ncol = 1)
  for (i in 1:(n - 1)) {
    deviance_obs[,1] = -2*(rep(loglike_prof[tau[[k]],1],
          (n - 1)) - loglike_prof)
  }
  cc_prob1[[k]] = pchisq(deviance_obs[-c(1:2,(n-3):(n-1)),1],2)
}
tau = matrix(unlist(tau), nrow = N, ncol = 1)
table(as.vector(tau))
cc_prob = matrix(unlist(cc_prob1), ncol = 1, byrow = TRUE)
cc_prob = matrix(cc_prob, ncol = N)
#### Performance for alpha values 0.5, 0.9, 0.95, 0.95
#performance_size = rowMeans(performance)
coverage = matrix(NA, nrow = 4, ncol = ncol(cc_prob))
for (i in 1:ncol(cc_prob)) {
  my_func <- function(x){
    ifelse(true.tau %in% x, 1, 0)
  }
  coverage[1, i] = my_func(which(cc_prob[, i] <= 0.5))
  coverage[2, i] = my_func(which(cc_prob[, i] <= 0.90))
  coverage[3, i] = my_func(which(cc_prob[, i] <= 0.95))
```

```
      coverage[4, i] = my_func(which(cc_prob[, i] <= 0.99))
    }
    coverage_size = rowMeans(coverage)
  all.cover = matrix(data= c(0.5,0.9,0.95,0.99,coverage_size),
       nrow = 4,ncol = 2,dimnames=list(c(),c("Alpha","Coverage")))
  }
  return(all.cover)
}
# Example
mylist = avg_func(10, 1, 1, 2, 40, 1.25, 1.25, 2.25, 10)
apply(simplify2array(mylist), c(1,2), mean)

# Average Size
library(FAdist)
library(bbmle)
rm(list = ls())
EPS = sqrt(.Machine$double.eps)  # 'epsilon' for very small numbers
llik.weibull <- function(shape, scale, thres, x) {
    # sum(dweibull(x - thres, shape, scale, log=T))
    sum(dweibull3(x, shape, scale, thres, log = TRUE))
}
thetahat.weibull <- function(x) {
    if (any(x <= 0))
        stop("x values must be positive")
    toptim <- function(theta) -llik.weibull(theta[1], theta[2],
        theta[3], x)
    mu = mean(log(x))
    sigma2 = var(log(x))
```

```
    shape.guess = 1.2/sqrt(sigma2)
    scale.guess = exp(mu + (0.572/shape.guess))
    # thres.guess = 1
    thres.guess = 0.5 * min(x)
    res = nlminb(c(shape.guess, scale.guess, thres.guess), toptim,
        lower = EPS)
    c(res$objective)
}
cc_prob1 = list()
N = 1000  # Number of interations
M = 1
tau = list()
all.cover = list()
p = 6
avg_func = function(nl, shapel, scalel, thresl, nr,
shaper, scaler, thresr, true.tau) {
    seeds2 = sample(1:1e+06, N, replace = FALSE)
    for (l in 1:M) {
        for (k in 1:N) {
            set.seed(seeds2[k])
            y = c(rweibull3(nl, shapel, scalel, thresl),
                rweibull3(nr, shaper, scaler, thresr))
            n = length(y)
            # Compute the log-likelihood
            loglike_prof = matrix(NA, nrow = (n - 1), 1)
            for (i in 3:(n - 3)) {
                loglike_prof[i, 1] = (thetahat.weibull(y[1:i]) +
```

```
            thetahat.weibull(y[(i + 1):n]))
    }
    # Compute Deviance
    deviance_obs = matrix(NA, nrow = (n - 1), ncol = 1)
    for (i in 1:(n - 1)) {
        deviance_obs[, 1] = -2 * (rep(loglike_prof[tau[[k]],
          1], (n - 1)) - loglike_prof)
    }
    cc_prob1[[k]] = pchisq(deviance_obs[-c(1:2, (n -
        3):(n - 1)), 1], 2)
}
tau = matrix(unlist(tau), nrow = N, ncol = 1)
table(as.vector(tau))
cc_prob = matrix(unlist(cc_prob1), ncol = 1, byrow = TRUE)
cc_prob = matrix(cc_prob, ncol = N)
#### Performance for alpha values 0.5, 0.9, 0.95, 0.95
performance = matrix(NA, nrow = 4, ncol = ncol(cc_prob))
for (i in 1:ncol(cc_prob)) {
    performance[1, i] = length(which(cc_prob[, i] <= 0.5))
    performance[2, i] = length(which(cc_prob[, i] <= 0.9))
    performance[3, i] = length(which(cc_prob[, i] <= 0.95))
    performance[4, i] = length(which(cc_prob[, i] <= 0.99))
}
performance_size = rowMeans(performance)
all.cover = matrix(data = c(0.5, 0.9, 0.95, 0.99,
    performance_size), nrow = 4, ncol = 2,
    dimnames = list(c(), c("Alpha", "Size")))
```

```
    }
    return(all.cover)
}
# Example
mylist = avg_func(10, 1, 1, 2, 40, 1.25, 1.25, 2.25, 10)
apply(simplify2array(mylist), c(1, 2), mean)
```