Nonparametric Tests for Change Points in Hazard Functions under Random Censorship in Survival Analysis

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern zur Verleihung des akademischen Grades Doktor der Naturwissenschaften (Doctor rerum naturalium, Dr. rer. nat.) genehmigte
Dissertation
von

Rumana Rois

Gutachten:
Prof. Dr. Jürgen Franke, Technische Universität Kaiserslautern
Prof. Dr. Claudia Kirch, Otto-von-Guericke-Universität Magdeburg

Datum der Disputation: 18.09.2017

D 386

Abstract

The thesis studies change points in absolute time for censored survival data with some contributions to the more common analysis of change points with respect to survival time. We first introduce the notions and estimates of survival analysis, in particular the hazard function and censoring mechanisms. Then, we discuss change point models for survival data. In the literature, usually change points with respect to survival time are studied. Typical examples are piecewise constant and piecewise linear hazard functions. For that kind of models, we propose a new algorithm for numerical calculation of maximum likelihood estimates based on a cross entropy approach which in our simulations outperforms the common Nelder-Mead algorithm.

Our original motivation was the study of censored survival data (e.g., after diagnosis of breast cancer) over several decades. We wanted to investigate if the hazard functions differ between various time periods due, e.g., to progress in cancer treatment. This is a change point problem in the spirit of classical change point analysis. Horváth (1998) proposed a suitable change point test based on estimates of the cumulative hazard function. As an alternative, we propose similar tests based on nonparametric estimates of the hazard function. For one class of tests related to kernel probability density estimates, we develop fully the asymptotic theory for the change point tests. For the other class of estimates, which are versions of the Watson-Leadbetter estimate with censoring taken into account and which are related to the Nelson-Aalen estimate, we discuss some steps towards developing the full asymptotic theory. We close by applying the change point tests to simulated and real data, in particular to the breast cancer survival data from the SEER study.

Abstract

In dieser Arbeit werden Changepoints in absoluter Zeit für zensierte Überlebensdaten betrachtet und zusätzlich Changepoints relativ zur Überlebenszeit. Wir betrachten zuerst die Begriffe und Schätzverfahren aus der Survivalanalysis, insbesondere die Hazardfunktion und Zensierungsmechanismen. In der Literatur werden in erster Linie Changepoints bzgl. der Überlebenszeit studiert. Typische Beispiele sind stückweise konstante bzw. stückweise lineare Hazardfunktionen. Für diese Art von Modellen schlagen wir einen neuen Algorithmus zur numerischen Berechnung der Maximum Likelihood-Schätzer vor, der auf einem Kreuzentropieansatz beruht und der in Simulationen bessere Ergebnisse als der übliche Nelder-Mead-Algorithmus liefert.

Unsere urspüngliche Motivation war die Untersuchung von zensierten Überlebensdaten (z.B. nach einer Brustkrebsdiagnose) über mehrere Jahrzehnte. Wir wollten untersuchen, ob die Hazardfunktionen in unterschiedlichen Zeitperioden verschieden sind, z.B. auf Grund von Fortschritten in der Krebstherapie. Dies ist ein Testproblem im Sinn der klassischen Changepointanalyse. Horváth (1998) hat einen passenden Changepointtest vorgeschlagen, der auf der kumulativen Hazardfunktion aufbaut. Als Alternative schlagen wir ähnliche Tests vor, die stattdessen nichtparametrische Schätzer der Hazardfunktion benutzen. Für eine Klasse von Tests, die Bezüge zu Kernschätzern von Wahrscheinlichkeitsdichten aufweisen, leiten wir eine vollständige asymptotische Theorie her. Für eine andere Klasse von Schätzern, die Versionen des Watson-Leadbetter-Schätzers unter Berücksichtigung der Zensierung der Daten benutzt und die mit den Nelson-Aalen-Schätzern verwandt sind, diskutieren wir Ansätze zur Herleitung einer vollständigen asymptotischen Theorie. Zum Abschluss wenden wir unsere Changepointtests auf simulierte und reale Daten, insbesondere auf die Brustkrebsüberlebensdaten der SEER-Studie, an.

Acknowledgements

Working on this thesis has been an extremely emotional journey full of challenges which I alone could not have navigated through. First and foremost, I am immensely grateful to my supervisor, Prof. Dr. Jürgen Franke, for his kind mentoring, extensive support, insightful suggestions, and continuous encouragement without which successful completion of this thesis would not have been a reality.

I am very grateful to my external reviewer Prof. Dr. Claudia Kirch from Otto-von-Guericke University Magdeburg for her insightful comments on my thesis and in-depth remarks for further study.

I am also extremely grateful to Prof. Dr. Jörn Sass for co-supervising me with continuous support and encouraging discussions especially during our weekly breakfast meetings. My gratitude goes to Prof. Dr. Claudia Redenbach for her deeply influential advice on contemporary research and useful comments after reading my thesis prior to submission. I thoroughly enjoyed collaborations with Dr. Stefanie Schwaar for her friendship and innovative ideas in change point analysis. I am quite grateful to Danilo Pezo for his friendship and thoughtful mathematical discussions.

I would like to thank to Dr. Falk Triebsch, Graduate School of Mathematics, and Frau Birgitt Skailes, Consultant, *Deutscher Akademischer Austausch Dienst* (DAAD), for their cordial administrative supports to accommodate my unusual personal circumstances. Special thanks to the DAAD for their funding which enabled me to pursue my research in Germany.

I give special thanks to Frau Beate Siegler, Secretary, Statistics Group, for her continuous support and cordial care making me emotionally strong, and for the proofreading of the manuscript. I thank to all my friends at the Statistics and Financial Mathematics groups at the University of Kaiserslautern for giving me a friendly and pleasant working environment.

More than all, I am grateful to my family for making undeniable sacrifice to let me accomplish this work and always believing in me (to my husband, Russell, for his love and supports; to my loved son, Ramin, for his sacrifice more than anyone at the hardest possible time and to my baby daughter, Riona, for her sacrifice of mothers touch at the age of 4 months; to my loving parents for their never-ending care).

To my husband and best friend, Md. Rofiqul Alam Russell

&

Our Ramin and Riona.

Contents

Abstract **i**

Acknowledgements **ii**

List of Figures **v**

List of Tables **xi**

1 Introduction **1**

1.1 Problem Overview 1

1.2 Motivation 4

1.3 State of the Art 6

1.4 Outline of the Thesis 10

2 Change Point in Hazard Functions with Censored Data **13**

2.1 Censored Data 14

2.2 Review of Hazard Functions Estimation 16

2.3 Change Point in Hazard Functions 22

2.3.1 Change Point Models and Methods for Hazard Function . . . 22

2.3.2 Change Point in Hazard using the Cross-Entropy (CE) Method 28

2.4 Simulations and Applications 32

2.4.1 Simulations 32

2.4.2 Real Data Applications 38

2.5 Discussion 43

3 Change Point Analysis with Missing Data **45**

3.1 Single Change Point Models 47

3.1.1 Single Change Point in Mean 47

3.1.2 Single Change Point in Linear Regression 56

3.2 Multiple Change Points Models 61

3.2.1 Multiple Change Points in Mean 61

3.2.2 Multiple Change Points in Linear Regression 64

3.3 Missing Data Imputation 64

3.3.1 Missing-Data Mechanism 65

3.3.2 Imputation Methods 67

3.3.3 Imputation in Cell Stimulus Response Data 69

3.4 Real Data Analysis . . . 73
3.4.1 Changes in Mean . . . 74
3.4.2 Changes in Regression Coefficients . . . 78
3.4.3 Changes in Distribution . . . 89
3.5 Discussion . . . 93
4 Change Point in Hazard Function using U-statistic for Censored Data 95
4.1 Test Hypothesis . . . 96
4.1.1 Horváth's Test for Changes based on Censored Data . . . 98
4.2 A Hazard-Function based Statistic . . . 99
4.2.1 Convergence of $\hat{\beta}_n^{(1)}$ to β . . . 100
4.2.2 Convergence of $\hat{\beta}_n^{(2)}$ to β . . . 107
4.2.3 Convergence of $\hat{\beta}_n^{(3)}$ to β . . . 110
4.3 Kernel-Based Change Point Test's . . . 114
4.3.1 Change Point Test using $\hat{\lambda}_2$ . . . 115
4.3.2 Change Point Tests with Different Weights . . . 120
4.3.3 Proof of Theorem 4.3.1 . . . 121
4.4 Consistency of the Change Point Test . . . 128
4.5 Test Statistic and Asymptotics under H_0 for Antisymmetric Kernels . 131
4.5.1 Change Point Tests with Different Weights . . . 133
4.6 Some Remarks on a Change Point Test using $\hat{\lambda}_1(x)$ . . . 134
4.7 Discussion . . . 136
5 Simulations and Applications 137
5.1 Simulations . . . 137
5.1.1 Asymptotic and Simulated Critical Values . . . 139
5.1.2 Simulated Powers . . . 150
5.1.3 Evaluation of Change Point Estimators . . . 155
5.2 Real Data Applications . . . 157
5.2.1 Cell Stimulus Response Data . . . 158
5.2.2 Breast Cancer Mortality Data . . . 162
5.3 Discussion . . . 166
6 Summary and Conclusions 167
A Breast Cancer Dataset 171
B Simulated Critical Values 172
Bibliography 179

List of Figures

1.1 Examples of Hazard Functions: the bathtub curve describes the process of human life in the *blue line*, a decreasing shape indicates the risk of infant mortality in the *red line*, the constant hazard function is the risk of healthy individuals between 18 and 40 years of age in the *green line*, an increasing hazard function observes the risk of cancer patients or old age individuals in the *garnet red line*, and an increasing and then decreasing hazard function is described by patients with tuberculosis after taking treatment in the *yellow line*. 2

2.1 (a) Right censoring with fixed censoring time, (b) Left censoring due to late study onset, and (c) Interval censoring due to discrete observation times. 15
2.2 Types of Point-Censored Observations 16
2.3 The estimated hazard functions of the Breast Cancer Patients using SEER Data estimated by Nelson-Aalen estimator (blue line), Product-Limit estimator (red line, overlapped by the blue line), and kernel hazard estimator (black line) from 1973 to 2010. 39
2.4 The estimated change points in the hazard functions for the model (2.29) of the Breast Cancer Patients diagnosed with the first tumor using SEER Data 1973-2010. 40
2.5 The estimated hazard functions for Cell Stimulus Response Data. . . 41
2.6 Estimated change points in the hazard functions for the Cell Stimulus Response using the model (2.25) with the CE method in (a), and Nelder-Mead Simplex Algorithm in (b). 42

3.1 Change point problems description in different cases. Multiple change points in means, variances, correlation strengths and probability distributions are displayed in (a) and changes in linear regression model exhibits in (b). 46

3.2 (a) Scatter Plot of the Cell Stimulus Response Data with missing observations in red dot. (b) Matrix Plot of the Cell Stimulus Response variable with missing values are highlighted by red, where each 25 cells of the data matrix is visualized by a blue rectangle. 70
3.3 Scatter Plots of Cell Stimulus Response data with MCAR mechanism, using four single imputation methods missing values imputed and highlighted in orange colour; *Mean Substitution* in (a), *Hot Deck* in (b), *Last and next average* in (c), and *Regression Imputation* in (d). . . . 72
3.4 Boxplot for different imputation methods of cell stimulus response data. Observed data in (1) and imputed data using *Mean substitution* in (2), *Hot Deck* in (3), *Last and next average* in (4) and *Regression imputation* in (5). 73
3.5 The Cell Stimulus Response Data. 74
3.6 CUSUM statistic process (3.4) for detecting multiple change points in the model (3.46) with the asymptotic (red) and bootstrap (blue) critical values are presented by horizontal lines, and location of change points are indicated by dotted vertical red lines in (a); and the estimated means are located in (b). 75
3.7 CUSUM R-type statistics process (3.14) for detecting multiple change points in the model (3.46) with the asymptotic (red) and bootstrap (blue) critical values are presented by horizontal lines, and location of change points are indicated by dotted vertical red lines in (a); and the estimated means are located in (b). 75
3.8 MOSUM process (3.20) for detecting multiple change points in the model (3.46) using bandwidth $G = 150$ and $\hat{\sigma^2}$ in (3.6) with the asymptotic (red) and bootstrap (blue) critical values are presented by horizontal lines, and location of change points are indicated by dotted vertical red lines in (a); and the estimated means are located in (b). . 76
3.9 MOSUM R-type statistics process (3.24) for detecting multiple change points in the model (3.46) using bandwidth $G = 150$ and $\hat{\sigma^2}$ in (3.6) with the asymptotic (red) and bootstrap (blue) critical values are presented by horizontal lines, and location of change points are indicated by dotted vertical red lines in (a); and the estimated means are located in (b). 77

3.10 Boxplot for different segments of cell stimulus response data and five segments are observed using CUSUM process (3.4) in (a), and three segments are observed using MOSUM process (3.20) in (b). 77
3.11 Model Adequacy Checking for the model (3.48): for *Independency*, current residuals versus lagged residuals are plotted in (a) and autocorrelation function (ACF) of residuals at different lags are in (b); *Homoscedasticity* assumption is verified by plotting residuals against predicted values in (c); and *Normality* assumption in (d). 79
3.12 Detection of change points in the model (3.48) using (a) OLS-CUSUM test statistic (3.37), (b) OLS-MOSUM test statistic (3.39) with bandwidth $h = 0.3$, (c) Recursive CUSUM test statistic (3.41), and (d) Recursive MOSUM test statistic (3.43) with bandwidth $h = 0.3$, and the location of change points are indicated by dotted vertical red lines. 80
3.13 Detection of change points in the model (3.48) 81
3.14 Model Adequacy Checking for the model (3.54): for *Independency*, current residuals versus lagged residuals are plotted in (a) and autocorrelation function (ACF) of residuals at different lags are in (b); *Homoscedasticity* assumption is verified by plotting residuals against fitted values in (c); and *Normality* assumption in (d). 83
3.15 Detection of change points in the model (3.54) using (a) OLS-CUSUM test statistic (3.37), (b) OLS-MOSUM test statistic (3.39), (c) Recursive CUSUM test statistic (3.41), and (d) Recursive MOSUM test statistic (3.43). 84
3.16 Change point detection in the model (3.54) deploying Gombay's CUSUM test (3.56) for dependent errors without smoothing the data in (a) and with smoothing the data in (b). 88
3.17 Detection of change points in the model (3.54) with the smoothed cell stimulus response data. 89
3.18 Wilcoxon-type rank test process (3.63) is proposed by Csörgo and Horváth [35] for detecting change point in the distribution for i.i.d. data with the asymptotic critical value, and the significant change point's locations are indicated by dotted vertical red lines in (a); and the estimated multiple change points are shown in (b). 91

3.19 Gombay and Liu [52] proposed Wilcoxon-type rank test process (3.64) for detecting change point in the distribution for censored data with the asymptotic critical value, and the significant change point is indicated by a dotted vertical red line in (a); and the estimated change point is shown in (b). 93

5.1 Detection of change point in the hazard distribution using the test $T_n^{(2)}$. (a) U-statistics process for $T_n^{(2)}$, simulated critical value, and the location of significant change point are presented; and (b) True (blue) and estimated (red) change points are located in the simulated survival data. 156

5.2 Detection of change point in the hazard distribution using the test $T_n^{(6)}$. (a) U-statistics process for $T_n^{(6)}$, simulated critical value, and the location of significant change point are presented; and (b) True (blue) and estimated (red) change points are located in the simulated survival data. 156

5.3 Detection of change point in the hazard distribution using the test $T_n^{(8)}$. (a) U-statistics process for $T_n^{(8)}$, simulated critical value, and the location of significant change point are presented; and (b) True (blue) and estimated (red) change points are located in the simulated survival data. 157

5.4 Detection of change point in the hazard distribution using the test statistic $T_n^{(3)}$ in (a), $T_n^{(4)}$ in (b), $T_n^{(7)}$ in (c), and $T_n^{(8)}$ in (d) with the optimal bandwidth $b = 3e^{-11}$, and the asymptotic critical values are presented by the red horizontal lines. 159

5.5 U-statistic process for the test statistic $T_n^{(4)}$ to detect multiple change points in the hazard distributions of cell stimulus responses with the asymptotic critical value is presented by the red horizontal line and the location of change points are indicated by a dotted vertical red line in (a); and the estimated multiple change points are indicated by the red vertical line in (b). 160

5.6 U-statistic process for the test statistic $T_n^{(8)}$ to detect multiple change points in the hazard distributions of cell stimulus responses with the asymptotic critical value is presented by the red horizontal line and the location of change points are indicated by a dotted vertical red line in (a); and the estimated multiple change points are indicated by the red vertical line in (b). 161
5.7 Detection of change point in the hazard distribution using test $T_n^{(7)}$ in (a), and test $T_n^{(8)}$ in (b) with the Sheather-Jones plug-in bandwidth $b = 0.05$, and the asymptotic critical values are presented by the red horizontal lines. 163
5.8 U-statistic process for the test statistic $T_n^{(8)}$ to detect multiple change points in the hazard distributions of breast cancer mortality with the asymptotic critical value is presented by the red horizontal line and the location of change points are indicated by a dotted vertical red line in (a); and the estimated multiple change points are indicated by the red vertical line in (b). 165
5.9 Estimated hazard functions using Nelson-Aalen estimator for the breast cancer mortality data during the time period 1973 to 1976 in (a); 1977 to 2004 in (b); and 2005 to 2010 in (c). 165

List of Tables

2.1 Piecewise Constant Model with Two Change Points based on 5,000 Simulations ($n = 500$) 34
2.2 Piecewise Linear Model with Two Change Points based on 5,000 Simulations ($n = 1000$) 37
2.3 Change points in the hazard functions detection using the Wald statistic (2.31) for the model (2.29) of the Breast Cancer Patients (SEER Data) diagnosed with the first tumor. 39
2.4 Change points in the hazard functions testing using the Wald statistic (2.28) for the model (2.25) of Cell Stimulus Responses under H_0 : $\alpha_{(k-1)} - \alpha_k = 0$ vs. $H_1 : \alpha_{(k-1)} - \alpha_k \neq 0$. 42

3.1 Single Imputation Producers 68
3.2 Performance of Single Imputation Methods for Parameter Estimates with MCAR mechanism 71

5.1 The Optimal Bandwidth for Epanechnikov kernel using the Sheather-Jones plug-in Method for Simulated Exponential Sample with Parameter μ and Sample Size n. 139
5.2 Some Selected Critical Values $c_\alpha^{(1)}(n)$ at various Level of Significance α and Sample Size n for Test Statistics $T_n^{(1)}$, $T_n^{(3)}$ and $T_n^{(9)}$. 140
5.3 Some Selected Critical Values $c_\alpha^{(2)}(n)$ at various Level of Significance α and Sample Size n for Test Statistics $T_n^{(2)}$, $T_n^{(4)}$ and $T_n^{(10)}$. 140
5.4 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(1)}(n)$ of Test Statistics $T_n^{(1)}$ and $T_n^{(3)}$ for $V = \frac{2}{\mu}$ and Optimal Bandwidth b. . . 141
5.5 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(2)}(n)$ of Test Statistics $T_n^{(2)}$ and $T_n^{(4)}$ for $V = \frac{2}{\mu}$ and Optimal Bandwidth b. . . 142
5.6 Selected Critical Values $c_\alpha^{(5)}(n)$ at various Level of Significance α and n for Test Statistic $T_n^{(5)}$ with $q(t) = (t(1-t))^{0.2}$. 143

5.7 Some Selected Critical Values $c_\alpha^{(6)}(n)$ at various Level of Significance α and n for Test Statistic $T_n^{(6)}$ with $q(t) = (t(1-t))^{0.2}$. 144
5.8 Some Selected Critical Values $c_\alpha^{(7)}(n)$ at various Level of Significance α and n for Test Statistic $T_n^{(7)}$. 144
5.9 Some Selected Critical Values $c_\alpha^{(8)}(n)$ at various Level of Significance α and n for Test Statistic $T_n^{(8)}$. 145
5.10 Some Selected Critical Values $c_\alpha^{(11)}(n)$ at various Level of Significance α and n for Test statistic $T_n^{(11)}$. 145
5.11 Some Selected Critical Values $c_\alpha^{(12)}(n)$ at various Level of Significance α and n for Test statistic $T_n^{(12)}$. 146
5.12 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(5)}(n)$ of Test Statistic $T_n^{(5)}$ with $q(t) = (t(1-t))^{0.2}$ and $V = \frac{2}{\mu}$ and Optimal Bandwidth b. 146
5.13 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(6)}(n)$ of Test Statistic $T_n^{(6)}$ with $q(t) = (t(1-t))^{0.2}$ and $V = \frac{2}{\mu}$ and Optimal Bandwidth b. 147
5.14 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(7)}(n)$ of Test Statistic $T_n^{(7)}$ with $q(t) = (t(1-t))^{0.2}$, $V = \frac{2}{\mu}$ and Optimal Bandwidth b. 148
5.15 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(8)}(n)$ of Test Statistic $T_n^{(8)}$ with $q(t) = (t(1-t))^{0.2}$, $V = \frac{2}{\mu}$ and Optimal Bandwidth b. 149
5.16 Simulated Powers of Test Statistics $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ for $V = \frac{2}{\mu_1}$, and Optimal Bandwidth b, using the Simulated Critical Value for $\mu_1 = \mu_2 = 1$. 151
5.17 Simulated Powers of Test Statistics $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ for $q(t) = (t(1-t))^{0.2}$, $V = \frac{2}{\mu_1}$, and Optimal Bandwidth b, using the Simulated Critical Value for $\mu_1 = \mu_2 = 1$. 152
5.18 Simulated Powers of Test Statistics $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ for $V = \frac{3}{\mu_1}$, and Optimal Bandwidth b, using the Simulated Critical Value for $\mu_1 = \mu_2 = 1$. 153
5.19 Simulated Powers of Test Statistics $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ for $V = \frac{3}{\mu_1}$, and Optimal Bandwidth b, using the Simulated Critical Value for $\mu_1 = \mu_2 = 1$. 154

5.20 Simulated Estimates of Change Point in the Hazard Distribution using Test Statistics $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ for $V = \frac{3}{\mu_1}$, and Optimal Bandwidth b. 155
5.21 Simulated Estimates of Change Point in the Hazard Distribution using Test Statistics $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ for $V = \frac{3}{\mu_1}$, and Optimal Bandwidth b. 155
5.22 Some Selected Critical Values of $T_n^{(7)}$ and $T_n^{(8)}$ Test Statistics at various α and n. 164

A.1 Variables descriptions for the Breast Cancer Patients using SEER Data from 1973-2010. 171

B.1 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(1)}(n)$ of Test Statistics $T_n^{(1)}$ and $T_n^{(3)}$ for $V = \frac{3}{\mu}$ and Optimal Bandwidth b. . . 172
B.2 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(2)}(n)$ of Test Statistics $T_n^{(2)}$ and $T_n^{(4)}$ for $V = \frac{3}{\mu}$ and Optimal Bandwidth b. . . 173
B.3 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(5)}(n)$ of Test statistic $T_n^{(5)}$ with $q(t) = (t(1-t))^{0.2}$ and $V = \frac{3}{\mu}$ and Optimal Bandwidth b. 174
B.4 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(6)}(n)$ of Test statistic $T_n^{(6)}$ with $q(t) = (t(1-t))^{0.2}$ and $V = \frac{3}{\mu}$ and Optimal Bandwidth b. 175
B.5 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(7)}(n)$ of Test statistic $T_n^{(7)}$ with $q(t) = (t(1-t))^{0.2}$, $V = \frac{3}{\mu}$ and Optimal Bandwidth b. 176
B.6 Comparison of Simulated and Approximated Critical Values $c_\alpha^{(8)}(n)$ of Test statistic $T_n^{(8)}$ with $q(t) = (t(1-t))^{0.2}$, $V = \frac{3}{\mu}$ and Optimal Bandwidth b. 177

Chapter 1: Introduction

1.1 Problem Overview

Several critical diseases like cancer, tuberculosis, diabetes, HIV/AIDS, brain stroke and heart disease cause heavy strain on scanty health-care resources and are major causes of mortality worldwide. So achieving medical advancement against these diseases remains the prime public health interest. Medical community exerts continuous endeavor and deploys significant amount of resources in different clinical trials on these diseases to assess impact on survival experience. Measuring and understanding the true impact of such medical breakthroughs, treatments, interventions and initiatives are possible through analyzing the survival trend of patient population as a whole. To explore the complete and accurate picture of survival trend we need to estimate the hazard function. Such an analysis is immensely valuable as it has implication in health-care policy and resource allocation decisions.

The hazard function is an important component of survival analysis as it describes the instantaneous risk of failure at a given time point. Survival analysis contains time-to-event data, which typically comprise an initiating event, say onset of a disease, and a terminating event, say death. Typical study interest of survival analysis revolves around the risk of failure to understand the true impact of newly developed medical breakthroughs and treatments, therefore, a careful study of the hazard function can be more helpful. This function is also known as 'failure rate function' in Engineering and as 'force of mortality function' in Demography. The hazard function may increase, decrease, remain constant, or indicate a more complicated structure. Figure 1.1 exhibits different kinds of hazard function for the risk of dying of human population at various stages. The *yellow line* in Figure 1.1 indicates, for instance, a complicated structure of hazard function, where an increasing and then decreasing hazard function is observed for patients with tuberculosis who have risks that increase initially then decrease after taking the treatment. Hence, the change in hazard

function reveals the significant impact of the treatment for tuberculosis, which is also true for any other critical diseases.

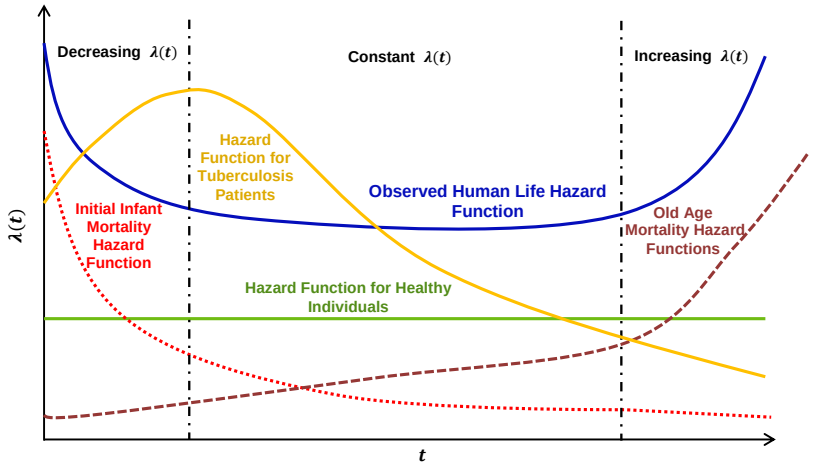


Figure 1.1: Examples of Hazard Functions: the bathtub curve describes the process of human life in the *blue line*, a decreasing shape indicates the risk of infant mortality in the *red line*, the constant hazard function is the risk of healthy individuals between 18 and 40 years of age in the *green line*, an increasing hazard function observes the risk of cancer patients or old age individuals in the *garnet red line*, and an increasing and then decreasing hazard function is described by patients with tuberculosis after taking treatment in the *yellow line*.

Although study of the hazard function provides a refined insight into the structural change of the risk pattern of diseases duration, common survival models are more concerned with the effects of the covariates on the hazard function and do not require explicit estimation of the hazard function, for instance, the Cox proportional hazards model. There are also several models where explicit estimation of the hazard function is required, for instance, change point models for hazard functions. These models assume a function with different hazard rates that change at a few time points. These time points are often referred to as the change points, and are unknown and need to be first detected and then estimated.

Change point models for hazard functions usually occur in medical follow-up studies after a major operation, treatments, and/or interventions, which is also known as change point hazard function models. Matthews and Farewell [93] first introduced this type of models to study data obtained in the treatment of leukaemia patients.

They considered the following type of model, also known as a piecewise constant hazard model, for the independent and identically distributed (i.i.d.) survival times $T_1, T_2, \ldots, T_n$,

$$\lambda(t) = \begin{cases} \lambda_1, & 0 \leq t < \tau \\ \lambda_2, & \tau \leq t, \end{cases} \tag{1.1}$$

where τ is the change point and $(\lambda_1, \lambda_2 \geq 0)$ are the value of the hazard functions before and after the change point τ. Chapter 2 is devoted to describing and discussing this type of models with their testing and estimation procedures for more detailed understanding of change point in the hazard functions. In such medical studies, there is usually a high initial risk and then the risk settles down to a lower constant long term risk, which indicates a change in the hazard functions. However, due to improvement in medical breakthroughs, treatments or diagnosis, there may be two or more changes in the hazard functions.

Furthermore, inevitable characteristic of survival analysis is censoring, which distinguishes survival analysis from other areas in statistics. Censoring occurs when incomplete information is observed about the survival time of some individuals in the study. Typically, survival data comprises an initiating event and a terminating event. Because of scanty health-care resources, a study generally recruits subjects who have experienced their initiating event, the so-called prevalent case, then the recruited samples are followed to a terminating event or censoring. Therefore, the survival time of recruited subjects is left truncated and will be right censoring in further followed-up.

Estimation of the change point in the hazard functions is insufficient without previous testing of the existence of this change along with inevitable censoring. Testing and estimation of change point in the hazard functions has been investigated in different settings. Most of the early work of testing and model fitting of one change point for the hazard functions in a piecewise constant hazard model focused on parametric setting, for instance, Matthews and Farewell [93], Nguyen et al. [104], Matthews et al. [94], Yao [150], Worsley [148], Henderson [59], and Loader [89], even without considering censoring. The semi-parametric approach was recently taken by Pons [114], Kosorok and Song [78], and Zhao et al. [153] with considering censoring, but only focused on estimation. Gijbels and Gürler [50] also investigated only the estimation of change point for the hazard functions based on the estimated cumulative hazard combines the least squared principle with the martingale approach with censoring,

but no theoretical results are available. Dupuy [42] used an exponential regression model with covariates based on right-censoring in testing the existence of a change in the parameters of hazard function as well as regressions, and proposed likelihood ratio type tests and constructed non-asymptotic bounds for the type II probability.

More recently Goodman et al. [54, 55] and Qian and Zhang [117] contributed in multiple change points hazard problems, the latter proposed an algorithm to fit both susceptible and long-term survivors with observed covariates through a grid search weighted least squares method assuming that all potential change points lie in a certain known interval, while the former developed sequential testing procedures with likelihood ratio test and Wald type test statistics in the piecewise constant hazard model and piecewise linear hazard model. Goodman et al. [54, 55] estimated multiple change points hazard using the Nelder-Mead Simplex algorithm. Those works are mainly based on simulations without adequate theoretical information.

In nonparametric setting, Antoniadis et al. [12] and Müller and Wang [98] studied the change point problem when observations are right censoring. Antoniadis et al. [12] used wavelets while Müller and Wang [98] developed a nonparametric alternative approach known as 'smooth approximation model' to approximate the piecewise constant hazard model by kernel smoothing estimator under right censoring. Authors proposed to detect the point of most rapid change in the smooth hazard function by finding the zero of the estimator of the second derivative hazard function.

Furthermore, despite the lack of evidence, the use of nonparametric classical change point methods in hazard problem is rather rare. In such circumstances, nonparametric classical change point methods are only observed in testing the change in distribution with censored data by Stute [136], Ferger [45], Horváth [62], Aly [2], Gombay and Liu [52], Hušková and Neuhaus [65], and Komárková [77].

1.2 Motivation

Early works on a change point in the hazard functions did not consider censored data, as they claimed that dropping those data did not affect significantly the outcome of the likelihood ratio test. In most of the subsequent work develops theory, hereafter, either by discarding censored data and only considering the observable survival time

variable or by modifying the likelihood function for censored data. Thereafter, many subsequent works though considered censoring but have some limitations. For instance, the most usable recent work, proposed by Goodman et al. [54, 55], testing and estimating multiple change points in the hazard functions in the piecewise constant hazard model and the piecewise linear hazard model, are capable of handling an unlimited number of covariates but restricted to an additive nature only and suffers for the number of censored observations. When censoring occurs near the change points, valid estimation of these points is not possible. Moreover, to avoid drawback issues surrounding the likelihood ratio test authors proposed a Wald type statistic, which requires to calculate the variance of the estimates by differentiating the likelihood in the denominator of the test statistic and is the biggest pitfall of this testing procedure. Additionally, the estimation procedure depends on an optimization technique which can only produce better results with significant technical insights.

Change point hazard models have been extensively investigated by many authors, however the literature on nonparametric classical change point problems is rather limited. In some instances, it is important to know the complete distribution pattern of the hazard function for an entire population, which allows researchers and clinicians a better understanding of how changing medical practice affects the survival experience for a patient population. Moreover, all the existing methods in the context of change point for the hazard functions are based on some models, which may face model misspecification errors. These change point models of the hazard functions also take into account the entire data set at a time, which is technically not that much convenient.

Hence, we are motivated to develop a nonparametric classical change point method to detect the change point in the hazard distribution. Such methodology has three types of benefits. Firstly, since this approach does not consider any model, so we do not need to take into account model misspecification errors. Secondly, this approach allows practitioners to divide the entire data into some homogeneous segments based on the significant change points of the hazard functions. Then we can estimate the hazard functions using either any nonparametric hazard estimator or any simple model, for instance, the Cox proportional hazard models. Thirdly, this approach allows any percentage of censored data to identify the significant change points of hazard distribution.

This research is also motivated by two real studies, the first is an interest in examining the breast cancer mortality rates among the recruited patients in the United States during 1973 to 2012. The second is examining the cell stimulus response rates among the sampled animal cells observed in an animal physiology study at the University of Kaiserslautern.

1.3 State of the Art

There are only a few classical change point papers dealing with detection of changes when only censored observations are available. In this section, we briefly summarize those works, as we are interested to develop a nonparametric classical change point method for identifying change point in hazard function with censored data. Typically, $T_1, T_2, \ldots, T_n$ is a sequence of independent non-negative survival times with distribution function F. The patient can be withdrawn from the study, hence committed censoring, due to an accidental death, a migration of human population, limited time of the study, etc. Therefore, the censoring random variables $C_1, C_2, \ldots, C_n$, which are assumed to be independent of T_i, have the distribution functions G. The observed right censored data are denoted by the pairs (X_i, δ_i), $i = 1, 2, \ldots, n$, where $X_i = \min\{T_i, C_i\}$ and $\delta_i = I\{T_i \leq C_i\}$. Here, I is an indicator function and δ_i is a censoring indicator variable. The observed data has a pdf $h(x)$ and a distribution function H defined by $1 - H = (1 - F)(1 - G)$.

Stute [136] considered an estimator of the change point based on U-statistics for randomly censored data. He suggested some estimators of the change point $\tau \in [0, 1]$, with $F_1 \neq F_2$, where $T_j \sim F_1$ for $1 \leq j \leq \lfloor \tau n \rfloor$ and $T_j \sim F_2$ for $\lfloor \tau n \rfloor + 1 \leq j \leq n$, based on U-statistics using antisymmetric kernels.

$$r_n(\lfloor \tau n \rfloor) = \frac{\lfloor \tau n \rfloor (n - \lfloor \tau n \rfloor)}{n^2} \int_{-\infty}^{V} \int_{-\infty}^{V} \frac{K(x, y) d\tilde{H}_{\lfloor \tau n \rfloor}(x) d\tilde{H}^*_{\lfloor \tau n \rfloor}(y)}{(1 - H_n(x-))\,(1 - H_n(y-))},$$

where V satisfying $F(V) < 1$ and $G(V) < 1$, and after dividing the entire sample into two subsamples up to and after the $\lfloor \tau n \rfloor$th observation, the sub-distribution functions are

$$\tilde{H}_{\lfloor \tau n \rfloor}(x) = \frac{1}{\lfloor \tau n \rfloor} \sum_{1 \leq i \leq \lfloor \tau n \rfloor} I\left\{X_i \leq x, \delta_i = 1\right\},$$

$$\tilde{H}^*_{\lfloor \tau n \rfloor}(x) = \frac{1}{n - \lfloor \tau n \rfloor} \sum_{\lfloor \tau n \rfloor < i \leq n} I\left\{X_i \leq x, \delta_i = 1\right\},$$

$$H_{\lfloor \tau n \rfloor}(x) = \frac{1}{\lfloor \tau n \rfloor} \sum_{1 \le i \le \lfloor \tau n \rfloor} I\left\{X_i \le x\right\}, \quad \text{and}$$
$$H^*_{\lfloor \tau n \rfloor}(x) = \frac{1}{n - \lfloor \tau n \rfloor} \sum_{\lfloor \tau n \rfloor < i \le n} I\left\{X_i \le x\right\},$$

and $K : \mathbb{R}^2 \to \mathbb{R}$ is a measurable mapping with the antisymmetry property $K(x, y) = -K(y, x)$. He proved that under the alternative hypothesis and as $n \to \infty$, $r_n(\lfloor \tau n \rfloor) - \tau = O\left(n^{-1} \log n\right)$ almost surely by assuming $\tau \in [0, 1]$ and equal censorship. His results were extended by Ferger [45] for antisymmetric kernel and Horváth [62] for antisymmetric as well as symmetric kernels. Ferger [45] considered a independent random sample $\xi_1 = (X_1, \delta_1), \ldots, \xi_n = (X_n, \delta_n)$ and assumed any change in the distribution function of the T_i usually results in a change in the distribution $\mathscr{L}(\xi_i)$ of ξ_i. Using the two sub-samples $\xi_1, \ldots, \xi_{\lfloor \tau n \rfloor}$ and $\xi_{\lfloor \tau n \rfloor + 1}, \ldots, \xi_n$ he proposed an estimator based on U-type statistic

$$\hat{\theta}_n = \frac{1}{n} \arg \max_{0 \le \lfloor \tau n \rfloor \le n-1} \omega\left(\frac{\lfloor \tau n \rfloor}{n}\right) \left| \sum_{i=\lfloor \tau n \rfloor + 1}^{n} \sum_{j=1}^{\lfloor \tau n \rfloor} K\left(\xi_i, \xi_j\right) \right|,$$

where $\omega : (0, 1) \to (0, \infty)$ is a weight function of the type $\omega(t) = t^{-a}(1 - t)^{-b}$, $0 \le a, b \le 1$ and K is antisymmetric kernel, for instance, $K(x, y) = x - y$. He also showed that for $K(x, y) = x - y$ and $\omega(t) = t^{-1}(1 - t)^{-1}$ the estimator takes the following form

$$\hat{\theta}_n = \frac{1}{n} \arg \max_{0 \le \lfloor \tau n \rfloor \le n-1} \left| \frac{1}{\lfloor \tau n \rfloor} \sum_{i=1}^{\lfloor \tau n \rfloor} \xi_i - \frac{1}{n - \lfloor \tau n \rfloor} \sum_{i=\lfloor \tau n \rfloor + 1}^{n} \xi_i \right|,$$

which means a successive comparison of the sub-sample means. And for $K(x, y) = x - y$ and $\omega(t) = t^{-1/2}(1 - t)^{-1/2}$ the estimator becomes

$$\hat{\theta}_n = \frac{1}{n} \arg \max_{0 \le \lfloor \tau n \rfloor \le n-1} \left| \frac{1}{\sqrt{\frac{1}{\lfloor \tau n \rfloor} + \frac{1}{n - \lfloor \tau n \rfloor}}} \left(\frac{1}{\lfloor \tau n \rfloor} \sum_{i=1}^{\lfloor \tau n \rfloor} \xi_i - \frac{1}{n - \lfloor \tau n \rfloor} \sum_{i=\lfloor \tau n \rfloor + 1}^{n} \xi_i \right) \right|,$$

which is the well-known Gaussian test-statistic for the hypothesis of equality of the means between the two sub-sample observations. Ferger [45] also proved that under alternative hypothesis and the assumption of equal censorship as $n \to \infty$, $|\hat{\theta}_n - \tau| = O(1/n)$ almost surely.

Horváth [62] studied test procedures based on U-statistics for antisymmetric as well as symmetric kernels using the following functional of

$$Q_n(\lfloor \tau n \rfloor) = \frac{\lfloor \tau n \rfloor \left(n - \lfloor \tau n \rfloor\right)}{n^{3/2}} \left(\hat{\theta}^{(3)}(\lfloor \tau n \rfloor) - \theta \right), \quad 1 \leq \lfloor \tau n \rfloor < n,$$

where

$$\hat{\theta}(\lfloor \tau n \rfloor) = \int_{-\infty}^{V} \int_{-\infty}^{V} K(x,y) \frac{d\tilde{H}_{\lfloor \tau n \rfloor}(x)}{1 - H_{\lfloor \tau n \rfloor}(x-)} \frac{d\tilde{H}^{*}_{\lfloor \tau n \rfloor}(y)}{1 - H^{*}_{\lfloor \tau n \rfloor}(y-)},$$

and the true parameter is

$$\theta = \int_{-\infty}^{V} \int_{-\infty}^{V} K(x,y) \frac{dH(x)}{1 - H(x-)} \frac{dH(y)}{1 - H(y-)}.$$

He developed the asymptotic distributions of different test statistics under null hypothesis to test the change in distribution with censored data, which are asymptotically distributed as Gumbel distribution, a sequence of Gaussian process, or a function of Brownian bridge. He also showed that his estimator $\hat{\theta}(\lfloor \tau n \rfloor)$ is asymptotically equivalent under null hypothesis, and hence constructed the asymptotic distribution of test statistics for antisymmetric kernels. The developed procedures of Horváth [62] will be elaborated in Chapter 4.

Gombay and Liu [52] proposed and investigated limit properties of a nonparametric test based on ranks related to the Gehan-Wilcoxon statistic that can be expressed as a U-statistic. Authors developed the following limit distribution under the no-change null hypothesis and assuming $G_1 = G_2$ as $n \to \infty$,

$$\max_{1 \leq k < n} \frac{|\sum_{i=1}^{k} U_i|}{\left(\sum_{i=1}^{n} U_i^2\right)^2} \xrightarrow{\mathscr{D}} \sup_{1<t<1} |B(t)|,$$

where $\{B(t), 0 \leq t \leq 1\}$ is a Brownian bridge, i.e., mean zero Gaussian process with covariance function $EB(s)B(t) = \min(t,s) - ts$ when $0 \leq t, s \leq 1$. The critical values are obtained from the well-known identity

$$P\left\{ \sup_{0<t<1} |B(t)| > b \right\} = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 b^2}, \quad b > 0,$$

which yields 1.63, 1.36 and 1.22 for $\alpha = 0.01, 0.05$ and 0.10, respectively. The generalized rank of (X_i, δ_i) is defined as

$$U_i = \sum_{j=1}^{n} \left\{ I\left(X_i > X_j, \delta_j = 1\right) - I\left(X_i < X_j, \delta_j = 1\right) \right\}, \quad i = 1, \ldots, n.$$

They used the theory of exchangeable variables to investigate its properties. Under the alternative hypothesis they proved $|\hat{\tau}_n - \tau| = O_p(1)$ as $n \to \infty$ and proposed the change point estimator as

$$\hat{\tau}_n = \lfloor \hat{\tau n} \rfloor = \arg \max_{1 \le k < n} \frac{|\sum_{i=1}^{k} U_i|}{\left(\sum_{i=1}^{n} U_i^2\right)^2}.$$

Extensive studies for such procedures with its weighted type forms were conducted in the doctoral thesis of Liu [87].

Aly [2] developed test based on quantile functions, first used by Csörgo and Horváth [35] in change point tests for uncensored data, and studied their limit behavior under the null hypothesis for right censored data. He separated the entire data into the uncensored $X_1^1, \ldots, X_{N_{1,k}}^1$ and the censored $X_1^2, \ldots, X_{N_{2,k}}^2$ observations, where the total number of uncensored observation is $N_{1,k}$ with the empirical process $\hat{F}_{1k}(x)$ and the quantile process $\hat{Q}_{1k}(y)$, and in the case of censored data that are $N_{2,k}$, $\hat{F}_{2k}(x)$, and $\hat{Q}_{2k}(y)$, respectively. The empirical and quantile processes was defined by

$$\hat{F}_{ik}(x) = \frac{1}{N_{i,k}} \sum_{j=1}^{N_{i,k}} I\left(X_j^i < x\right), \quad i = 1, 2,$$
$$\hat{Q}_{ik}(y) = \sup\left\{x : \hat{F}_{ik}(x) \le y\right\}, \quad i = 1, 2.$$

He used the following process to develop different kind of tests

$$Y_n^i(s,t) = \frac{1}{\sqrt{N_{i,n}}} \sum_{j=1}^{N_{i,[ns]}} \Psi_t\left(X_j^i - \hat{Q}_{in}(t)\right), \quad s,t \in (0,1), \quad i = 1, 2,$$

where

$$\Psi_t(x) = \begin{cases} -(1-t) & x < 0, \\ t & x \ge 0. \end{cases}$$

[2] developed nine kinds of test statistics using the process $Y_n^i(s,t)$ and assuming $G_1(t) = G_2(t)$ to detect the change in distribution with censored random observations and also derived their asymptotic distributions under the null hypothesis, which follows either a Brownian bridge, or a two-parameter Gaussian process, or the standard normal distribution. Another nonparametric rank change point test was proposed by Buhamra et al. [25] using the modified ranks of Albers and Akritas [1] and the modified ranks of Gehan, Gilbert and Mantel.

All these papers considered censoring variables to be independent and identically

distributed (i.i.d.), but Hušková and Neuhaus [65] developed a test procedure as a generalization of two-sample rank tests under random censoring, where censoring variables were assumed to be independent but not necessarily identically distributed. They assume that the censoring variables C_i's can change at some unknown time point that need not coincide with an eventual change point of censored observations. In this circumstance, Komárková [77] also proposed various rank test statistics as two-sample, max-type and MOSUM-type statistics along the lines of the developed tests of Hušková and Neuhaus [65], and developed the change point estimator corresponding to max-type test statistics.

All of the aforementioned tests focus on a change only in the distribution, but not for the change in the hazard functions when the variables of interest are subject to censoring. Hence, we are motivated to develop a nonparametric classical change point test to detect the change in hazard distribution under the right censoring. Since any change in the distribution function F results in a change in the distribution function H by assuming i.i.d. censoring variables, which is a reasonable assumption with a practical point of view. Hence, we are not considering the change in the censoring variable's distribution, i.e., $G_1 \neq G_2$. More precisely, we are interested to extend the methodology developed by Horváth [62] in the context of hazard distribution for symmetric and antisymmetric kernels under the variables of interest are subject to right censoring.

1.4 Outline of the Thesis

Let us conclude this introduction with an outline of the rest of the topics covered in this thesis. Starting with the classical nonparametric change point analysis with missing data, we have extended our effort on the change point in the hazard functions with censored data and then finally, develop different types of classical nonparametric change point tests to detect change point in the hazard distribution for censored data based on U-statistics using symmetric and antisymmetric kernels.

In Chapter 2, we review various estimators of the hazard function along with different change point models for the hazard functions in the case of i.i.d. data as well as censored data. We also define the concept and meaning of censoring, which is a nearly universal feature of survival data. Existing testing and estimation procedures

of the change point models for the hazard functions are discussed comprehensively in this chapter to get a deeper understanding of this long-standing problem of survival analysis. Subsequently, we develop and use the Cross-Entropy (CE) algorithm in Goodman et al. [54, 55] methodology to estimate multiple change points in the hazard functions after detecting those by Wald type test. The performance of the CE algorithm along with existing counterpart is evaluated by simulation and two real data examples; breast cancer mortality, and cell stimulus responses. The first data set is a well-known survival data set and the later one is a cognitive data set. The cell stimulus response data set is collected from the animal physiology group at the University of Kaiserslautern and the breast cancer mortality data is taken from the 1973-2010 Surveillance, Epidemiology, and End Results (SEER) program data (for details see Appendix A).

Chapter 3 reveals different classical nonparametric change point procedures for detecting and estimating change point(s) in the location and regression models, along with different missing data mechanisms and various imputation methods. Thereafter, we apply our understanding to analyze the Cell Stimulus Response data set, which contains 15 censored observations out of 3000 observations, but we considered those as missing to illustrate multiple change point models in mean and regression structures with missing data feature. Additionally, we present a brief introduction and a demonstration to the change point methods for detecting a change in the distribution of the observations under either missingness or random censorship in Section 3.4.3. Therefore, this review attempts to raise the awareness of the missing data problem in change point analysis. Nevertheless, all missing observations suppress as censored in survival analysis.

Chapter 4 is devoted to developing the necessary theory for our proposed tests based on U-statistic process to detect the change point in the hazard distribution. First, we focus our efforts to find an equivalent estimator of $\hat{\theta}(k)$ in Horváth [62], afterwards, to develop the limit distribution of four types of unweighted as well as another four types of weighted test statistics under the null hypothesis to detect the change in the hazard functions using a symmetric kernel function. We further show that our procedure has an asymptotic power of 1. We also derive the asymptotic distributions of some weighted and unweighted test statistics under the null hypothesis of no change in the hazard functions for an antisymmetric kernel.

Chapter 5 illustrates all the simulation results based on Monte Carlo experiments to describe the asymptotic behavior of our developed tests in Chapter 4 with their corresponding power performances. We demonstrate the usefulness of our proposed approach on two real data examples; breast cancer mortality with 78.9% of censored observations, and cell stimulus responses with 0.5% of censored observations. Hence, we investigate the performances of our developed methodologies with very high as well as with very low percentages of censored scenarios. Finally, Chapter 6 summarizes all of our findings and also pointing out some directions for future research.

Chapter 2: Change Point in Hazard Functions with Censored Data

The hazard function is an important component of survival analysis as it describes the instantaneous risk of failure at a given time point. Although common survival methods such as the Cox proportional hazards model do not require explicit estimation of the hazard function but there are several situations where explicit estimation of the hazard function is useful. One such case is change point hazard rate models. These models assume a function with different hazard rates that change at a few time points. These time points are often referred to as the change points, and are unknown and need to be estimated. To explore the complete and accurate picture of survival trend we need to estimate the change point in the hazard functions. Such an analysis is immensely valuable as it has implication in health-care policy and resource allocation decisions.

The goal of this chapter is to make a short review of different hazard estimators along with various existing methodologies for testing and estimation of change point(s) in the hazard functions. Although the emphasis of this chapter is on the hazard function and its change point analysis, one cannot describe this topic without considering censoring characteristic. We pay considerable attention to describing and discussing hazard function estimators and change point in the hazard functions with i.i.d. and censored data features. Moreover, we also develop and utilize an optimization algorithm using the Cross-Entropy method for estimating multiple change points in hazard analysis.

The rest of this chapter is structured as follows. Section 2.1 explains the censoring mechanism to understand different kinds of censored data in survival analysis. A brief review of different types of parametric and nonparametric hazard estimators

in the case of both i.i.d. and censored data are discussed in Section 2.2. Extensive literature on change points in hazard functions is summarized in Section 2.3. In that section, we also propose an algorithm using the Cross-Entropy (CE) method to estimate multiple change points in hazard functions. Section 2.4 illustrates simulation results along with real data application to evaluate the performance of CE method along with existing counterpart. Finally, we conclude with a general discussion in Section 2.5.

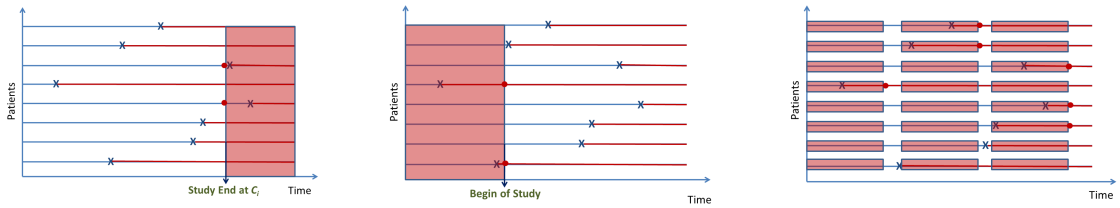
2.1 Censored Data

Time-to-event is a common feature in survival analysis, which introduces censoring. The term 'lifetime', 'survival time' or 'time-to-event' denotes the time until the occurrence of an event of interest, e.g., recurrence of a disease event, discharge from hospital, time to complete a task (such as PhD thesis), patients being followed to a cancer event, etc. The survival time can be defined by two time points: *the time of origin*, i.e., the time at which an original event, such as an infection, occurs and *the time of failure*, i.e., the time at which the final event, such as death, occurs. A subject is said to be at risk if the original event has occurred, but the final event has not.

Censoring occurs when incomplete information is available due to random cause about the survival time of some individuals. Point censoring (implies right censoring and left censoring) and interval censoring are the common types of censoring. In right censoring, a sample has been followed for some time (study starting to ending time), but we don't know the exact time of occurrence of the event rather the time is known as greater than the study ending time. In left censoring, the time of origin is reported as less than the study starting time, but it is not known exactly when it occurred. Whereas, the exact time of event occurs is not known precisely in interval censoring, but an interval bounding this time is known. Figure 2.1 illustrates these three types of censoring situations.

Censoring also can be classified based on reasons or levels of censoring as Type I censoring and Type II censoring. When a study is conducted over a specified period of time, then Type I censoring is concerned. Type II censoring arises when a study progresses until the failure of the first r individuals (number of individuals are fixed in advance), which is frequently done in industrial quality assessment. A sample is

randomly censored when both the number of censored observations and the censoring levels are random outcomes. This type of censoring commonly arises in medical time-to-event studies. For a good discussion on censoring with various examples and different types of censoring mechanisms, we refer to Leung et al. [83].



(a) (b) (c)

Figure 2.1: (a) Right censoring with fixed censoring time, (b) Left censoring due to late study onset, and (c) Interval censoring due to discrete observation times.

Truncation is another feature of time-to-event data, which makes researcher confused with censoring. Truncated values are those that are not reported if the value exceeds some limit. To illustrate truncation, consider an analytical laboratory reporting the concentration of atrazine in a ground water sample and the laboratory equipment can detect the presence of atrazine if it is reported equal or greater than 0.05 ppb. Suppose, { 0.02 ppb, 5 ppb, 2 ppb} is our data set. Our example data set would become: { <0.05 ppb, 5 ppb, 2 ppb} in censoring and { 5 ppb, 2 ppb} in truncation. The practical difference between censored and truncated data is that the number of censored values is known, but the number of truncated values is not. Observed and censored survival data are much more common than truncated data.

For giving a more detailed information regarding right and left censoring, we constructed Figure 2.2 from Leung et al. [83].Figure 2.2 represents a study which begins at time T_0 and ends at T_1, and observing seven subjects (A, B, C, D, E, F and G). Each subject's follow-up time ended with either by 'X' or by 'red dot', where 'X' indicates that event of interest occurs at that point and 'red dot' indicates that due to other reasons subject become *loss-to-follow-up* or early termination of a study.

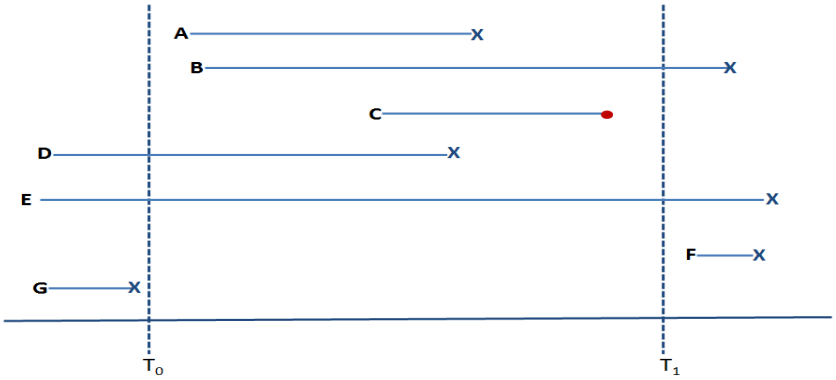


Figure 2.2: Types of Point-Censored Observations

For Subject A, the entire risk period falls within the observation period and the time of occurrence of the event is known, hence there is no censoring for A. The risk period starts during the observation period for Subject B and the event occurs after follow-up is terminated at T_1. The observation of Subject B is therefore right censored due to study termination at T_1, also known as *end-of-study* censoring. Subject C also represents a case with right censoring, specially known as *loss-to-follow-up* censoring as it is lost due to other reasons. A left truncation case is observed for Subject D, where the event starts before at time T_0 and ends within follow-up time. Subject E represents a special case of censoring known as *doubly censored*, in which the observation is both left and right censored. Completely right censored and completely left censored is observed in Figure 2.2 for Subjects F and G, respectively.

2.2 Review of Hazard Functions Estimation

Estimation of hazard functions have been studied extensively in many fields, for instance, reliability theory, engineering, geophysics, actuarial science, medical statistics or broadly speaking survival analysis, etc. The hazard function is also known as 'failure rate function' in engineering and 'force of mortality' in demography . The hazard function can be estimated based on independent and identically distributed (i.i.d.) data or censored data. Censored data frequently observed in survival studies, and this feature makes the estimation of hazard function an attractive and challenging topic in survival analysis. Nonparametric methods are more common in estimating hazard function due to their flexible, model-free and data-driven features. Different

nonparametric approaches are available to estimate hazard functions for grouped lifetime data observed at certain time interval and for continuously observed lifetime data, for details confer Wang [143], which was first published in Wang [142].

This thesis concerns with continuously distributed data to estimate hazard functions, which estimation procedure is conceptually close to density estimation. Kernel estimators, spline estimators, local smooth estimators, and ratio-type estimators are usually discussed in nonparametric estimation of the hazard function. As we are interested in the ratio-type hazard and the kernel type hazard estimators, so a brief review of these two types of estimators based on i.i.d. and censored data will be given in this section.

Case 1: i.i.d. Data

Suppose, $T_1, T_2, \ldots, T_N$ be i.i.d. survival times with probability density function (pdf) $f(t)$ and distribution function $F(t)$. Therefore, the survival function can be defined as $S(t) = P(T_i \geq t) = 1 - F(t)$, which implies the probability of being alive at time t; and the hazard function, which specifies the instantaneous rate of death or failure at time t, given that the individual survives until time t is defined by

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T_i < t + \Delta t \mid T_i \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}. \tag{2.1}$$

In particular, $\lambda(t)\Delta t$ is the approximate probability of death in $[t, t + \Delta t)$ for those individuals who survive until time t. The corresponding cumulative hazard function is $\Lambda(t) = \int_0^t \lambda(y)dy$ and also can be defined in terms of survival function as $\Lambda(t) = -\log(S(t))$. Hence, $\lambda(t)$ for a continuous lifetime distribution possesses the properties: $\lambda(t) \geq 0$ and $\int_0^\infty \lambda(t)dt = \infty$. The hazard $\lambda(t)$ must be non-negative but does not necessarily have an upper bound. The cumulative hazard function $\Lambda(t)$ must be non-negative, nondecreasing, and unbounded.

We can elaborate the property $\lambda(t) \geq 0$ of hazard functions with an example. For instance, we are studying the time until a patient gets the flu (influenza), and we measured time in months and we got a hazard rate of .10, that is, a person is expected to get the flu .10 times per month assuming the hazard remains constant during that month. We could just as well measure the time in years (12 months), and we would get a hazard rate of 1.20, i.e., a person is expected to get flu 1.20 times per year.

In parametric hazard function estimation, using the relationship (2.1) one can calculate various types of hazard functions: the hazard function is a constant function in t for the exponential distribution; for the extreme value, logistic and the Weibull distributions hazard functions are strictly increasing and continuous in t; for more details confer Kalbfleisch and Prentice [71].

The nonparametric estimation of the cumulative hazard function leads to the Product-Limit estimator or, sometimes, Kaplan-Meier estimator [73] of the survival function

$$\hat{S}(t) = \prod_{i:T_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right), \tag{2.2}$$

where d_i is the number of deaths at $T_{(i)}$, n_i is the number of alive just before time $T_{(i)}$, and $T_{(i)}$'s are ordered and have no ties, i.e., $T_{(1)} < T_{(2)} < \cdots < T_{(n)}$. Hence, n is the number of different values in $T_1, T_2, \ldots, T_N$. Hereafter, we can estimate the cumulative hazard function by $\hat{\Lambda}(t) = -\log\left(\hat{S}(t)\right)$.

The Nelson-Aalen estimator [103] estimates $\Lambda(t)$ directly by

$$\hat{\Lambda}_n(t) = \sum_{j=1}^{i} \frac{d_j}{n_j}, \quad T_{(i-1)} < t \leq T_{(i)}. \tag{2.3}$$

Note that (2.2) and (2.3) cover the general case, where survival times do not necessarily have a density. In case of a density, $d_i = 1$, $i = 1, \ldots, n$, with probability 1, and $n_i = n - i + 1$.

The kernel density estimation of hazard function has been discussed by many authors. Asymptotic properties of kernel estimators of the hazard function were investigated by Watson and Leadbetter [144, 145], Murthy [101], Rice and Rosenblatt [120], Singpurwalla and Wong [132], Ramlau-Hansen [118], Burke and Horváth [26] and Patil [109] for i.i.d. data. Watson and Leadbetter [144, 145] studied three estimators (2.5), (2.8) and (2.9) of the hazard function and Rice and Rosenblatt [120] reviewed those estimators in finding the bias and covariance properties, by considering a bounded, bandlimited, symmetric sequence of smooth functions $W_n(u) = \frac{1}{b_n} W\left(\frac{u}{b_n}\right)$ approaching the Dirac delta-function for large n. This delta-sequence method is quite general and covers various types of smoothing methods, including the kernel method with

$$W_n(u) = \frac{1}{b_n} K\left(\frac{u}{b_n}\right),$$

with $\int W_n(u)du = 1$ and by assuming

$$b_n \to 0, \quad \text{with} \quad nb_n \to \infty \quad \text{as} \quad n \to \infty. \tag{2.4}$$

Hence, they defined the ratio-type hazard estimator as

$$\hat{\lambda}_{1,n}(t) = \frac{f_n(t)}{1 - F_n(t)}, \tag{2.5}$$

where $f_n(t)$ is an estimate of the density $f(t)$ by assuming (2.4), and defined by

$$f_n(t) = \int W_n\left(t - u\right) d\hat{F}_n(u), \tag{2.6}$$

and $\hat{F}_n(t)$ is the usual empirical distribution function

$$\hat{F}_n(t) = \frac{1}{N}\sum_{j=1}^{N} 1_{[0,t]}(T_j) = \frac{1}{N}\sum_{i=1}^{n} d_i 1_{[0,t]}(T_{(i)}),$$

where $N = \sum_{i=1}^{n} d_i$ is the total number of deaths in the sample. Moreover, let $F_n(t)$ be the smoothed empirical distribution function as an estimate of $F(t)$, defined as

$$F_n(t) = \int_0^t f_n(u)du. \tag{2.7}$$

Another two estimators of the hazard functions based on the delta-sequence smoothing introduced by Watson and Leadbetter [144, 145] and Rice and Rosenblatt [120] are

$$\hat{\lambda}_{2,n}(t) = \int W_n\left(t - u\right) \frac{dF_n(u)}{1 - F_n(u)} = \sum_{i=1}^{n} W_n\left(t - T_{(i)}\right) \frac{1}{(n - i + 1)}, \tag{2.8}$$

$$\hat{\lambda}_{3,n}(t) = \int W_n\left(t - u\right) d\hat{\Lambda}_n(u) = \sum_{i=1}^{n} W_n\left(t - T_{(i)}\right) \log\left[1 + \frac{1}{(n - i + 1)}\right], \tag{2.9}$$

where $\hat{\Lambda}_n(u)$ is an estimate of the cumulative hazard function, i.e., $\hat{\Lambda}_n(u) = -\log(1 - \hat{F}_n(u))$. It was shown in Rice and Rosenblatt [120] that $\hat{\lambda}_{2,n}(t) - \hat{\lambda}_{3,n}(t) = O_p(n^{-1})$, all estimators $\hat{\lambda}_{1,n}(t), \hat{\lambda}_{2,n}(t)$, and $\hat{\lambda}_{3,n}(t)$ have the same asymptotic variance, but $\hat{\lambda}_{1,n}(t)$ has a different asymptotic bias.

Case 2: Censored Data

For a sample of n independent individuals, let $T_1, T_2, \ldots, T_n$ be the i.i.d. survival times with probability density function (pdf) $f(t)$, which are rightly censored by the

i.i.d. censoring random variables, $C_1, C_2, \ldots, C_n$, assumed to be independent of T_i. Let F and G be the distribution functions of T_i and C_i, respectively. The observed right censored data are denoted by the pairs (X_i, δ_i), $i = 1, 2, \ldots, n$, where

$$X_i = \min\{T_i, C_i\}, \quad \delta_i = I\{T_i \leq C_i\} = I\{T_i = X_i\}.$$

Here, I is an indicator function and δ_i is a censoring indicator variable. The observed data has a distribution function H defined by $1 - H = (1 - F)(1 - G)$. $H_n(x)$ is the empirical distribution function of the random variables X_i. For censored variables it is defined as

$$H_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\left\{X_i \leq x, \delta_i = 1\right\}.$$

In the censored case, the Kaplan-Meier estimator of the survival function becomes

$$\hat{S}(x) = \prod_{i:X_{(i)} \leq x} \left(1 - \frac{1}{n - i + 1}\right)^{\delta_{(i)}}, \tag{2.10}$$

where we assume that $T_1, T_2, \ldots, T_n$ are all different survival times which, in case of a density, holds with probability 1. Then, the X_i with $\delta_i = 1$ are also all different, and the probability that a X_i with $\delta_i = 1$ coincides with a censored X_i, $\delta_i = 0$, is 0 too. $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ again denote the ordered values, where ties can only occur between censored observations, and $\delta_{(i)}$ is the indicator of $X_{(i)}$. Hereafter, we can estimate the cumulative hazard function by $\hat{\Lambda}(x) = -\log\left(\hat{S}(x)\right)$.

The Nelson-Aalen estimator $\hat{\Lambda}_n(x)$ for the cumulative hazard function $\Lambda(x)$, which is instrumental in survival analysis for censored data, is defined by

$$\hat{\Lambda}_n(x) = \int_{-\infty}^{x} \frac{dH_n(u)}{1 - \hat{H}_n(u)} = \sum_{i:X_{(i)} \leq x} \frac{\delta_{(i)}}{n - i + 1}, \tag{2.11}$$

where

$$\hat{H}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\left\{X_i \leq x\right\},$$

is the usual empirical distribution function of $X_1, \ldots, X_n$. Smoothing the increments of the random step function $\hat{\Lambda}_n(x)$ and differentiation are used to obtain hazard estimators. Properties of $\hat{\Lambda}_n(x)$ have been studied extensively, for details confer Section IV.1 in Andersen et al. [3].

For censored data Yandell [149], Tanner and Wong [137], Schafer [125], Diehl and

Stute [41], Lo et al. [88], Müller and Wang [99], and Müller and Wang [100] have discussed the kernel hazard estimators with their properties. Tanner and Wong [137] proposed a kernel hazard function estimator for censored data

$$\hat{\lambda}_{4,n}(x) = \sum_{i=1}^{n} K_b\left(x - X_{(i)}\right) \frac{\delta_{(i)}}{(n-i+1)}, \tag{2.12}$$

where K is a symmetric nonnegative kernel, $K(t) = o(t^{-1})$ as $t \to \infty$, $\int K(t)dt = 1$, $K_b(u) = b^{-1}K(u/b)$. The point of interest x is assumed fixed throughout the study and satisfying $0 < H(x) < 1$.

Lo et al. [88] studied an estimator of the hazard function given by

$$\hat{\lambda}_{5,n}(x) = \frac{f_n(x)}{\bar{\Gamma}_n(x)}, \tag{2.13}$$

where the density estimate is defined as $f_n(x) = \frac{1}{b_n}\int K\left(\frac{x-u}{b_n}\right) d\Gamma_n(u)$, assuming K is symmetric, compactly supported, continuous and having bounded variation kernel and the bandwidth sequence $\{b_n\}$ follows Assumption (2.4) along with Assumptions (b2), (b3) and (b4) in [88], and $\bar{\Gamma}_n(x) = 1 - \Gamma_n(x)$, $\Gamma_n(x)$ is the modified version of the Kaplan-Meier estimator defined as

$$\Gamma_n(x) = \begin{cases} 1 - \prod_{X_{(i)} \le x}^{n} \left(\frac{n-i+1}{n-i+2}\right)^{\delta_{(i)}}, & \text{if } x \le X_{(n)}; \\ \Gamma_n\left(X_{(n)}\right), & \text{if } x > X_{(n)} \text{ and the largest observation is uncensored.} \end{cases}$$

Müller and Wang [99] investigated the properties of kernel based hazard estimator with local bandwidth choice by using the convolution of the Nelson [103] estimator with a kernel function, and considered the following general type of estimate

$$\hat{\lambda}_n^{(\nu)}(x) = \frac{1}{b^{(\nu+1)}} \int K_\nu\left(\frac{x-u}{b}\right) d\Lambda_n(u) = \frac{1}{b^{(\nu+1)}} \sum_{i=1}^{n} K_\nu\left(\frac{x - X_{(i)}}{b}\right) \frac{\delta_{(i)}}{(n-i+1)}, \tag{2.14}$$

where $b = b(n)$ is a sequence of bandwidths for which

$$b \to 0, \quad nb^{2\nu+1} \to \infty, \quad \frac{nb}{(\log n)^2} \to \infty, \quad \text{as } n \to \infty. \tag{2.15}$$

Further, K_ν is a kernel function of bounded variation and is of order (ν, k) with support $[-1, 1]$, where $k \ge \nu$ and $\nu \ge 0$. Assuming that the ν'th derivative of λ is k times continuously differentiable on $[0, T]$ with $H(T) < 1$, $\hat{\lambda}_n^{(\nu)}(x)$ estimates $\lambda^{(\nu)}(x)$ with a rate depending on k.

In hazard estimation, symmetric kernel functions are commonly used such as Uniform, Epanechnikov, Biweight and Gaussian, with the following expressions

$$\text{Uniform Kernel:} \qquad K(x) = \frac{1}{2} 1_{\{|x| \leq 1\}}, \qquad -1 \leq x \leq 1, \tag{2.16}$$

$$\text{Epanechnikov Kernel:} \quad K(x) = \frac{3}{4}(1 - x^2) 1_{\{|x| \leq 1\}}, \quad -1 \leq x \leq 1, \tag{2.17}$$

$$\text{Gaussian Kernel:} \qquad K(x) = \frac{1}{\sqrt{(2\pi)}} e^{-x^2/2}, \qquad -\infty < x < \infty, \tag{2.18}$$

$$\text{Biweight Kernel:} \qquad K(x) = \frac{15}{16}(1 - x^2)^2 1_{\{|x| \leq 1\}}, \quad -1 \leq x \leq 1. \tag{2.19}$$

Wang [143] recommended to use either the Epanechnikov kernel or the Gaussian kernel, moreover, Müller [96] found that the Epanechnikov kernel has certain optimal properties.

2.3 Change Point in Hazard Functions

The hazard function is a frequently used function for modeling and evaluating the time related events. Therefore, this function may have one or more change points, or it remains constant. Testing for the existence of a change point(s) and its estimation is the prime concern of change point analysis in hazard functions. There are different hazard models in change point methodology, for instance, parametric change point model, nonparametric change point model, smooth approximation model etc. Good discussion about different methodologies of the change point in hazard functions can be found in the Monograph paper of Müller and Wang [97], and few recent papers Qian and Zhang [117], Bhore and Huque [19] and Anis [6].

Note that a change point in this context has a different meaning than in classical change point analysis. There, a sequence of observations is considered, and at some times their distribution changes. Here, we consider an individual observation of a survival time where the hazard changes after some time more or less suddenly. In this chapter, we are talking about the latter type of change point.

2.3.1 Change Point Models and Methods for Hazard Function

Estimation and testing in a piecewise constant model with one change point have been extensively investigated by many authors (see for example, Matthews and Farewell

[93], Nguyen et al. [104], Matthews et al. [94], Yao [150], Worsley [148], Henderson [59], Loader [89], Pons [114], Gijbels and Gürler [50], and Zhao et al. [153]) in parametric and semi-parametric approaches using likelihood-ratio type tests, score test and Bayesian test. Matthews and Farewell [93] first noted the existence of the change point for the hazard functions in a parametric piecewise constant model when analyzing the failure times of nonlymphoblastic leukemia patients. They considered the following model for the i.i.d. survival times T_i,

$$\lambda(t) = \begin{cases} \alpha_1, & 0 \leq t < \tau, \\ \alpha_2, & \tau \leq t, \end{cases} \tag{2.20}$$

where τ is the change point and (α_1, α_2) are the values of the hazard functions. They used likelihood ratio type tests for detecting τ. Nevertheless, nonlymphoblastic patients' data had 24 censored observations at 182 days, Matthews and Farewell [93] did not consider those censored data in their analysis. They claimed that dropping those data did not affect significantly the outcome of the likelihood ratio test. Hereafter, most of the subsequent work develops theory either by discarding censored data and only considering the observable survival time variable, or by modifying the likelihood function for censored data. The unboundedness feature of the likelihood function when the change point approaches the maximum observation of the failure times is discussed by Nguyen et al. [104]. Matthews et al. [94] considered tests based on the maximal score statistic and showed that the asymptotic limiting process of the normalized score process is related to the OrnsteinUhlenbeck process and the standard Brownian bridge. Yao [150] assumed $T_{(n-1)}$ is the second largest observation and suggested to maximize the log-likelihood function in the change point over $\left[0, T_{(n-1)}\right]$ and gave the asymptotic properties of the estimators for both the change point and the piecewise hazard functions. Worsley [148] derived the exact critical values of the maximum likelihood estimator over three intervals: (i) $\left[0, T_{(n-1)}\right]$, (ii) $[p$th sample quantile, $(1-p)$th sample quantile], and artificially censored the largest observation so that the likelihood function in the change point is finite. Loader [89] explored the model (2.20) with and without censoring for the i.i.d. event times to derive a likelihood-ratio test and gave with the approximate confidence regions and joint confidence regions for the change point and the size of change over another interval.

Müller and Wang [98] developed a nonparametric alternative approach known as 'smooth approximation model' to approximate model (2.20) by a kernel smoothing

estimator (2.14). They proposed to detect the point of most rapid change in the smooth hazard function by finding the zero of the estimator (2.14) with $\nu = 2$ of the second derivative hazard function.

A semiparametric extension of the model (2.20) proposed by Liang et al. [84] incorporates covariates and was defined as

$$\lambda\left(t;\mathbf{Z},\mathbf{x}\right)=\begin{cases}\lambda_0(t)\exp\left((\beta+\theta)\mathbf{Z}+\gamma^{'}\mathbf{x}\right), & t\leq\tau,\\ \lambda_0(t)\exp\left(\beta\mathbf{Z}+\gamma^{'}\mathbf{x}\right), & t>\tau,\end{cases} \tag{2.21}$$

where the change point τ is within a known range $[a, b]$, $\mathbf{Z}$ is a scalar and expressing the risk factor by 0 and 1 to the early and late onset of data, and $\mathbf{x}$ is a $p \times 1$ vector of covariates that vary from subject to subject. The authors proposed an extension of the score test [94] for the change point for testing $H_0 : \theta = 0$, however, they gave more effort to construct the confidence intervals for parameters in (2.21).

Another semi-parametric approach is studied by Chang et al. [29] and Gijbels and Gürler [50] assuming that the unknown change point τ belongs to a certain known interval $[0, B]$, which is a hybrid martingale based method. [29] combines the score function with the martingale approach, while [50] combines the least squared principle with the martingale approach. Gijbels and Gürler [50] considered the following function

$$Y(x)=\frac{\Lambda(x)}{x}. \tag{2.22}$$

The empirical average hazard rate process $Y_n(x)$ can be found by replacing $\Lambda(x)$ with its Nelson-Aalen estimator (2.11). Again, they have considered a simple structure for the average hazard function and defined as

$$Y^{'}(x)=\beta+\theta\left(1-\frac{\tau}{x}\right)I_{\{\tau<x\}}, \tag{2.23}$$

it remains constant up to time τ and from τ on starts increasing (in case $\theta > 0$) or decreasing (in case $\theta < 0$) as a function of $1/x$. Then the splitting point which gives the best least square fit between $Y_n(x)$ and $Y^{'}(x)$ over a set of prefixed grid points is defined as the estimator of τ.

Dupuy [42] used an exponential regression model with covariates based on right-censoring in testing the existence of a change in the parameters, and proposed likelihood ratio type tests. He also constructed non-asymptotic bounds for the type II

probability. He considered the following *exponential regression model*

$$f_{X|Z}\left(t_i \mid z_i; \mu, \beta\right) = \begin{cases} \mu_1 \exp\left(-\mu_1 t_i \exp(\beta_1 z_i) + \beta_1 z_i\right), & i = 1, \ldots, \tau_1 - 1, \\ \mu_2 \exp\left(-\mu_2 t_i \exp(\beta_2 z_i) + \beta_2 z_i\right), & i = \tau_1, \ldots, n, \end{cases} \tag{2.24}$$

where z_i is a measured covariate and τ_1 is the unknown change point, $\theta = (\mu_1, \beta_1)$ and $\phi = (\mu_2, \beta_2)$ are the parameters before and after the change, respectively. Precisely, he considered testing of existence of a change in both hazard and regression parameters of the model (2.24). Most of the previously cited literature on change point problems in the exponential survival model focuses only on the baseline hazard function, and consider the regression parameter as fixed, i.e., $\theta = (\mu_1, 0)$ and $\phi = (\mu_2, 0)$.

There is extensive literature on single change point hazard problem but the literature on multiple change point problems is rather small. So far our knowledge there are only two published works by Goodman et al. [54, 55] and Qian and Zhang [117] in multiple change points hazard problems. Qian and Zhang [117] proposed an algorithm to fit both susceptibles and long-term survivors with observed covariates through a grid search weighted least squared method assuming that all potential change points lie in a certain known interval $[B_1, B_2]$ for detecting the number of change point in the hazard functions and estimating those. This is a simulation study without any theoretical contents, moreover, assuming known interval for change points makes the application even more impractical.

Goodman et al. [54, 55] proposed a methodology for estimation of multiple change points using the Nelder-Mead Simplex algorithm and a model selection approach using sequential testing with likelihood ratio test and Wald type test statistics in the *piecewise constant hazard* model and *piecewise linear hazard* model. The latter model will easily accommodate the addition of covariates. The estimation of multiple change points using the Nelder-Mead Simplex algorithm is capable of handling an unlimited number of covariates but restricted to an additive nature only. This method also suffers from the number of censored observations. When censoring occurs near the change points valid estimation of these points is not possible.

Goodman et al. [54, 55] considered the random right censored data (X_i, δ_i), which is

described in Section 2.1, in the *piecewise constant hazard* model

$$\lambda(x) = \begin{cases} \alpha_1, & 0 \le x < \tau_1, \\ \alpha_2, & \tau_1 \le x < \tau_2, \\ \vdots & \\ \alpha_{k+1}, & \tau_k \le x, \end{cases} \tag{2.25}$$

where α_i are the hazard values, τ_i are the change points with $0 < \tau_1 < \cdots < \tau_k$. When $\alpha_{k-1} \neq \alpha_k$, the log-likelihood function is

$$\begin{aligned} \log L &\equiv \log L\left(\alpha_1, \ldots, \alpha_k, \tau_1, \ldots, \tau_{k-1}\right) \\ &= D(\tau_1)\log\alpha_1 + [D(\tau_2) - D(\tau_1)]\log\alpha_2 + \cdots + [n_u - D(\tau_{k-1})]\log\alpha_k \\ &- \alpha_1 \sum_{i=1}^{n} (X_i \wedge \tau_1) - \alpha_2 \sum_{i=1}^{n} (X_i \wedge \tau_2 - \tau_1) I\left(X_i > \tau_1\right) - \cdots \\ &- \alpha_k \sum_{i=1}^{n} (X_i - \tau_{k-1}) I\left(X_i > \tau_{k-1}\right), \end{aligned} \tag{2.26}$$

where n_u is the total number of non-censored events, and $D(x) = \sum_{i=1}^{n} I\left(X_i < x\right)\delta_i$ denotes the number of deaths observed up to time x. For fixed τ_j's, the maximum likelihood estimates (MLE's) of the parameters $\alpha_1, \ldots, \alpha_k$ are given by

$$\begin{aligned} \hat{\alpha}_1 &= \frac{D(\tau_1)}{\sum_{i=1}^{n} (X_i \wedge \tau_1)} \\ \hat{\alpha}_2 &= \frac{D(\tau_2) - D(\tau_1)}{\sum_{i=1}^{n} (X_i \wedge \tau_2 - \tau_1)\, I(X_i > \tau_1)}, \ldots, \\ \hat{\alpha}_{k-1} &= \frac{D(\tau_{k-1}) - D(\tau_{k-2})}{\sum_{i=1}^{n} (X_i \wedge \tau_{k-1} - \tau_{k-2})\, I(X_i > \tau_{k-2})} \\ \hat{\alpha}_k &= \frac{n_u - D(\tau_{k-1})}{\sum_{i=1}^{n} (X_i - \tau_{k-1})\, I(X_i > \tau_{k-1})}. \end{aligned} \tag{2.27}$$

Estimates of the parameters in the model (2.25) are calculated by minimizing the negative log-likelihood function (2.26) using the optimization function Nelder-Mead Simplex algorithm evaluated at the maximum likelihood estimates of the α_j (2.27) and finding those values of the τ_j that minimize the function. The maximum likelihood estimates of the τ_j are those values returned by the optimization function that minimize the negative log-likelihood (2.26).

Goodman et al. [54, 55] used a sequential analysis problem in testing for the number of change points, where they performed a hypothesis test and if the null hypothesis is rejected they will continue on to the next hypothesis test. If they failed to reject

the null hypothesis, they stop and conclude that they have found the final model. Although their proposed method is a multiple testing procedure, this is a stepwise process. Therefore, they only need to test one hypothesis at a time.

$$H_0 : \alpha_{k-1} - \alpha_k = 0, \quad \text{against}$$
$$H_1 : \alpha_{k-1} - \alpha_k \neq 0.$$

They used a Wald type test statistic of the form

$$W = \frac{\left(\hat{\alpha}_{k-1} - \hat{\alpha}_k\right)^2}{Var\left(\hat{\alpha}_{k-1} - \hat{\alpha}_k\right)} \sim \chi_1^2. \tag{2.28}$$

To approximate the variance of estimates in the denominator of the above test statistic they used a partitioned Hessian matrix containing only those parameters of $\theta' = (\alpha_1, \ldots, \alpha_k, \tau_1, \ldots, \tau_{k-1})$ which are available in the test statistic (2.28).

The *piecewise linear hazard model* considers $\eta = \log \lambda$, where η is a piecewise linear spline function with knots at $\tau_1, \ldots, \tau_k$, and defined as

$$\eta(x) = \begin{cases} \alpha_0 + \alpha_1 x + \mathbf{Z}'\beta, & 0 \leq x < \tau_1, \\ \alpha_0 + \alpha_1 x + \alpha_2 (x - \tau_1)_+ + \mathbf{Z}'\beta, & \tau_1 \leq x < \tau_2, \\ \vdots & \\ \alpha_0 + \alpha_1 x + \alpha_2 (x - \tau_1)_+ + \cdots + \alpha_{k+1}(x - \tau_k)_+ + \mathbf{Z}'\beta, & \tau_k \leq x, \end{cases} \tag{2.29}$$

for fixed k, where $x_+ \equiv \max(0, x)$, $\mathbf{Z}$ is the covariate vector and β is a vector of the parameters for the effects of the covariates (cf. Cai et al. [27]). The log-likelihood is

$$\log L \equiv \log L\left(\alpha_0, \alpha_1, \ldots, \alpha_{k+1}, \tau_1, \ldots, \tau_k, \beta\right) = \sum_{i=1}^{n} \left\{ \delta_i \eta(X_i) - \int_0^{X_1} e^{\eta(u)} du \right\}. \tag{2.30}$$

The Wald test for testing $H_0 : \alpha_{k+1} = 0$ versus $H_1 : \alpha_{k+1} \neq 0$ to verify the existence of change point τ_k, takes the form

$$W' = \frac{\hat{\alpha}_{k+1}^2}{Var\left(\hat{\alpha}_{k+1}\right)} \sim \chi_1^2. \tag{2.31}$$

Estimation in the *piecewise linear hazard* model (2.29) was conducted by minimizing the negative log-likelihood function (2.30) using the optimization function Nelder-Mead Simplex algorithm and find those values of α_j, τ_j and β_j which optimize the likelihood function. They assumed that the changes only affect parameters of the

baseline hazard function and regression coefficients β_j's are fixed. Properties of the suggested test and estimators of the change points are investigated via only simulations, but no theoretical results are available.

We are interested in testing and estimation of multiple change points in the hazard distributions. Such methodology has twofold benefits: since this approach does not consider any model so we do not need to concern about model misspecification errors; and after getting some significant change points in the hazard functions, this approach allows practitioners to use a homogeneous segment of data to estimate the hazard function rather using the entire data set. There is extensive literature on changes in parameter for hazard functions. However, as far our knowledge, there is no literature for a change point in the hazard distribution with standard change point analysis technique. We will develop theoretical content for detecting and estimating a change point in the hazard distribution using a standard change point analysis technique, i.e., U-statistic process, in Chapter 4. Before doing that we are also interested to simulate the multiple change point in the hazard functions with another optimization procedure, and hence compare that results with the Nelder-Mead Simplex algorithm's results. In the next section, we develop an algorithm to estimate multiple change points in the hazard function using an optimization technique named the Cross-Entropy (CE) method.

2.3.2 Change Point in Hazard using the Cross-Entropy (CE) Method

In this section, we propose to use the Cross-Entropy (CE) method, which is developed by Rubinstein [122] and Rubinstein and Kroese [123], for estimating the multiple change points in the *piecewise constant hazard* model as well as *piecewise linear hazard* model to handle the additive covariates with any number of censored observations. More specifically, this is an extension of Goodman et al. [54, 55] work.

The cross-entropy (CE) method is a new generic approach to combinatorial and multi-extremal optimization and rare event simulation based on a Kullback-Leibler (also called cross-entropy) minimization technique. This method has proven to be very successful in solving wide range of difficult optimization and estimation problems. Good discussions of this method can be found in the CE monograph by Rubinstein

and Kroese [123] and a gentle tutorial by De Boer et al. [40]. The CE method is an iterative optimization method that involves the following two phases:

1. Generation of a set of random samples (trajectories, vectors, etc.) according to a specified random mechanism.

2. Updating the parameters of the random mechanism based on the best samples generated in the previous phase. This phase involves the Kullback-Leibler (or the cross-entropy) minimization.

We are interested in developing an algorithm based on the CE method in estimating the multiple change points in the hazard functions for the *piecewise constant hazard* model as well as the *piecewise linear hazard* model by minimizing the negative log-likelihood functions (2.26) and (2.30), respectively. Before presenting our algorithm, we provide some necessary concepts of the usual CE method from Rubinstein and Kroese [124] and Benham et al. [18]. Let $\mathscr{X}$ be a set of states and S be a real-valued performance function on $\mathscr{X}$. The goal is to find the minimum of S over $\mathscr{X}$, say γ^*, and the state(s), minimizer say $\mathbf{x}^*$, corresponding to this value. Thus

$$S\left(\mathbf{x}^*\right) = \gamma^* = \min_{\mathbf{x}\in\mathscr{X}} S\left(\mathbf{x}\right). \tag{2.32}$$

The first step of the CE method is to turn the optimization problem (2.32) into a meaningful estimation problem of the probability $\ell = \mathbb{P}\left(S(\mathbf{X}) \leq \gamma\right)$, where $\mathbf{X}$ has some probability density $f\left(\mathbf{x};\mathbf{u}\right)$ on $\mathscr{X}$ depending on a parameter $\mathbf{u}$ and a *level* γ. Hence, for optimization problems randomness is purposely introduced in order to make the model stochastic. If γ is chosen close to the unknown γ^*, then ℓ is typically a rare-event probability. One of the most effective ways to estimate rare-event probabilities is to use *importance sampling*. Hence, the importance sampling estimator of $\ell = \mathbb{P}\left(S(\mathbf{X}) \leq \gamma\right)$ is

$$\hat{\ell} = \frac{1}{M}\sum_{i=1}^{M} \frac{f(\mathbf{X_i})}{g(\mathbf{X_i})} I\left\{S(\mathbf{X_i}) \leq \gamma\right\},$$

where $\mathbf{X}_1, \ldots, \mathbf{X}_M$ is an i.i.d. sample from a well-chosen importance sampling density g. The optimal importance sampling density is in this case $g^*(\mathbf{x}) = f(\mathbf{x})I\left\{S(\mathbf{x}) \leq \gamma\right\}/\ell$, which gives a zero-variance estimator, but depends on the unknown quantity ℓ. Benham et al. [18] summarized that the main idea behind the CE method for estimation is to adaptively determine an importance sampling pdf $f\left(\mathbf{x};\mathbf{v}^*\right)$ - hence within the same family as the original distribution - that is close to g^* in the Kullback-Leibler

sense, where the Kullback-Leibler 'distance' (divergence) or the cross-entropy distance between two densities g and h is defined as

$$\mathcal{D}(g,h) = \mathbb{E}_g\left[\log\frac{g(\mathbf{X})}{h(\mathbf{X})}\right] = \int g(\mathbf{x})\log g(\mathbf{x})d\mathbf{x} - \int g(\mathbf{x})\log h(\mathbf{x})\,d\mathbf{x}.$$

Notice that $\mathcal{D}(g,h)$ is not a 'distance' between g and h, since in general $\mathcal{D}(g,h) \neq \mathcal{D}(h,g)$. Nevertheless, it is often useful to think of $\mathcal{D}(g,h)$ as a distance as $\mathcal{D}(g,h) \geq 0$ and $\mathcal{D}(g,h) = 0$ if and only if $g(x) = h(x)$ *a.e.*

In the CE method, specifically, a parameter $\mathbf{v}^*$ is sought that minimizes the cross-entropy distance

$$\mathcal{D}(g^*, f(.;\mathbf{v})) = \mathbb{E}_{g^*}\left[\log\frac{g^*(\mathbf{X})}{f(\mathbf{X};\mathbf{v})}\right] = \int g^*(\mathbf{x})\log g^*(\mathbf{x})d\mathbf{x} - \int g^*(\mathbf{x})\log f(\mathbf{x};\mathbf{v})\,d\mathbf{x},$$

which is equivalent to maximizing, with respect to $\mathbf{v}$

$$\int f(\mathbf{x};\mathbf{u})\,I\{S(\mathbf{x}) \leq \gamma\}\log f(\mathbf{x},\mathbf{v})\,d\mathbf{x} = \mathbb{E}_{\mathbf{u}}\left[I\{S(\mathbf{X}) \leq \gamma\}\log f(\mathbf{X};\mathbf{v})\right],$$

which can be estimated by maximizing the sample average

$$\frac{1}{M}\sum_{i=1}^{M}\left[I\{S(\mathbf{X}_i) \leq \gamma\}\log f(\mathbf{X}_i,\mathbf{v})\right], \tag{2.33}$$

where $\mathbf{X}_1,\ldots,\mathbf{X}_M$ is an i.i.d. sample from $f(\mathbf{x};\mathbf{u})$. Especially, maximizing (2.33) gives the maximum likelihood estimator of $\mathbf{v}$ based on only the samples $\mathbf{X}_1,\ldots,\mathbf{X}_M$ that have a function value less than or equal to γ. These are so-called *elite samples*. The *elite sample* is defined as the proportion of the sample based on the performance function S and using a predefined rarity parameter ρ, which is a real number between 0 and 1. For a random sample $\mathbf{X}_1,\ldots,\mathbf{X}_M$ let $S_{(1)} \leq \cdots \leq S_{(M)}$ be the performances of $\{S(\mathbf{X}_i)\}$ ordered from smallest to largest. Thus, $S_{(j)}$ is the j-th *order statistic* of the sequence $S(\mathbf{X}_1),\ldots,S(\mathbf{X}_M)$. Hence, the *elite sample* is chosen using $\gamma_t = S_{(\lceil\rho M\rceil)}$ for each iteration, where t is the counter in this iterative approach.

The relevance to optimization is that when γ is close to the (usually unknown) minimum γ^*, then the importance sampling density g^* concentrates most of its mass in the vicinity of the minimizer $\mathbf{x}^*$. Sampling from such a distribution thus produces a sequence of levels $(\gamma_t)_{t=1}^T$ and reference parameters $(\mathbf{v}_t)_{t=1}^T$ determined from (2.33) such that the former tends to the optimal γ^* and the latter to the optimal reference vector $\mathbf{v}^*$, where $f(\mathbf{x};\mathbf{v}^*)$ corresponds to the point mass at $\mathbf{x}^*$ (cf. Rubinstein and

Kroese [124], Page 251).

The algorithm for estimating multiple change points in the hazard functions for the *piecewise constant hazard* model and the *piecewise linear hazard* model by respectively minimizing the negative log-likelihood functions (2.26) and (2.30) with censored data is illustrated in Algorithm 1. Here, we assume that we fit the *piecewise constant hazard* model (2.25) or the *piecewise linear hazard* model (2.29) to the data resulting in a density $f(\mathbf{x};\mathbf{v})$ of $\mathbf{X}$. However, we allow for misspecification, i.e., the true density does not have to be of the form $f(.;\mathbf{v})$.

Algorithm 1 CE Algorithm for estimating the change point in the hazard functions

1: Initialize parameter vectors $\tilde{\mathbf{v}}_0 = (\boldsymbol{\alpha}_0, \boldsymbol{\tau}_0, \boldsymbol{\beta}_0)$. Set sample size M, rarity parameter ρ, smoothing parameter c and $t = 1$.

2: Generate the observed survival times $\mathbf{X}_1, \ldots, \mathbf{X}_M$, where $\mathbf{X}_m = (X_{m1}, \ldots, X_{mn})$ for $m = 1, \ldots, M$, from the the density $f(.;\mathbf{v}_{t-1})$, which can be done by generating the survival times $\mathbf{T}_1, \ldots, \mathbf{T}_M$ from the density (2.35) and (2.41), and the censoring times $\mathbf{C}_1, \ldots, \mathbf{C}_M$ from (2.38) and Exp(b) for some b, respectively, for the models (2.25) and (2.29). Hence, we get $\mathbf{X}_{m,i} = \min(\mathbf{T}_{m,i}, \mathbf{C}_{m,i})$ and $\boldsymbol{\delta}_{m,i} = I_{\{\mathbf{T}_{m,i} \leq \mathbf{C}_{m,i}\}}$ for $m = 1, \ldots, M$ and $i = 1, \ldots, n$.

3: Calculate the negative log-likelihood functions (2.26) and (2.30) respectively for the models (2.25) and (2.29) as $S(\mathbf{X}_i) = -\log L$ for all i, and order them from the smallest to the largest, i.e., $S_{(1)} \leq \cdots \leq S_{(M)}$. Let γ_t be the sample ρ-quantile of performances, i.e., the number of elite samples $\gamma_t = S_{(\lceil \rho M \rceil)}$.

4: Use the same sample $\mathbf{X}_1, \ldots, \mathbf{X}_M$ and solve the stochastic program

$$\hat{\mathbf{v}}_t = \max_{\mathbf{v}} \sum_{k=1}^{M} I\{S(\mathbf{X}_k) \leq \gamma_t\} \log f(\mathbf{X}_k, \mathbf{v}).$$

5: Smooth, $\tilde{\mathbf{v}}_t = c\hat{\mathbf{v}}_t + (1-c)\,\tilde{\mathbf{v}}_{t-1}$, where c is a smoothing parameter.

6: If for some $t \geq k$, say $k = 5$, $\gamma_t = \gamma_{t-1} = \ldots = \gamma_{t-k}$ then stop. Otherwise set $t = t + 1$ and go to Step 2.

For the *piecewise constant hazard* model, the initialize parameter vectors $\tilde{\mathbf{v}}_0 = (\alpha_0, \tau_0)$. To run the algorithm, we need to provide the class of sampling densities $\{f(.;\mathbf{v})\}$, the initial vector $\tilde{\mathbf{v}}_0$, the sample size M, rarity parameter ρ, smoothing parameter c and the stopping criterion.

Although we are using the CE method as an optimization technique to estimate

multiple change points in the hazard functions and compare the accuracy of the estimates along with the Nelder-Mead Simplex algorithm, it can be used as a detecting method for identifying multiple change points in the hazard functions, for related works confer Evans et al. [44], Priyadarshana et al. [116] and Polushina and Sofronov [113].

2.4 Simulations and Applications

2.4.1 Simulations

Monte Carlo simulations were carried out to compare the accuracy of estimates of multiple change points in the hazard functions using the CE method along with the Nelder-Mead Simplex algorithm. The classic Nelder-Mead method, developed by Nelder and Mead [102], evaluates the function at the vertices of a simplex and then iteratively shrinks the simplex as better points are found until some desired bound is obtained. This algorithm requires the user to provide initial starting values for the parameter estimates. Moreover, this method is formulated for unconstrained optimization problems only. We performed 5,000 replications to estimate multiple change points in the hazard functions in the *piecewise constant model* (2.25) as well as the *piecewise linear model* (2.29) and present the results in Tables 2.1 and 2.2, respectively. In each model, we have estimated the average estimated parameter value and the standard error (which is the standard deviation of these estimates) from all 5,000 simulation iterations for all parameters.

Piecewise Constant Multiple Change Point Hazard Model

A simulation in the *piecewise constant model* for two change points (τ_1, τ_2) hazard functions was conducted for $n = 500$ with various values of the parameters $(\alpha_1, \alpha_2, \alpha_3, \tau_1, \tau_2)$ and different percentages of censoring in Table 2.1. For two change points in the hazard functions the model (2.25) can be defined as

$$\lambda(t) = \begin{cases} \alpha_1, & 0 \leq t < \tau_1, \\ \alpha_2, & \tau_1 \leq t < \tau_2, \\ \alpha_3, & \tau_2 \leq t, \end{cases} \tag{2.34}$$

having the probability density function (pdf)

$$f(t) = \begin{cases} \alpha_1 \exp\left(-\alpha_1 t\right), & 0 \leq t < \tau_1, \\ \alpha_2 \exp\left(-\alpha_1 \tau_1 - \alpha_2 (t - \tau_1)\right), & \tau_1 \leq t < \tau_2, \\ \alpha_3 \exp\left(-\alpha_1 \tau_1 - \alpha_2 (\tau_2 - \tau_1) - \alpha_3 (t - \tau_2)\right), & \tau_2 \leq t. \end{cases} \tag{2.35}$$

Survival times T_i from the model (2.34) can be simulated either by the inversion of the cumulative hazard, which is called the inverse transformation method (or so-called inverse cumulative distribution function (CDF)) method or by the composition method. Here, we used the inverse CDF method. Since

$$\begin{aligned} F(t) &= 1 - \exp\left(-\Lambda(t)\right) \\ U &= 1 - \exp\left(-\Lambda(T)\right), \end{aligned}$$

where $U \sim \mathcal{U}(0,1)$ and hence,

$$T = \Lambda^{-1}\left(-\log(1-U)\right).$$

As $-\log(1-U) \sim \text{Exp}(1)$, we can apply the inverse cumulative hazard to an exponential random variable. The cumulative hazard of the model (2.34) becomes

$$\Lambda(t) = \begin{cases} \alpha_1 t, & 0 \le t < \tau_1, \\ \alpha_1\tau_1 + \alpha_2(t-\tau_1), & \tau_1 \le t < \tau_2, \\ \alpha_1\tau_1 + \alpha_2(\tau_2-\tau_1) + \alpha_3(t-\tau_2), & \tau_2 \le t. \end{cases} \tag{2.36}$$

Therefore, the inverse of the cumulative hazard takes the form

$$\Lambda^{-1}(u) = \begin{cases} \frac{u}{\alpha_1}, & 0 \le u < \alpha_1\tau_1, \\ \tau_1 + \frac{(u-\alpha_1\tau_1)}{\alpha_2}, & \alpha_1\tau_1 \le u < \alpha_1\tau_1 + \alpha_2(\tau_2-\tau_1), \\ \tau_2 + \frac{(u-\alpha_1\tau_1-\alpha_2(\tau_2-\tau_1))}{\alpha_3}, & u > \alpha_1\tau_1 + \alpha_2(\tau_2-\tau_1). \end{cases} \tag{2.37}$$

Now, generate an exponential random variable with rate 1, and plug it into Λ^{-1} in (2.37). Censoring times C_i are generated by using the uniform distribution on $[0,\tau_1]$, $[\tau_1,\tau_2]$, and $[\tau_2, \max(T_i)]$, respectively for the corresponding ranges of survival times in (2.37). Thus

$$C = \begin{cases} U_1, & 0 \le Y < \alpha_1\tau_1, \\ U_2, & \alpha_1\tau_1 \le Y < \alpha_1\tau_1 + \alpha_2(\tau_2-\tau_1), \\ U_3, & Y > \alpha_1\tau_1 + \alpha_2(\tau_2-\tau_1), \end{cases} \tag{2.38}$$

where $U_1 \sim \mathcal{U}(0,\tau_1)$, $U_2 \sim \mathcal{U}(\tau_1,\tau_2)$, $U_3 \sim \mathcal{U}(\tau_2, \max(T_i))$ and $Y \sim \text{Exp}(1)$. Generated censoring times' censoring proportion can be specified by rearranging the interval of the generated uniform variate, e.g., generate U_1, U_2, and U_3 on $[0, \tau_1+5]$, $[\tau_1, \tau_2+5]$, and $[\tau_2, \max(T_i)]$ respectively in (2.38) to generate 15% censored observations. Then, we calculated observed data $X_i = \min(T_i, C_i)$ and also identified δ_i.

To run our proposed Algorithm 1, at first we need to initialize input parameter's

value. Hence, we used the initial vector $\tilde{\mathbf{v}}_0$ as $\tau_0 = (1.0, 3.0)$, the rarity parameter $\rho = 0.20$, the smoothing parameter $c = 0.98$, and the sample size $M = 100$ for conducting simulation with the CE method. We performed *CEoptim* function from the *CEoptim* package in R to estimate the change points in the hazard function based on our proposed Algorithm 1, where we specified the arguments *mean=c(1, 3)*, *sd=c(10,10)*, and a linear constrains on the change points using the *conMat* and *conVec* arguments as *conMat = rbind(diag(2),-diag(2))* and *conVec = c(3,5,0,-3)*. To implement the Nelder-Mead Simplex algorithm in this context, we used *optim* function in R. Although the classic Nelder-Mead method is formulated for unconstrained optimization problems only, we applied the initial value of the change points $\tau_1 = 1.0$ and $\tau_2 = 3.0$ with the bound $[0.01, 3.0]$ and $[3.0, 5.0]$, respectively. After estimating the change points τ_1 and τ_2 using the CE method or the Nelder-Mead Simplex algorithm, we estimated the parameters α_1, α_2 and α_3 using (2.27).

Table 2.1: Piecewise Constant Model with Two Change Points based on 5,000 Simulations ($n = 500$)

Censor	**Parameters**	**Parameter value**	**Nelder-Mean Simplex**		**Cross Entropy**	
			Estimated Value	**Standard Error**	**Estimated Value**	**Standard Error**
0%	α_1	**0.10**	0.1239	0.0295	**0.0993**	0.0106
	α_2	**0.30**	0.3614	0.0963	**0.3010**	0.0227
	α_3	**0.75**	0.7443	0.0897	**0.7585**	0.0515
	τ_1	**2.00**	2.4140	0.5191	**2.0075**	0.0519
	τ_2	**4.00**	4.3268	0.6241	**4.0066**	0.0536
15%	α_1	**0.15**	**0.1608**	0.0284	0.1321	0.0142
	α_2	**0.25**	0.2856	0.0480	**0.2214**	0.0182
	α_3	**0.80**	0.7825	0.1091	**0.8087**	0.0625
	τ_1	**1.50**	2.1703	0.5996	**1.5112**	0.1284
	τ_2	**4.00**	4.3285	0.3972	**4.0106**	0.0270
50%	α_1	**0.15**	**0.1015**	0.0266	0.1001	0.0006
	α_2	**0.45**	0.2551	0.0388	**0.3003**	0.0017
	α_3	**0.95**	0.9572	0.1241	**0.9483**	0.0586
	τ_1	**2.00**	2.3745	0.3743	**2.0095**	0.0138
	τ_2	**4.00**	4.4300	0.1862	**4.0211**	0.0331

Table 2.1 reveals the mean estimated values of parameters $(\alpha_1, \alpha_2, \alpha_3, \tau_1, \tau_2)$ with their standard errors (SE) using the CE method along with the Nelder-Mean Simplex algorithm for estimating multiple change points in the hazard functions in model (2.25)

for $n = 500$ and censoring percentages 0%, 15% and 50%. When there is no censoring (0%) we observed that parameter estimation using the CE optimization method is more accurate than that of the Nelder-Mean Simplex algorithm. For instance, simulated estimates of baseline hazard $\alpha_1 = 0.10$ is 0.0993 in the CE method with SE 0.0106 and is 0.1239 using the Nelder-Mean Simplex algorithm with SE 0.0295. In case of moderate (15%) and higher (50%) censoring percentages the CE method performed better than the Nelder-Mean Simplex algorithm, specially for the change points (τ_1, τ_2). For example, for the change point $\tau_2 = 4.00$ we found the closer one $\hat{\tau}_2 = 4.0211$ using the CE method whereas $\hat{\tau}_2 = 4.4300$ is observed in the Nelder-Mean Simplex algorithm for the case of (50%) censoring proportion. The standard error is always minimum in all cases for the CE method, which implies a more consistent estimator.

Piecewise Linear Multiple Change Point Hazard Model

Simulation in the *piecewise linear model* was also conducted for 5,000 iterations in a similar manner to that of the *piecewise constant model*. We generated survival times T_i using the inverse CDF method for the *piecewise linear model* (2.29) with two change points (τ_1, τ_2) and two covariates, one dichotomous and one continuous. Using $\lambda(t) = \exp(\eta(t))$ the hazard function of (2.29) is defined as

$$\lambda(t) = \begin{cases} \exp\left(\alpha_0 + \alpha_1 t + \mathbf{Z}'\beta\right), & 0 \le t < \tau_1, \\ \exp\left(\alpha_0 + \alpha_1 t + \alpha_2(t-\tau_1) + \mathbf{Z}'\beta\right), & \tau_1 \le t < \tau_2, \\ \exp\left(\alpha_0 + \alpha_1 t + \alpha_2(t-\tau_1) + \alpha_3(t-\tau_2) + \mathbf{Z}'\beta\right), & \tau_2 \le t. \end{cases} \tag{2.39}$$

Hence, the cumulative hazard function is

$$\Lambda(t) = \begin{cases} \frac{1}{\alpha_1}\left[\exp\left(\alpha_0 + \alpha_1 t + \mathbf{Z}'\beta\right) - \exp\left(\alpha_0 + \mathbf{Z}'\beta\right)\right], & 0 \le t < \tau_1, \\ A + \frac{1}{a_1}\left[\exp\left(\alpha_0 - \alpha_2\tau_1 + a_1 t + \mathbf{Z}'\beta\right)\right] & \\ -\frac{1}{a_1}\left[\exp\left(\alpha_0 + \alpha_1\tau_1 + \mathbf{Z}'\beta\right)\right], & \tau_1 \le t < \tau_2, \\ A + B + \frac{1}{a_2}\left[\exp\left(\alpha_0 - \alpha_2\tau_1 - \alpha_3\tau_2 + a_2 t + \mathbf{Z}'\beta\right)\right] & \\ -\frac{1}{a_2}\left[\exp\left(\alpha_0 - \alpha_2\tau_1 + a_1\tau_2 + \mathbf{Z}'\beta\right)\right], & \tau_2 \le t. \end{cases} \tag{2.40}$$

where

$$\begin{aligned} a_1 &= \alpha_1 + \alpha_2, \\ a_2 &= \alpha_1 + \alpha_2 + \alpha_3, \\ A &= \frac{1}{\alpha_1}\left[\exp\left(\alpha_0 + \alpha_1\tau_1 + \mathbf{Z}'\beta\right) - \exp\left(\alpha_0 + \mathbf{Z}'\beta\right)\right], \\ B &= \frac{1}{\alpha_1 + \alpha_2}\left[\exp\left(\alpha_0 - \alpha_2\tau_1 + \alpha_1\tau_2 + \alpha_2\tau_2 + \mathbf{Z}'\beta\right) - \exp\left(\alpha_0 + \alpha_1\tau_1 + \mathbf{Z}'\beta\right)\right]. \end{aligned}$$

Therefore, the inverse of the cumulative hazard function becomes

$$\Lambda^{-1}(u) = \begin{cases} \frac{1}{\alpha_1}\left[\log\left(\alpha_1 u + \exp\left(\alpha_0 + \mathbf{Z}'\beta\right)\right) - \alpha_0 - \mathbf{Z}'\beta\right], & 0 \le u < A, \\ \frac{1}{a_1}\left[\log\left(a_1(u-A) + \exp\left(\alpha_0 + \alpha_1 t + \mathbf{Z}'\beta\right)\right)\right] & \\ +\frac{1}{a_1}\left[-\alpha_0 + \alpha_2\tau_1 - \mathbf{Z}'\beta\right], & A \le u < A+B, \\ \frac{1}{a_2}\log\left[a_2(u-A-B) + \exp\left(\alpha_0 - \alpha_2\tau_1 + a_1\tau_2 + \mathbf{Z}'\beta\right)\right] & \\ +\frac{1}{a_2}\left[-\alpha_0 + \alpha_2\tau_1 + \alpha_3\tau_2 - \mathbf{Z}'\beta\right], & A+B \le u. \end{cases} \tag{2.41}$$

Hence, survival times T_i are generated for the *piecewise linear model* (2.29) using an exponential random variable with rate 1 and then plug it into Λ^{-1} in (2.41). Censoring times C_i are generated by using the exponential distribution with rate b. Generated censoring times' censoring proportion can be specified by redefining the rate of the generated exponential variate, e.g., using $C_i \sim \text{Exp}(0.45)$ we generated 15% censored observations for the *piecewise linear model* (2.29). Then, we calculated observed data $X_i = \min(T_i, C_i)$ and also identified corresponding non-censoring indicator δ_i. We generated the dichotomous covariate by using the binomial random variable with parameters $n = 1$ and $p = 0.7$. The continuous covariate was generated by using the uniform distribution on $[0, 1]$.

To run our proposed Algorithm 1 for the *piecewise linear model* (2.29), at first we need to initialize input parameter's value. Hence, we initialized the input vector $\tilde{\mathbf{v}}_0$ with $\alpha_0 = (0.5, 0.5, 0.5, 0.5)$, $\tau_0 = (0.5, 1.0)$, and $\beta_0 = (0.5, 0.5)$, and used the rarity parameter $\rho = 0.20$, the smoothing parameter $c = 0.98$ and the sample size $M = 100$ for conducting simulation with the CE method. We performed *CEoptim* function from the *CEoptim* package in R to estimate the change points in the hazard function based on our proposed Algorithm 1, where we specified the arguments *mean=c(0.5,1,0.5,0.5,0.5,0.5,0.5,0.5)*, *sd=c(.5,1,.5,.5,.5,.5,.5,.5)*, and a linear constrains on the change points using the *conMat* and *conVec* arguments as *conMat = rbind(diag(8),-diag(8))* and *conVec = c(.8,1.8,.2,.3,.5,.8,1,1,0,0,0,0,0,0,0,0)*. To implement the Nelder-Mead Simplex algorithm in this context, we used *optim* function in R, where we considered the starting value of the parameters in the *par* argument as $\alpha_0 = (0.5, 0.5, 0.5, 0.5)$, $\tau_0 = (0.5, 1.0)$, and $\beta_0 = (0.5, 0.5)$. Although the classic Nelder-Mead method is formulated for unconstrained optimization problems only, we applied the bound arguments with *lower=c(0.01,0.5,0,0.05,0.1,0.2,0,0)* and *upper=c(0.8,1.8,.2,.3,.5,.8,1,1)*. All the parameters were directly estimated by these two algorithms and are presented in Table 2.2.

Table 2.2: Piecewise Linear Model with Two Change Points based on 5,000 Simulations ($n = 1000$)

Censor	Parameters	Parameter value	Nelder Mean Simplex		Cross Entropy	
			Estimated Value	**Standard Error**	**Estimated Value**	**Standard Error**
0%	α_0	**0.05**	0.0572	0.0633	**0.0524**	0.0478
	α_1	**0.08**	0.0624	0.0301	**0.0735**	0.0465
	α_2	**0.10**	0.1253	0.0480	**0.1247**	0.0737
	α_3	**0.50**	0.4718	0.2358	**0.4976**	0.2095
	τ_1	**0.50**	0.5190	0.2299	**0.5057**	0.2403
	τ_2	**1.50**	1.2655	0.3216	**1.3999**	0.2857
	β_1	**0.50**	**0.4971**	0.0660	0.4921	0.0628
	β_2	**0.75**	**0.7417**	0.0998	0.7260	0.0893
15%	α_0	**0.10**	**0.0987**	0.0715	0.1042	0.0543
	α_1	**0.15**	0.1033	0.0750	**0.1144**	0.0666
	α_2	**0.20**	**0.1917**	0.1140	0.2180	0.1144
	α_3	**0.50**	0.4515	0.2496	**0.5077**	0.2649
	τ_1	**0.50**	0.4171	0.2038	**0.4405**	0.2381
	τ_2	**1.50**	1.0648	0.3099	**1.3054**	0.3542
	β_1	**0.50**	0.5029	0.0726	**0.5006**	0.0688
	β_2	**0.75**	0.7557	0.1061	**0.7470**	0.0963

Table 2.2 illustrates the mean estimated values of hazard parameters $(\alpha_0, \alpha_1, \alpha_2, \alpha_3)$, change points (τ_1, τ_2), and regression parameters (β_1, β_2) with their standard errors (SE) using the CE method as well as the Nelder-Mean Simplex algorithm for estimating multiple change points in the hazard functions in the piecewise linear model (2.29) for $n = 1000$ and censoring percentages 0%, and 15%. In case of no censoring (0%) we observed that hazard and change points estimates are more accurate using the CE optimization method than the Nelder-Mean Simplex algorithm, but regression parameters are more accurate in the Nelder-Mean Simplex algorithm. Nevertheless, only estimates of baseline hazard α_0 and 2nd segment hazard α_2 functions are giving closer result for the Nelder-Mean Simplex algorithm when the data contains 15% censoring percentages. Moreover, change points estimates are more precise with the CE optimization method than that of the Nelder-Mean Simplex algorithm, for instance, when $\tau_1 = 0.50$ and $\tau_2 = 1.50$ we observed $\hat{\tau}_1 = 0.5057$ and $\hat{\tau}_2 = 1.3999$ in the CE method, and $\hat{\tau}_1 = 0.5190$ and $\hat{\tau}_2 = 1.2655$ in the Nelder-Mean Simplex algorithm for the case of (15%) censoring proportion.

2.4.2 Real Data Applications

We apply the CE method and Goodman et al. [54, 55]'s methodology to two data examples: Cell Stimulus Response data and Breast Cancer mortality data. The Cell Stimulus Response data described in Section 3.4 of Chapter 3 and the Breast Cancer mortality data will be explained in the following paragraphs and in Appendix A. To avoid ties, we restrict the change points to be larger than the first survival time and smaller than the second to last survival time, assuming these are non-censored time points, i.e., $T_{(1)} < \tau_1 < \ldots < \tau_k < T_{(n-1)}$, (cf. Yao [150] and Müller and Wang [97]).

Application 1: Breast Cancer Mortality Data
We use the Surveillance, Epidemiology, and End Results (SEER) Program Public-Use Data (1973-2010), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2013, based on the November 2012 submission. This data contains cancer incidence and survival for cases diagnosed from 1973 to 2010, follow-up continued until December 31, 2012. Our focus is on the breast cancer mortality data only. Calculation of hazard rate will be based on the breast cancer patient's risk of death.

For the purpose of analysis we excluded patients with unknown follow-up time and not a case of first tumor. We define an event as death from breast cancer. If a patient dies from another cause they are censored at the time of their death. We found that there are 576250 observations among them 455196 are censored, hence, 78.9% of the observations being censored. Table A.1 in Appendix A shows all the necessary variables information in a summarized structure for the Breast Cancer data to complete our analysis.

It appears from Figure 2.3 that estimated hazard functions using the Product-Limit estimator and the Nelson-Aalen estimator are nearly coinciding for the breast cancer SEER data, whereas a smooth exponential decreasing curve is observed due to the kernel hazard estimator. To calculate the hazard rates we convert monthly survival times ($SurvM$) into yearly survival times and use $DeathCause$ variable as the censoring indicator variable.

Figure 2.3 represents the estimated hazard functions for the entire Breast Cancer Patients of SEER data. Now, we consider only those patients who are diagnosed with the first tumor. Therefore, the second tumor cases or any other cases are discarded

from the data set. For conducting the test to identify change point in hazard functions we need to use the Wald statistic (2.31), accordingly, $\hat{Var}(\hat{\alpha}_2)$, $\hat{Var}(\hat{\alpha}_3), \ldots, \hat{Var}(\hat{\alpha}_k)$ are required to estimate. This is actually the biggest pitfall of this testing procedure. Nevertheless, we can approximate these variances using the R software package *numDeriv* with the function *hessian*. Hence, $\hat{Var}(\hat{\alpha}_2)$, $\hat{Var}(\hat{\alpha}_3)$, and $\hat{Var}(\hat{\alpha}_4)$ are estimated as $2.088586e^{-04}$, $3.460357e^{-05}$, $3.315643e^{-11}$, respectively.

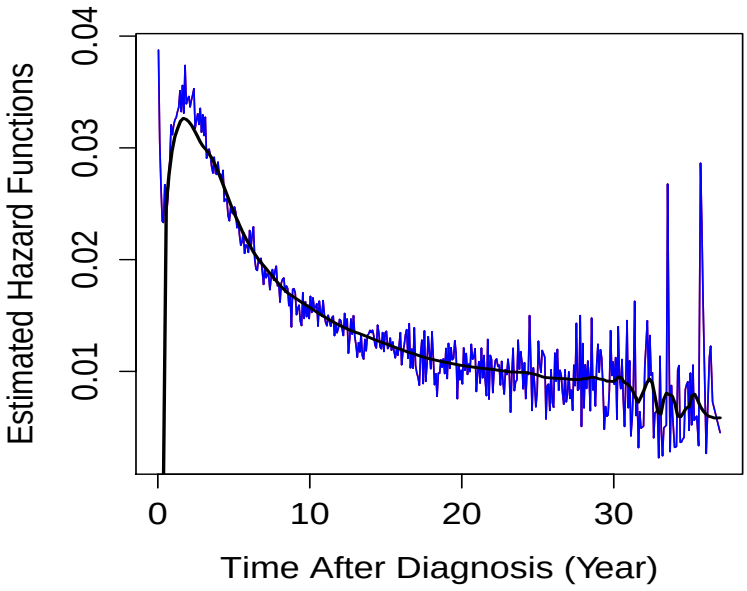


Figure 2.3: The estimated hazard functions of the Breast Cancer Patients using SEER Data estimated by Nelson-Aalen estimator (blue line), Product-Limit estimator (red line, overlapped by the blue line), and kernel hazard estimator (black line) from 1973 to 2010.

Table 2.3: Change points in the hazard functions detection using the Wald statistic (2.31) for the model (2.29) of the Breast Cancer Patients (SEER Data) diagnosed with the first tumor.

Hypotheses	**Wald Statistic**	**Critical Value**	**Decision**
$H_0 : \alpha_2 = 0$ vs. $H_1 : \alpha_2 \neq 0$	253.2810	3.8415	Significant τ_1
$H_0 : \alpha_3 = 0$ vs. $H_1 : \alpha_3 \neq 0$	184.9520	3.8415	Significant τ_2
$H_0 : \alpha_4 = 0$ vs. $H_1 : \alpha_4 \neq 0$	0.7540	3.8415	Insignificant τ_3

Table 2.3 recognizes two change points in the model (2.29) for the Breast Cancer Patients (SEER Data) diagnosed with the first tumor. Figure 2.4 displays those estimated change points in the hazard functions, where the black line is an estimate of the hazard function using the product-limit method of *muhaz* package in R. The red line is the estimated hazard function using the Nelder-Mead algorithm used by [54, 55] and the blue line is the estimated hazard using CE method. The hazard function diagnosed with breast cancer between 1973 and 1976 increases until the first change point, and then begins to decrease sharply until the second change point, followed by a exponential decline until the end of the follow-up period.

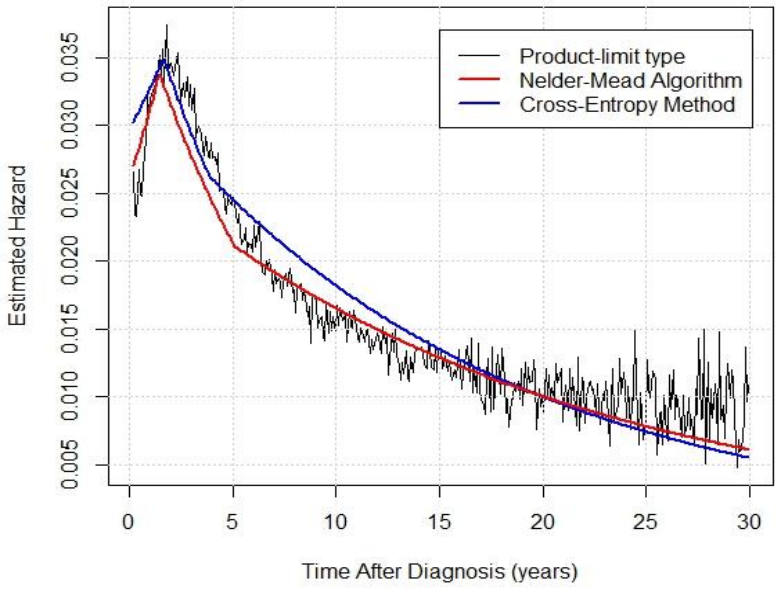


Figure 2.4: The estimated change points in the hazard functions for the model (2.29) of the Breast Cancer Patients diagnosed with the first tumor using SEER Data 1973-2010.

The estimated log hazard function using CE method has two change points (namely, at 1.67 and 3.87), which is shown in equation (2.43). Using Nelder-Mead algorithm we also found two change points in log hazard function at 1.47 and at 5.12, which is defined in (2.42). In a previous study of recovery from breast cancer, it has been observed by Langlands et al. [81] that the maximum mortality occurs after about

three years of diagnosis and then it decreases slowly over a fixed period of time, which also reflected on the findings of the CE method, i.e., significant change points of the hazard function are found at 1.67 and 3.87 years after diagnosis.

$$\eta(t)_{NM} = -3.65 + 0.1t - 0.31(t - 1.47)_+ + 0.08(t - 5.12)_+, \quad t \geq 0, \tag{2.42}$$

$$\eta(t)_{CE} = -3.52 + 0.1t - 0.23(t - 1.67)_+ + 0.08(t - 3.87)_+, \quad t \geq 0. \tag{2.43}$$

Application 2: Cell Stimulus Response Data

Cell Stimulus Response data explained in the Chapter 3 contains 0.5% censored data. We have calculated the hazard functions for this data using three different methods and explored in Figure 2.5, where red line indicates hazard estimated by Product-Limit estimator, blue line indicates the Nelson-Aalen estimator and black line implies kernel hazard estimator.

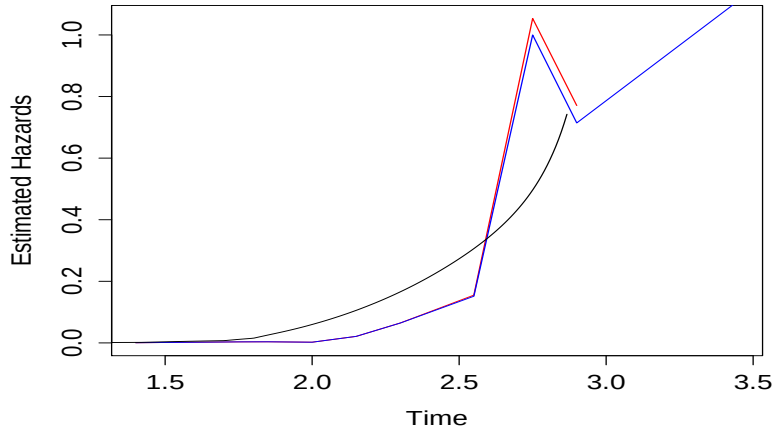


Figure 2.5: The estimated hazard functions for Cell Stimulus Response Data.

Using the *hessian* function in the R software package *numDeriv* we can approximate $\hat{Var}(\hat{\alpha}_1)$, $\hat{Var}(\hat{\alpha}_2)$, $\hat{Var}(\hat{\alpha}_3)$, and $\hat{Var}(\hat{\alpha}_4)$ by $9.517272e^{-08}$, $1.334135e^{-05}$, $8.157135e^{-02}$, and $1.603287e^{-01}$, respectively. Therefore, to find out the significant change points in the model (2.25) we calculate the Wald statistic (2.28) and summarize the hypothesis tests in Table 2.4.

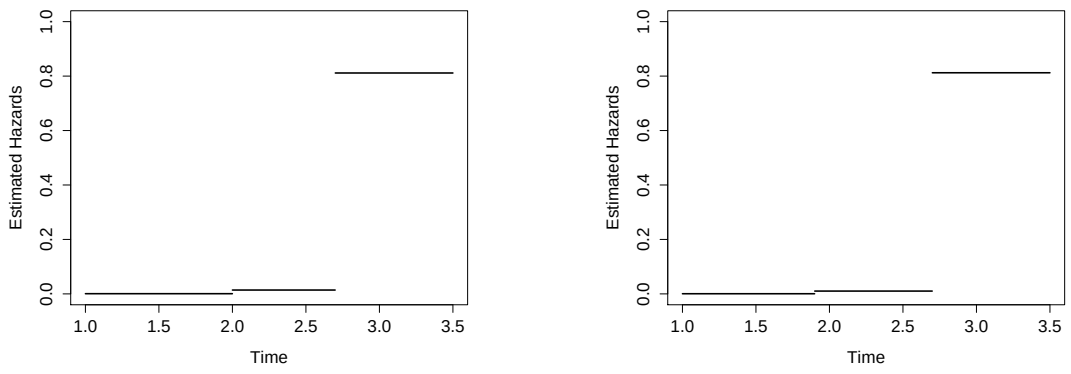
Table 2.4: Change points in the hazard functions testing using the Wald statistic (2.28) for the model (2.25) of Cell Stimulus Responses under $H_0 : \alpha_{(k-1)} - \alpha_k = 0$ vs. $H_1 : \alpha_{(k-1)} - \alpha_k \neq 0$.

Hypotheses	**Wald Statistic**	**Critical Value**	**Decision**
$H_0 : \alpha_1 - \alpha_2 = 0$	7.0963	3.8415	Significant τ_1
$H_0 : \alpha_2 - \alpha_3 = 0$	7.8823	3.8415	Significant τ_2
$H_0 : \alpha_3 - \alpha_4 = 0$	0.0003	3.8415	Insignificant τ_3

Table 2.4 reveals inevitable two change points in the hazard functions in the model(2.25) for the Cell Stimulus Responses. These two change points have observed at 1.999 and 2.698 using the CE method and defined in (2.44), and using the Nelder-Mead Simplex algorithm in (2.45). Estimated multiple change points in the hazard functions are plotted in Figure 2.6.

$$\lambda(t) = \begin{cases} 0.00069, & 0 \leq t < 1.999, \\ 0.01408, & 1.999 \leq t < 2.698, \\ 0.81140, & 2.698 \leq t, \end{cases} \tag{2.44}$$

$$\lambda(t) = \begin{cases} 0.00052, & 0 \leq t < 1.899, \\ 0.01029, & 1.899 \leq t < 2.699, \\ 0.81221, & 2.699 \leq t. \end{cases} \tag{2.45}$$



(a) (b)

Figure 2.6: Estimated change points in the hazard functions for the Cell Stimulus Response using the model (2.25) with the CE method in (a), and Nelder-Mead Simplex Algorithm in (b).

2.5 Discussion

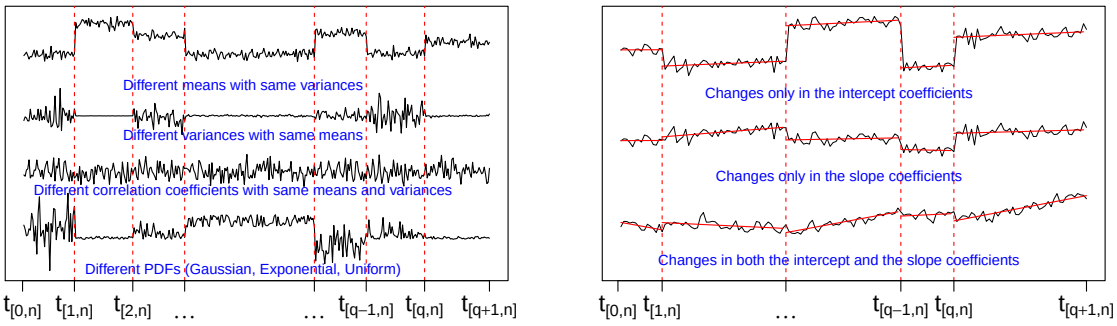
As there are some public health examples that suggest that, due to improvement in treatments or diagnosis, there may be two or more changes in the hazard rate and censoring may also be occurred near these change points. In this context, we have reviewed different hazard estimators in Section 2.2 and different existing methodologies for testing and estimating change points in the hazard functions in Section 2.3. Moreover, we developed an algorithm to estimate multiple change points for the hazard functions in the *piecewise constant model* (2.25) as well as the *piecewise linear model* (2.29) using the Cross-Entropy method and methodology from [54, 55], and its performances illustrated by simulations and real data application. In most of the cases, the CE method performed better than the existing counterpart Nelder-Mead Simplex algorithm with minimum SE. Although we have used the CE method as an optimization technique in estimating multiple change points in the hazard functions and found this method performed more reasonably than its existing counterpart, it can be used as a detecting method for identifying multiple change points in the hazard functions in further research.

Chapter 3: Change Point Analysis with Missing Data

A change point causes heterogeneity or abrupt changes in data. This heterogeneity of data should be properly checked before estimation of parameters of statistical models and decision making procedures. Therefore, statistical models with change points are highly relevant in both theory and practice of many fields. Change points analysis has its origins in quality control Page [106, 107] but has since become an integral part of a wide variety of fields, among them economics [110], finance [4], climatology [119], engineering [135], signal processing [15], time series of counts [47], biological sequences [116] and epidemiology [133]. There is extensive literature on change point problems for references Brodsky and Darkhovsky [23], Basseville and Nikiforov [15], Carlstein et al. [28], Csörgö and Horváth [37], and Chen and Gupta [30, 31]. The change point problem can be considered as one of the central problems of statistical inference, linking together statistical control theory, the theory of estimation and testing hypotheses, classical and Bayesian approaches, and fixed sample and sequential procedures [23].

Application of change points increases as more data sets are collected. Based on the method of data acquisition, there are two main types of change point problems: posteriori (retrospective, offline) and sequential (quickest, online) change point problems. The posteriori problems arise when all observations have been already received, whereas, the sequential problems include situations when the data have been observed sequentially and the future observations are unknown. Change point analysis can also be categorized by the contrast of parametric versus nonparametric and frequentist versus Bayesian detection. In all the situations, it is of interest to know where, when and how many changes arise in the model. There could be a single or many change points in the data set. The change point methods greatly depend on the data nature; whether independent or not, missing or complete, censored or fully observed, etc. In this chapter, we focus our efforts on review different offline nonparametric change

point tests for independent observations with missing data features.



(a) (b)

Figure 3.1: Change point problems description in different cases. Multiple change points in means, variances, correlation strengths and probability distributions are displayed in (a) and changes in linear regression model exhibits in (b).

Figure 3.1 represents various change point problems in multiple change points cases. In the 1st plane of Figure 3.1(a), only mean values of the different segments are changed but variances are unchanged, whereas, variances are changed with the same means in the 2nd plane. In the 3rd plane only the correlation strengths are changed and the 4th plane expresses when the whole shape of their probability distribution functions (PDFs) have been changed. Multiple change points in linear regression model are exhibited in Figure 3.1(b): different intercept coefficients with the same slope coefficients are presented in the 1st plane, different slope coefficients with the same intercept coefficients are in the 2nd plane, and the 3rd plane shows changes in both the intercept and slope coefficients. The multiple change points, $t_{[1,n]}, t_{[2,n]}, \ldots t_{[q,n]}$, are indicated by dotted vertical red lines.

The chapter is structured as follows. Section 3.1 contains different offline nonparametric tests procedure for detecting and estimating single change point problem in the location and regression models. In Section 3.2, we explain multiple change point models in mean and regression structures with illustration of hypothesis testing and estimation of those multiple change points. A review of different missing data mechanisms with various imputation methods is presented in Section 3.3. For real data

analysis we consider the cell stimulus response data. Section 3.4 illustrates the application of change point analysis to this data set with imputed missing values for multiple structural changes in mean and linear regression models, and for multiple change points in the distribution. Finally, the chapter concludes with a discussion in Section 3.5.

3.1 Single Change Point Models

The problem of abrupt changes in general arises in quite a variety of models within time-ordered observations, for instance, changes in mean with known or unknown starting value and variance, changes in variance, changes in mean and variance, changes in location and/or scale, changes in correlation coefficient, changes in regression coefficient. This section deals with At-Most-One-Change (AMOC) mean and regression models with independent observations in offline inference with missing data feature.

3.1.1 Single Change Point in Mean

Model and Assumptions

The classical change point model with single change in mean is defined by

$$X_i = \begin{cases} \mu + e_i, & 1 \leq i \leq m, \\ \mu + \delta + e_i, & m < i \leq n, \end{cases} \tag{3.1}$$

where μ, δ and $m \leq n$ are unknown, n is the total number of observations and known, and m is called the change point. In the nonparametric settings, we assume that the errors $\{e_i \colon i = 1, \ldots, n\}$ are i.i.d. (independent and identically distributed) but non-observable with

$$E(e_i) = 0, \quad 0 < E(e_i^2) = \sigma^2 < \infty, \quad E|e_i|^\nu < \infty \quad \text{for some} \quad \nu > 2. \tag{3.2}$$

Most of the statistics were originally developed for independent normal errors, but it can be shown that the statistics work for all non-degenerate sequences of i.i.d. errors as long as the ν^{th} moment ($\nu > 2$) exists. Nevertheless, for details confer Csörgö and Horváth [37]. Moreover, Antoch et al. [8] and Horváth [61] described that the statistics also work for dependent errors that follow a linear process.

After specifying the model of interest in a change point analysis, then we focus on the detection of the change point, while the change point is estimated in the next

step. In the detection or testing of change point, we decide whether the sequence of observations is homogeneous or not with the following test hypothesis

$$\begin{aligned} &H_0 : \mu_1 = \cdots = \mu_n = \mu, \quad \text{against} \\ &H_1 : \mu_1 = \cdots = \mu_m = \mu; \quad \mu_{m+1} = \cdots = \mu_n = \mu + \delta, \end{aligned} \tag{3.3}$$

where $m = 1, \ldots, n-1$ is an unknown index of change point, the initial value μ and δ (either $\delta > 0$ or $\delta \neq 0$) are also unknown.

As mentioned, the decision whether the sequence of observations is stationary or not is based on the test statistics. In this subsection, we want to make a brief review on different types of test statistics in offline nonparametric inferences to detect change in the AMOC mean model.

Test Statistics

The most common statistics usually applied for the location model in nonparametric inferences are the *pseudo maximum-likelihood* method and the *pseudo-Bayesian* method, which are the base types for the *max-type* statistics and the *sum-type* statistics, respectively. To derive *pseudo maximum-likelihood* statistics, a common practice is to consider i.i.d. standard normal errors first and then prove that the statistic derived under Assumption 3.2 still gives valid results for different distribution. This is reasonable owing to the central limit theorem. Whereas, the pioneering work of change point detection in offline inferences within a *pseudo-Bayesian* framework by Chernoff and Zacks [32] and Kander and Zacks [72], is based on the assumption that the unknown mean μ and unknown change point m are independent random variables such that the *prior* distribution of m is uniform and μ is distributed as normal with mean zero and constant variance. In particular, we are interested in the well known cumulative sum (CUSUM) and moving sum (MOSUM) statistics to apply to testing change in the AMOC mean model with missing data. As outliers are also likely to appear, we give a short introduction of CUSUM and MOSUM statistics along with their robust counterparts.

Weighted CUSUM Statistic

The CUSUM test statistic was initially proposed by Page [106, 107, 108] in the context of quality control. Hereafter, various authors used that statistic in different contexts with different types. We focus on the weighted CUSUM for details confer [37], [7], [11] and [74]. To detect the change point in the AMOC mean model (3.1) with the

hypothesis (3.3), at first we assume that the errors are independent and normally distributed with $N(0, \sigma^2)$ and m is known. For calculating the likelihood ratio test statistic, suppose $m = k$, $a = \mu$ under H_0 and $b = \mu + \delta$ under H_1. Hence, the log likelihood ratio with the density function $\phi(x)$ and the distribution function $\Phi(x)$ is

$$\begin{aligned}\log \lambda_k &= \sup_{a,b} \log \frac{\prod_{i=1}^{k} \phi(X_i - a) \prod_{i=k+1}^{n} \phi(X_i - b)}{\prod_{i=1}^{n} \phi(X_i - a)} \\ &= \log \frac{\prod_{i=1}^{k} \phi(X_i - \bar{X}_k) \prod_{i=k+1}^{n} \phi(X_i - \bar{X}_k^*)}{\prod_{i=1}^{n} \phi(X_i - \bar{X}_n)},\end{aligned}$$

where $\bar{X}_k = \frac{1}{k} \sum_{j=1}^{k} X_j$, $\bar{X}_k^* = \frac{1}{n-k} \sum_{i=k+1}^{n} X_i$.

$$\begin{aligned}\log \lambda_k &= \frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 - \sum_{i=1}^{k} (X_i - \bar{X}_k)^2 - \sum_{i=k+1}^{n} (X_i - \bar{X}_k^*)^2 \right) \\ &= \frac{1}{2\sigma^2} \left(\frac{1}{k} \left(\sum_{i=1}^{k} X_i \right)^2 + \frac{1}{n-k} \left(\sum_{i=k+1}^{n} X_i \right)^2 - \frac{1}{n} \left(\sum_{i=1}^{n} X_i \right)^2 \right) \\ &= \frac{1}{2\sigma^2} \left(\frac{\left(\sum_{i=1}^{k} X_i \right)^2}{k} + \frac{\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{k} X_i \right)^2}{(n-k)} - \frac{\left(\sum_{i=1}^{n} X_i \right)^2}{n} \right) \\ &= \frac{1}{2\sigma^2} \frac{1}{nk(n-k)} \left(n^2 \left(\sum_{i=1}^{k} X_i \right)^2 + 2nk \sum_{i=1}^{n} X_i \sum_{i=1}^{k} X_i - k^2 \left(\sum_{i=1}^{n} X_i \right)^2 \right) \\ &= \frac{1}{2\sigma^2} \frac{n}{k(n-k)} \left(\sum_{i=1}^{k} X_i - \frac{k}{n} \sum_{i=1}^{n} X_i \right)^2 \\ &= \frac{1}{2\sigma^2} \frac{n}{k(n-k)} \left(\sum_{i=1}^{k} (X_i - \bar{X}_n) \right)^2 .\end{aligned}$$

Since, k is unknown, so the maximally selected likelihood ratio is natural to use, which leads to the test statistic

$$T_n^{(1)} = \max_{1 \le k < n} \left\{ \frac{1}{\sigma} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} (X_i - \bar{X}_n) \right| \right\}. \tag{3.4}$$

Test statistic (3.4) is used for the two-sided alternative with $\delta \neq 0$, while for the one-sided alternative with $\delta > 0$ is,

$$T_n^{(1)} = \max_{1 \le k < n} \left\{ \frac{1}{\sigma} \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} (X_i - \bar{X}_n) \right\}. \tag{3.5}$$

The error variance can be estimated by using any of the following forms:

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2, \tag{3.6}$$

$$\hat{\sigma^2} = \frac{1}{n-2}\left(\sum_{i=1}^{k}(X_i - \bar{X}_k)^2 + \sum_{i=k+1}^{n}(X_i - \bar{X}_k^*)^2\right). \tag{3.7}$$

One will get a more powerful test by using the estimated variance (3.7) than (3.6) in the max-type statistic (3.4) or (3.5). For large n and normally distributed random variables, approximate critical values of statistic (3.4) with estimated variance (3.7) were suggested by James et al. [68].

In order to use this test in testing for a change point, we need to find the critical value of this test statistic. That means, we need to know the distribution of the test statistic under H_0. The exact distribution of the test statistic depends on the distribution of errors, and in practice it is so complex and not known even for known error distribution with small n. Therefore, we need to know either the limit distribution of the test statistic under H_0 or approximate the critical values for further testing purpose. Using Bonferroni inequality one can approximate an upper estimate of the critical value at the significance level α, which is good enough for small sample but too conservative for large n, confer Worsley [146, 147]. The approximate critical values can be calculated for large n from the asymptotic behavior of the probabilities under H_0 in (3.3) using Darling et al. [38]

$$P\left(T_n^{(1)}A(\log n) \le t + D(\log n)\right) \approx \exp(-2e^{-t}), \tag{3.8}$$

for $t \in R^1$, and where $A(x) = \sqrt{2\log x}$, $D(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log\pi$, and $T_n^{(1)}$ is defined in (3.4), also suitable for (3.5) as well. It is assumed that $\max\limits_{1\le k<n}|\hat{\sigma^2} - \sigma^2| = o_p\left((\log\log n)^{-n}\right)$, for proofs and details confer Theorems 1.3.1 and 1.4.2 in Csörgö and Horváth [37]. Hence, the asymptotic critical value at α level is

$$c_\alpha = \frac{1}{A(\log n)}\left(-\log\left(-\frac{\log(1-\alpha)}{2}\right) + D(\log n)\right). \tag{3.9}$$

Moreover, it is a known fact that the convergence of extreme-value-type statistics is rather slow. Another approach to get the asymptotic distribution of the test statistic under H_0, convergence in distribution for large sample sizes is appealed to, i.e.,

$$\sup_{0<t<1}\frac{1}{\sigma}\frac{T_n(t)}{q(t)} \quad \xrightarrow{\text{D}} \quad \sup_{0<t<1}\frac{B(t)}{q(t)}, \tag{3.10}$$

where $q(t)$ is a general weight function, replacing $\sqrt{t(1-t)}$ at $t = \frac{k}{n}$ in (3.4), which is positive on $(0,1)$, and $\{B(t), 0 \leq t \leq 1\}$ is a Brownian bridge, i.e., B is a continuous Gaussian process with $EB(t) = 0$ and $EB(t)B(s) = \min(t,s) - ts$, for proof confer Chapter 2 of Csörgö and Horváth [37]. Csörgö and Horváth [37] illustrated through a simulation study that an asymptotic critical value coming from (3.8) is giving a conservative rejection region, since the rate of convergence to extreme value distributions is slow, and not performing better than the asymptotic critical value coming from (3.10). In recent years the trimmed max-type test statistic became popular, since they do not suffer so much from the convergence problems for extreme values. For two-sided alternative the test statistic is

$$\tilde{T}_n^{(1)} = \max_{\lfloor \beta n \rfloor \leq k \leq \lfloor (1-\beta)n \rfloor} \left\{ \frac{1}{\sigma} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} (X_i - \bar{X}_n) \right| \right\}, \tag{3.11}$$

where β is a small positive constant less than one, typically, $\beta \in [0.01, 0.1]$, and $\lfloor t \rfloor$ denotes the integer part of t. However, one should reasonably be sure that no change occurred in the stretch $X_1, \ldots, X_{\lfloor \beta n \rfloor}$ and $X_{\lfloor n - \beta n \rfloor}, \ldots, X_n$.

After having a positive decision about the existence of the change point in detection procedures, then the next concern is to estimate the location of the change point. Of course, other parameters of the model also have to be estimated. Approximations to the distributions of change point estimators and the construction of the interval estimators for the change points are also interesting topics in the change point analysis, but we limit our discussion to the estimation of the change point.

Suppose that the observations $X_1, \ldots, X_n$ are observed at the ordered time moments $t_1 < \cdots < t_n$ in the model (3.1). After rejecting the null hypothesis (3.3), the AMOC model (3.1) indicates that there is just one change after the m^{th} observation, and the corresponding time moment is known as the *change point*. Moreover, our focus is to estimate the integer valued t_m. We assume that the change point m can occur neither at the beginning nor at the very end of the observational period and satisfies

$$m = \lfloor n\gamma \rfloor, \quad \gamma \in (0,1), \tag{3.12}$$

where $\lfloor n\gamma \rfloor$ denotes the integer part of $n\gamma$. To estimate the unknown parameters m, μ and δ, one can use some general estimation methods, for instance, maximum likelihood estimation(MLE) or least square (LS) method. LS gives simple solutions, whereas, MLE requires the distribution of errors to be known in the model (3.1).

The LS estimators $\hat{m}_{LS}$, $\hat{\mu}_{LS}$ and $\hat{\delta}_{LS}$ of the parameters m, μ and δ, respectively, are estimated in such way that the sum of squared residuals is minimal, while, the change point estimators $\hat{m}_{LS}$ can be obtained from the maximization problem

$$\max\left\{\sqrt{\frac{n}{k(n-k)}}\left|\sum_{i=1}^{k}\left(X_i-\bar{X_n}\right)\right|;k\in\{1,\ldots,n-1\}\right\}.$$

Since, the solution of the maximization problem needs not to be unique. Using a convention, proposed by Antoch et al. [7, 11], the estimator for m_{LS} is

$$\hat{m}_{LS}=\arg\max\left\{\sqrt{\frac{n}{k(n-k)}}\left|\sum_{i=1}^{k}\left(X_i-\bar{X_n}\right)\right|;k\in\{1,\ldots,n-1\}\right\}. \tag{3.13}$$

CUSUM R-type Statistic

Handling of missing data may introduce outliers and as a result, robust-type of statistics would be a good choice in this context. Among robust methods the M-procedure, rank-type R-procedure, U-statistics and Kolmogorov-Smirnov type test statistics are the most common procedures for detection of changes in location models as well as regression models. Antoch and Hušková [9], Antoch et al. [7], and Antoch et al. [11] deal with various robust-types of change point tests and estimators. So far we have focused on the rank-type R-procedure, more specifically, CUSUM R-type statistic, for testing changes in mean, which is based on the simple linear rank statistics. The CUSUM R-type statistic is defined by

$$T_n^{(2)}=\max_{1\le k\le n-1}\left\{\frac{1}{\sigma_{n,R}}\left|\sqrt{\frac{n}{k(n-k)}}\sum_{i=1}^{k}\left(a_n\left(R_i\right)-\bar{a_n}\right)\right|\right\}, \tag{3.14}$$

where $(R_1,\ldots,R_n)$ is the vector of ranks corresponding to the observations $X_1,\ldots,X_n$; the scores $a_n(1),\ldots,a_n(n)$ are typically defined either by Wilcoxon scores or van der Waerden scores with $\bar{a_n}=\frac{1}{n}\sum_{i=1}^{n}a_n(i)$, here Wilcoxon scores $(a_n(i)=i/(n+1),i=1,\ldots,n)$ are used to define $a_n(i)$; and the scale $\sigma_{n,R}$ is defined by

$$\sigma_{n,R}^2=\frac{1}{(n-1)}\sum_{i=1}^{n}\left(a_n(i)-\bar{a_n}\right)^2. \tag{3.15}$$

Under H_0 the approximate critical values can be calculated for large n from

$$P\left(T_n^{(2)}A(\log n)\le t+D(\log n)\right)\approx\exp(-2e^{-t}), \tag{3.16}$$

for $t \in R^1$, and where $A(x) = \sqrt{2 \log x}$, $D(x) = 2 \log x + \frac{1}{2} \log \log x - \frac{1}{2} \log \pi$. And the estimator for m_{LS} by

$$\hat{m}_{LS} = \arg\max \left\{ \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} \left(a_n \left(R_i \right) - \bar{a_n} \right) \right| ; k \in \{1, \ldots, n-1\} \right\}. \tag{3.17}$$

MOSUM Statistic

There is another important type of test statistic based on moving sums, which is known as MOSUM statistic. The MOSUM statistic was first proposed but not analyzed mathematically by Antoch et al. [7]. Many authors have already investigated MOSUM statistics in the context of testing, most influential are Bauer and Hack [16], Chu et al. [33], Hušková [63], Hušková and Slabỳ [67], and Preuss et al. [115], and in the context of estimation of number and location of change points, Kirch and Muhsal [75] have investigated asymptotic properties of change point estimators based on MOSUM statistic. The performance of MOSUM statistic crucially depends on a bandwidth choice, which ideally should be chosen as the minimum distance between two neighboring change points. However, this method does not require to fix an upper bound for the number of changes, is not computationally expensive and the overall significance level is controlled [75]. The MOSUM statistic is defined by

$$T_n^{(3)}(G) = \max_{G<k\leq n} \frac{1}{\sqrt{G}} \frac{1}{\hat{\sigma_n}} \left| S_k - S_{k-G} \right|, \tag{3.18}$$

where $S_k = \sum_{i=1}^{k} (X_i - \bar{X_n})$ and $G < n$. We assume that G/n is small, typically, $G/n \sim 0.05$ or 0.10 (this is meant as a rule of thumb not in the correct asymptotic mathematical sense). Another MOSUM type statistic, which is especially suitable for more than one change and as a diagnostic tool (confer Antoch et al. [7], Chapter 4.1.3), uses the second order difference of S_k's. Whereas, MOSUM statistic (3.18) is the first order difference of S_k's.

$$\tilde{T}_n^{(3)}(G) = \max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G}} \frac{1}{\hat{\sigma_n}} \left| S_{k+G} - 2S_k + S_{k-G} \right|, \tag{3.19}$$

which also can be written as

$$\tilde{T}_n^{(3)}(G) = \max_{G\leq k\leq n-G} \frac{1}{\sqrt{2G}} \frac{1}{\hat{\sigma_n}} \left| \sum_{i=k+1}^{k+G} X_i - \sum_{i=k-G+1}^{k} X_i \right|. \tag{3.20}$$

For large n and small G/n we can use the following approximation to the distribution of (3.18) and (3.19) under null hypothesis, i.e.,

$$P\left(A_{n,G} T_n^{(3)}(G) > t + B_{n,G}(G) \right) \approx 1 - exp\left(-2e^{-t} \right), \tag{3.21}$$

$$P\left(A_{n,G}\tilde{T}_n^{(3)}(G) > t + B_{n,G}(G) - \log(\frac{2}{3})\right) \approx 1 - exp\left(-2e^{-t}\right), \tag{3.22}$$

where $t \in R^1$, $A_{n,G} = \sqrt{2\log\frac{n}{G}}$ and $B_{n,G} = 2\log\frac{n}{G} + \frac{1}{2}\log\log\frac{n}{G} - \frac{1}{2}\log\pi$. Based on the MOSUM the estimator $\hat{m}_{LS}(G)$ is defined as

$$\hat{m}_{LS}(G) = \arg\max\left\{\left|S_{k+G} - 2S_k + S_{k-G}\right|; k \in \{G, \ldots, n-G\}\right\}. \tag{3.23}$$

MOSUM R-type Statistic

The MOSUM type M- and R-estimators are also most usable statistics in robust case. The rank based MOSUM R-type statistic is defined analogously by

$$T_n^{(4)}(G) = \max_{G\le k\le n-G} \frac{1}{\sqrt{2G}} \frac{1}{\hat{\sigma_{n,G}}} \left|\sum_{i=k+1}^{k+G} a_n(R_i) - \sum_{i=k-G+1}^{k} a_n(R_i)\right|. \tag{3.24}$$

where $(R_1, \ldots, R_n)$ is the vector of ranks corresponding to the observations $X_1, \ldots, X_n$; the scores $a_n(1), \ldots, a_n(n)$ are Wilcoxon scores $(a_n(i) = i/(n+1), i = 1, \ldots, n)$; and the scale $\sigma_{n,R}$ is defined in (3.15). Under H_0 the approximate critical values can be calculated for large n from

$$P\left(A_{n,G}T_n^{(4)}(G) > t + B_{n,G}(G) - \log(\frac{2}{3})\right) \approx 1 - exp\left(-2e^{-t}\right), \tag{3.25}$$

where $t \in R^1$, $A_{n,G} = \sqrt{2\log\frac{n}{G}}$ and $B_{n,G} = 2\log\frac{n}{G} + \frac{1}{2}\log\log\frac{n}{G} - \frac{1}{2}\log\pi$. Based on the MOSUM R-type statistic (3.24) the estimator $\hat{m}_{LS}(G)$ is defined as

$$\hat{m}_{LS}(G) = \arg\max\left\{\frac{1}{\sqrt{2G}}\left|\sum_{i=k+1}^{k+G} a_n(R_i) - \sum_{i=k-G+1}^{k} a_n(R_i)\right|; G \le k \le n-G\right\}. \tag{3.26}$$

Bootstrapping

Finding good approximations to the critical values is one of the leading problems in hypothesis testing. In change point literature, the distributions of test statistics are very complex even in those cases where the assumption of normal errors is fulfilled, so that they can be determined explicitly only for small sample sizes. Hence, critical values based on limit distributions of the test statistics under null hypothesis are mostly recommended. However, in some cases the convergence is rather slow, for instance, if the limit distribution belongs to the extreme value type (see e.g., Antoch and Hušková [10], Hušková and Picek [66] and Kirch [74]), or its explicit form is unknown or it depends on unknown parameters. Consequently, the asymptotic critical

values are only good approximations for large sample sizes, otherwise they fail (see Kirch [74]).

Therefore, resampling becomes one of the reasonable possibilities to approximate the critical values. In change point context this approach was first suggested by Antoch and Hušková [10] and later pursued by others (see e.g., Hušková and Picek [66], Hušková [64], Kirch and Steinebach [76] and Kirch [74], etc.). Permutation and bootstrap tests have been developed in such approximations. This section focuses on bootstrap approach.

The bootstrap procedure simply resamples from the empirical distribution defined by one sample. If the original sample size is n, then the sampling is done at random and with replacement and resamples are all of size n. Whereas, bootstrap without replacement is coinciding with permutation principles. There are many types of bootstrap, e.g., percent bootstrap, naive bootstrap, block bootstrap, residual bootstrap etc. A good discussion of such resampling procedures can be found in the books by Efron and Tibshirani [43], Davison and Hinkley [39], and Good [53].

Bootstrap Critical Value
An algorithm for finding bootstrap critical values in the weighted CUSUM test (3.4) is illustrated in Algorithm 2.

Algorithm 2 Algorithm for approximating bootstrap critical value

1: Draw a bootstrap sample $X_j^* = (x_1^*, \ldots, x_n^*)$ from $(x_1, \ldots, x_n)$ with replacement.
2: Calculate the test statistic $T_{ni}^{(1)*}$.
3: Repeat step 1. and step 2. for $i = 1, \ldots, B$, where $B = 10,000$ (or 1000 times).
4: Calculate the empirical critical value, α-quantile $= c$, such that

$$\frac{1}{B}\sum_{i=1}^{B} 1_{\left\{T_{ni}^{(1)*} \leq c\right\}} \geq (1-\alpha).$$

5: Reject the null hypothesis if $T_n^{(1)} > c$.

Algorithm 2 can be carried out for finding the bootstrap critical values using CUSUM R-type test (3.14), MOSUM test (3.20) and MOSUM R-type test (3.24).

3.1.2 Single Change Point in Linear Regression

A common phenomenon in time-series regression models is parameter instability, as it is more likely for underlying data-generating mechanism to be disturbed over a longer horizon by various factors that fostered the changes in relationship among some variables. Even such changes are present in the simple linear regression model, where only the relationship between two variables is studied with explained variable depending linearly on the explanatory one. There are many terminologies to describe a change point regression, such as: segmented [82], broken-line [138], structural change, structural break or smoothing transition [13], in which the relationship between the response and explanatory variable is linear. The change point detection problem in the simple linear regression arises with many features, e.g., either one or both parameters (intercept, slope) can change, the starting parameters before the change point are either known or unknown, either the continuity of the regression function at the change point is assumed or there can be discontinuity etc., [7]. In this section, we concentrate on reviewing single change point detection in both the parameters of the simple linear regression model with unknown starting values.

Model and Assumptions

$$Y_i = \begin{cases} \boldsymbol{x_i'}\boldsymbol{\beta} + e_i, & 1 \leq i \leq m, \\ \boldsymbol{x_i'}(\boldsymbol{\beta} + \boldsymbol{\delta}) + e_i, & m < i \leq n, \end{cases} \tag{3.27}$$

where $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ and m are unknown, $\boldsymbol{x_i} = (1, x_{2i}, \ldots, x_{ki})'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$, $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_k)'$. The model is called either a random or a fixed design depending on x_i values being random or deterministic, respectively. Moreover, assume that errors $\{e_i\}$ are independent of explanatory variables and i.i.d. random variables with

$$E(e_i) = 0, 0 < E(e_i^2) = \sigma^2 < \infty, \quad E|e_i|^\nu < \infty \quad \text{for some} \quad \nu > 2. \tag{3.28}$$

The variance, σ^2, is supposed to be known. If it is unknown, it can be replaced by its usual estimator

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{x_i}\hat{\boldsymbol{\beta}} \right)^2, \tag{3.29}$$

where k is the number of regressors, n is the number of observations, and $\hat{\boldsymbol{\beta}}$ is the least squares estimators of the parameters of regression coefficients under the null hypothesis and can be estimated by using the data vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ and the

design matrix $\boldsymbol{X} = (X_1, \ldots, X_n)'$ as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}, \tag{3.30}$$

which is also the maximum-likelihood estimator (MLE) for normal errors. The values of independent variables has an important role in finding the limit behavior of test statistics. The design points $\boldsymbol{x_i}$ are either may random, e.g., they represent a realization of a sequence of independent random variables, or may generally be a realization of a stationary time series, or equally spaced, i.e., $x_i = i/n, i = 1, \ldots, n$. Here, we consider a fixed design for the model (3.27) with assumptions

$$\boldsymbol{X}\boldsymbol{X}' \quad \text{is invertible,} \tag{3.31}$$

i.e., the vectors $(1, x_{2i}, \ldots, x_{ki})$ are linearly independent. And

$$\left\| \frac{1}{n}\boldsymbol{X}\boldsymbol{X}' - \boldsymbol{C} \right\|_{\infty} = O\left(n^{-1/2}\right), \tag{3.32}$$

for some positive definite matrix $\boldsymbol{C}$. At this moment, we are interested to test the model (3.27) with equally spaced design points for changes in both intercept and slope coefficients using the following test hypothesis

$$\begin{aligned} &H_0 : m = n, \quad \text{against,} \\ &H_1 : m < n, \boldsymbol{\delta} \neq 0. \end{aligned} \tag{3.33}$$

Test Statistics

Change points detection in linear regression model is such a large area of research, which in itself attests the importance of the problem. Therefore, different types of test procedure have been suggested in this literature, among them the generalized fluctuation test framework Kuan and Hornik [80] (which includes different types of CUSUM and MOSUM type tests), F-type test statistics (Andrews and Ploberger [5], Bai and Perron [14]), and Bayesian sum-type tests (Stephens [134], Green [56]) are most common to use. Within this framework, starting from the Recursive CUSUM test (Brown et al. [24]), a variety of fluctuation type tests has been introduced, for instance, OLS-CUSUM test (based on ordinary least squares residuals) of Ploberger and Krämer [111], OLS-MOSUM test of Chu et al. [33] and Recursive MOSUM test (based on recursive residuals) by Bauer and Hackl [17], Recursive-estimates test of Ploberger et al. [112], Moving-estimates test of Chu et al. [34], and more recently generalized M-type test procedures by Marušiaková and Hušková [92] based on functionals of weighted M-residuals. Zeileis et al. [151, 152] also discuss several tests with

implementation in R package *strucchange* based on independent errors. Within this section we discuss OLS-CUSUM, Recursive CUSUM, OLS-MOSUM and Recursive-MOSUM tests which will be applied to detect change points in the linear regression model with missing data.

All of these four tests are based on the sums of estimated residuals, which are estimated either by ordinary least squares (OLS) residuals or by recursive residuals. The OLS residuals can be estimated by

$$\hat{e_i} = \boldsymbol{Y}_i - \boldsymbol{x_i}'\hat{\boldsymbol{\beta}}, \tag{3.34}$$

where $\hat{\boldsymbol{\beta}}$ is estimated from (3.30) and the estimated error variance $\hat{\sigma}_n^2$ can be estimated by (3.29). The recursive residuals are the standardized residuals from the regression of each observation. Let $\boldsymbol{X}_r = (x_1, x_2, \ldots, x_r)'$ be a matrix of dimension $r \times k$ and $\boldsymbol{Y}_r = (Y_1, Y_2, \ldots, Y_r)'$ be a matrix of dimension $r \times 1$. Also, suppose $\hat{\boldsymbol{\beta}}_r = \left(\boldsymbol{X}'_r\boldsymbol{X}_r\right)^{-1}\boldsymbol{X}'_r\boldsymbol{Y}_r$ be the ordinary least squares estimator at time r and then the r^{th} recursive residual has the form

$$\tilde{e_r} = \frac{\boldsymbol{Y}_r - \boldsymbol{x_r}'\hat{\boldsymbol{\beta}}_{r-1}}{\sqrt{1 + \boldsymbol{x_r}'(\boldsymbol{X}'_{r-1}\boldsymbol{X}_{r-1})^{-1}\boldsymbol{x_r}}}, \quad r = k+1, \ldots, n. \tag{3.35}$$

Under the null hypothesis the recursive residuals are uncorrelated with zero mean and constant variance and are therefore independent under the normality assumption. The corresponding variance estimate is

$$\tilde{\sigma}^2 = \frac{1}{n-k}\sum_{i=k+1}^{n}\left(\tilde{e_i} - \bar{\tilde{e}}\right)^2. \tag{3.36}$$

OLS-CUSUM Test Statistic

The OLS-CUSUM test suggested by Ploberger and Krämer [111] is based on the estimated OLS residuals (3.34) and using the test statistic

$$W_n^o = \max_{1\le t\le n}\frac{1}{\hat{\sigma}\sqrt{n}}\left|\sum_{i=1}^{t}\hat{e}_i\right|. \tag{3.37}$$

In real field applications such as finance, quality control, econometrics, etc., one is primarily interested in the mean change in a particular direction. In these cases, it is more expected to use the one-sided tests. Hence, the one-sided OLS-CUSUM test statistic is

$$W_n^o = \max_{1\le t\le n}\frac{1}{\hat{\sigma}\sqrt{n}}\sum_{i=1}^{t}\hat{e}_i. \tag{3.38}$$

Ploberger and Krämer [111] showed that the asymptotic distribution of W_n^o is determined by the probability of a Brownian bridge $B(t)$, crossing a pair of constant boundaries.

OLS-MOSUM Test Statistic

Chu et al. [33] introduced OLS-MOSUM test statistic in the light of Bauer and Hackl [17], and Ploberger and Krämer [111], where the test statistic does not contain the sum of all residuals for the entire data but the sum of fixed number of residuals in a data window moving across the whole sample and the size of the data window is determined by the bandwidth parameter $h \in (0, 1)$. Hence, the two-sided OLS-MOSUM test statistic is

$$M_{n,h}^o = \max_{0 \le j \le n - \lfloor nh \rfloor} \frac{1}{\hat{\sigma} \lfloor nh \rfloor^{\frac{1}{2}}} \left| \sum_{i=j+1}^{j+\lfloor nh \rfloor} \hat{e}_i \right|, \quad 0 < h < 1, \tag{3.39}$$

where $\lfloor nh \rfloor$ denotes the integer part of nh. The performance of moving sum statistics depends on the choice of bandwidth h. Moving sums with large h are not very sensitive to parameter variation, since each moving sum includes 'too many' residuals and only a few moving sums are available to detect possible parameter changes. On the other hand, if h is small, the sample variation in the moving sums is likely to be large, and the limit distribution may not be a good approximation [33]. In practice, one-sided OLS-MOSUM test statistic is also applicable and defined as

$$M_{n,h}^o = \max_{0 \le j \le n - \lfloor nh \rfloor} \frac{1}{\hat{\sigma} \lfloor nh \rfloor^{\frac{1}{2}}} \sum_{i=j+1}^{j+\lfloor nh \rfloor} \hat{e}_i, \quad (0 < h < 1). \tag{3.40}$$

As the representations (3.39) and (3.40) indicate, the limit process for the OLS-MOSUM processes are the increments of a Brownian bridge, for more details see [33].

Recursive CUSUM Test Statistic

Brown et al. [24] suggested to consider cumulative sums of recursive residuals in construction of the recursive CUSUM test Statistic, i.e.,

$$W_n^r = \max_{k+1 \le t \le n} \frac{1}{\tilde{\sigma}\sqrt{\tau}} \left| \sum_{i=k+1}^{t} \tilde{e}_i \right|, \tag{3.41}$$

where $\tau = n - k$ is the number of recursive residuals. One-sided recursive-CUSUM test statistic can be defined by

$$W_n^r = \max_{k+1 \le t \le n} \frac{1}{\tilde{\sigma}\sqrt{\tau}} \sum_{i=k+1}^{t} \tilde{e}_i. \tag{3.42}$$

The asymptotic distribution of the test statistic (3.41) or (3.42) is determined by the probability of a Wiener Process (or a Standard Brownian Motion). Krämer et al. [79] showed that the main properties of the CUSUM test remain the same even under weaker assumptions and also extended the test to detect the change in a dynamic model.

Recursive MOSUM Test Statistic

The recursive-MOSUM test statistic, introduced by Bauer and Hackl [17], also considers the recursive residuals (3.35). This test is differs from the CUSUM test in that a fixed number, $\lfloor \tau h \rfloor$, of residuals are available in each moving sum, whereas cumulated sum incorporate more and more residuals. Two-sided statistic has the form

$$M^r_{n,h} = \max_{0 \le j \le \tau - \lfloor \tau h \rfloor} \frac{1}{\tilde{\sigma} \lfloor \tau h \rfloor^{\frac{1}{2}}} \left| \sum_{r=k+j+1}^{k+j+\lfloor \tau h \rfloor} \tilde{e}_i \right|, \quad (0 < h < 1). \tag{3.43}$$

Accordingly, the one-sided test statistic is

$$M^r_{n,h} = \max_{0 \le j \le \tau - \lfloor \tau h \rfloor} \frac{1}{\tilde{\sigma} \lfloor \tau h \rfloor^{\frac{1}{2}}} \sum_{r=k+j+1}^{k+j+\lfloor \tau h \rfloor} \tilde{e}_i, \quad (0 < h < 1). \tag{3.44}$$

Bauer and Hackl [17] determined the critical values by incorrectly ignoring correlations of moving sums, which was corrected by Chu et al. [33] who characterised the limiting process of moving sums of recursive residuals in terms of the increments of a standard Wiener Process. As this limiting process has a constant variance, the asymptotic critical values are determined by the probability that this process crosses a pair of constant boundaries [33].

When the null hypothesis is rejected and there is an evidence of change point in the regression model (3.27), then we need to estimate the change point m. We assume that the change point m fulfills (3.12). Hence, the estimator of m is based on OLS-CUSUM is defined as

$$\hat{m} = \arg\max \left\{ \frac{1}{\hat{\sigma}\sqrt{n}} \left| \sum_{i=1}^{k} \hat{e}_i \right| ; k \in \{1, \ldots, n\} \right\}. \tag{3.45}$$

The estimator is consistent under H_0 as well as H_1 and the arguments of the maxima of recursive CUSUM process, OLS-MOSUM process and recursive MOSUM process are the estimates of the change points in respective tests.

3.2 Multiple Change Points Models

Several authors identified multiple change points in sequence of observations from time series, which is related to different fields, for instance, finance, climatology, genomics, epidemiology, and model validation, etc. Therefore, methods for identifying and estimating those change points are highly required for both methodological and practical reasons. There is considerable literature regarding multiple changes in the location and the regression models. The very first article, to the best of our knowledge, on multiple changes was by Vostrikova [141]. She proposed Binary Segmentation methods in a stochastic process setting and showed consistency of this method for the number and locations of change points for a finite sample size. Initially, Binary Segmentation method uses the entire data via any test designed for AMOC, particularly CUSUM type test, for detecting a single change point (in a sequence of independent observations). Once a change point is estimated if the test is significant, the data are split into two segments defined by the detected change point, and then the procedure is repeated for each segment until it is no longer significant. Having advantages of simplicity and computer efficiency, one drawback of this method is the inability to control the overall significance level. Specially, the power can suffer greatly in case of multiple changes, which can be overcome by introducing an extra randomization step in selection of the segment to be tested, proposed by Fryzlewicz et al. [48]. In this section, multiple changes in structure of mean and regression model is discussed with their offline nonparametric detection and estimation aspects.

3.2.1 Multiple Change Points in Mean

Model and Assumptions

The model with multiple change points in mean, assuming an observed sequence of independent random variables $X_1, \ldots, X_n$, is defined by

$$X_i = \mu_i + e_i, \qquad \lfloor n\vartheta_{j-1} \rfloor < i \leq \lfloor n\vartheta_j \rfloor, \quad j = 1, \ldots, q+1, \tag{3.46}$$

where $0 = \vartheta_0 < \vartheta_1 \leq \ldots \leq \vartheta_q \leq \vartheta_{q+1} = 1$, and the change points $\lfloor n\vartheta_1 \rfloor, \ldots, \lfloor n\vartheta_q \rfloor$, the number of change points $q \in N$ as well as the expected values $\mu_1, \ldots, \mu_{q+1} \in R$ with $\mu_j \neq \mu_{j+1}$, $j = 1, \ldots, q$, are unknown. The errors $e_1, \ldots, e_n$ are i.i.d. random

variables with assumptions (3.2). For testing the model (3.46) concerned hypothesis is

$$\begin{aligned} &H_0 : \mu_1 = \cdots = \mu_{q+1}, \quad \text{against} \\ &H_1 : \mu_j \neq \mu_{j+1}, \quad j = 1, \ldots, q; q \geq 1. \end{aligned} \tag{3.47}$$

Test Statistics

Among all nonparametric test statistics our focus is on the CUSUM type statistics considering Binary Segmentation procedure and MOSUM type statistics to identify the multiple changes in the mean model with independent errors. We apply weighted CUSUM test (3.4) and CUSUM R-type test (3.14) in Binary segmentation procedure to find out significant multiple changes in the model (3.46) under the assumptions (3.2). Usually, the most popular tests in the context of multiple change situation are MOSUM type statistics, e.g., the MOSUM test statistic (3.20) and MOSUM R-type test (3.24) introduced in Section 3.1.1.

Weighted CUSUM Statistics

Binary Segmentation procedure uses directly any type of CUSUM test statistic (3.4) in multiple change situation and is still consistent. Details on the consistency of this approach for estimating the true change point locations under various conditions are found in [141, 140]. Binary segmentation procedure can be used to extend any single change point method to multiple change points. The stepwise algorithm for weighted CUSUM statistic (3.4) is illustrated in Algorithm 3.

Algorithm 3 The generic Binary Segmentation algorithm to detect multiple change points using the Weighted CUSUM test (3.4).

1: Set the data for testing $X_1, \ldots, X_n$.
2: Calculate the test statistic $T_n^{(1)}$ and critical value c_α (3.9) at α level of significance; if $T_n^{(1)} > c_\alpha$ then select $\tilde{m}_1^{(1)} := \arg\max |T_n^{(1)}|$.
3: Split the data into two segments, i.e., $X_1, \ldots, X_{\tilde{m}_1}$ and $X_{\tilde{m}_1+1}, \ldots, X_n$ and repeat step 1 and 2 for each segment until no significant change points are detected.
4: Obtain $\tilde{m}_1^{(1)}, \ldots, \tilde{m}_q^{(1)}$.

Using the binary segmentation procedure it is difficult to obtain an appropriate variance estimator which is necessary for the test. Since the number of change points q is unknown, so the variance estimator becomes larger if changes are not taken into account, which returns smaller test statistic. As a result, we loose power for small

samples. The overall significance level can become quite large, which is not controllable.

CUSUM R-type Statistics

As this chapter deals with the missing data feature of cell stimulus response data in change point analysis, so our data set may also have outliers. A robust type test seems more appropriate in this situation. Hence, Binary Segmentation Algorithm 3 with CUSUM R-type Statistic (3.14) and estimator (3.17) can also be used in multiple change points cases.

MOSUM Statistics

MOSUM test (3.20) and the estimator (3.23) can be applied directly to the entire data without using the binary segmentation procedure to test and estimate the multiple change points in the location model with the critical value from (3.22). The steps of MOSUM procedure is explained in Algorithm 4.

Algorithm 4 MOSUM procedure using test statistic (3.20).

1: Find all pairs of indices ν_j, ω_j, $j = 1, \ldots, q$, such that $\omega_j - \nu_j \geq \tau_0 G$ for $\tau_0 > 0$ arbitrary but fixed.
2: Calculate the test statistic $\tilde{T}^{(3)}_{n,k}(G)$ in (3.20) for $\nu_j < k < \omega_j$ and compare that with the critical value $c_\alpha(G)$ from (3.22) at α level of significance.
3: Select $\hat{m}_{LS}(G)$ using (3.23) for those segments, where $\tilde{T}^{(3)}_{n,k}(G) > c_\alpha(G)$.
4: Obtain $\tilde{m}_1(G), \ldots, \tilde{m}_q(G)$.

MOSUM R-type Statistics

The robust test, MOSUM R-type test (3.24), and the estimator (3.26) can also be used to test hypothesis (3.47) for the model (3.46) and estimate significance multiple change points with the approximate critical value from (3.25).

3.2.2 Multiple Change Points in Linear Regression

Model and Assumptions

Linear regression models with multiple changes in coefficients can be considered as

$$Y_i = \begin{cases} \boldsymbol{x_i}'\boldsymbol{\beta_1} + e_i, & 1 \le i \le m_1, \\ \boldsymbol{x_i}'\boldsymbol{\beta_2} + e_i, & m_1 < i \le m_2, \\ \cdots & \\ \boldsymbol{x_i}'\boldsymbol{\beta_{q+1}} + e_i, & m_q < i \le n, \end{cases} \tag{3.48}$$

where regression coefficients $\boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_q}$ and change points $m_1, \ldots, m_q$ are unknown, $\boldsymbol{x_i} = (1, x_{2i}, \ldots, x_{ki})'$, $\boldsymbol{\beta_1} = (\beta_{11}, \ldots, \beta_{k1})'$, and Y is a $n \times 1$ response variable. The model is called either a random or a fixed design depending on x_i values whether random or deterministic, respectively. Moreover, assume that errors $\{e_i\}$ are independent of explanatory variables and i.i.d. random variables with the assumptions (3.28), (3.31), and (3.32). In testing multiple change points in regression coefficients of the model (3.48), test hypothesis is

$$\begin{aligned} &H_0 : \boldsymbol{\beta_1} = \cdots = \boldsymbol{\beta_{q+1}}, \quad \text{against} \\ &H_1 : \boldsymbol{\beta_j} \neq \boldsymbol{\beta_{j+1}}, q \ge 1, \quad j = 1, \ldots, q. \end{aligned} \tag{3.49}$$

Test Statistics

OLS-CUSUM test (3.37) and recursive CUSUM test (3.41) can be used incorporating Binary Segmentation method described in section 3.2.1 for identifying significant multiple changes in the model (3.48), whereas, OLS-MOSUM test (3.39) and recursive MOSUM test (3.43) can be directly used to the entire data for detecting the multiple changes in the regression model (3.48). When we reject the null hypothesis, estimate the number of changes and their locations is the prime concerned at this stage of change point analysis.

3.3 Missing Data Imputation

Missing data (or missing values) is defined as the data values that are not available or stored for a variable in the observation of interest. The necessity of missing data analysis naturally emerged from the applied work conducted by researchers in various fields. Reasons for such concern with missing data include: data are difficult to collect so better to account all collected data; failing to adequately address issues of missing data can lead to biased estimates, incorrect standard errors, imprecise confidence intervals, distorted statistical power, and invalid conclusions. Therefore, it is

required to use the optimum strategy for handling missing data. Good discussions of such procedures can be found in the books by Little and Rubin [86, 85], Schafer [126], and Schafer and Graham [127].

Appropriately dealing with missing data can be challenging as it requires a careful examination of the data to identify the type and pattern of missingness. Little and Rubin [86] distinguished the *missing-data pattern* and the *missing-data mechanism* from each another; the *missing-data pattern* describes which values are observed in the data matrix and which values are missing, whereas, the *missing-data mechanism* (or mechanisms) concerns the relationship between missingness and the values of variables in the data matrix. As a result, the impact of the missing data analysis depends on the missing-data mechanism. Correspondingly, a clear understanding of how the different imputation methods works also another important consideration in missing data analysis. In this section, a brief discussion about the missing-data mechanism and some imputation methods are presented to identify the best imputation method for the cell stimulus response data.

3.3.1 Missing-Data Mechanism

To decide how to handle missing data, it is helpful to know why they are missing. More formally, Rubin [121] defines the missing-data mechanism in three distinct types: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Each term refers to the probability of missing values, given information about (1) the variable(s) with the missing data, (2) associated variables (including covariates), and (3) a hypothetical mechanism underlying the missing data. Adopting a generic notation to characterize the missing-data mechanism, let us denote the complete data matrix as $Y = (y_{ij})$, where y_{ij} is the value of the j^{th} variable for individual i, which contains the observed data as Y_{obs} and missing data as Y_{mis}, i.e., $Y = (Y_{\text{obs}}, Y_{\text{mis}})$; the missing-data indicator matrix as $M = (m_{ij})$ which defines the pattern of missing data, such that $m_{ij} = 1$ if observation y_{ij} is missing and 0 otherwise; and ϕ denotes unknown parameters. Hence, the missing-data mechanism is defined as the conditional distribution of M given Y, say $P(M|Y, \phi)$.

Missing completely at random (MCAR)

If missingness does not depend on the values of the data set, i.e.,

$$P(M \mid Y, \phi) = P(M \mid \phi) \quad \text{for all} \quad Y_{\text{obs}}, \phi, \tag{3.50}$$

then the data are MCAR. Therefore, this assumption does not mean that the pattern itself is random, but rather missingness does not depend on either the observed (Y_{obs}) or the unobserved (Y_{mis}) values of data matrix Y. For instance, n individuals had their blood pressure measured and a random sample of size $\tilde{n} < n$ also had their Body Mass Index (BMI). That illustrates, if the probability that BMI is missing is the same for all individuals, regardless of their measured blood pressure or BMI.

Missing at random (MAR)

A more realistic assumption under these circumstances is MAR mechanism, which specifies missingness depends only on the components Y_{obs} of Y, and not on the components that are missing. Hence, MAR is defined as

$$P(M \mid Y, \phi) = P(M \mid Y_{\text{obs}}, \phi) \quad \text{for all} \quad Y_{\text{mis}}, \phi. \tag{3.51}$$

For example, n individuals had their blood pressure measured and only those individuals with high blood pressure also had their BMI. This implies, if the probability of measured BMI is missing varies according to the high measured blood pressure of the respondent but does not vary according to the BMI of respondents.

Missing Not at Random (MNAR)

In situations, when missingness depends on the unobserved values of the data set, the data are MNAR. Mathematically, this can be expressed as

$$P(M \mid Y, \phi) = P(M \mid Y_{\text{mis}}, \phi) \quad \text{for all} \quad Y_{\text{obs}}, \phi. \tag{3.52}$$

For example, n individuals had their blood pressure measured but only overweight BMI individuals also had their BMI. Which indicates, if the probability that BMI is recorded varies according to overweight BMI individuals but does not vary according to the measured blood pressure of the respondent. Under these circumstances, the nonresponse is said to be informative. However, the analysis based on the responding subsample is generally biased for the parameters of the distribution of Y if the data are MNAR.

The MCAR assumption may be more plausible if the missing data are *missing by design*, whereas, according to mechanism characteristics MAR is called *ignorable*

nonresponse, and MNAR is called *nonignorable*. Another important feature of data, which is censoring and more likely to appear in the time-to-event data, also can be illustrated as a missing data mechanism, e.g., non-informative censoring can reveal as *ignorable missing* which implies that a suitable assumption is either MCAR or MAR mechanism, and informative censoring can be expressed as *nonignorable missing* where MNAR is more plausible. Particularly, left censoring or right censoring always is related with the MNAR mechanism.

The missing data mechanisms are not characteristics of an entire data set, but the mechanisms are merely assumptions which apply to specific analysis methods. Any attempt to identify the missing data mechanism and correct for selective nonresponse will typically represent an improvement in the accuracy of results over making no attempt at all. Nevertheless, the significance of MCAR, MAR and MNAR assumptions about the missing data mechanism also depends somewhat on the objective of the analysis.

3.3.2 Imputation Methods

The practice of filling in missing values with plausible values is known as imputation. Imputation is a general and desired analysis for handling missing-data problems rather than discard the unit entirely. This method helps to prevent loss of power resulting from a diminished sample size, also can make use of imputed information and maintain high precision. However, it has pitfalls, particularly in multivariate settings, since it may distort data distributions and relationships

There are many different imputation methods among them single imputation, multiple imputation (MI), Expectation Maximization (EM) Algorithm and full-maximum likelihood methods. Single imputation uses to fill in one value for each missing item, multiple imputation imputes more than one value for each missing item and incorporates the uncertainty. EM technique iteratively goes through the data to impute a value (Expectation) then checks whether that is the value most likely (Maximization) while still preserving the covariance structure of the data. Maximum likelihood method does not replace or impute missing values rather the missing data is handled within the analysis model, which model is estimated by a full information maximum likelihood method, that way all available information is used to estimate the model.

As the rate of missing data in cell stimulus response data is 0.5%, which is very small, so we focus on different single imputation methods only.

Table 3.1: Single Imputation Producers

General Procedure	Specific Procedure	Type of Data
Constant	Mean substitution	Continuous normal
	ML mean substitution	Continuous normal
	Median substitution	Continuous
	Zero imputation	Categorical or continuous
Random		
Data-based	Hot Deck	Any
	Cold Deck	Any
Model-based	Bayesian (MCMC)	Any
	ML	Continuous normal
Nonrandom		
One condition	Group mean	Continuous with groups
	Group median	Continuous with groups
	Last observation carried forward (LOCF)	Longitudinal
	Next observation carried backward (NOCB)	Longitudinal
Multiple condition	Mean previous observations	Longitudinal
	Mean subsequent observations	Longitudinal
	Last and next average	Longitudinal
	Regression	Multivariate continuous
	Regression with error	Multivariate continuous

NOTE: ML = maximum likelihood, MCMC = Markov chain Monte Carlo.

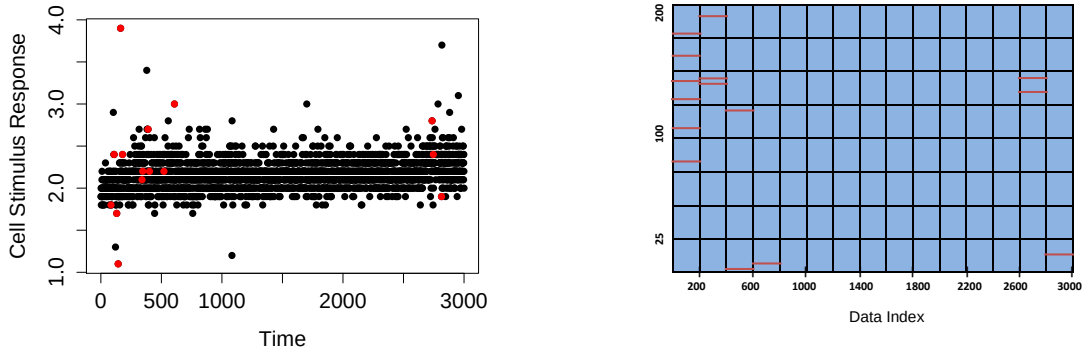
Table 3.1 outlines the single imputation procedures that replace the missing value with either (1) a constant, (2) a randomly selected value, or (3) a nonrandomly derived value discussed by McKnight et al. [95]. In brief summary of these methods, the general method of constant replacement involves replacing a constant to fill in the missing value according to their data types, e.g., imputing for continuous normal data either by mean (known as *mean substitution*) or by the estimated population mean (known as *ML mean*), and for continuous data by *median substitution* and *zero imputation* in case of continuous and categorical data. Randomly derived values can come from either data-based procedures which includes *Hot Deck* and *Cold Deck* methods, or model-based procedures that involves *Bayesian (MCMC)* and *ML* methods. All

the methods can be applied to any type of data except *ML* requires continuous normal. Finally, nonrandomly obtained values can be assigned using *Group mean* and *Group median*; *LOCF*, *NOCB*, *Mean previous*, *Mean subsequent* and *Last and next average*; and *Regression* imputation methods depending on continuous with groups, longitudinal, and multivariate continuous data types, respectively.

In *mean substitution*, missing values are replaced by the average of the observed values for that item. The problem with this is that it reduces the variance and the absolute value of the covariance. *Hot Deck* imputation method fills in missing value by a random draw from the observed values. It has no parametric model and solves the problem of understanding uncertainty. The method still distorts correlations and other measures of association. *Last and next average* imputes the average of last observed and next observed values of missing item. It provides nonrandom replacing values for the missing values but not constant. *Regression imputation* replaces predicted values from a regression of the missing items on items observed for the unit, usually it uses only the observed set of data for calculation. This method is also called *conditional mean imputation*. This method is not recommended for analyses of covariances or correlations. If there is no association between two variables then the method reduces to ordinary mean substitution [127].

3.3.3 Imputation in Cell Stimulus Response Data

The brain is a stimulus-based system. When the brain comes to functioning in the real world, it does provide itself with certain intense stimuli. The brain responds to different stimuli in different ways. Neurons send signals to other cells as electrochemical waves traveling along thin fibers (called axons), which cause chemicals (called neurotransmitters) to be released at junctions (called synapses). Excited synapses, inhibited synapses, or otherwise modulated synapses are different kind of reactions or signals that a cell receives from a neuron. In this study, one brain cell's reaction or response to a stimulus is observed at 50 Hz, i.e., 50 dta per second. This data set contains only one variable, i.e., cell stimulus response, which is observed for 1 minute, i.e., the total sample size is $60 \times 50 = 3000$. Therefore, the data has time-series feature. According to different data types mentioned in the Table 3.1, we can classify the cell stimulus response data as longitudinal data.



(a) (b)

Figure 3.2: (a) Scatter Plot of the Cell Stimulus Response Data with missing observations in red dot. (b) Matrix Plot of the Cell Stimulus Response variable with missing values are highlighted by red, where each 25 cells of the data matrix is visualized by a blue rectangle.

The distribution of the missing values in the data is very important to identify the missing data mechanism and desirable imputation method for the analysis. Figure 3.2 reveals this important feature using scatter plot and matrix plot. Timewise cell stimulus response are explored in Figure 3.2(a) using a scatter plot with missing values in red, which emphasizes that missing values are not constant and have negligible correlation between two variables. Consequently, this implies a MCAR mechanism as the occurrence of missing values does not depend on the time (observed variable) and not even on cell stimulus response (missing variable). It appears from Figure 3.2(b) that there is not large amount of accumulation points of missing values in a certain value interval of cell stimulus response, for instance, maximum two missing values are observed in a rectangle interval (which contains 25 data points). Hence, missing data distribution shows an almost arbitrary pattern except in the 4th to 8th rectangle intervals. Moreover, notice that missing responses to the cell stimulus arises at time 84, 109, 131, 143, 163, 179, 340, 346, 390, 402, 521, 608, 2736, 2746, and 2814 seconds. Hence, there is 15 missing values out of 3000 data values. As a result, only 0.5% of the data values are missing.

Having MCAR mechanism and longitudinal data features in the cell stimulus response data set, we prefer to impute certain nonrandom values. Accordingly, *Last and next average* seems more reasonable for our data structure. *Last and next average*

imputation method is considered as the appropriate one to apply in the cell stimulus response data along with *Mean Substitution*, *Hot Deck* and *Regression Imputation* methods to make an evaluation on their performances.

Table 3.2: Performance of Single Imputation Methods for Parameter Estimates with MCAR mechanism

Parameters	**Mean Substitution**	**Hot Deck**	**Last and next Average**	**Regression Imputation**
μ	2.15719 **(0.16064)**	2.15713 (0.16126)	2.15695 **(0.16101)**	2.15701 (0.89722)
$\beta_{Y\|X}$	0.14477 **(0.00981)**	0.14661 (0.00984)	0.14731 **(0.00982)**	0.14626 **(0.00981)**
$\beta_{X\|Y}$	0.46769 (0.03170)	0.46998 **(0.03155)**	0.47371 **(0.03159)**	0.47217 (0.03166)

NOTE: Estimate's standard errors are in parentheses and boldface type indicates smallest two standard errors.

In Figure 3.3, *Mean substitution* causes all the imputed values of cell stimulus response to fall on a horizontal line, whereas *Regression imputation* causes them to fall on a regression line. *Hot Deck* produces a random cloud with too little correlation and *Last and next average* method produces a reasonable point cloud with slightly better correlation. As per our expectation, nonrandom values are imputed only by *Hot Deck* and *Last and next average* methods. Table 3.2 shows estimated values with their standard errors for the parameters mean, regression coefficients Y (cell stimulus response) on X (response time) and regression coefficients X on Y using four imputation methods. *Last and next average* performs consistently better than all other counterparts in all parameters cases.

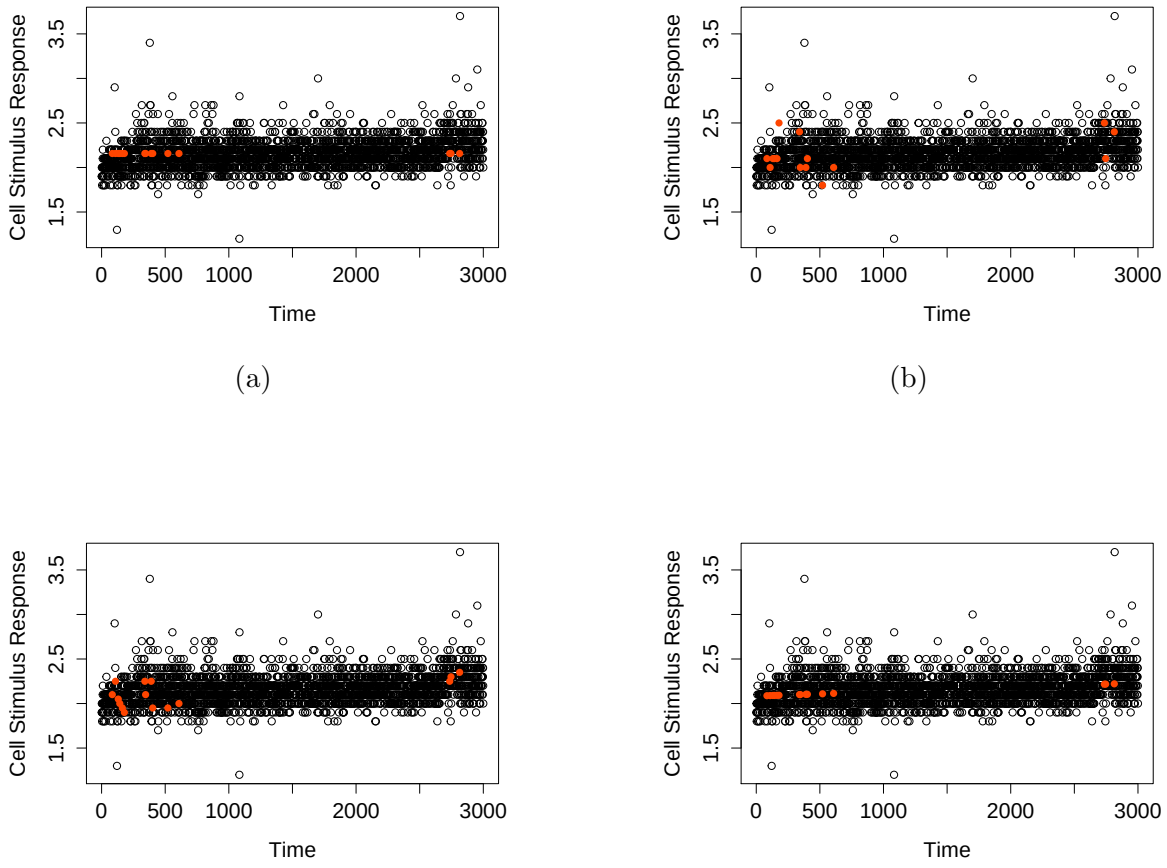


Figure 3.3: Scatter Plots of Cell Stimulus Response data with MCAR mechanism, using four single imputation methods missing values imputed and highlighted in orange colour; *Mean Substitution* in (a), *Hot Deck* in (b), *Last and next average* in (c), and *Regression Imputation* in (d).

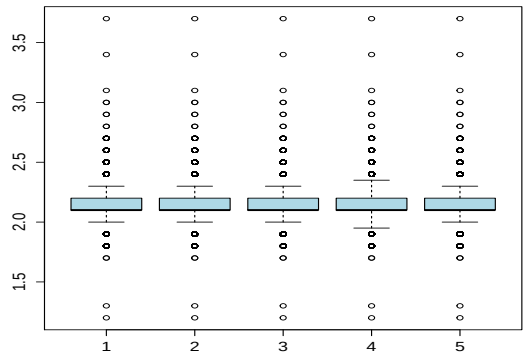


Figure 3.4: Boxplot for different imputation methods of cell stimulus response data. Observed data in (1) and imputed data using *Mean substitution* in (2), *Hot Deck* in (3), *Last and next average* in (4) and *Regression imputation* in (5).

It appears from Figure 3.4 that all imputation methods are not imputed missing values correctly, hence their boxplots behave as the same as the observed data when deleted the missing values, whereas *Last and next average* method respond differently as it includes imputed missing values correctly. It is also evident the presence of outliers, as shown in Figure 3.4.

3.4 Real Data Analysis

In this section, we apply all of the aforementioned tests to the cell stimulus response data for detecting change points in the mean and regression structures. The cell stimulus response data were provided by the animal physiology group of the University of Kaiserslautern. The data contains the reaction of one brain cell, related to the processing of acoustic signals, to repeated electrical stimuli. In this experiment, the stimulus was applied 50 times per second (frequency 50 Hz) over a period of 1 minute, i.e., 3000 times in total. We only use the delays, i.e., the waiting times between stimulus and the onset of the response. A change point analysis of the complete response modelled as functional data, has been done by Nyarige [105]. The delay already contains important information on the cell's reaction to the stimulus. In particular, it would be plausible if the delays become larger over time as the cell becomes accustomed to always the same stimulus and/or there is onsetting 'fatigue' over the course of the experiment which, with such a high frequency, puts considerable stress onto the cell.

The data have a sample size of 3000 with two variables: *Cell Stimulus Response*, exhibited by the delay, and *Time*. As a time unit we use observation number, i.e., the unit time is $\frac{1}{50}$ second. The data contains 15, i.e., few, missing values which is rather good for such delicate physiological experiments on the individual all level. The 15 missing values are imputed by the *Last and next average* method, which is the most appropriate method according to missing data mechanism and distribution described in the Section 3.3.3.

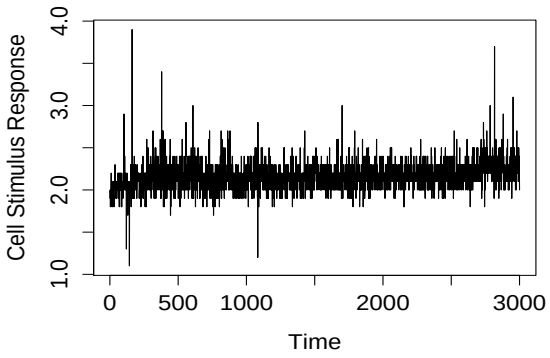
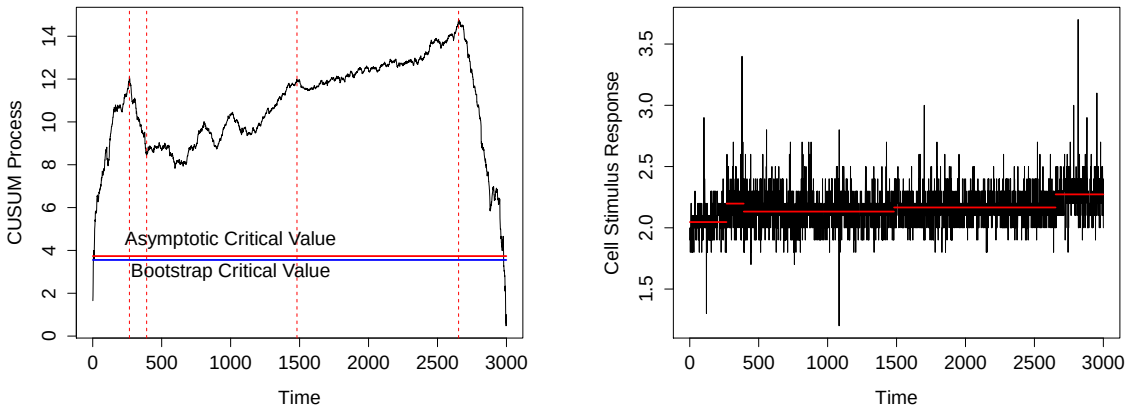


Figure 3.5: The Cell Stimulus Response Data.

There is no consistent trend (upward or downward) over the entire time span and the data sequence appears on a first glance stationary with constant variance, as shown in Figure 3.5. The Shapiro-Wilk normality test, for details see [129], gives 0.92363 with p-value $< 2.2e-16$ and suggests that the data is non-Gaussian. An offline nonparametric change detector hence seems an appropriate tool to use for analysis.

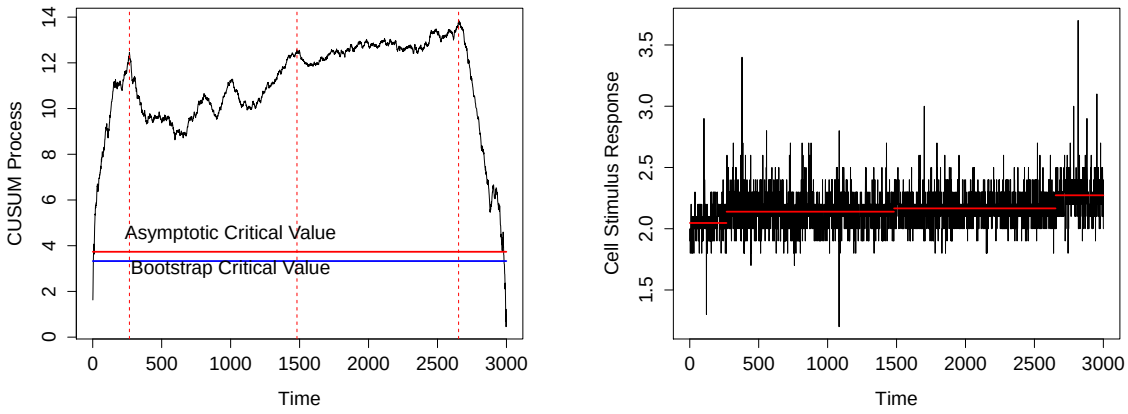
3.4.1 Changes in Mean

The classical change point model (3.46) with multiple change points in mean, assumes an observed sequence of independent random variables $X_1, \ldots, X_n$. In cell stimulus response data, consider X_i as the *Cell Stimulus Response* variable. Therefore, multiple change points in mean can be detected by the CUSUM test (3.4) and CUSUM R-type test (3.14) integrated within the binary segmentation method, MOSUM test (3.20), and MOSUM R-type test (3.24).



(a) (b)

Figure 3.6: CUSUM statistic process (3.4) for detecting multiple change points in the model (3.46) with the asymptotic (red) and bootstrap (blue) critical values are presented by horizontal lines, and location of change points are indicated by dotted vertical red lines in (a); and the estimated means are located in (b).

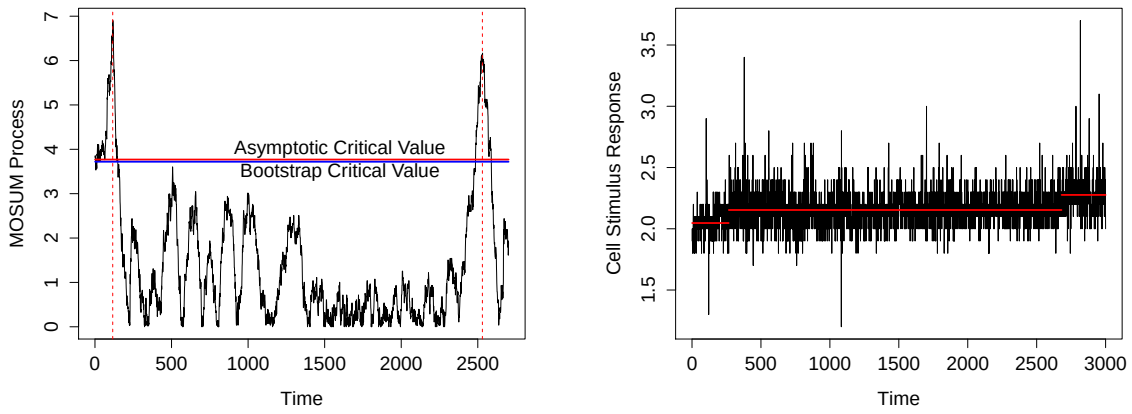


(a) (b)

Figure 3.7: CUSUM R-type statistics process (3.14) for detecting multiple change points in the model (3.46) with the asymptotic (red) and bootstrap (blue) critical values are presented by horizontal lines, and location of change points are indicated by dotted vertical red lines in (a); and the estimated means are located in (b).

Figure 3.6 represents *four* change points using the CUSUM test statistic (3.4) and binary segmentation. The first change point is located at 2653, hereafter, the detected change points are observed at 266, 1481 and 391 data points, respectively. On

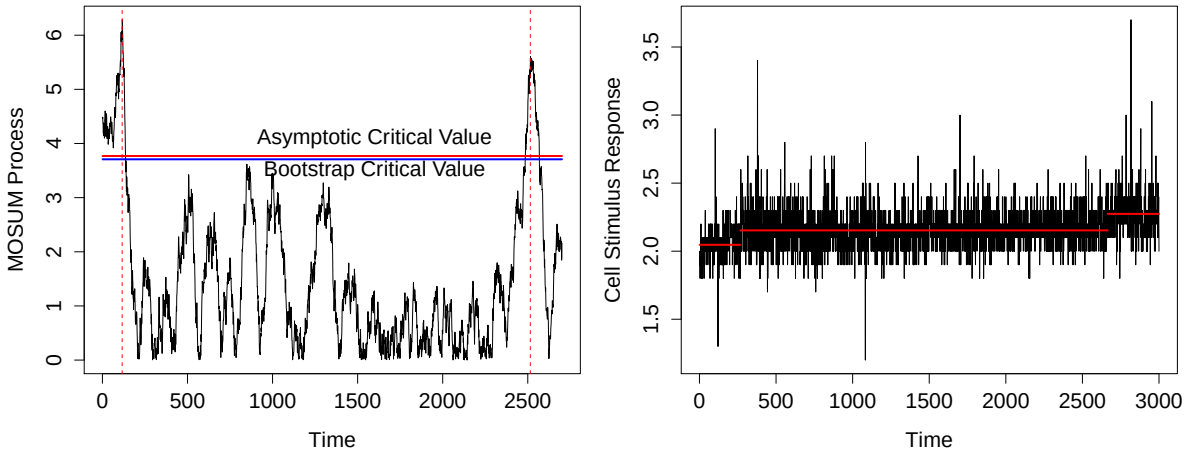
the other hand, Figure 3.7 shows the sequence of observations along with the time at which *three* change points are detected by the robust test CUSUM R-type statistic (3.14). The first detected change in mean at 2653, the second change detected in mean at 266 and the third change at 1481 data points. Hence, the CUSUM test overestimates changes in mean for the model (3.46).



(a) (b)

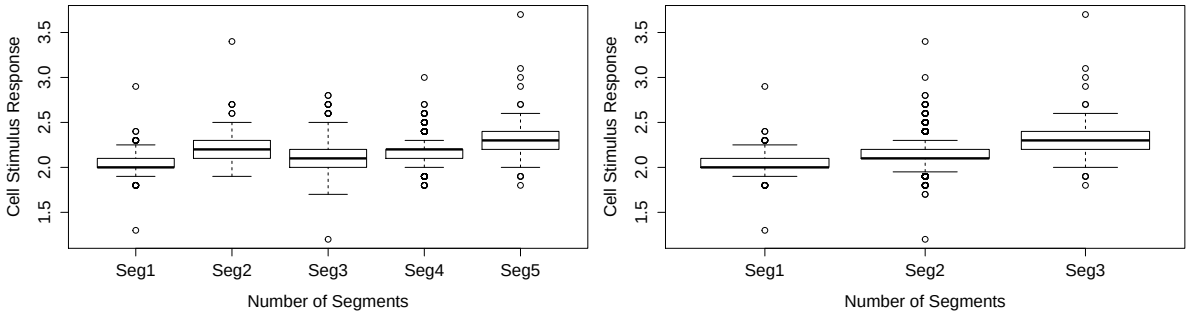
Figure 3.8: MOSUM process (3.20) for detecting multiple change points in the model (3.46) using bandwidth $G = 150$ and $\hat{\sigma^2}$ in (3.6) with the asymptotic (red) and bootstrap (blue) critical values are presented by horizontal lines, and location of change points are indicated by dotted vertical red lines in (a); and the estimated means are located in (b).

Deploying the MOSUM test (3.20) on the sequence of cell stimulus response, the sequence of random variables which undergoes a change in mean after the 266th and 2680th observations, as shown in Figure 3.8. Figure 3.9 illustrates three types of change in structure of mean with the associated *two* change points identified by using MOSUM R-type statistic (3.24). The first identified change point is at 2665, hereafter, at 266.



(a) (b)

Figure 3.9: MOSUM R-type statistics process (3.24) for detecting multiple change points in the model (3.46) using bandwidth $G = 150$ and $\hat{\sigma^2}$ in (3.6) with the asymptotic (red) and bootstrap (blue) critical values are presented by horizontal lines, and location of change points are indicated by dotted vertical red lines in (a); and the estimated means are located in (b).



(a) (b)

Figure 3.10: Boxplot for different segments of cell stimulus response data and five segments are observed using CUSUM process (3.4) in (a), and three segments are observed using MOSUM process (3.20) in (b).

From Figure 3.10, it is apparent that there is more heterogeneity in structure of means using the MOSUM test than that of the CUSUM test. Finally, the analysis reveals that two change points $\{266, 2665\}$ detected by the MOSUM R-type statistic (3.24)

for the model (3.46) are the best identified change points according to the resulting segments characteristics. Hence, the estimated model is,

$$\hat{X}_i = \begin{cases} 2.0468, & 1 \leq i \leq 266, \\ 2.1528, & 266 < i \leq 2665, \\ 2.2740, & 2665 < i \leq n. \end{cases} \tag{3.53}$$

This result is quite in line with a careful visual inspection of the data which implies a rather constant level over the major part of the time series with lower values at the beginning and higher values at the end. A biological explanation for the changes would be that after a starting phase, the cell becomes accustomed to the repeated stimuli and does not react so fast. At the end, fatigue sets in, which leads to another increase in delay.

3.4.2 Changes in Regression Coefficients

It is evident to apply model adequacy checking before using any model for inference and prediction purposes. Therefore, we look at regression diagnostics of the model (3.48) prior to use any test for detecting change points in that model.

Figure 3.11 exhibits the diagnostics for regression model (3.48). Figure 3.11(a) shows residuals are independent and it is also apparent from Figure 3.11(b) that there is no significant spike in the ACF plot. Linearity and homoscedasticity assumptions can be diagnosed by a plot of residuals versus fitted values, which is shown in Figure 3.11(c), where the points are distributed around horizontal line with a roughly constant variance. Therefore, assumptions of the linearity relationship between dependent and independent variables, and homoscedasticity (constant variance) of the errors are fulfilled. In Figure 3.11(d), notice the points fall along a line in the middle of the Normal Q-Q plots but curve off in the extremities, which indicates a heavy-tailed residuals distribution, i.e., significant non-normality. As a result, an offline nonparametric test seems appropriate in this case.

Therefore, we are motivated to use nonparametric classical change point methods OLS-CUSUM (3.37), OLS-MOSUM (3.39), Recursive CUSUM (3.41) and Recursive MOSUM (3.43) tests to identify the change point in the regression model (3.48).

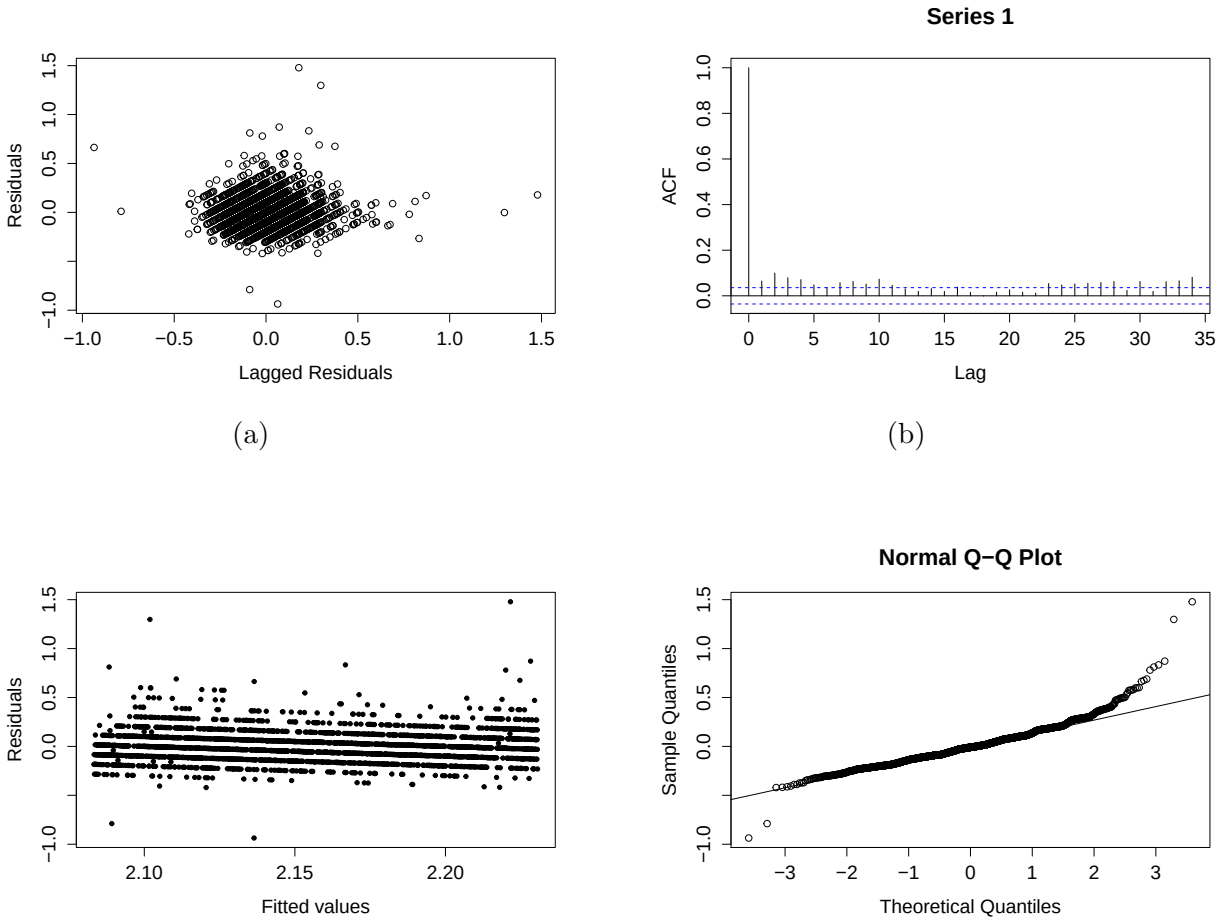


Figure 3.11: Model Adequacy Checking for the model (3.48): for *Independency*, current residuals versus lagged residuals are plotted in (a) and autocorrelation function (ACF) of residuals at different lags are in (b); *Homoscedasticity* assumption is verified by plotting residuals against predicted values in (c); and *Normality* assumption in (d).

In Figure 3.12, the nonparametric test processes illustrate the significance of change points in the model (3.48) with the asymptotic (red horizontal line) and bootstrap (blue horizontal line) critical values, the significant change points are also indicated by dotted vertical red lines. In all the cases bootstrap critical values yield small values than the asymptotic critical values, for instance, the asymptotic critical value for the OLS-CUSUM (3.37) test is 1.358 and the bootstrap one is 1.196. The best estimate of the change point location is the maximum value of test statistic. OLS-CUSUM (3.37) provides two significant change points at $\hat{m}_1 = 2653$ and $\hat{m}_2 = 266$ (in order of significance), OLS-MOSUM (3.39) test also indicates two change points with

bandwidth $h = 0.3$ but in different locations, i.e., 267 and 1753. Recursive CUSUM (3.41) unable to detect a reasonable number of change points, it detects more than 28 changes, e.g., at 2649, 2451, 2447, 2445, 2441, etc., and Recursive MOSUM (3.43) identifies two changes at 532 and 1147 data location. According to the significance of estimated regression coefficients in each segments, the OLS-CUSUM (3.37) test seems more reasonable in identifying the changes. In particulars, it results in similar change points as in (3.53).

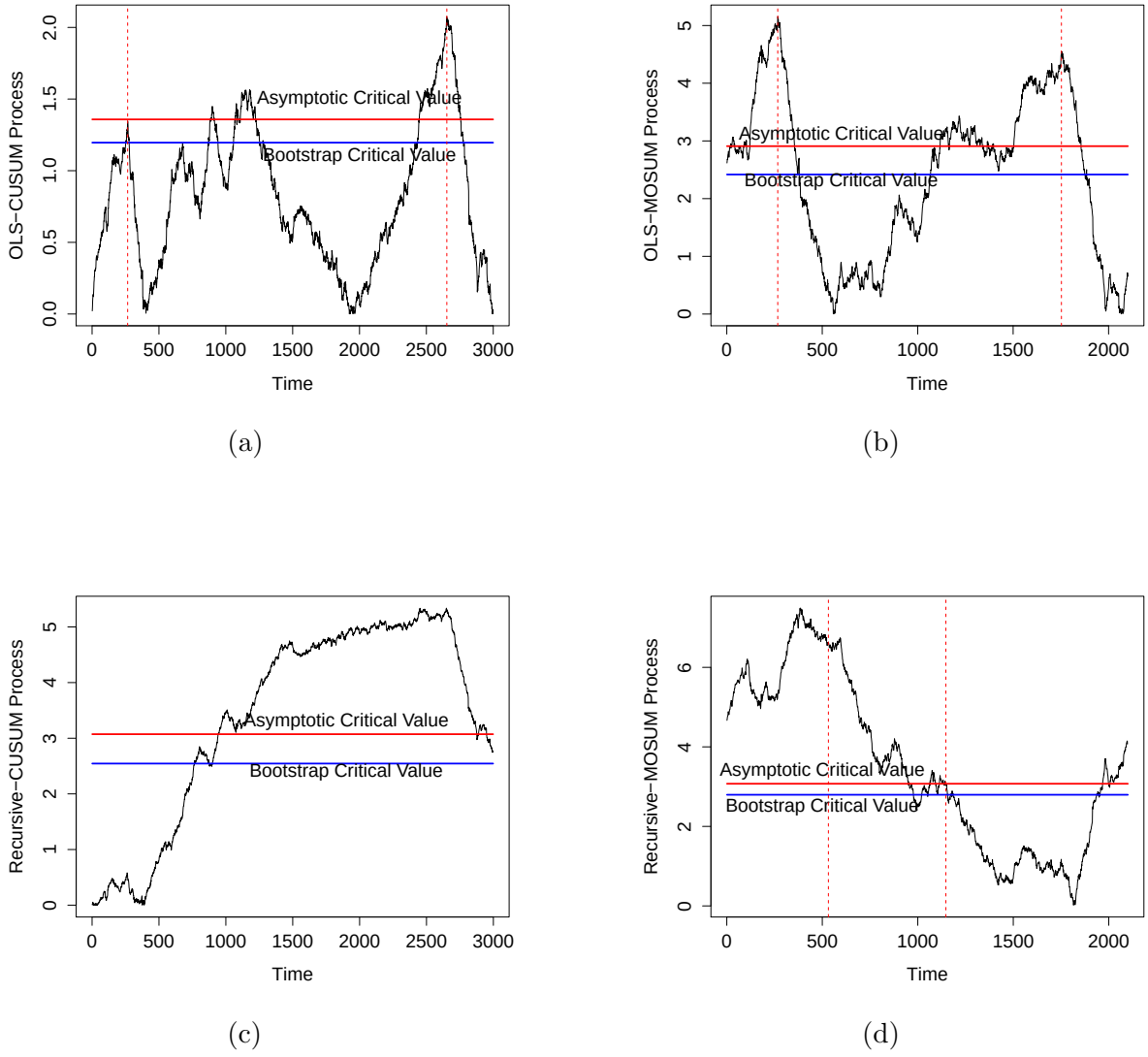


Figure 3.12: Detection of change points in the model (3.48) using (a) OLS-CUSUM test statistic (3.37), (b) OLS-MOSUM test statistic (3.39) with bandwidth $h = 0.3$, (c) Recursive CUSUM test statistic (3.41), and (d) Recursive MOSUM test statistic (3.43) with bandwidth $h = 0.3$, and the location of change points are indicated by dotted vertical red lines.

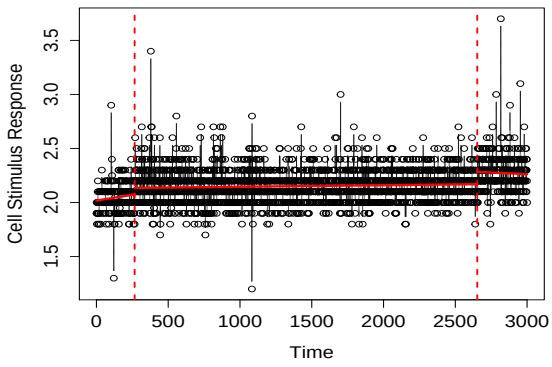


Figure 3.13: Detection of change points in the model (3.48)

Figure 3.13 visualizes three fitted regression lines in red colour due to two estimated change points at $\hat{m}_1 = 2653$ and $\hat{m}_2 = 266$ (in order of significance) using the OLS-CUSUM (3.37) test. The intercept and slope coefficients are estimated as $\hat{\beta}_{11} = 2.018$ and $\hat{\beta}_{21} = 0.643$ before the change point $\hat{m}_2 = 266$, in the 2nd segment the estimates are $\hat{\beta}_{12} = 2.131$ and $\hat{\beta}_{22} = 0.044$, and finally after the change point $\hat{m}_1 = 2653$ estimated coefficients are $\hat{\beta}_{13} = 2.438$ and $\hat{\beta}_{23} = -0.176$. All the coefficients in each segments are significant with p-value less than 0.01 except coefficient $\hat{\beta}_{23}$ is insignificant.

From the analysis of the more complicated model (3.48), we get, at least with some methods, a similar result as far the simple change-in-the-mean setting. In the central part of the data, the mean is rather constant (note that $\hat{\beta}_{22} \approx 0$), whereas we have a different behaviour at the start and end of the experiment. This corresponds to the visual impression presented by the data. The positive slope $\hat{\beta}_{21} = 0.643$ at the start may indicate that the effect of the cell getting accustomed to the stimuli seems to be gradual and not sudden as implied by a pure change-in-the-mean model. After the initial phase a state of saturation is achieved, which only is left at the end of the experiment. Whether the negative slope in that part of the data is reliable seems doubtful. It may be an effect of a few very large observations soon after the second change point. As a next step, we also have tried to find the changes in the first

difference regression model (3.48).

$$\begin{aligned} Y_i - Y_{i-1} &= \beta_0 + \beta_1 x_i + e_i - \beta_0 - \beta_1 x_{i-1} - e_{i-1} \\ (Y_i - Y_{i-1}) &= \beta_1 (x_i - x_{i-1}) + (e_i - e_{i-1}) \\ \Delta Y_i &= \beta_1 \Delta x_i + \Delta e_i. \end{aligned}$$

Here, $\Delta e_i = e_i - e_{i-1}$ is a moving average process of order 1, i.e., MA(1). Therefore, it may have dependency structure. Since the error e_i is unobserved, so we can consider their differences also as a new unobservable error $u_i = \Delta e_i$. At first, we assume that new errors u_i are also independent of explanatory variables and i.i.d. random variables with assumption (3.28) in the model (3.54).

$$\Delta Y_i = \begin{cases} \beta_1 \Delta x_i + u_i, & 1 \leq i \leq m, \\ \beta_1^* \Delta x_i + u_i, & m < i \leq (n-1). \end{cases} \tag{3.54}$$

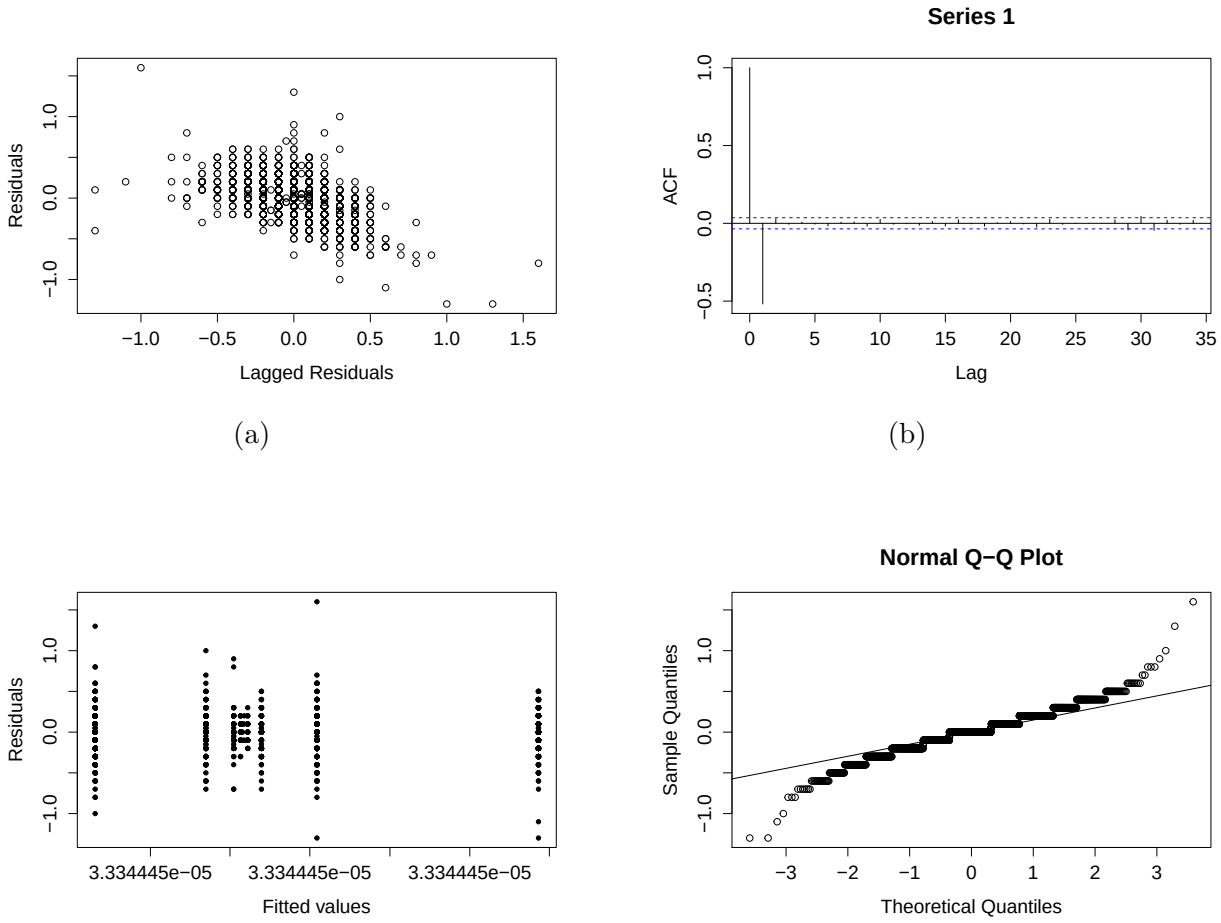


Figure 3.14: Model Adequacy Checking for the model (3.54): for *Independency*, current residuals versus lagged residuals are plotted in (a) and autocorrelation function (ACF) of residuals at different lags are in (b); *Homoscedasticity* assumption is verified by plotting residuals against fitted values in (c); and *Normality* assumption in (d).

It is apparent from Figure 3.14 that autocorrelation arises in the residuals of the model (3.54) with homoscedastic and non normality characteristics. One significant spike of ACF is observed in Figure 3.14(b). Hence, there is a violation of the assumption (3.28). At this stage nonparametric change point tests mentioned in Section 3.1.2 are no longer feasible for dependent error structure. Figure 3.15 reflects that infeasible conditions of those tests for the model (3.54).

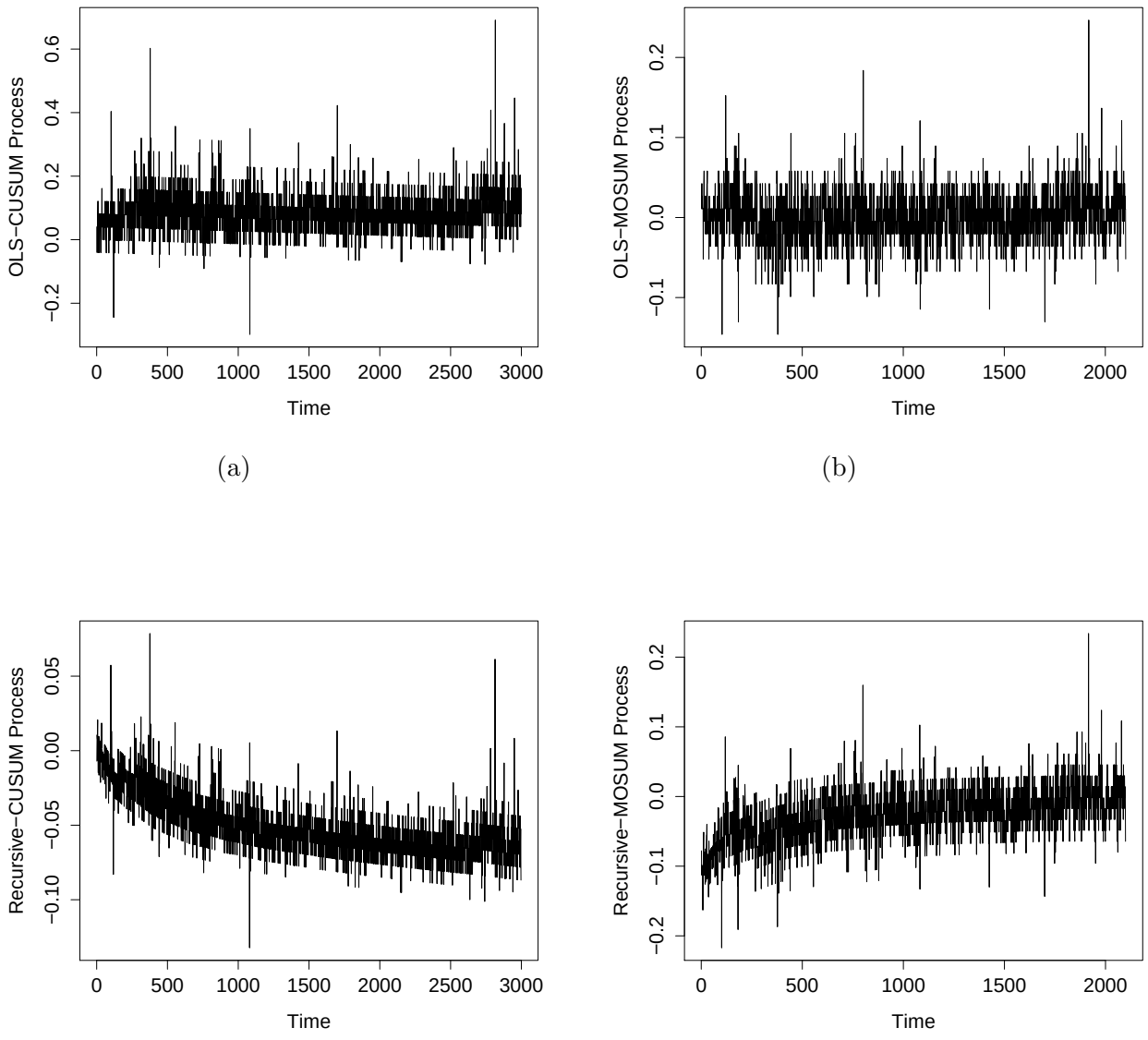


Figure 3.15: Detection of change points in the model (3.54) using (a) OLS-CUSUM test statistic (3.37), (b) OLS-MOSUM test statistic (3.39), (c) Recursive CUSUM test statistic (3.41), and (d) Recursive MOSUM test statistic (3.43).

Therefore, we are motivated to use a CUSUM test for dependent errors proposed by Gombay [51]. She assumes $\{u_i\}$ are dependent errors in the model (3.54) and described by the linear relationship,

$$u_i = \sum_{j=0}^{\infty} a_i \eta_{i-j}, \quad i = 1, 2, \ldots, n, \tag{3.55}$$

where η_j are independent identically random variables with mean zero, variance σ_η^2, $0 < \sigma_\eta^2 < \infty$, and for some $\nu > 2$ satisfying the condition $E|\eta|^\nu < \infty$; and for the $\{a_j\}$ sequence of constants satisfying $a_j = O(\gamma^j)$, $j \to \infty$ for some γ, $0 < \gamma < 1$.

She proposed the following CUSUM test statistic

$$Z_n = \max_{1<k\leq n} \left| \left(\frac{n}{k(n-k)} \right)^{-1/2} \sum_{i=1}^{k} \left(y_i - \bar{y}_i - \hat{\beta}_n (x_i - \bar{x}_n) \right) \right|, \tag{3.56}$$

with the limit distribution of (3.56) is

$$\lim_{n\to\infty} P \left\{ \frac{1}{\sigma_n} a(\log n) Z_n \leq t + b(\log n) \right\} = exp\left(-2e^{-t}\right), \tag{3.57}$$

for some $\sigma_n > 0$, where $t \in R^1$, $a(x) = (2\log t)^{1/2}$ and $b(x) = 2\log t + \frac{1}{2}\log\log t - \frac{1}{2}\log\pi$, for more details see [51]. The estimation of σ_n^2 in (3.57) can be explained by the long-run variance estimation concept. A long-run variance is a measure of the standard error of the sample mean when there is serial dependence. The long-run variance of a time series measures its total serial dependence and heterogenity, and it typically enters into the limiting distribution of a statistic in many testing procedures.

Gombay [51] proposed a long run variance estimator of σ_n for the error process approximated by a moving average of order q MA(q) by,

$$\hat{\sigma}_n^2 = \text{var}\left(\sum_{i=1}^{n} \hat{u}_i \right) \cong \sum_{i=1}^{n} \hat{\gamma}_0 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{(i+q)\bigwedge n} \hat{\gamma}_{i-j}, \tag{3.58}$$

where $\hat{\gamma}_h$ covariances are estimated by the usual formula for $h = 0, \ldots, q$ as

$$\hat{\gamma}_h = \frac{1}{n} \sum_{k=1}^{n-h} (\hat{u}_k - \bar{\hat{u}}_n)(\hat{u}_{k+h} - \bar{\hat{u}}_n), \quad \bar{\hat{u}}_n = \frac{1}{n} \sum_{k=1}^{n} \hat{u}_k.$$

The estimator (3.58) is very close to Bartlett's estimator with uniform weights. However, the long run variance estimator (3.58) gives negative value, e.g., -4.99, for the model (3.54) using cell stimulus response data. Hence, we are interested to express the long-run variance in the context of spectral density.

The spectral density can be used to represent $\{\gamma(j) : j = 0, 1, \ldots\}$, the sequence of autocovariances, of a covariance stationary process. The spectral density is defined

as

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-i\lambda j} \\ &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) \cos(\lambda j), \end{aligned}$$

where $i = \sqrt{-1}$. The spectral density is real valued, since $\gamma(j) = \gamma(-j)$. The spectral density is symmetric around zero, since $\cos(\lambda j) = \cos(-\lambda j)$. Furthermore, since cos is a periodic function with the period 2π, the range of values of the spectral density is determined by the values of $f(\lambda)$ for $0 \leq \lambda \leq \pi$.

An autocovariance function with $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ can be expressed by means of the spectral density,

$$\gamma(j) = \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda j} d\lambda.$$

Hence,

$$\gamma(0) = \int_{-\pi}^{\pi} f(\lambda) d\lambda.$$

Thus, the area under the spectral density function of X_t between $-\pi$ and π gives the variance of X_t. The argument λ of $f(\lambda)$ is called the angular frequency. Notice, if $\{X_t\}$ is covariance stationary with absolutely summable autocovariances, the long-run variance is determined by the spectral density at the zero frequency.

$$\begin{aligned} \omega_X &= \lim_{n \to \infty} Var \left(\frac{\sum_{t=1}^{n} X_t}{\sqrt{n}} \right) \\ &= \sum_{h=-\infty}^{\infty} \gamma(h) \\ &= 2\pi f(0). \end{aligned}$$

In the following, we usually exclude the degenerate situation of $\sum_{h=-\infty}^{\infty} \gamma_X(h) = 0$, without mentioning it explicitly. Theorem 3.4.1 (confer Theorem 2 of Marmer [91]) illustrates how linear (MA) transformations of a covariance stationary process affect the spectral density and long-run variance.

Theorem 3.4.1. *Let $\{X_t\}$ be a covariance stationary process with the autocovariance function γ_X such that $\sum_{j=-\infty}^{\infty} |\gamma_X(j)| < \infty$. Define $Y_t = \sum_{j=0}^{\infty} c_j X_{t-j}$, where $\sum_{j=0}^{\infty} c_j^2 < \infty$. Then $\{Y_t\}$ is covariance stationary and its density is given by $f_Y(\lambda) = \left| \sum_{j=0}^{\infty} c_j e^{-i\lambda j} \right|^2 f_X(\lambda)$, where f_X is the spectral density of $\{X_t\}$.*

Proof of Theorem 3.4.1 is available in [91], here we only give a sketch of the proof. At first, we need to show that $cov(Y_t, Y_{t-h})$ is independent of t. Then this covariance can be shown as bounded using the assumptions $\sum_{j=0}^{\infty} c_j^2 < \infty$ and $\sum_{j=0}^{\infty} |\gamma_X(j)| < \infty$. Therefore, $\{Y_t\}$ is covariance stationary. Now, using the value of $\sum_{j=0}^{\infty} c_j e^{-i\lambda j}$ and its complex conjugate, we get $\left(\sum_{j=0}^{\infty} c_j e^{-i\lambda j}\right)\left(\sum_{j=0}^{\infty} c_j e^{i\lambda j}\right) = \left|\sum_{j=0}^{\infty} c_j e^{-i\lambda j}\right|^2$. Hence, we can establish $f_Y(\lambda) = \left|\sum_{j=0}^{\infty} c_j e^{-i\lambda j}\right|^2 f_X(\lambda)$.

Suppose that $\{X_t\}$ is covariance stationary and purely in-deterministic. Then it has the MA(∞) representation

$$X_t = \sum_{t=0}^{\infty} a_j \varepsilon_{t-j},$$

where $\{\varepsilon_t\}$ is white noise, and $\sum_{t=0}^{\infty} a_j^2 < \infty$. Let $\text{Var}(\varepsilon_t) = \sigma^2$. Since the spectrum of a white noise process is flat

$$f(\lambda) = \frac{\sigma^2}{2\pi} \text{ for all } \lambda.$$

Using Theorem 3.4.1, the spectral density of $\{X_t\}$ exists and satisfies

$$f_X(\lambda) = \frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{\infty} a_j e^{-i\lambda j}\right|^2,$$

and the long-run variance of $\{X_t\}$ is,

$$\begin{aligned} \omega_X &= 2\pi f_X(0) \\ &= \sigma^2 \left(\sum_{j=0}^{\infty} a_j\right)^2. \end{aligned} \tag{3.59}$$

Now, for the MA(1) process,

$$X_t = \varepsilon_t - \varepsilon_{t-1}, \tag{3.60}$$

we have $a_0 = 1$ and $a_1 = -1$. Hence, the long run variance from (3.59) is

$$\begin{aligned} \omega_X &= \sigma^2(1-1)^2 \\ &= 0. \end{aligned}$$

This corresponds to the degenerate case mentioned above as we have $\gamma_x(0) = 2\sigma^2$, $\gamma_x(1) = cov(\varepsilon_t - \varepsilon_{t-1}, \varepsilon_{t-1} - \varepsilon_{t-2}) = -\sigma^2 = \gamma_x(-1)$ and $\gamma_x(j) = 0, |j| \geq 2$. The long-run variance is plausibly 0 as

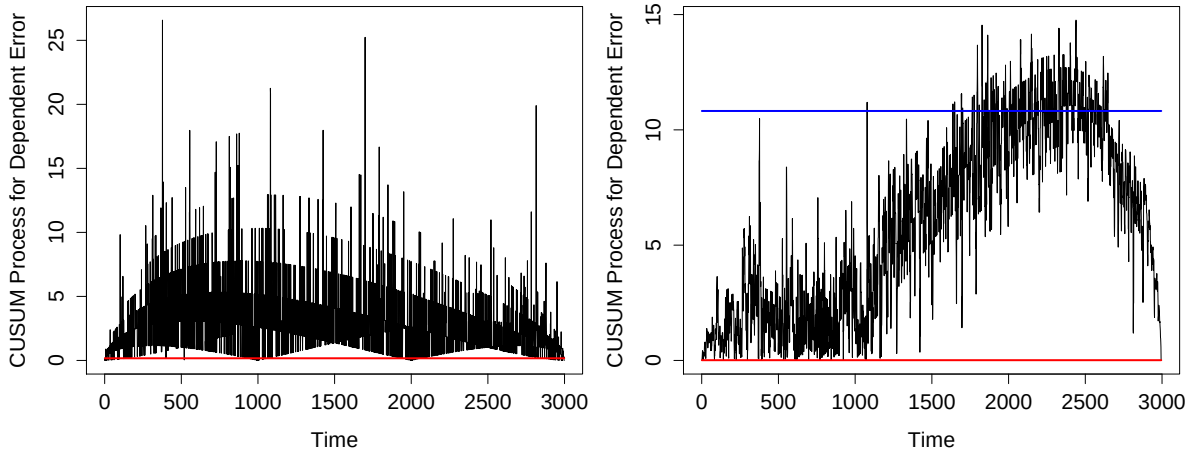
$$\begin{aligned}\sum_{t=1}^{n} X_t &= (\varepsilon_n - \varepsilon_{n-1}) + (\varepsilon_{n-1} - \varepsilon_{n-2}) + \cdots + \varepsilon_1 - \varepsilon_0 \\ &= \varepsilon_n - \varepsilon_0,\end{aligned}$$

which has the variance $2\sigma^2$ independently of n. Due to Brockwell and Davis [22] the spectral density for MA(1) process, $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, is

$$f(\lambda) = \frac{\sigma^2}{2\pi}(1 + 2\theta\cos\lambda + \theta^2). \tag{3.61}$$

Using (3.61) the spectral density for the MA(1) process (3.60) with $\theta = -1$ satisfies also $f(0) = 0$.

Gombay [51] also introduced a smoothing of the process when the time series is dominated by high frequencies, using a moving average filter. In this smoothing y_i, $i = 1, \ldots, n$ is replaced by $(y_i + y_{i-1} + y_{i-2} + y_{i-3})/4$, $i = 4, \ldots, n$. This transformation has no effect on the linear regression parameters [51].



(a) (b)

Figure 3.16: Change point detection in the model (3.54) deploying Gombay's CUSUM test (3.56) for dependent errors without smoothing the data in (a) and with smoothing the data in (b).

Figure 3.16(a) represents a process with high frequencies and the result is far too many

rejections, also in this case the long run variance (3.58) does not exist, so only the critical value is calculated using least squares variance from (3.57) and shown in the red line. Figure 3.16(b) shows that smoothing reduces high frequencies and the long run variance (3.58) produces critical value 10.8189 from (3.57), which is illustrated in the blue line. There is evident of significant change points, as shown in Figure 3.16(b).

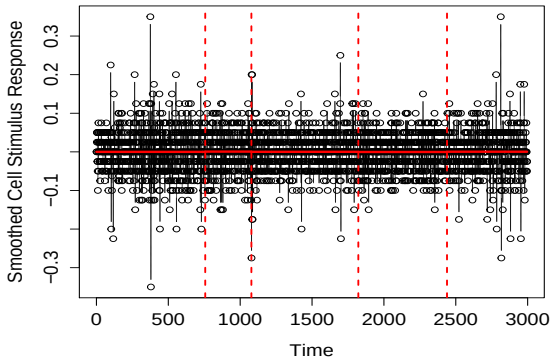


Figure 3.17: Detection of change points in the model (3.54) with the smoothed cell stimulus response data.

Five fitted regression lines are presented by red lines in Figure 3.17, which are estimated from the five segments of the smoothed cell stimulus response data due to detected four significant change points at $\hat{m}_1 = 2439$, $\hat{m}_2 = 1822$, $\hat{m}_3 = 1078$ and $\hat{m}_4 = 757$ (in order of significance). The slope coefficients are estimated for the model (3.54) with the smoothed data as $\hat{\beta}_1 = -4.324e-14$, $\hat{\beta}_2 = 1.869$, $\hat{\beta}_3 = 0.403$, $\hat{\beta}_4 = -0.851$, and $\hat{\beta}_5 = 1.077$, respectively, in each segment. However, all the coefficients are insignificant.

3.4.3 Changes in Distribution

Testing whether a change in the distribution occurs within the sample is analytically beneficial, as this procedure does not consider any model. Therefore, this testing procedure is free from the model misspecification problem. Once we detect a change occurs in the distribution, then we need to estimate the change point instant after which the distribution of the observations switches from one distribution to another different distribution. Wilcoxon-type statistics and process based on U-statistics

are commonly used to identify the change in distribution in nonparametric context. Csörgo and Horváth [35] developed a nonparametric test based on Wilcoxon-type statistics using quantile score statistic to detect a change in distribution. They considered $X_1, X_2, \ldots, X_n$ as independent observations and defined the test hypothesis as

$$\begin{aligned} &H_0 : X_1, \ldots, X_n \quad \text{are i.i.d. random variable with a continuous distribution function} \\ &\text{against the alternative} \\ &H_1 : \text{there is an unknown integer } k^*, 1 \leq k^* < n, \text{ such that } P\{X_1 \leq t\} = \cdots = P\{X_{k^*} \leq t\} \\ &\quad \text{and } P\{X_{k^*+1} \leq t\} = \cdots = P\{X_n \leq t\} \text{ for all } t, \text{ and } P\{X_{k^*} \leq t_0\} \neq P\{X_{k^*+1} \leq t_0\} \\ &\quad \text{for some } t_0. \end{aligned} \tag{3.62}$$

They used quantile score function to develop the test statistic for detecting change in distribution, where the empirical quantile function was defined by

$$Q_n(y) = \inf\{t : F_n(t) \geq y\}, \quad 0 < y < 1,$$

using the empirical distribution function

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq t), \quad -\infty < t < \infty.$$

They also defined

$$\Psi_t(x) = \begin{cases} -(1-t) & x \leq 0, \\ t & x > 0, \end{cases}$$

and then used the following process to develop different kind of tests

$$Y_n(s,t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{[ns]} \Psi_t(X_i - Q_n(t)), \quad 0 < s, t < 1.$$

Csörgo and Horváth [35] developed eight kinds of test statistics using the process $Y_n(s,t)$ to detect the change in distribution with i.i.d. observations and also derived their asymptotic distributions under the null hypothesis, which follows either a Brownian bridge, or a two-parameter Gaussian process, or the standard normal distribution. In this section, we use the following test statistic among eight developed tests of [35].

$$(t_0(1-t_0))^{-1/2} \sup_{0<s<1} |Y_n(s,t_0)| \xrightarrow{\mathscr{D}} \sup_{0<s<1} |B(s)|, \tag{3.63}$$

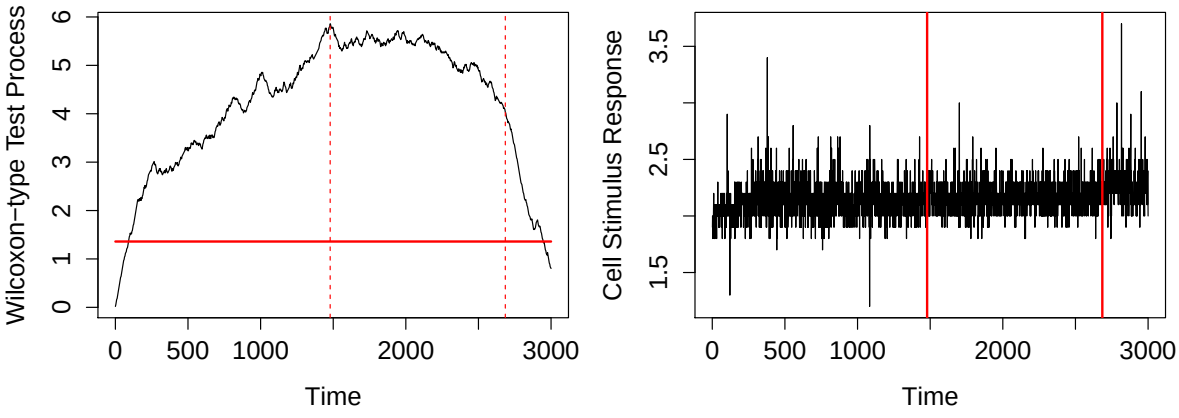
where $\{B(s), 0 \leq s \leq 1\}$ is a Brownian bridge and we assume $t_0 = 0.5$. The critical values are obtained from the well-known identity

$$P\left\{\sup_{0<s<1}|B(s)| > b\right\} = 2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2b^2}, \quad b > 0,$$

which yields 1.63, 1.36 and 1.22 for $\alpha = 0.01, 0.05$ and 0.10, respectively. When the null hypothesis of (3.62) is rejected then we need to estimate the significant change point by

$$\frac{\hat{k}}{n} = \inf\left\{s : \sup_{0<s<1}|Y_n(s,t)| = \sup_{0<u<1}\sup_{0<t<1}|Y_n(u,t)|\right\},$$

where $\hat{k}$ is the estimate of the change point k^*. Now, we are in such a position to use the Wilcoxon-type rank test of [35] for detecting change point in the distribution of observations with missing data. Multiple change points can also be detected using Binary Segmentation Algorithm 3 for the test statistic (3.63).



(a) (b)

Figure 3.18: Wilcoxon-type rank test process (3.63) is proposed by Csörgo and Horváth [35] for detecting change point in the distribution for i.i.d. data with the asymptotic critical value, and the significant change point's locations are indicated by dotted vertical red lines in (a); and the estimated multiple change points are shown in (b).

Figure 3.18 illustrates that two significant change points are observed at 1479 and

2685 seconds for the cell stimulus responses at the 5% level of significance when we consider the missing data feature. Note that the procedure detects the change at the end of the signal, but not the one at the start. Instead, a change in the center is detected. We believe that this is an artificial and a weakness of binary segmentation. Recall that we had also a change point around 1500 in the change-in-the-mean model (confer Figure 3.7), but there the power of the test against a simple change-in-the-mean alternative at least enables the procedure to detect the change around 260, whereas the test of this section with broad change-in-distribution alternative has a too small power to achieve this. We conclude that the simultaneous multiple change point detection procedures like MOSUM in Figure 3.8 are preferable, at least for this particular data set.

Censored Data

As censoring is the inevitable characteristic of survival analysis, which is our main concern. Hence, we wish to test the change point in the distribution under censorship for the cell stimulus response data by considering all the missing data as censored. In this context, Gombay and Liu [52] developed a Wilcoxon-type rank test using the score function of Gehan [49] and Mantel [90] to detect the change in distribution for censored data, which we have discussed along with other tests in Section 1.3 of Chapter 1. Gombay and Liu [52] considered the test statistic as

$$\max_{1\leq k<n} \frac{|\sum_{i=1}^{k} U_i|}{\left(\sum_{i=1}^{n} U_i^2\right)^2} \xrightarrow{\mathscr{D}} \sup_{1<t<1} |B(t)|, \tag{3.64}$$

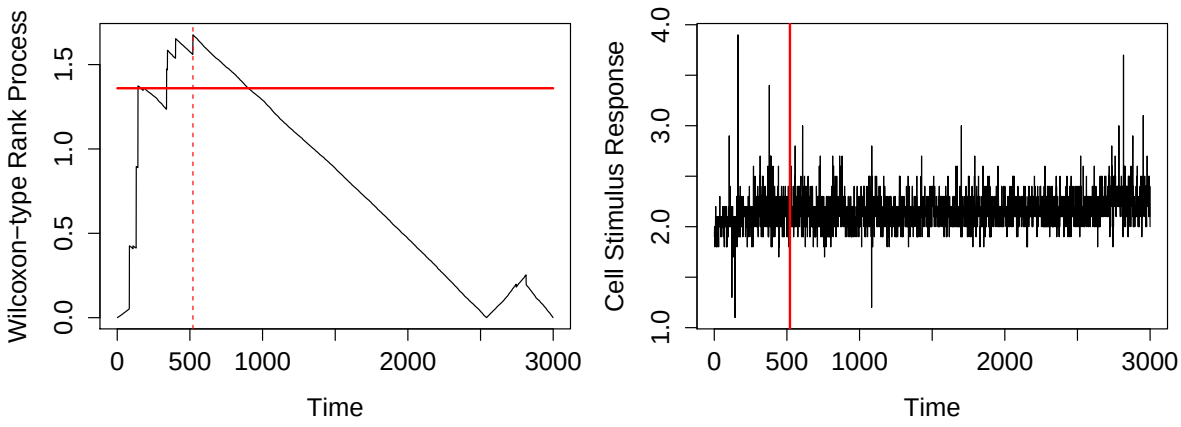
where the generalized rank of (X_i, δ_i) is defined as

$$U_i = \sum_{j=1}^{n} \left\{ I\left(X_i > X_j, \delta_j = 1\right) - I\left(X_i < X_j, \delta_j = 1\right) \right\}, \quad i = 1, \ldots, n.$$

They defined the estimator of the change point k^* as

$$\hat{\tau}_n = \hat{k^*} = \arg\max_{1\leq k<n} \frac{|\sum_{i=1}^{k} U_i|}{\left(\sum_{i=1}^{n} U_i^2\right)^2}.$$

The limiting distribution of the test statistic (3.64) follows a Brownian bridge, so the critical values are 1.63, 1.36 and 1.22 for $\alpha = 0.01, 0.05$ and 0.10, respectively, similar to the test (3.63). At this moment, we are interested for testing the significant change point in the distribution of the cell stimulus responses by considering the missing observations as censored using the test (3.64).



(a) (b)

Figure 3.19: Gombay and Liu [52] proposed Wilcoxon-type rank test process (3.64) for detecting change point in the distribution for censored data with the asymptotic critical value, and the significant change point is indicated by a dotted vertical red line in (a); and the estimated change point is shown in (b).

We conducted the test (3.64) at the 5% level of significance and found the significant change point at 521 seconds of the cell stimulus responses. Though we have utilized this test with the Binary Segmentation Algorithm 3 for detecting multiple change points, but found only one significant change point. Figure 3.19 deploys this testing procedure with the significant change point at 521 seconds.

3.5 Discussion

We have reviewed a number of offline nonparametric procedures for detecting change points with their application in presence of missing data feature in the mean and regression models. The cell stimulus response data have MCAR mechanism with time series pattern, so the most appropriate imputation method is *Last and next average* which was justified in Section 3.3.3. In the application of multiple change points detection in mean, we found that the CUSUM test (3.4) overestimated change points compared to the MOSUM test (3.20), whereas robust counterparts of those tests provide reasonable significant change points. In all cases the asymptotic critical values are bigger than the residuals bootstrap critical values. Similar results found in OLS-CUSUM, recursive CUSUM, OLS-MOSUM and recursive MOSUM tests for

multiple changes in linear regression models in case of independent errors. However, these tests were unable to identify the changes in first difference regression model (3.54) due to dependent errors. Using Gombay's CUSUM test (3.56) with smoothing technique we have observed four significant change points in the regression model (3.54). In Section 3.4.3, we also investigated change in the distribution for missing and censored data feature with the Wilcoxon-type rank test (3.63) and (3.64), respectively. Two significant change points are observed for missing data feature whereas only one significant change point is found with censored data feature. Hence, this is important to know and use the right data feature for investigating change point correctly. Therefore, this chapter provides a framework how to identify change points in mean, regression coefficients and distribution with missing data.

Chapter 4: Change Point in Hazard Function using U-statistic for Censored Data

Change point hazard models have been extensively investigated by many authors, but the literature on nonparametric classical change point problems with censoring is rather small. There are only a few classical change point papers dealing with detection of changes when only censored observations are available. Stute [136] considered an estimator of the change point based on U-statistics. Horváth [62] studied test procedures based on U-statistics too. Gombay and Liu [52] proposed and investigated limit properties of a nonparametric test based on ranks related to the Gehan-Wilcoxon statistic that can be expressed as a U-statistic. Aly [2] developed tests based on quantile functions and studied their limit behavior under the null hypothesis. All these papers considered censoring variables to be i.i.d. Hušková and Neuhaus [65] developed a test procedure as a generalization of two-sample rank tests under random censoring, where censoring variables were assumed to be independent but not necessarily identically distributed. A MOSUM type test was proposed by Komárková [77] considering i.i.d. censoring variables. All of the aforementioned tests focus on the change in distribution only, but not on the change in the hazard functions.

In Chapter 2, we considered changes of the hazard functions in individual survival time, i.e., all test purposes or objects have the same hazard function, but this shows some change at certain points after the time of origin. This is a suitable framework for data collected over moderate time span. The change points here represent different phases in, e.g., a disease with, e.g., initial high risk, then moderate risk in the later stages of the disease and finally remains small risk after recovery from the disease.

However, there may be a different kind of changes in the hazard function, in particular in data from long-term studies like the Breast Cancer data. Here, a change in

the survival chances may happen due to medical progress, e.g., due to a new drug. This would simultaneously affect all patients in the study independently of how much time has passed since their individual times of origin. This study is in line with classical change point analysis outside of survival analysis. We call such changes **change points in absolute time**, in contrast, to **change points in survival time**, which correspond to the setting of Chapter 2.

We are interested in testing and estimation of change points in absolute time of the hazard functions. There are few papers studying such a problem with censored data, and, as far our knowledge, all of them are based on specific parametric models. Therefore, we are motivated to develop a test based on U-statistics to detect change points in the hazard distribution which does not require any model, hence, is not affected by model misspecification errors.

The chapter is organized as follows. Section 4.1 structures the test hypothesis to detect change point in the hazard functions in absolute time with a brief introduction of Horváth [62]'s test for changes in distribution based on censored data. For finding an equivalent estimator of Horváth [62]'s estimator all the efforts are summarized in Section 4.2. Section 4.3 illustrates the asymptotic distributions of our test statistics under H_0 using symmetric kernels, where we also derive the weighted asymptotic of our test statistics in Section 4.3.2. In Section 4.4, we explain the asymptotic behavior of the proposed test statistics under H_1. Section 4.5 contains information regarding the limit distributions of our test statistics under H_0 for antisymmetric kernels. All of our developed test statistics are based on the estimator $\hat{\lambda}_2$ for symmetric and antisymmetric kernels, although we have found estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ as the equivalent estimators of Horváth's estimator. Therefore, some remarks are made on a change point test using $\hat{\lambda}_1$ in Section 4.6. Finally, the chapter concludes with some discussions in Section 4.7.

4.1 Test Hypothesis

For a sample of n independent individuals, let $T_1, T_2, \ldots, T_n$ be the independent and identically distributed (i.i.d.) survival times with probability density function (pdf) $f(t)$, which are right censored by the i.i.d. censoring random variables, $C_1, C_2, \ldots, C_n$, which are assumed to be independent of T_i. Let F and G be the distribution functions

of T_i and C_i, respectively. The observed right censored data are denoted by the pairs (X_i, δ_i), $i = 1, 2, \ldots, n$, where

$$X_i = \min\{T_i, C_i\}, \quad \delta_i = I\{T_i \leq C_i\} = I\{T_i = X_i\}. \tag{4.1}$$

Here, I is an indicator function and δ_i is a non-censoring indicator variable. The observed data has a pdf $h(x)$ and a distribution function H given by $1 - H = (1 - F)(1 - G)$.

Now we assume that the individuals enter the study successively, i.e., the corresponding times of origin t_i, $i = 1, \ldots, n$, are ordered $t_1 \leq t_2 \leq \cdots \leq t_n$. We are interested in testing if the survival times T_i are really identically distributed or if there is a change at some t_{k^*}. We want to construct a test which is based on estimates of the hazard function.

We are aware of the work of Stute [136], Horváth [62], Aly [2], and Gombay and Liu [52] only, where the authors investigated the change point in censored data by considering any change in the distribution function F which results in a change in the distribution function H and also incorporated the information contained in the δ's for censored observations. One might argue that a method designed for detecting a change from a sequence of completely observable data could be applied to the X's as well to detect a change in the F's. This is true in principle. On the other hand, such a procedure would necessarily not incorporate the information contained in the δ's and therefore lead to an inefficient procedure. However, they developed the procedure for detecting the change point in survival times distribution's but not in the hazard functions.

In survival analysis, it is more plausible to look for a change in hazard functions based on the survival times as well as censoring times. Nevertheless, most of the developed tests for change point in the hazard function considered theory either by discarding censored data and only considering the observable survival times or by modifying the likelihood function for censored data. Hence, we are interested to the following hypothesis for change in hazard functions

$$\begin{aligned} &H_0 : \lambda_1(t) = \lambda_2(t) = \cdots = \lambda_n(t), \quad \text{for all } t; \quad \text{against} \\ &H_1 : \text{there exist an unknown } k^*, 1 \leq k^* < n, \text{ such that } \lambda_1(t) = \cdots = \lambda_{k^*}(t), \\ &\quad \lambda_{k^*+1}(t) = \cdots = \lambda_n(t), \text{ for all } t, \text{ and } \lambda_{k^*}(t) \neq \lambda_{k^*+1}(t) \text{ for some } t. \end{aligned} \tag{4.2}$$

4.1.1 Horváth's Test for Changes based on Censored Data

Horváth [62] developed a test for detecting change in distribution of survival variables which is focused on the cumulative hazard function $\Lambda(x)$. We like to modify this test by using estimates for the hazard function $\lambda(x)$ directly. In this situation we first describe Horváth's approach. In contrast to Horváth, who considers arbitrary random variables, here $T_1, \ldots, T_n$ are survival times, and hence $T_i \geq 0$. Therefore, all integrals should start at 0, not $-\infty$. He considered the following functional of $\Lambda(x)$ with a notation $h(t-) = \lim_{s \uparrow c} h(s)$ which denotes the limit from the left-hand side

$$\theta = \int_0^V \int_0^V \tilde{K}(x,y) d\Lambda(x) d\Lambda(y) = \int_0^V \int_0^V \tilde{K}(x,y) \frac{dF(x)}{1-F(x-)} \frac{dF(y)}{1-F(y-)}, \quad (4.3)$$

where he assume that $\tilde{K}(x,y) = \tilde{K}(y,x)$ is a symmetric kernel function uniformly bounded on $[0,V]^2$, and $1 - H(t) = (1-F(t))\,(1-G(t))$, for some V satisfying

$$F(V) < 1 \quad \text{and} \quad G(V) < 1. \quad (4.4)$$

To test for a change in distribution, Horváth considered for all $1 \leq k < n^{\text{th}}$ statistic

$$\begin{aligned} \hat{\theta}(k) &= \int_0^V \int_0^V \tilde{K}(x,y) d\hat{\Lambda}_k(x) d\hat{\Lambda}^*_{k^*}(y) \\ &= \int_0^V \int_0^V \tilde{K}(x,y) \frac{dH_k(x)}{1-\hat{H}_k(x-)} \frac{dH^*_k(y)}{1-\hat{H}^*_k(y-)}, \end{aligned} \quad (4.5)$$

where H_k and $\hat{H}_k$ are the empirical distribution functions of respectively the censored and uncensored observations for all $X_1, \ldots, X_k$, and, H^*_k and $\hat{H}^*_k$ are the corresponding quantities based on the second subsample $X_{k+1}, \ldots, X_n$, which are defined as

$$\begin{aligned} H_k(x) &= \frac{1}{k} \sum_{1 \leq i \leq k} I\left\{X_i \leq x, \delta_i = 1\right\}, \\ H^*_k(x) &= \frac{1}{n-k} \sum_{k < i \leq n} I\left\{X_i \leq x, \delta_i = 1\right\}, \\ \hat{H}_k(x) &= \frac{1}{k} \sum_{1 \leq i \leq k} I\left\{X_i \leq x\right\}, \quad \text{and} \\ \hat{H}^*_k(x) &= \frac{1}{n-k} \sum_{k < i \leq n} I\left\{X_i \leq x\right\}. \end{aligned}$$

Under H_0, $\hat{\theta}(k)$ estimates θ for all $1 \leq k < n$. Note that

$$\hat{\Lambda}_k(x) = \int_0^x \frac{dH_k(t)}{1-\hat{H}_k(t-)}, \quad (4.6)$$

is the Nelson-Aalen estimator for the cumulative hazard function

$$\Lambda(x) = \int_0^x \frac{dF(t)}{1-F(t-)} = \int_0^x \frac{f(t)}{1-F(t-)}dt = \int_0^x \lambda(t)dt, \tag{4.7}$$

based on $X_1, \ldots, X_k$, and analogously $\hat{\Lambda}_k^*(x)$ for $X_{k+1}, \ldots, X_n$.

We intend to modify Horváth's test by using estimates of the hazard function $\lambda(x)$ instead of the Nelson-Aalen estimate for $\Lambda(x)$. In the next section, we first study the one-sample analogue of θ and $\hat{\theta}(k)$.

4.2 A Hazard-Function based Statistic

In this section, we only consider one sample $X_1, \ldots, X_n$ of i.i.d. censored data. We define analogously to (4.3) and (4.5) respectively as

$$\beta = \int_0^V \tilde{K}(x)\frac{dH(x)}{1-H(x-)} \tag{4.8}$$

$$\hat{\beta}_n^{(1)} = \int_0^V \tilde{K}(x)\frac{dH_n(x)}{1-\hat{H}_n(x-)} \tag{4.9}$$

with the kernel $\tilde{K}(x)$ satisfying

$$\sup_{0<x\leq V} |\tilde{K}(x)| < \infty, \tag{4.10}$$

where V is defined in (4.4). We want to replace $\hat{\beta}_n^{(1)}$ using the following two kernel type estimators $\hat{\beta}_n^{(2)}$ and $\hat{\beta}_n^{(3)}$.

$$\begin{aligned}\hat{\beta}_n^{(2)} &= \int_0^V \tilde{K}(x)\hat{\lambda}_1(x)dx \\ &= \int_0^V \tilde{K}(x)\frac{1}{b}\sum_{i=1}^n K\left(\frac{x-X_{(i)}}{b}\right)\frac{\delta_{(i)}}{n-i+1}dx,\end{aligned} \tag{4.11}$$

where the kernel hazard estimator $\hat{\lambda}_1(x)$ is defined by

$$\hat{\lambda}_1(x) = \frac{1}{b}\int K\left(\frac{x-u}{b}\right)\frac{dH_n(u)}{1-\hat{H}_n(u-)} = \frac{1}{b}\sum_{i=1}^n K\left(\frac{x-X_{(i)}}{b}\right)\frac{\delta_{(i)}}{n-i+1}, \tag{4.12}$$

where $X_{(i)}$ are the ordered observed observations, and $H_n(x)$, $\hat{H}_n(x)$ are empirical distribution functions based on $X_1, \ldots, X_n$ defined as

$$H_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \leq x, \delta_i = 1\right), \quad \text{and} \quad \hat{H}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \leq x\right). \tag{4.13}$$

The second kernel type estimator is defined as

$$\hat{\beta}_n^{(3)} = \int_0^V \tilde{K}(x)\hat{\lambda}_2(x)dx = \int_0^V \tilde{K}(x)\frac{h_n(x)}{1-\hat{H}_n(x-)}dx, \tag{4.14}$$

where the kernel hazard estimator $\hat{\lambda}_2(x)$ is defined as

$$\hat{\lambda}_2(x) = \frac{h_n(x)}{1-\hat{H}_n(x-)}, \tag{4.15}$$

with the empirical distribution function $\hat{H}_n(x)$, which is defined in (4.13), and the kernel density estimator $h_n(x)$ is defined as

$$h_n(x) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{x-X_i}{b}\right)\delta_i. \tag{4.16}$$

The bandwidths b satisfy

$$b \to 0, \quad nb \to \infty, \quad \text{as } n \to \infty. \tag{4.17}$$

We assume that the kernel K satisfies the following common assumptions

(K) K is bounded, symmetric probability density with support $[-1, +1]$.

Examples are the Uniform, Epanechnikov, and Biweight kernel functions described in Section 2.2 of Chapter 2. The finiteness of the support of K is only assumed for convenience. The results will also hold for sufficiently fast decreasing kernels like the Gaussian.

4.2.1 Convergence of $\hat{\beta}_n^{(1)}$ to β

In this section, we focus to establish the convergence of $\hat{\beta}_n^{(1)}$ to β, which contain the Nelson-Aalen estimator for the cumulative hazard function $\hat{\Lambda}_n(x)$ and the cumulative

hazard function $\Lambda(x)$, respectively. Therefore, we need the asymptotic representation of the convergence of the Nelson-Aalen estimator for the cumulative hazard function. Using martingale and counting process concept many authors proved that $\hat{\Lambda}_n(x)$ converges uniformly to $\Lambda(x)$. We summarize the asymptotically uniformly convergence of $\hat{\Lambda}_n(x)$ and $\Lambda(x)$ in Theorem 4.2.1 due to Hobbs [60]. For details and notations, compare the proof given by Hobbs [60].

Let the counting process $N_{\cdot}(t)$ and at-risk process $Y_{\cdot}(t)$ be defined as

$$N_{\cdot}(t) = \sum_{i=1}^{n} N_i(t) = \sum_{i=1}^{n} I\left(X_i \leq t, \delta_i = 1\right), \quad Y_{\cdot}(t) = \sum_{i=1}^{n} Y_i(t) = \sum_{i=1}^{n} I\left(X_i \geq t\right).$$

where

Theorem 4.2.1. *Under the conditions* $\inf_{t\in(0,\tau]} Y_{\cdot}(t) \to \infty$ *in probability as* $n \to \infty$, *and* $\Lambda(\tau) < \infty$, *for continuous* $F(t)$,

$$\sup_{0\leq t\leq\tau} \left|\hat{\Lambda}_n(t) - \Lambda(t)\right| \xrightarrow{P} 0.$$

Proof. The Nelson-Aalen estimator $\hat{\Lambda}_n(t)$ in (4.6), which is the instrumental in survival analysis for censored data, is in terms of $N_{\cdot}(t)$ and $Y_{\cdot}(t)$

$$\hat{\Lambda}_n(t) = \int_0^t \frac{I\left(Y_{\cdot}(s) > 0\right)}{Y_{\cdot}(s)} dN_{\cdot}(s), \quad 0 \leq t \leq \tau, \tag{4.18}$$

and the hazard function can be defined as

$$\lambda(t) = \frac{-d\left[\log S(t)\right]}{dt}, \tag{4.19}$$

where the survival function is $S(t) = 1 - F(t)$ and distribution function is $F(t) = P\left(T_i \leq t\right)$. By using $I\left(Y_{\cdot}(s) > 0\right)$ and defining $\frac{0}{0} = 0$, we can avoid mathematical clumsier situation for the integrand when $Y_{\cdot}(s) = 0$ (confer the convention in Fleming and Harrington [46] and Andersen et al. [3]). The corresponding compensator process

$$A_{\cdot}(t) = \sum_{i=1}^{n} A_i(t) = \int_0^t I\left(X \geq s\right) \lambda(s) ds.$$

Let $G(c) = P(C > c)$ and $G(c-) = \lim_{s \uparrow c} G(s)$ for the limit from the left-hand side. Therefore,

$$\begin{aligned} E(N_{\cdot}(t)) &= P(T \leq t, T \leq C) = \int_0^t G(c-)dF(c) \\ &= \int_0^t G(c-)S(c)\frac{dF(c)}{S(c)} = \int_0^t P(X \geq c)\lambda(c)dc \\ &= E\int_0^t I(X \geq c)\lambda(c)dc = E(A_{\cdot}(t)). \end{aligned}$$

Suppose, a martingale $M_{\cdot}(.)$ is define as

$$M_{\cdot}(t) = N_{\cdot}(t) - A_{\cdot}(t) = \sum_{i=1}^{n} I(X_i \leq t, \delta_i = 1) - \int_0^t I(X \geq s)\lambda(s)ds.$$

Furthermore, suppose that $F_s = \sigma\{N(c), I(X \leq c, \delta = 0) : 0 \leq c \leq s\}$ is the filtration for the process $M_{\cdot}(t)$ and F_{s-} is the information in $N(c)$ and $I(X \leq c, \delta = 0)$ up to, but not including time s. Both $\{X < s\}$ and $\{X \geq s\}$ are F_{s-}-measurable since $\{X < s\} = \bigcup_{n=1}^{\infty}\{X \leq s - \frac{1}{n}\}$. Furthermore, $dN(s)$ is a Bernoulli random variable such that

$$E[dN_{\cdot}(s) \mid F_{s-}] = I(X \geq s)\lambda(s)ds = dA_{\cdot}(s).$$

When we have independent censoring it follows that

$$P\{s \leq T < s + ds \mid T \geq s\} = P\{s \leq T < s + ds \mid T \geq s, C \geq s\},$$

and

$$\begin{aligned} E[dA_{\cdot}(s) \mid F_{s-}] &= E[I(X \geq s)\lambda(s)ds \mid F_{s-}] \\ &= I(X \geq s)\lambda(s)ds = dA_{\cdot}(s). \end{aligned}$$

Therefore, the change in $M_{\cdot}(t) = N_{\cdot}(t) - A_{\cdot}(t)$ over an infinitesimal interval $(s-ds, s]$, $dM_{\cdot}(s) = dN_{\cdot}(s) - dA_{\cdot}(s)$ has expectation 0 given F_{s-}. Hence, $M_{\cdot}(t)$ is a martingale with respect to F_s.

Suppose, $L(s) = \frac{I(Y_{\cdot}(s)>0)}{Y_{\cdot}(s)}$ and define a martingale $Q(.)$ with $E(Q(t)) = 0$, since

$dM_{\cdot}(s)$ has expectation 0, by

$$\begin{aligned} Q(t) &= \int_0^t L(s) dM_{\cdot}(s) \\ &= \int_0^t L(s) d\left(N_{\cdot}(s) - A_{\cdot}(s)\right) \\ &= \int_0^t L(s) dN_{\cdot}(s) - \int_0^t L(s) dA_{\cdot}(s) \\ &= \hat{\Lambda}_n(t) - \int_0^t L(s) Y_{\cdot}(s) \lambda(s) ds \\ &= \hat{\Lambda}_n(t) - \int_0^t I\left(Y_{\cdot}(s) > 0\right) \lambda(s) ds \end{aligned}$$

$$\begin{aligned} Q(t) &= \hat{\Lambda}_n(t) - \int_0^t \left[1 - 1 + I\left(Y_{\cdot}(s) > 0\right)\right] \lambda(s) ds \\ &= \hat{\Lambda}_n(t) - \int_0^t \lambda(s) ds + \int_0^t \lambda(s) \left[1 - I\left(Y_{\cdot}(s) > 0\right)\right] ds \\ &= \hat{\Lambda}_n(t) - \Lambda(t) + \int_0^t \lambda(s) \left[1 - I\left(Y_{\cdot}(s) > 0\right)\right] ds \\ &= \hat{\Lambda}_n(t) - \Lambda(t) + D(t), \end{aligned} \tag{4.20}$$

where

$$D(t) := \int_0^t \lambda(s) \left[1 - I\left(Y_{\cdot}(s) > 0\right)\right] ds,$$

is a non-negative and nondecreasing integral for $0 \leq t \leq \tau$. Now, to show that $D(t)$ converges to zero in probability, we need to examine this as $n \to \infty$. Note that the bracketed term in the integrand of $D(t)$ is just $I\left(Y_{\cdot}(s) = 0\right)$ and suppose there exist $\epsilon \in (0, 1)$ such that

$$\begin{aligned} P\left[I\left(Y_{\cdot}(s) = 0\right) > \epsilon\right] &= P\left[Y_{\cdot}(s) = 0\right] \\ &= P\left[Y_1(s) = 0, Y_2(s) = 0, \ldots, Y_n(s) = 0\right] \\ &= \prod_{i=1}^{n} P\left(Y_i(s) = 0\right) \\ &= \prod_{i=1}^{n} \left(1 - P\left(Y_i(s) = 1\right)\right) \\ &= \prod_{i=1}^{n} \left\{1 - \left(1 - F(s)\right)\left(1 - G(s)\right)\right\} \\ &= \left\{1 - \left(1 - F(s)\right)\left(1 - G(s)\right)\right\}^n. \end{aligned} \tag{4.21}$$

Here, we assume that τ lies in the interior of the support of $F(.)$ and $G(.)$, such that

$$1 - F(t) > 0 \quad \text{and} \quad 1 - G(t) > 0 \quad \text{for} \quad 0 \leq t \leq \tau. \tag{4.22}$$

Hence, (4.21) is strictly less than one, and goes to zero as $n \to \infty$. Since, $\lambda(t)$ is continuous on the compact interval $[0, \tau]$, so bounded there. Therefore, the integrand in $D(t)$ converges in probability to zero. At this stage, we need Theorem 4.2.2 and 4.2.3 due to the Proposition II.5.2 and II.5.3 of Andersen et al. [3], respectively.

Theorem 4.2.2. *Consider a sequence $X_n(.)$ of stochastic process such that*

$$\lim_{C \uparrow \infty} \sup_n E\left[|X_n(t)| I\left(|X_n(t)| \geq C\right)\right] = 0$$

for all t, i.e., the sequence $X_n(t)$ is uniformly integrable, and

$$E|X_n(t)| \leq k(t)$$

for all t and n with $\int_o^\tau k(t)dt < \infty$. Suppose that

$$X_n(t) \xrightarrow{P} h(t) \quad \text{as } n \xrightarrow{P} \infty$$

for every $t \in [0, \tau]$. Then it follows that

$$E\left(\sup_{t \in [0,\tau]} \left| \int_0^t X_n(s)ds - \int_0^t h(s)ds \right| \right) \xrightarrow{P} 0. \tag{4.23}$$

Theorem 4.2.3. *Consider a sequence $X_n(.)$ of stochastic process and such that for all $\delta > 0$ there exists k_δ with $\int_0^\tau k_\delta < \infty$ such that*

$$\lim_{n \to \infty} \inf P\left(|X_n(t)| \leq k_\delta(t) \text{ for all } t\right) \geq 1 - \delta.$$

Suppose that

$$X_n(t) \xrightarrow{P} h(t) \quad \text{as } n \to \infty$$

for every $t \in [0, \tau]$ and suppose that

$$\int_0^t |h(t)|\, dt < \infty.$$

Then it follows that

$$\sup_{t \in [0,\tau]} \left| \int_0^t X_n(s)ds - \int_0^t h(s)ds \right| \xrightarrow{P} 0. \tag{4.24}$$

As the integrand in $D(t)$ converges in probability to zero, so it follows from Theorem 4.2.2 that

$$E\left(D(t)\right) \xrightarrow{P} 0, \tag{4.25}$$

and from Theorem 4.2.2

$$D(t) \xrightarrow{P} 0. \tag{4.26}$$

Thus, from (4.20) we have

$$E\left(\hat{\Lambda}_n(t)\right) \to \Lambda(t) \quad \text{as} \quad n \to \infty. \tag{4.27}$$

Since, $E\left(Q(t)\right) = 0$ and

$$\begin{aligned} \text{Var}\left(Q(t)\right) &= E\int_0^t H^2(s)dA_{\cdot}(s) \\ &= E\int_0^t \frac{I\left(Y_{\cdot}(s) > 0\right)\lambda(s)}{Y_{\cdot}(s)}ds. \end{aligned} \tag{4.28}$$

As we are assuming that (4.22) holds, so (4.21) goes to zero as $n \to \infty$. Also, $\lambda(t)$ is continuous on the compact interval $[0, \tau]$, so bounded there. Therefore, the integrand in (4.28) converges in probability to zero. Again, following Theorem 4.2.2 we have as $n \to \infty$

$$\text{Var}\left(Q(t)\right) = E\int_0^t \frac{I\left(Y_{\cdot}(s) > 0\right)\lambda(s)}{Y_{\cdot}(s)}ds \to 0. \tag{4.29}$$

So, $Q(t) \xrightarrow{P} 0$. Also, we have $D(t) \xrightarrow{P} 0$. Thus, (4.20) guarantees that

$$\hat{\Lambda}_n(t) \xrightarrow{P} \Lambda(t). \tag{4.30}$$

Therefore, $\hat{\Lambda}_n(t)$ is consistent provided (4.22) holds. The above result shows the pointwise consistency of $\Lambda(t)$. Since, $D(t)$ is non-negative and nondecreasing for $0 \leq t \leq \tau$, so considering

$$\sup_{0 \leq t \leq \tau} |D(t)| = D(\tau),$$

so that for any $\epsilon > 0$

$$P\left(\sup_{0 \leq t \leq \tau} |D(t)| > \epsilon\right) = P\left(D(\tau) > \epsilon\right) \to 0.$$

Thus, $D(.) \xrightarrow{P} 0$ over $[0, \tau]$. Due to the corollary of Lenglart's inequality (see Fleming and Harrington [46], p. 113) for any $\epsilon > 0$ and large n

$$P\left(\sup_{t\in[0,\tau]} |Q(t) - 0| > \epsilon\right) < \epsilon.$$

So, $Q(.) \xrightarrow{P} 0$. Therefore, using Slutsky's lemma for random variables we have

$$\sup_{0\le t\le\tau} \left|\hat{\Lambda}_n(t) - \Lambda(t)\right| \le \sup_{0\le t\le\tau} |Q(t)| + \sup_{0\le t\le\tau} |D(t)| \xrightarrow{P} 0. \tag{4.31}$$

Hence, $\hat{\Lambda}_n(t)$ converges uniformly to $\Lambda(t)$ as $n \to \infty$. □

Now, using Theorem 4.2.1 of the asymptotically uniformly convergence of $\hat{\Lambda}_n(t)$ to $\Lambda(t)$, we get

Corollary 4.2.4. *Under the assumptions of Theorem 4.2.1, let $\tilde{K}(x)$ be continuous. Then*

$$\hat{\beta}_n^{(1)} \xrightarrow{P} \beta \quad \text{as } n \to \infty. \tag{4.32}$$

Proof. Recall that $d\Lambda(x) = \frac{1}{1-F(x)} dF(x)$. Hence, Λ defines a finite measure on $[0, V]$ as, due to monotonicity of F

$$\int_0^V d\Lambda(x) \le \frac{1}{1 - F(V)} \int_0^V dF(x) = \frac{F(V)}{1 - F(V)} < \infty,$$

by our choice of V. From Theorem 4.2.1, we have

$$\sup_{0<x\le V} \left|\hat{\Lambda}_n(x) - \Lambda(x)\right| \xrightarrow{P} 0,$$

i.e., the measure on $[0, V]$ characterized by $\hat{\Lambda}_n$ converges weakly to the measure characterized by Λ. We can even make them to probability measures on $[0, V]$ by setting

$$\hat{\mu}_n(B) = \frac{1}{\hat{\Lambda}_n(V)} \int_B d\hat{\Lambda}_n(x), \quad \mu(x) = \frac{1}{\Lambda(V)} \int_B d\Lambda(x),$$

and as $\hat{\Lambda}_n(V) \to \Lambda(V)$, we also have convergence in distribution of $\hat{\mu}_n$ to μ. Then, by standard results of probability theory (e.g., Proposition 8.12 of Breiman [21]), using continuity of $\tilde{K}$

$$\int_0^V \tilde{K}(x) d\hat{\mu}_n(x) \to \int_0^V \tilde{K}(x) d\mu(x),$$

which, together with $\hat{\Lambda}_n(V) \to \Lambda(V)$ implies

$$\hat{\beta}_n^{(1)} - \beta = \int_0^V \tilde{K}(x) d\hat{\Lambda}_n(x) - \int_0^V \tilde{K}(x) d\Lambda(x) \xrightarrow{P} 0.$$

□

4.2.2 Convergence of $\hat{\beta}_n^{(2)}$ to β

In this section, we first discuss convergence of

$$\hat{\lambda}_1(x) = \frac{1}{b}\sum_{i=1}^{n} K\left(\frac{x - X_{(i)}}{b}\right)\frac{\delta_{(i)}}{n-i+1},$$

to the hazard rate $\lambda(x)$. We need the following notation

$$\bar{\lambda}_1(x) = \frac{1}{b}\int K\left(\frac{x-u}{b}\right)\lambda(u)du.$$

The asymptotic behaviour of $\hat{\lambda}_1(x)$ has been investigated, in particular in view of uniform convergence, by Diehl and Stute [41]. They decompose the estimation error $\hat{\lambda}_1(x)-\lambda(x)$ into a random part $\hat{\lambda}_1(x)-\bar{\lambda}_1(x)$ and a basic part $\bar{\lambda}_1(x)-\lambda(x)$. Then, they show that $\hat{\lambda}_1(x) - \bar{\lambda}_1(x)$ behaves like a sum of independent terms which essentially is

$$h_n(x) = \frac{1}{b}\int K\left(\frac{x-u}{b}\right)dH_n(u) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{x-X_i}{b}\right)\delta_i,$$

which is the basis of our estimate $\hat{\lambda}_2(x)$. More precisely, Theorem 2 of Diehl and Stute [41] state in our terminology.

Theorem 4.2.5. *Let the kernel K satisfy (K) and, additionally, be continuously differentiable on $[-1,+1]$. Let F and G have bounded densities f and g on $[0, V+\Delta]$ for some $\Delta > 0$, where V satisfies $F(V) < 1$ and $G(V) < 1$. Then*

$$\sup_{0\le x\le V}\sqrt{nb}\left|\hat{\lambda}_1(x) - \bar{\lambda}_1(x) - \frac{h_n(x) - Eh_n(x)}{1-H(x)}\right| = O_{a.s.}\left(\frac{\log\log n}{\sqrt{nb}} + \sqrt{b\log\log n}\right).$$

If we only are interested in convergences in probability, then, from the same Theorem of Diehl and Stute [41] we would have a rate $O_p\left(\frac{1}{\sqrt{nb}}\right) + O_p\left(\sqrt{b}\right)$. As in Corollary 2 of Diehl and Stute [41], we conclude

Corollary 4.2.6. *Under the assumptions of Theorem 4.2.5 and if, additionally, $f(x) \ge c > 0$ for $0 \le x \le V + \Delta$ and some $c > 0$, we have for $n \to \infty$, $b \to 0$ such that $nb \to \infty$ and*

$$\frac{1}{nb}\log\frac{1}{b} \to 0, \quad \frac{1}{\log\log n}\log\frac{1}{b} \to \infty,$$

$$\sqrt{\frac{nb}{2\log\frac{1}{b}}}\sup_{0\le x\le V}\frac{1}{\sqrt{\lambda(x)}}|\hat{\lambda}_1(x) - \bar{\lambda}_1(x)| \xrightarrow{a.s.} \sqrt{\int K^2(u)du}.$$

Note that under the conditions of the corollary for all $0 \leq x \leq V$

$$\lambda(x) = \frac{f(x)}{1 - F(x)} \geq f(x) \geq c > 0$$
$$\lambda(x) \leq \frac{f(x)}{1 - F(V)} \leq \frac{c}{1 - F(V)} < \infty,$$

as we have assumed boundedness of f on $[0, V + \Delta]$. So, the corollary also implies uniform convergence of $\hat{\lambda}_1(x) - \bar{\lambda}_1(x)$ to 0 on $[0, V]$ with the same rate.

To get uniform consistency of $\hat{\lambda}_1(x)$ as an estimate of $\lambda(x)$, we have to investigate the bias term $\bar{\lambda}_1(x) - \lambda(x)$. Its convergence behaviour depends on smoothness assumptions about $\lambda(x)$, i.e., about $f(x)$. If, e.g., $f(x)$ is differentiable on $[0, V + \Delta]$ with derivative $f^{'}(x)$ which is Hölder continuous of order α, $0 < \alpha \leq 1$, i.e.,

$$|f^{'}(x) - f^{'}(y)| \leq C_H |x - y|^{\alpha}, \quad \text{for all } x, y \in [0, V],$$

for some $C_H > 0$, then $\lambda(x)$ is also differentiable with derivative

$$\lambda^{'}(x) = \frac{f^{'}(x)}{1 - F(x)} + \frac{f^2(x)}{(1 - F(x))^2},$$

and a straight-forward calculation, using $1 - F(V) \leq 1 - F(x) \leq 1$, $c \leq f(x) \leq C$ and boundedness of $f^{'}(x)$ for $0 \leq x \leq V$, show that $\lambda^{'}(x)$ is also Hölder continuous of order α. From this, we get

Lemma 4.2.7. *If under the assumptions of Corollary 4.2.6, additionally, $f(x)$ is differentiable and $f^{'}(x)$ is Hölder continuous of order α on $[0, V]$, then*

$$\sup_{0 \leq x \leq V} \left| \bar{\lambda}_1(x) - \lambda(x) \right| = O\left(b^{1+\alpha}\right).$$

Proof. For $0 \leq x \leq V$ and small enough b, by the mean-value theorem,

$$\lambda(x - b\nu) - \lambda(x) = -\lambda^{'}(x - \theta b\nu) b\nu,$$

for some $0 \leq \theta \leq 1$. From the Hölder continuity of $\lambda^{'}$, we have for suitable C_H

$$\left| \lambda^{'}(x - \theta b\nu) - \lambda^{'}(x) \right| \leq C_H |\theta b\nu|^{\alpha} \leq C_H |b\nu|^{\alpha} \leq C_H b^{\alpha},$$

uniformly in $0 \leq x \leq V$, $-1 \leq \nu \leq 1$. As K is a probability density with support $[-1, +1]$, we conclude, substituting $u = x - b\nu$,

$$\begin{aligned}
\left|\bar{\lambda}_1(x) - \lambda(x)\right| &= \left|\frac{1}{b}\int K\left(\frac{x-u}{b}\right)\lambda(u)du - \lambda(x)\right| \\
&= \left|\int K(\nu)\left(\lambda(x-b\nu) - \lambda(x)\right)d\nu\right| \\
&= \left|\int K(\nu)\lambda'(x)b\nu\left(1 + O\left(b^{\alpha}\right)\right)d\nu\right| \\
&= \left|\int \nu K(\nu)d\nu\lambda'(x)b + O\left(b^{1+\alpha}\right)\right| \\
&= O\left(b^{1+\alpha}\right), \quad \text{uniformly in } 0 \leq x \leq V,
\end{aligned}$$

as, due to symmetry of K, $\int \nu K(\nu)d\nu = 0$. □

Combining the previous results we get, using $\left|\hat{\lambda}_1 - \lambda\right| \leq \left|\hat{\lambda}_1 - \bar{\lambda}_1\right| + \left|\bar{\lambda}_1 - \lambda\right|$,

Theorem 4.2.8 (Uniform Consistency of $\hat{\lambda}_1$)**.** *Under the conditions of Corollary 4.2.6 , we have for $n \to \infty$, $b \to 0$ such that $nb \to \infty$,*

$$\frac{1}{nb}\log\frac{1}{b} \to 0, \quad \frac{1}{\log\log n}\log\frac{1}{b} \to \infty,$$

$$\sup_{0\leq x\leq V}\left|\hat{\lambda}_1(x) - \lambda(x)\right| = O_{a.s.}\left(\frac{1}{nb}\log\frac{1}{b} + O\left(b^{1+\alpha}\right)\right).$$

As an immediate consequence, we have the desired consistency of $\hat{\beta}_n^{(2)}$.

Corollary 4.2.9. *Under the assumptions of Theorem 4.2.8, we have for $\tilde{K}$ satisfying (4.10)*

$$\hat{\beta}_n^{(2)} \xrightarrow{a.s.} \beta.$$

Proof.

$$\begin{aligned}
\left|\hat{\beta}_n^{(2)} - \beta\right| &= \left|\int_0^V \tilde{K}(x)\left(\hat{\lambda}_1(x) - \lambda(x)\right)dx\right| \leq \sup_{0\leq x\leq V}\left|\tilde{K}(x)\right|\int_0^V \left|\hat{\lambda}_1(x) - \lambda(x)\right|dx \\
&\leq V \sup_{0\leq x\leq V}\left|\tilde{K}(x)\right| \sup_{0\leq x\leq V}\left|\hat{\lambda}_1(x) - \lambda(x)\right|,
\end{aligned}$$

and the assertion follows from Theorem 4.2.8. □

4.2.3 Convergence of $\hat{\beta}_n^{(3)}$ to β

We first want to prove some results illustrating that $\hat{\lambda}_2(x)$ is a reasonable estimate of the hazard function $\lambda(x)$. Recall that

$$\hat{\lambda}_2(x) = \frac{h_n(x)}{1 - \hat{H}_n(x-)} = \frac{1}{1 - \hat{H}_n(x-)} \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right) \delta_i.$$

Theorem 4.2.10. *Let F have a density f, and let $(1 - G(x))\, f(x)$ satisfy a Hölder condition*

$$|(1 - G(x))\, f(x) - (1 - G(y))\, f(y)| \leq C_H |x - y|^{\alpha},$$

for some $C_H > 0$, $0 < \alpha \leq 1$. Let the kernel K satisfy (K). Then

a) *For $n \to \infty$, $b \to 0$ such that $nb \to \infty$*

$$\sqrt{nb}\,(h_n(x) - E h_n(x)) \xrightarrow{d} \mathcal{N}(0, \sigma_x^2),$$

with

$$\sigma_x^2 = (1 - G(x))\, f(x) \int K^2(\nu) d\nu,$$

b) $h_n(x) \xrightarrow{P} (1 - G(x))\, f(x)$ *and, in particular,*

$$mse\ h_n(x) = E\left[h_n(x) - (1 - G(x))\, f(x)\right]^2 = O\left(\frac{1}{nb}\right) + O\left(b^{2\alpha}\right).$$

Proof. We prove a) by applying the Lyapounov central limit theorem (cf. Billingsley [20]). For this purpose, we write

$$h_n(x) = \frac{1}{n} \sum_{j=1}^{n} Z_{jn}, \quad Z_{jn} = \frac{1}{b} K\left(\frac{x - X_j}{b}\right) \delta_j.$$

For the first three moments of Z_{jn}, we have

$$\begin{aligned}
\mu_{jn} = E Z_{jn} &= E \frac{1}{b} K\left(\frac{x - T_j}{b}\right) I\left(T_j \leq C_j\right) \\
&= \int \int \frac{1}{b} K\left(\frac{x - t}{b}\right) I\left(t \leq z\right) f(t) g(z) dz dt \\
&= \int \frac{1}{b} K\left(\frac{x - t}{b}\right) (1 - G(t))\, f(t) dt \\
&= \int K\left(u\right) (1 - G(x - bu))\, f(x - bu) du,
\end{aligned}$$

substituting $t = x - bu$. Using the Hölder condition and $\int K(u)du = 1$, we have

$$\begin{aligned}|\mu_{jn} - (1 - G(x))\, f(x)| &\leq \int K(u)\, |(1 - G(x - bu))\, f(x - bu) - (1 - G(x))\, f(x)|\, du \\ &\leq C_H b^\alpha \int |u|^\alpha K(u)\, du = O\left(b^\alpha\right),\end{aligned}$$

uniformly in j, as $Z_{1,n}, \ldots, Z_{n,n}$ are i.i.d., i.e., we have

$$\mu_{jn} = (1 - G(x))\, f(x) + O\left(b^\alpha\right).$$

$$\begin{aligned}EZ_{jn}^2 &= \int\int \frac{1}{b^2} K^2\left(\frac{x - t}{b}\right) I\left(t \leq z\right) f(t) g(z) dz dt \\ &= \frac{1}{b} \int K^2(u)\, (1 - G(x - bu))\, f(x - bu) du.\end{aligned}$$

As K is bounded, we get as for μ_{jn} that we may replace $(1 - G(x - bu))\, f(x - bu)$ in the integral by $(1 - G(x))\, f(x)$ up to an error of order $O\left(b^\alpha\right)$. Hence,

$$EZ_{jn}^2 = \frac{1}{b} (1 - G(x))\, f(x) \int K^2(u)\, du + O\left(\frac{b^\alpha}{b}\right),$$

and, as $\mu_{jn} = O(1)$,

$$\sigma_{jn}^2 = EZ_{jn}^2 - \mu_{jn}^2 = \frac{1}{b} (1 - G(x))\, f(x) \int K^2(u)\, du + O\left(b^{\alpha - 1}\right).$$

For the third moment, we get by exactly the same kind of argument

$$\gamma_{jn} = E|Z_{jn} - \mu_{jn}|^2 = \frac{1}{b^2} (1 - G(x))\, f(x) \int K^3(u)\, du + O\left(b^{\alpha - 2}\right).$$

Hence,

$$\begin{aligned}\lim_{n\to\infty} \frac{\sum_{j=1}^n \gamma_{jn}}{(\sum_{j=1}^n \sigma_{jn}^2)^{3/2}} &= \lim_{n\to\infty} \frac{\frac{n}{b^2} (1 - G(x))\, f(x) \int K^3(u)\, du}{(\frac{n}{b})^{3/2} \left[(1 - G(x))\, f(x) \int K^2(u)\, du\right]^{3/2}} \\ &= \lim_{n\to\infty} \frac{\text{const}}{\sqrt{nb}} = 0,\end{aligned}$$

i.e., the Lyapounov condition is fulfilled, and a) follows, where the form of σ_x^2 follows from

$$\text{var } h_n(x) = \frac{1}{n^2} \sum_{j=1}^n \text{var } Z_{jn} = \frac{1}{n} \sigma_{1n}^2 = \frac{1}{nb} \sigma_x^2 + o\left(\frac{1}{nb}\right).$$

For the bias we have, as $Eh_n(x) = \mu_{jn}$,

$$Eh_n(x) - (1 - G(x))\, f(x) = O\left(b^\alpha\right),$$

such that, the mse expression follows as it is the sum of variance and squared bias. The consistency of $h_n(x)$ follows, as the mse converges to 0. □

Corollary 4.2.11. *Under the assumptions of Theorem 4.2.10*

$$\hat{\lambda}_2(x) \xrightarrow{P} \lambda(x) = \frac{f(x)}{1 - F(x)} \quad \textit{for } n \to \infty.$$

Proof. The results follows immediately from the last theorem and from $1-\hat{H}_n(x-) \xrightarrow{P} 1 - H(x) = (1 - F(x))(1 - G(x))$ as a consequence of the law of large numbers. □

As for $\hat{\lambda}_1(x)$, we need however uniform consistence over bounded intervals. As a first step, we remark that the Glivenko-Cantelli Theorem can be strengthened to a law of the iterated logarithm such that

$$\sup_x \left| \hat{H}_n(x-) - H(x) \right| = O_{a.s.} \left(\sqrt{\frac{\log \log n}{n}} \right). \tag{4.33}$$

(confer, Van der Vaart [139] p. 268). So, it remains to investigate uniform convergence of $h_n(x)$.

For the proof of this result we use for convenience a uniform consistency result of Györfi et al. [57], which has been formulated for uniformly $(\varphi-)$ mixing random variables (Z_j, Y_j), $j = 1, \ldots, n$, which of course also applies to the independent data (T_j, δ_j), $j = 1, \ldots, n$, respectively $(T_j, 1)$, $j = 1, \ldots, n$.

Note that, as $\delta_i = I\left(T_i \leq C_i\right)$,

$$h_n(x) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - T_i}{b} \right) \delta_i.$$

We also consider the Nadaraya-Watson kernel estimate

$$\hat{h}_n(x) = \frac{h_n(x)}{\frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - T_i}{b} \right)} = \frac{h_n(x)}{\hat{f}_n(x)},$$

which estimates

$$E\left\{ \delta_i \mid T_i = x \right\} = \mathbb{P}\left(C_i \geq x \mid T_i = x \right) = 1 - G(x),$$

as T_i and C_i are independent. The denominator of $\hat{h}_n(x)$ is the usual kernel estimate of $f(x)$. We do not really need to consider $\hat{h}_n(x)$ and $\hat{f}_n(x)$, but the results which we want to use are formulated in terms of these quantities. Alternatively, we would have to mimic the lengthy proof of the theorem of Györfi et al. [57] and show the result for $h_n(x)$ directly. We now formulate the assumptions on the distribution of T_j and C_j.

(D) T_j and C_j, $j = 1, \ldots, n$, are independent with distribution functions respectively F and G. For V satisfying $F(V) < 1$ and $G(V) < 1$, F and G have densities respectively f and g on $[0, V]$, $f(x) > 0$ *a.e.* on $[0, V]$ and f is differentiable with derivative f' on $[0, V]$. f' and g are the Hölder continuous of order α on $[0, V]$ for some $0 < \alpha \leq 1$.

Note that (D) does not exclude point masses of the censoring distribution G beyond V which may be appropriate for planned studies with a given maximal duration.

Proposition 4.2.12. *Let the kernel K satisfy (K) and, additionally, be also Hölder continuous of order α, and let (D) be satisfied. Then, for $n \to \infty$, $b \to 0$ such that $\frac{\log n}{nb} \to 0$*

$$\sup_{0 \leq x \leq V} \left| \hat{h}_n(x) - (1 - G(x)) \right| = O_{a.s.} \left(b^{1+\alpha} + \sqrt{\frac{\log n}{nb}} \right)$$

$$\sup_{0 \leq x \leq V} \left| \hat{f}_n(x) - f(x) \right| = O_{a.s.} \left(b^{1+\alpha} + \sqrt{\frac{\log n}{nb}} \right).$$

Proof. We apply Theorem 3.3.2 of Györfi et al. [57] to the pairs (T_j, δ_j) and $(T_j, 1)$, respectively. Their conditions (A.1)-(A.4) and (K.1)-(K.6) respectively follow from (D), using also $|\delta_j| \leq 1$, and the assumptions on K. Note that we consider the special case $k = 1$. As our data are independent, i.e., the φ-mixing coefficients are all 0, we may choose $m_n = 1$ in (3.3.1). As δ_j and 1 are respectively bounded by 1, we do not have to use the truncation argument in the formulation of Theorem 3.3.2 and can choose $M_n = 1$, too. □

Corollary 4.2.13. *Under the assumptions of Proposition 4.2.12,*

$$\sup_{0 \leq x \leq V} |h_n(x) - (1 - G(x)) f(x)| = O_{a.s.} \left(b^{1+\alpha} + \sqrt{\frac{\log n}{nb}} \right).$$

Proof. As $h_n(x) = \hat{h}_n(x) \hat{f}_n(x)$, we have

$$|h_n(x) - (1 - G(x)) f(x)| \leq \left| \hat{h}_n(x) - (1 - G(x)) \right| \hat{f}_n(x) + (1 - G(x)) \left| \hat{f}_n(x) - f(x) \right|$$
$$\leq \left| \hat{h}_n(x) - (1 - G(x)) \right| \left| \hat{f}_n(x) - f(x) \right| + \left| \hat{h}_n(x) - (1 - G(x)) \right| f(x) + (1 - G(x)) \left| \hat{f}_n(x) - f(x) \right|,$$

and the assertion follows from Proposition 4.2.12, as f is bounded on $[0, V]$. □

Theorem 4.2.14 (Uniform Consistency of $\hat{\lambda}_2$)**.** *Under the assumptions of Proposition 4.2.12, we have*

$$\sup_{0\le x\le V}\left|\hat{\lambda}_2(x)-\lambda(x)\right| = O_{a.s.}\left(b^{1+\alpha}+\sqrt{\frac{\log n}{nb}}\right).$$

Proof. We decompose

$$\begin{aligned}\left|\hat{\lambda}_2(x)-\lambda(x)\right| &= \left|\frac{h_n(x)}{1-\hat{H}_n(x-)}-\frac{f(x)}{1-F(x)}\right| \\ &\le \frac{|h_n(x)-(1-G(x))f(x)|}{|1-\hat{H}_n(x-)|}+(1-G(x))f(x)\left|\frac{1}{1-\hat{H}_n(x-)}-\frac{1}{1-H_n(x)}\right| \\ &\le \frac{1}{\Delta}\left|h_n(x)-(1-G(x))f(x)\right|+\frac{C}{\Delta^2}\left|\hat{H}_n(x-)-H(x)\right|,\end{aligned}$$

where we have chosen n_0, Δ such that $1-H(V)$, $1-\hat{H}_n(V)\ge\Delta>0$ for all $n\ge n_0$, using (4.33), and C such that $(1-G(x))f(x)\le C$ for all $0\le x\le V$.

The first term on the right-hand side of the last inequality converges uniformly in $[0,V]$ to 0 with rate at least $b^{1+\alpha}+\sqrt{\frac{\log n}{nb}}$ by Corollary 4.2.13, and the second term has the even small upper bound $\sqrt{\frac{\log\log n}{n}}$ on the rate by (4.33). □

Analogously to Corollary 4.2.13, we conclude from Theorem 4.2.14.

Corollary 4.2.15. *Under the assumptions of Theorem 4.2.14, we have for $\tilde{K}$ satisfying (4.33)*

$$\hat{\beta}_n^{(3)} \xrightarrow{a.s.} \beta.$$

4.3 Kernel-Based Change Point Test's

In this section, we modify Horváth's test described in Section 4.1.1 for a change point by replacing the Nelson-Aalen estimate for the cumulative hazard function by the estimates $\hat{\lambda}_1$ and $\hat{\lambda}_2$, respectively, for the hazard functions describe in Section 4.2. We prove that the test statistic has the same asymptotic distribution under the hypothesis H_0 of no change point as Horváth's test statistic. Recall that the testing problem is given in (4.2).

We first study the test based on $\hat{\lambda}_2$, as the statistic there is rather similar to the one considered by Horváth, and we can use many arguments from Horváth's proof.

4.3.1 Change Point Test using $\hat{\lambda}_2$

Recall that the estimate $\hat{\lambda}_2$ of the hazard function $\lambda(x)$ is given by

$$\hat{\lambda}_2(x) = \hat{\lambda}_{2,n}(x) = \frac{h_n(x)}{1 - \hat{H}_n(x-)} = \frac{1}{1 - \hat{H}_n(x-)} \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right) \delta_i.$$

We now stress the dependence on n, as for given k, $1 \leq k < n$, we split the data into two subsamples $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_n$, and $\hat{\lambda}_2(x) = \hat{\lambda}_{2,k}(x)$, then, is the hazard function estimate based on the first subsample only. The corresponding estimate based on the second subsample is

$$\hat{\lambda}_2^*(x) = \hat{\lambda}_{2,k}^*(x) = \frac{h_k^*(x)}{1 - \hat{H}_k^*(x-)} = \frac{1}{1 - \hat{H}_k^*(x-)} \frac{1}{(n-k)b} \sum_{i=k+1}^{n} K\left(\frac{x - X_i}{b}\right) \delta_i,$$

where $\hat{H}_k^*$ denotes the empirical distribution function of the sample $X_{k+1}, \ldots, X_n$. Now, for convenience, we drop again the index k in $\hat{\lambda}_2(x)$ and $\hat{\lambda}_2^*(x)$.

Analogously to (4.5), we define

$$\begin{aligned}
\hat{\theta}_3(k) &= \int_0^V \int_0^V \tilde{K}(x,y) \hat{\lambda}_2(x) \hat{\lambda}_2^*(y) dx dy \\
&= \int_0^V \int_0^V \tilde{K}(x,y) \frac{h_k(x)}{1 - \hat{H}_k(x-)} \frac{h_k^*(y)}{1 - \hat{H}_k^*(y-)} dx dy.
\end{aligned}$$

Following Horváth, we get

$$Q_n(k) = \frac{k\,(n-k)}{n^{3/2}} \left(\hat{\theta}_3(k) - \theta\right),$$

where θ is given by (4.3). Then, we have the following analogue to Theorem 2.1 of Horváth [62]

Theorem 4.3.1. *Assume that $\tilde{K}(x,y) = \tilde{K}(y,x)$ for all x, y, and that $|\tilde{K}(x,y)| \leq \tilde{C}_k$ for all x, y for some constant $\tilde{C}_k$. Moreover, let $\tilde{K}(x,y)$ be Lipschitz continuous in one coordinate, i.e., for some $\tilde{C} > 0$.*

$$\left|\tilde{K}(x,y) - \tilde{K}(x,z)\right| \leq \tilde{C}\,|y - z|\,, \quad \textit{for all } x, y, z \in [0, V],$$

where V again satisfies $F(V), G(V) < 1$. Let K satisfy assumption (K). Let the density f of F be bounded, and let G be continuous. Moreover, let $\tilde{h}(x) = (1 - G(x))f(x)$ satisfy a Lipschitz condition

$$\left|\tilde{h}(x) - \tilde{h}(y)\right| \leq C_H \left|x - y\right| \quad \text{for all } x, y \in [0, V].$$

Then, under the hypothesis H_0, for $n \to \infty$, $b = O\left(n^{\nu-1}\right)$ and $nb \to \infty$ for some $0 < \nu \leq \frac{1}{2}$, we have

$$\max_{1 \leq k < n} \left| Q_n(k) - \sigma \Gamma_n \left(\frac{k}{n}\right) \right| \left(\frac{k}{n}\left(1 - \frac{k}{n}\right)\right)^{-\nu} = O_p\left(n^{\nu - \frac{1}{2}}\right), \tag{4.34}$$

where $\Gamma_n(t)$, $0 \leq t \leq 1$, is a sequence of Gaussian processes distributed as $(1-t)W(t) + t\{W(1) - W(t)\}$, $0 \leq t \leq 1$, with a standard Wiener process $W(t)$, $0 \leq t \leq 1$. The scale parameter σ is given by

$$\sigma^2 = \int_0^V \left(\int_0^V \tilde{K}(x, y) \lambda(y) dy \right)^2 \frac{\lambda(x)}{1 - H(x)} dx.$$

We postpone the proof to the end of section 4.3 and discuss some immediate consequences of the main theorem first. Note that, as pointed out by Horváth [62] the limit process $\Gamma_n(t)$ already appeared in an earlier paper of Csörgő and Horváth [36] about U-statistics-type processes without censoring. Their special form is, therefore, not related to censoring.

Note that the bandwidth rates we need for (4.34) and, then, for change point tests, are much faster than the *mse* optimal rate if we would just be interested in estimating $\lambda(x)$. The latter would be $\left(n^{-\frac{1}{3}}\right)$ if we only assume Lipschitz continuity of the function to be estimated whereas for (4.34) we need $O\left(n^{-\frac{1}{2}}\right)$ for $\nu = \frac{1}{2}$, which is the common weight or even small for $\nu < \frac{1}{2}$. We need this to keep the bias of the kernel estimates under control.

As an immediate consequence, we get analogously to Corollary 4.2.15 that $\hat{\theta}_3(k)$ converges to θ if k grows with n. If we avoid the boundary regions of the sample, we even have uniform consistency with rate $\frac{1}{\sqrt{n}}$.

Corollary 4.3.2. *Under the assumptions of Theorem 4.3.1, we have for any $0 < \varepsilon < \frac{1}{2}$*

$$\max_{\varepsilon n \leq k \leq (1-\varepsilon)n} \left| \sqrt{n}\left(\hat{\theta}_3(k) - \theta\right) - \sigma \Gamma_n\left(\frac{k}{n}\right) \right| = O_p\left(n^{\nu - 1/2}\right).$$

Proof. The result follows immediately from the theorem, the definition of $Q_n(k)$ and from $\varepsilon^2 \leq \frac{k(n-k)}{n^2} \leq (1-\varepsilon)^2$ for all $\varepsilon n \leq k \leq (1-\varepsilon)n$. □

Now, we are in a position to construct a change point test. Intuitively, under H_0 $\hat{\theta}_3(k) \approx \theta$ for all k, and under the alternative $|\hat{\theta}_3(k) - \theta|$ will be large for k near the change point k^*. $\hat{\theta}_3(k)$ is unreliable for $\frac{k}{n} \approx 0$ or ≈ 1 as respectively $\hat{\lambda}_2$ and $\hat{\lambda}_2^*$ will depend on few data only, usually $|\hat{\theta}_3(k) - \theta|$ is down-weighted in those cases. The common weight function is $q(t) = (t(1-t))^{1/2}$ which corresponds to the case $\nu = \frac{1}{2}$ in Theorem 4.3.1. The test statistic is

$$\max_{1 \leq k < n} \left| \frac{Q_n(k)}{q\left(\frac{k}{n}\right)} \right| = \sqrt{n} \max_{1 \leq k < n} q\left(\frac{k}{n}\right) \left| \hat{\theta}_3(k) - \theta \right|. \tag{4.35}$$

Its asymptotic distribution under H_0 follows from

Corollary 4.3.3. *If the conditions of Theorem 4.3.1 are satisfied with $\nu = \frac{1}{2}$, then if H_0 holds*

$$\lim_{n \to \infty} P\left\{ a(\log n) \frac{1}{\sigma} \max_{1 \leq k < n} \frac{Q_n(k)}{\left(\frac{k}{n}\left(1 - \frac{k}{n}\right)\right)^{\frac{1}{2}}} \leq t + d(\log n) \right\} = \exp(-e^{-t}), \tag{4.36}$$

and

$$\lim_{n \to \infty} P\left\{ a(\log n) \frac{1}{\sigma} \max_{1 \leq k < n} \frac{|Q_n(k)|}{\left(\frac{k}{n}\left(1 - \frac{k}{n}\right)\right)^{\frac{1}{2}}} \leq t + d(\log n) \right\} = \exp(-2e^{-t}), \tag{4.37}$$

for all t, with

$$a(x) = (2 \log x)^{\frac{1}{2}},$$

and

$$d(x) = 2 \log x + \frac{1}{2} \log \log x - \frac{1}{2} \log \pi.$$

This follows immediately from Theorem 4.3.1 as Horváth [62] concluded his Corollary 2.2 from his Theorem 2.1.

The test statistic and the asymptotic distribution under H_0 depend on θ and σ which both are usually not known, but they can be replaced by appropriate estimates. Let us first consider the scale parameter σ. Any estimate $\hat{\sigma}_n^2$ satisfying $\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$ and $\hat{\sigma}_n^2 \geq \varepsilon$ for some $\varepsilon > 0$ and large enough n may be used to replace σ^2 without

changing the asymptotic distribution in Corollary 4.3.3. For instance, we would use Horváth's estimate

$$\hat{\sigma}_n^2 = \int_0^V \left(\int_0^V \tilde{K}(x,y) \frac{1}{1-\hat{H}_n(y-)} dH_n(y) \right)^2 \frac{1}{\left(1-\hat{H}_n(x-)\right)^2} dH_n(x). \tag{4.38}$$

for which we know (compare (2.11) of Horváth [62])

$$|\hat{\sigma}_n^2 - \sigma^2| = O_p\left(\frac{(\log n)^2)}{n^{1/2}} \right), \tag{4.39}$$

or we could use

$$\begin{aligned} \tilde{\sigma}_n^2 &= \int_0^V \left(\int_0^V \tilde{K}(x,y) \hat{\lambda}_{2,n}(y) dy \right)^2 \frac{\hat{\lambda}_{2,n}(x)}{1-\hat{H}_n(x-)} dx \\ &= \int_0^V \left(\int_0^V \tilde{K}(x,y) \frac{h_n(y)}{1-\hat{H}_n(y-)} dy \right)^2 \frac{h_n(x)}{\left(1-\hat{H}_n(x-)\right)^2} dx. \end{aligned} \tag{4.40}$$

which is also a consistent estimate of σ^2 from Corollary 4.2.13 and the Glivenko-Cantelli Theorem.

So, we can consider the following test statistics with asymptotic distribution under H_0 given by Corollary 4.3.3

$$T_n^{(1)} = \max_{1 \le k < n} \frac{1}{\hat{\sigma}_n} \frac{Q_n(k)}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\frac{1}{2}}},$$

and

$$T_n^{(2)} = \max_{1 \le k < n} \frac{1}{\hat{\sigma}_n} \frac{|Q_n(k)|}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\frac{1}{2}}}.$$

The approximated critical values of the test can be given by (4.36) for $T_n^{(1)}$ and by (4.37) for $T_n^{(2)}$. Let $c_\alpha^{(1)}(n)$ and $c_\alpha^{(2)}(n)$ are the $(1-\alpha)$-quantile of the corresponding asymptotic distribution of $T_n^{(1)}$ and $T_n^{(2)}$, respectively. Which are defined as,

$$c_\alpha^{(1)}(n) = \frac{1}{a(\log n)} \left\{ -\log(-\log(1-\alpha)) + d(\log n) \right\}, \tag{4.41}$$

and

$$c_\alpha^{(2)}(n) = \frac{1}{a(\log n)} \left\{ -\log\left(-\frac{\log(1-\alpha)}{2} \right) + d(\log n) \right\}. \tag{4.42}$$

Therefore, the tests are defined by rejecting H_0 if $T_n^{(i)} > c_\alpha^{(i)}(n)$ for $i = 1, 2$. We can calculate the asymptotic critical value by using (4.41) and also simulate finite sample

size critical values.

For unknown θ, we follow Horváth [62] and replace it by the estimate under H_0 $\hat{\theta}_3\left(\lfloor\frac{n}{2}\rfloor\right)$, i.e., we replace $Q_n(k)$ by

$$Q_n^*(k) = \frac{k\,(n-k)}{n^{3/2}}\left(\hat{\theta}_3(k) - \hat{\theta}_3\left(\lfloor\frac{n}{2}\rfloor\right)\right), \quad 1 \le k < n.$$

Then, we have immediately from Theorem 4.3.1

Theorem 4.3.4. *If under H_0, the assumptions of Theorem 4.3.1 are satisfied*

$$\max_{1\le k<n} \frac{\left|Q_n^*(k) - \sigma\left(\Gamma_n\left(\frac{k}{n}\right) - 4\frac{k}{n}\left(1-\frac{k}{n}\right)\Gamma_n\left(\frac{1}{2}\right)\right)\right|}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\nu}} = O_p(n^{\nu-1/2}),$$

where

$$\left\{\Gamma_n(t) - 4t(1-t)\Gamma_n\left(\frac{1}{2}\right), 0 \le t \le 1\right\} \stackrel{\mathcal{D}}{=} \left\{(1-2t)B(t), 0 \le t \le 1\right\},$$

for each n, where $\{B(t), 0 \le t \le 1\}$ is a standard Brownian bridge.

The last relation follows immediately from $\Gamma_n(t) \stackrel{\mathcal{D}}{=} (1-t)W(t) + t\,(W(1) - W(t))$ and $B(t) \stackrel{\mathcal{D}}{=} W(t) - tW(1)$ for a standard Wiener process $W(t)$, $0 \le t \le 1$.

Corollary 4.3.3 is still true for $Q_n^*(k)$ replacing $Q_n(k)$ by the argument given in Horváth [62], such that

Corollary 4.3.5. *If the conditions of Theorem 4.3.1 are satisfied with $\nu = \frac{1}{2}$, then under H_0*

$$\lim_{n\to\infty} P\left\{a(\log n)\frac{1}{\sigma}\max_{1\le k<n}\frac{Q_n^*(k)}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\frac{1}{2}}} \le t + d(\log n)\right\} = \exp(-e^{-t}), \tag{4.43}$$

and

$$\lim_{n\to\infty} P\left\{a(\log n)\frac{1}{\sigma}\max_{1\le k<n}\frac{|Q_n^*(k)|}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\frac{1}{2}}} \le t + d(\log n)\right\} = \exp(-2e^{-t}), \tag{4.44}$$

for all t, with

$$a(x) = (2\log x)^{\frac{1}{2}},$$

and

$$d(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log\pi.$$

The asymptotic distribution of the test statistic under the null hypothesis of no change in hazard functions, when θ is unknown, can be found from Corollary 4.3.5. Thus we obtain the test statistic from (4.43) and (4.44), respectively, as

$$T_n^{(3)} = \max_{1\leq k<n} \frac{1}{\hat{\sigma}_n} \frac{Q_n^*(k)}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\frac{1}{2}}},$$

and

$$T_n^{(4)} = \max_{1\leq k<n} \frac{1}{\hat{\sigma}_n} \frac{|Q_n^*(k)|}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\frac{1}{2}}}.$$

The asymptotic critical values are $c_\alpha^{(1)}(n)$ and $c_\alpha^{(2)}(n)$, which are defined in (4.41) and (4.42), may be used for $T_n^{(3)}$ and $T_n^{(4)}$, respectively.

4.3.2 Change Point Tests with Different Weights

We briefly discuss alternative tests using other weight functions $q_\nu(t) = (t(1-t))^\nu$ for $0 < \nu < \frac{1}{2}$ instead of the standard one with $\nu = \frac{1}{2}$ discussed in the last section. We follow the exposition of Horváth [62]. The mathematical derivations are the same as they are only based on Theorem 2.1 of Horváth [62] and our analogous Theorem 4.3.1, respectively.

Corollary 4.3.6. *Let the assumptions of Theorem 4.3.1 be satisfied for some* $0 < \nu < \frac{1}{2}$, $q_\nu(t) = (t(1-t))^\nu$. *If* H_0 *holds*

$$\frac{1}{\sigma} \sup_{0<t<1} \frac{Q_n(\lceil nt \rceil)}{q_\nu(t)} \xrightarrow{\mathscr{D}} \sup_{0<t<1} \frac{(1-t)W(t) + t(W(1)-W(t))}{q_\nu(t)}, \tag{4.45}$$

and

$$\frac{1}{\sigma} \sup_{0<t<1} \frac{|Q_n(\lceil nt \rceil)|}{q_\nu(t)} \xrightarrow{\mathscr{D}} \sup_{0<t<1} \frac{|(1-t)W(t) + t(W(1)-W(t))|}{q_\nu(t)}. \tag{4.46}$$

Note that, for our choice of $q_\nu(t)$ with $\nu < \frac{1}{2}$, the integrability condition of Corollary 2.1 of Horváth [62] is automatically satisfied. Now, we can consider change point tests based on the statistics

$$T_n^{(5)} = \sup_{0<t<1} \frac{1}{\hat{\sigma}_n} \frac{Q_n(\lceil nt \rceil)}{q_\nu(t)},$$

$$T_n^{(6)} = \sup_{0<t<1} \frac{1}{\hat{\sigma}_n} \frac{|Q_n(\lceil nt \rceil)|}{q_\nu(t)}.$$

The asymptotic critical values are $c_\alpha^{(5)}(n)$ and $c_\alpha^{(6)}(n)$ for test statistics $T_n^{(5)}$ and $T_n^{(6)}$, respectively, where $c_\alpha^{(5)}(n)$ is the $(1-\alpha)$-quantile of $\sup_{0<t<1} \frac{(1-t)W(t)+t(W(1)-W(t))}{q(t)}$ and $c_\alpha^{(6)}(n)$ is the $(1-\alpha)$-quantile of $\sup_{0<t<1} \frac{|(1-t)W(t)+t(W(1)-W(t))|}{q(t)}$.

If θ is unknown, we can consider in analogy to $T_n^{(3)}$ and $T_n^{(4)}$ the following test statistics

$$T_n^{(7)} = \sup_{0<t<1} \frac{1}{\hat{\sigma}_n} \frac{Q_n^*(\lceil nt \rceil)}{q_\nu(t)},$$

and

$$T_n^{(8)} = \sup_{0<t<1} \frac{1}{\hat{\sigma}_n} \frac{|Q_n^*(\lceil nt \rceil)|}{q_\nu(t)}.$$

The critical values can be calculated by simulation from a reasonable approximating-model. Instead, we could also use the asymptotic values based on Theorem 4.3.4 following the discussion of Horváth [62] at the end of his Section 2. Suppose, $c_\alpha^{(7)}(n)$ and $c_\alpha^{(8)}(n)$ are the approximate critical values respectively for test statistics $T_n^{(7)}$ and $T_n^{(8)}$, where $c_\alpha^{(7)}(n)$ is the $(1-\alpha)$-quantile of $\sup_{0<t<1} \frac{\{(1-2t)B(t)\}}{q_\nu(t)}$ and $c_\alpha^{(8)}(n)$ is the $(1-\alpha)$-quantile of $\sup_{0<t<1} \frac{|(1-2t)B(t)|}{q_\nu(t)}$.

4.3.3 Proof of Theorem 4.3.1

We mainly follow the proof of Theorem 2.1 of Horváth [62] and, for ease of reference, use the same notation. First, we split

$$\begin{aligned}
\hat{\theta}_3(k) - \theta &= A_1(k) + A_2(k) + A_3(k), \quad \text{with} \\
A_1(k) &= \int_0^V \int_0^V \frac{\tilde{K}(x,y)}{(1-H(x))\,(1-H(y))} h_k(x) h_k^*(y) dx dy - \theta \\
&= \int_0^V \int_0^V \tilde{K}_1(x,y) h_k(x) h_k^*(y) dx dy - \theta \\
A_2(k) &= \int_0^V \int_0^V \frac{\tilde{K}(x,y)\left(\hat{H}_k^*(y-) - H(y)\right)}{(1-H(x))\,(1-H(y))\left(1-\hat{H}_k^*(y-)\right)} h_k(x) h_k^*(y) dx dy \\
A_3(k) &= \int_0^V \int_0^V \frac{\tilde{K}(x,y)\left(\hat{H}_k(x-) - H(x)\right)}{(1-H(x))\left(1-\hat{H}_k(x-)\right)\left(1-\hat{H}_k^*(y-)\right)} h_k(x) h_k^*(y) dx dy.
\end{aligned}$$

Note that, as $1-H(x) \geq 1-H(V) > 0$ for $0 \leq x \leq V$, the kernel $\tilde{K}_1(x,y)$ is also symmetric and uniformly bounded as $\tilde{K}$ itself. In comparison to Horváth's proof, also note that we have assumed that $H(x)$ is continuous in $[0,V]$, such that $H(x-) = H(x)$ etc.

$A_1(k)$ will be the dominant part of $\hat{\theta}_3(k) - \theta$. We first show in a) that the other two terms are of small order, and, then, in b) investigate the asymptotic behaviour of $A_1(k)$.

a) As $X_1, \ldots, X_n$ are i.i.d., we can apply the law of the iterated logarithm (compare, e.g., Section 1.10 of Serfling [128]) to replace $\hat{H}_k$, $\hat{H}_k^*$ in the denominators of $A_2(k)$, $A_3(k)$ by their limits. More precisely, we have as in (4.2) and (4.3) of Horváth [62]

$$\max_{1\leq k<n} |A_2(k) - A_4(k)| \left(\frac{\log\log k}{k}\right)^{-1} = O_p(1)$$
$$\max_{1\leq k<n} |A_3(k) - A_5(k)| \left\{\left(\frac{\log\log k}{k}\right)^{\frac{1}{2}} \left[\left(\frac{\log\log k}{k}\right)^{\frac{1}{2}} + \left(\frac{\log\log(n-k)}{(n-k)}\right)^{\frac{1}{2}}\right]\right\}^{-1} = O_p(1)$$

with

$$A_4(k) = \int_0^V \int_0^V \tilde{K}_4(x,y) \left(\hat{H}_k^*(y-) - H(y)\right) h_k(x) h_k^*(y) dxdy$$
$$A_5(k) = \int_0^V \int_0^V \tilde{K}_5(x,y) \left(\hat{H}_k(x-) - H(x)\right) h_k(x) h_k^*(y) dxdy,$$

where

$$\tilde{K}_4(x,y) = \frac{\tilde{K}(x,y)}{(1-H(x))(1-H(y))^2}, \quad \tilde{K}_5(x,y) = \frac{\tilde{K}(x,y)}{(1-H(x))^2(1-H(y))}.$$

Recall from Section 4.2.3 that $h_n(x) \xrightarrow{P} \tilde{h}(x) = (1-G(x))f(x)$ for $n \to \infty$. As the next step, we want to replace $h_k(x)$ and $h_k^*(y)$ in $A_4(k)$ by respectively $h(x)$ and $h(y)$. For that purpose, we split $A_4(k) = A_6(k) + A_7(k) + A_8(k)$ and show that the last two terms are negligible, where

$$A_6(k) = \int_0^V \int_0^V \tilde{K}_4(x,y) \left(\hat{H}_k^*(y-) - H(y)\right) \tilde{h}(x)\tilde{h}(y) dxdy$$
$$A_7(k) = \int_0^V \int_0^V \tilde{K}_4(x,y) \left(\hat{H}_k^*(y-) - H(y)\right) \left(h_k(x) - \tilde{h}(x)\right) h_k^*(y) dxdy$$
$$A_8(k) = \int_0^V \int_0^V \tilde{K}_4(x,y) \left(\hat{H}_k^*(y-) - H(y)\right) \tilde{h}(x) \left(h_k^*(y) - \tilde{h}(y)\right) dxdy.$$

First, we consider $A_7(k)$. Uniformly in $0 \leq y \leq V$, $1 \leq k < n$

$$\int_0^V \tilde{K}_4(x,y) \left(Eh_k(x) - \tilde{h}(x)\right) dx = O(b), \tag{4.47}$$

as, from our assumptions of $\tilde{K}$ and V, $\tilde{K}_4$ is uniformly bounded in x and y, and as $Eh_k(x) - \tilde{h}(x) = O(b)$ uniformly in x and k from the proof of Theorem 4.2.5 with $\alpha = 1$. As

$$\begin{aligned}&\int_0^V \tilde{K}_4(x,y)\left(h_k(x) - Eh_k(x)\right)dx\\ &= \frac{1}{b}\sum_{i=1}^{k}\tilde{K}_4(x,y)\frac{1}{b}\left(K\left(\frac{x-X_i}{b}\right)\delta_i - EK\left(\frac{x-X_i}{b}\right)\delta_i\right)dx,\end{aligned}$$

is a mean of i.i.d. bounded, zero-mean random variables, we get from Hoeffding's inequality (compare, e.g., Section 2.3.3 of Serfling [128]) similar to (4.5) of Horváth [62]

$$P\left(\sqrt{k}\left|\int_0^V \tilde{K}_4(x,y)\left(h_k(x) - Eh_k(x)\right)dx\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\tilde{C}_4}\right), \tag{4.48}$$

where $\tilde{C}_4$ is a uniform bound on $|\tilde{K}_4(x,y)|$ and where we have used

$$\left|\int_0^V \tilde{K}_4(x,y)\frac{1}{b}K\left(\frac{x-X_i}{b}\right)\delta_i dx\right| \leq \tilde{C}_4\int\frac{1}{b}K\left(\frac{x-X_i}{b}\right)dx = \tilde{C}_4,$$

as K is a probability density. However, we need the above exponential inequality for

$$\begin{aligned}\tilde{D}_k(y) &= \int_0^V \tilde{K}_4(x,y)\left(h_k(x) - \tilde{h}(x)\right)dx = \int_0^V \tilde{K}_4(x,y)\left(h_k(x) - Eh_k(x)\right)dx + O(b)\\ &= D_k(y) + O(b),\end{aligned}$$

uniformly in y and k by (4.47). From (4.48), we get for some $c > 0$

$$\begin{aligned}P\left(\sqrt{k}\left|\tilde{D}_k(y)\right| > t\right) &\leq P\left(\sqrt{k}\left|D_k(y)\right| > t - \sqrt{k}O(b)\right)\\ &\leq 2\exp\left(-\frac{(t-\sqrt{k}O(b))^2}{2\tilde{C}_4}\right)\\ &\leq 2\exp\left(-\frac{t^2}{2\tilde{C}_4} + \frac{\sqrt{n}O(b)t}{\tilde{C}_4}\right) \leq 2\exp\left(-ct^2\right),\end{aligned} \tag{4.49}$$

for all large enough n and large t, as from $b = O\left(n^{\nu-1}\right)$, we have $nb^2 = O(1)$. We need an upper bound for the supremum over y. We use the standard approach and consider an equidistant gird $y_l = (l - \frac{1}{2})\frac{V}{m}$, $l = 1, \ldots, m$, such that for every y in $[0, V]$ there is a y_l with $|y - y_l| \leq \frac{V}{m}$. Then, uniformly in k,

$$\begin{aligned}\sup_{0\leq y\leq V}|\tilde{D}_k(y)| &= \sup_{1\leq l\leq m}\sup_{|y-y_l|\leq\frac{V}{m}}|\tilde{D}_k(y)| = \sup_{1\leq l\leq m}\sup_{|y-y_l|\leq\frac{V}{m}}|\tilde{D}_k(y_l) + \tilde{D}_k(y) - \tilde{D}_k(y_l)|\\ &\leq \sup_{1\leq l\leq m}|\tilde{D}_k(y_l)| + \tilde{L}\frac{V}{m},\end{aligned} \tag{4.50}$$

where we have used that $\tilde{D}_k(y)$ is Lipschitz continuous with some constant $\tilde{L}$. The latter follows from Lipschitz continuity of $\tilde{K}(x,y)$ and our assumption that F and G have bounded densities on $[0,V]$, which implies that H has a bounded density on $[0,V]$ too. Noting that $1-H(x) \geq 1-H(V) > 0$ on $[0,V]$, we get that $\tilde{K}_4(x,y)$ is Lipschitz too, i.e., for some constant L_4

$$\left|\tilde{K}_4(x,y) - \tilde{K}_4(x,z)\right| \leq L_4|y-z|, \qquad \text{for all } x,y,z \in [0,V],$$

such that

$$\begin{aligned}\left|\tilde{D}_k(y) - \tilde{D}_k(z)\right| &\leq \int_0^V L_4 \left|y-z\right| \left|h_k(x) - \tilde{h}(x)\right| dx \\ &\leq L_4 \int_0^V \left(h_k(x) + \tilde{h}(x)\right) dx \left|y-z\right| \leq 2L_4 \left|y-z\right|\end{aligned}$$

where we have used $\tilde{h}(x) \leq f(x)$ and K is a probability density and

$$\int_0^V h_k(x)dx \leq \frac{1}{k}\sum_{i=1}^{k} \int_0^V \frac{1}{b} K\left(\frac{x-X_i}{b}\right) dx \leq 1. \tag{4.51}$$

Using the subadditivity of probabilities, we get from(4.49)

$$P\left(\max_{1\leq k<n} \max_{1\leq l\leq m} \sqrt{k}\left|\tilde{D}_k(y_l)\right| > t\right) \leq 2nm \exp\left(-ct^2\right),$$

and, therefore, with $n=m$

$$\max_{1\leq k<n} \max_{1\leq l\leq n} \sqrt{k}\left|\int_0^V \tilde{K}_4(x,y_l)\left(h_k(x) - \tilde{h}(x)\right) dx\right| = O_p\left(\sqrt{\log n}\right),$$

using (4.50), we immediately have also

$$\max_{1\leq k<n} \max_{0\leq y\leq V} \sqrt{k}\left|\int_0^V \tilde{K}_4(x,y)\left(h_k(x) - \tilde{h}(x)\right) dx\right| = O_p\left(\sqrt{\log n}\right).$$

As, using (4.51) for the last factor of the first line

$$\begin{aligned}|A_7(k)| &\leq \sup_y \left|\hat{H}_k^*(y-) - H(y)\right| \sup_y \left|\int_0^V \tilde{K}_4(x,y)\left(h_k(x) - \tilde{h}(x)\right) dx\right| \int_0^V h_k^*(y)dy \\ &\leq \sup_y \left|\hat{H}_k^*(y-) - H(y)\right| \sup_y \left|\int_0^V \tilde{K}_4(x,y)\left(h_k(x) - \tilde{h}(x)\right) dx\right|,\end{aligned}$$

we get again from the law of iterated logarithm applied to $\hat{H}_k^*$ in analogy to (4.6) of Horváth [62]

$$\max_{1\leq k<n} |A_7(k)| \sqrt{k(n-k)} = O_p\left(\sqrt{\log n}\sqrt{\log\log n}\right).$$

$A_8(k)$ is of even smaller order. We can apply the same arguments as Horváth [62] in deriving his relation (4.7) to derive

$$\max_{1\le k<n} |A_8(k)| \frac{n-k}{\log(n-k)} = O_p(1),$$

by using Lemma B in Section 5.2.2 of Serfling [128] on U-statistics. Note, in particular, that $A_8(k) = \int_0^V \tilde{h}(x)L(x)dx$, where $\tilde{h}(x) \le f(x)$, and

$$\begin{aligned} L(x) &= \int_0^V \tilde{K}_4(x,y)\left(\hat{H}_k^*(y-) - H(y)\right)\left(h_k^*(y) - \tilde{h}(y)\right)dy \\ &= \frac{1}{(n-k)^2}\sum_{i,j=k+1}^{n}\int_0^V \tilde{K}_4(x,y)\left(I(X_i<y) - H(y)\right)\left(\frac{1}{b}K\left(\frac{x-X_j}{b}\right)\delta_j - \tilde{h}(y)\right)dy. \end{aligned}$$

Omitting the diagonal $i=j$, this becomes a U-statistic

$$L^{'}(x) = \frac{2}{(n-k)^2}\sum_{k+1\le i<j\le n} r^{'}\left((X_i,\delta_i),(X_j,\delta_j)\right) = \frac{1}{(n-k)^2}\sum_{k+1\le i<j\le n} r\left((X_i,\delta_i),(X_j,\delta_j)\right)$$

where

$$\begin{aligned} r^{'}\left((X,\delta),(X^{'},\delta^{'})\right) &= \int_0^V \tilde{K}_4(x,y)\left(I(X<y) - H(y)\right)\left(\frac{1}{b}K\left(\frac{y-X^{'}}{b}\right)\delta^{'} - \tilde{h}(y)\right)dy \\ r\left((X,\delta),(X^{'},\delta^{'})\right) &= \frac{1}{2}\left[r^{'}\left((X,\delta),(X^{'},\delta^{'})\right) + r^{'}\left((X^{'},\delta^{'}),(X,\delta)\right)\right]. \end{aligned}$$

Note that, r is symmetric and, moreover, as (X_i,δ_i), (X_j,δ_j) are independent for $i\neq j$ and $EI(X_i>y) = H(y)$ due to continuity of H,

$$Er\left((X_i,\delta_i),(X_j,\delta_j)\right) = 0, \qquad \text{for } i\neq j,$$

i.e., $EL^{'}(x) = 0$, too. Also, for fixed b, r is a bounded random variable, as, in particular, $\tilde{K}$ and K are bounded.

Combining the results on $A_2(k)$, $A_4(k)$ and the components of $A_4(k)$, we get as in Horváth [62]

$$\max_{1\le k<n}\frac{k(n-k)}{n^{3/2}}|A_2(k) - A_6(k)|\left[\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-\nu} = O_p\left(n^{\nu-\frac{1}{2}}\right). \tag{4.52}$$

Analogously, we get as in (4.9) of Horváth [62]

$$\max_{1\le k<n}\frac{k(n-k)}{n^{3/2}}|A_3(k) - A_9(k)|\left[\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-\nu} = O_p\left(n^{\nu-\frac{1}{2}}\right), \tag{4.53}$$

with

$$A_9(k) = \int_0^V \int_0^V \tilde{K}_5(x,y) \left(\hat{H}_k(x-) - H(x) \right) \tilde{h}(x)\tilde{h}(y)dxdy.$$

b) To make the bias induced by kernel smoothing explicit, we write

$$\begin{aligned} A_1(k) &= \frac{1}{k(n-k)} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \int_0^V \int_0^V \tilde{K}_1(x,y) \frac{1}{b^2} K\left(\frac{x - X_i}{b} \right) K\left(\frac{y - X_j}{b} \right) \delta_i \delta_j dxdy - \theta \\ &= \frac{1}{k(n-k)} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(K^*(Z_i, Z_j) - \theta^* \right) + (\theta^* - \theta) = U_k + (\theta^* - \theta), \end{aligned}$$

where $Z_i = (T_i, \delta_i)$, $i = 1, \ldots, n$, and for $z = (z_1, z_2)$, $z' = (z_1', z_2')$

$$K^*(z, z') = \int_0^V \int_0^V \tilde{K}_1(x,y) \frac{1}{b^2} K\left(\frac{x - z_1}{b} \right) K\left(\frac{y - z_1'}{b} \right) z_2 z_2' dxdy$$

$\theta^* = EK^*(Z_i, Z_j)$. Then, U_k is a U-statistic with symmetric kernel K^*, mean 0 and finite second moment, as $\tilde{K}_1$ and K are bounded. So, we can conclude as Horváth [62] from Hall [58]'s invariance principle for U-statistics that

$$\begin{aligned} \max_{1 \le k < n} \Bigg| \sum_{i=1}^{k} \sum_{j=k+1}^{n} K^*(Z_i, Z_j) - k(n-k) \Bigg\{ & \int_0^V \int_0^V \tilde{K}_1(x,y) \left(h_k(x) - \tilde{h}(x) \right) \tilde{h}(y) dxdy \\ & + \int_0^V \int_0^V \tilde{K}_1(x,y) \left(h_k^*(y) - \tilde{h}(y) \right) \tilde{h}(x) dxdy \Bigg\} \Bigg| \\ & = O_p(n). \end{aligned}$$

As in Horváth [62], this relation together with (4.52) and (4.53) implies

$$\max_{1 \le k < n} \left| \frac{k(n-k)}{n^{3/2}} \left(\hat{\theta}_3(k) - \theta^* \right) - \tilde{Q}_n^*(k) \right| \left(\frac{k}{n} \left(1 - \frac{k}{n} \right) \right)^{-\nu} = O_p\left(n^{\nu - \frac{1}{2}} \right). \qquad (4.54)$$

with

$$\begin{aligned} \tilde{Q}_n^*(k) = \frac{k(n-k)}{n^{3/2}} \{ & \int_0^V \int_0^V \tilde{K}_1(x,y) \left(h_k(x) - \tilde{h}(x) \right) \tilde{h}(y) dxdy \\ & + \int_0^V \int_0^V \tilde{K}_1(x,y) \left(h_k^*(y) - \tilde{h}(y) \right) \tilde{h}(x) dxdy \\ & + \int_0^V \int_0^V \tilde{K}_4(x,y) \left(\hat{H}_k^*(y-) - H(y) \right) \tilde{h}(x)\tilde{h}(y) dxdy \\ & + \int_0^V \int_0^V \tilde{K}_5(x,y) \left(H_k(x-) - H(x) \right) \tilde{h}(x)\tilde{h}(y) dxdy \}. \end{aligned}$$

$\tilde{Q}_n^*(k)$ is quite similar to $Q_n^*(k)$ in Horváth [62]. We would like to replace $\tilde{Q}_n^*(k)$ by $Q_n^*(k)$ in (4.54) to make immediate use of the final part of the proof of Theorem 2.1 of Horváth [62].

First, note that the 3rd and 4th summands of $\tilde{Q}_n^*(k)$ and $Q_n^*(k)$ are identical. The 1st and 2nd summands have exactly the same structure, as $\tilde{K}_1$ is symmetric, so it is sufficient to consider the first one only. Looking at the difference between the first two summands of $\tilde{Q}_n^*(k)$ and $Q_n^*(k)$, we get

$$\left|\int_0^V \int_0^V \tilde{K}_1(x,y)\tilde{h}(y)dy\,(dH_k(x) - h_k(x)dx)\right| = \left|\int_0^V L_1(x)\,(dH_k(x) - h_k(x)dx)\right|$$

with

$$L_1(x) = I\,(x \le V)\int_0^V \tilde{K}_1(x,y)\tilde{h}(y)dy = I\,(x \le V)\,J_1(x).$$

Note that, by symmetry and Lipschitz continuity of $\tilde{K}$

$$\left|\tilde{K}(x,y) - \tilde{K}(x^{'},y)\right| \le \tilde{C}\left|x - x^{'}\right|, \quad x, x^{'}, y \in [0, V+\Delta],$$

and as $1-H(x)$ also satisfies a Lipschitz condition on $[0, V+\Delta]$ due to the boundedness assumption on the densities of F and G, we have that

$$\left|\tilde{K}_1(x,y) - \tilde{K}_1(x^{'},y)\right| \le \tilde{C}_1\left|x - x^{'}\right|, \quad x, x^{'}, y \in [0, V+\Delta],$$

too for a suitable constant $\tilde{C}_1$ and, hence $|J_1(x) - J_1(x^{'})| \le \tilde{C}_1\left|x - x^{'}\right|$ too as $0 \le \tilde{h}(y) \le f(y)$. Therefore, for $x \le x^{'}$

$$\begin{aligned}\left|L_1(x) - L_1(x^{'})\right| &\le \left|I\,(x \le V) - I\left(x^{'} \le V\right)\right| |J_1(x)| + \left|J_1(x) - J_1(x^{'})\right| \\ &\le C_J I\left(x \le V < x^{'}\right) + \tilde{C}_1\left|x - x^{'}\right|,\end{aligned}$$

where C_J is an upper bound for $|J_1(x)|$, which exists due to the boundedness of $\tilde{K}_1$. Using this inequality and substitution $x = bu + T_i$, we have, as $\int K(u)du = 1$, and $\delta_i \in [0,1]$

$$\begin{aligned}&\left|\int_0^\infty L_1(x)\,(dH_k(x) - h_k(x)dx)\right| = \left|\frac{1}{k}\sum_{i=1}^k \delta_i\left[L_1(T_i) - \frac{1}{b}\int_0^\infty L_1(x)K\left(\frac{x-T_i}{b}\right)dx\right]\right| \\ &= \left|\frac{1}{k}\sum_{i=1}^k \delta_i \int [L_1(T_i) - L_1(T_i+bu)]\,K(u)du\right| \\ &\le \frac{1}{k}\sum_{i=1}^k \int \left[C_J I\,(T_i \le V < T_i + bu) + C_J I\,(T_i + bu \le V < T_i) + \tilde{C}_1 bu\right] K(u)du \\ &= O_p(b),\end{aligned}$$

from Markov's inequality as, e.g.,

$$EI\left(T_i \leq V < T_i + bu\right) = F(V) - F(V - bu) = O(b),$$

as F has a bounded density f.

We conclude, as $k(n-k)$ becomes maximal for $k \approx \frac{n}{2}$

$$\begin{aligned}\max_{1\leq k<n} \left|\tilde{Q}_n^*(k) - Q_n^*(k)\right| \left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{-\nu} &= \max_{1\leq k<n} \frac{k(n-k)}{n^{3/2}} O_p(b) \left(\frac{k(n-k)}{n^2}\right)^{-\nu} \\ &= O_p\left(\sqrt{n}b\right) = O_p\left(n^{\nu-\frac{1}{2}}\right), \qquad (4.55)\end{aligned}$$

if $b = O\left(n^{\nu-1}\right)$. Hence, we may replace $\tilde{Q}_n^*(k)$ by $Q_n^*(k)$ in (4.54).

Finally, we show that we may replace θ^* by θ in (4.54). Recall that $\tilde{h}(x) = (1 - G(x))f(x)$ and

$$\begin{aligned}\theta &= \int_0^V \int_0^V \tilde{K}(x,y)\lambda(x)\lambda(y)dxdy = \int_0^V \int_0^V \tilde{K}_1(x,y)\tilde{h}(x)\tilde{h}(y)dxdy \\ \theta^* &= \int_0^V \int_0^V \tilde{K}_1(x,y)E\left(h_k(x)h_k^*(y)\right)dxdy = \int_0^V \int_0^V \tilde{K}_1(x,y)Eh_k(x)Eh_k^*(y)dxdy,\end{aligned}$$

as $h_k(x)$ depends only on (X_i, δ_i), $1 \leq i \leq k$, and $h_k^*(y)$ on (X_j, δ_j), $k+1 \leq j \leq n$, and hence are independent. From the proof of Theorem 4.2.5 with $\alpha = 1$, we have $Eh_k(x) - \tilde{h}(x) = O(b)$ uniformly in x and k. Using the boundedness of $\tilde{K}_1(x,y)$ on $[0,V] \times [0,V]$, we immediately get $\theta^* - \theta = O(b)$. As in (4.55), we get that replacing θ^* by θ in (4.54) has a negligible effect. Finally, we conclude that, from (4.54), we have

$$\max_{1\leq k<n} \left|Q_n(k) - Q_n^*(k)\right| \left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{-\nu} = O_p\left(n^{\nu-\frac{1}{2}}\right),$$

where $Q_n^*(k)$ is defined exactly as on p.239 of Horváth [62], and our result follows from the rest of the proof of Theorem 2.1 of Horváth [62]. Note that our formula looks slightly different as we consider only nonnegative T_j and C_j, and as under our assumptions $H(x)$ is continuous. □

4.4 Consistency of the Change Point Test

In this section, we derive the consistency of the proposed tests under alternatives, which means that the test have asymptotic power one. For instance, to show consistency of a test statistic $T_n^{(i)}$ having an asymptotic critical value $c_\alpha^{(i)}(n)$ it suffices to

show

$$P\left(T_n^{(i)} > c_\alpha^{(i)}(n)\right) \to 1, \quad \text{under } H_1,$$

as $n \to \infty$.

To show consistency of the change point test, we first have to investigate the behaviour of $\hat{\lambda}_2(x)$ under the alternative. Let H_1 be satisfied for some $k^* < n$. We consider the case where $\frac{k^*}{n} \to \tau$, $0 < \tau < 1$, for $n \to \infty$. Let us first consider the numerator of $\hat{\lambda}_2$

$$\begin{aligned} h_n(x) &= \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right) \delta_i \\ &= \frac{k^*}{n} \frac{1}{k^*} \sum_{i=1}^{k^*} \frac{1}{b} K\left(\frac{x - X_i}{b}\right) \delta_i + \frac{n - k^*}{n} \frac{1}{n - k^*} \sum_{i=k^*+1}^{n} \frac{1}{b} K\left(\frac{x - X_i}{b}\right) \delta_i. \end{aligned}$$

As from our assumption on k^*, we immediately have from Corollary 4.2.9

$$h_n(x) \xrightarrow{a.s.} \tau \tilde{h}(x) + (1 - \tau)\tilde{h}^*(x), \quad \text{uniformly in } 0 \le x \le V,$$

where $\tilde{h}(x) = (1 - G(x))f(x)$, $\tilde{h}^*(x) = (1 - G^*(x))f^*(x)$, and F, G and H denote the distribution functions of T_i, C_i and X_i before the change point and F^*, G^* and H^* after the change point, respectively. Analogously, we write f and f^* for densities of T_i, and λ and λ^* for the hazard functions respectively of the two subsamples, i.e., before and after the change point.

For the denumerator of $\hat{\lambda}_2$, we have

$$\begin{aligned} 1 - \hat{H}_n(x-) = \frac{1}{n} \sum_{i=1}^{n} I\left(X_i \ge x\right) &= \frac{k^*}{n} \frac{1}{k^*} \sum_{i=1}^{k^*} I\left(X_i \ge x\right) + \frac{n - k^*}{n} \frac{1}{n - k^*} \sum_{i=k^*+1}^{n} I\left(X_i \ge x\right) \\ &\xrightarrow{a.s.} \tau(1 - H(x)) + (1 - \tau)(1 - H^*(x)), \quad \text{uniformly in } x, \end{aligned}$$

by the Glivenko-Cantelli Theorem. Therefore, again uniformly in $0 \le x \le V$

$$\begin{aligned} \hat{\lambda}_2(x) = \frac{\hat{h}_n(x)}{1 - \hat{H}_n(x-)} &\xrightarrow{a.s.} \frac{\tau \tilde{h}(x) + (1 - \tau)\tilde{h}^*(x)}{\tau(1 - H(x)) + (1 - \tau)(1 - H^*(x))} \\ &= r_\tau \lambda(x) + (1 - r_\tau)\lambda^*(x), \end{aligned}$$

with

$$r_\tau = \frac{\tau(1 - H(x))}{\tau(1 - H(x)) + (1 - \tau)(1 - H^*(x))}.$$

To consider a specific test, we use the statistic $T_n^{(4)}$, which is based on Corollary 4.3.3

$$T_n^{(4)} = \max_{1\le k<n} \frac{1}{\hat{\sigma}_n}\sqrt{\frac{k}{n}\left(1-\frac{k}{n}\right)}\sqrt{n}\left(\hat{\theta}_3(k) - \hat{\theta}_3(\lfloor\frac{n}{2}\rfloor)\right).$$

Let $k = \lceil \gamma n \rceil$, $\tau < \gamma < 1$. Then, all data contributing to $h_k^*(x)$ and $\hat{H}_k^*(x-)$ belong to the second subsample, and $\hat{\lambda}_2^*(y) \xrightarrow{a.s.} \lambda^*(y)$ uniformly in $0 \le y \le V$. $\hat{\lambda}_2(x)$ will be a mixture of $\lambda(x)$ and $\lambda^*(x)$ asymptotically as discussed above.

$$\hat{\lambda}_2(x) \xrightarrow{a.s.} r_\tau(x)\lambda(x) + (1 - r_\tau(x))\lambda^*(x) = \lambda_\gamma(x),$$

again uniformly in $0 \le x \le V$. Hence, we have

$$\hat{\theta}_3(k) = \int_0^V \int_0^V \tilde{K}(x,y)\hat{\lambda}_2(x)\hat{\lambda}_2^*(y)dxdy \xrightarrow{a.s.} \int_0^V \int_0^V \tilde{K}(x,y)\lambda_\gamma(x)\lambda^*(y)dxdy.$$

If, e.g., $\tau < \frac{1}{2}$, then analogously for specifically $\gamma = \frac{1}{2}$, we have

$$\hat{\theta}_3(k) - \hat{\theta}_3\left(\lfloor\frac{n}{2}\rfloor\right) \xrightarrow{a.s.} \int_0^V \int_0^V \tilde{K}(x,y)\left[\lambda_\gamma(x) - \lambda_{\frac{1}{2}}(x)\right]\lambda^*(y)dxdy = \Delta_\gamma.$$

As, additionally, the weight $\left[\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{\frac{1}{2}} \to \gamma(1-\gamma)$ for $n \to \infty$, we have that for any $\varepsilon > 0$ and all large enough n

$$D_n(k) = \sqrt{\frac{k}{n}\left(1-\frac{k}{n}\right)}\left|\hat{\theta}_3(k) - \hat{\theta}_3\left(\lfloor\frac{n}{2}\rfloor\right)\right| \ge |\Delta_\gamma| - \varepsilon,$$

and

$$\max_{1\le k<n} D_n(k) \ge \max_{\tau<\gamma<1} |\Delta_\gamma| - \varepsilon.$$

As $\hat{\sigma}_n^2$ will be $O_p(1)$ and bounded away from 0 under H_1 too, we get that $T_n^{(4)} = \frac{\sqrt{n}}{\hat{\sigma}_n}\max_{1\le k<n} D_n(k) \xrightarrow{P} \infty$ with rate $\sqrt{n}$ if $|\Delta_\gamma| > 0$ for some $\tau < \gamma < 1$, i.e., the power of the test will converge to 1 for $n \to \infty$, and the test is consistent. The case $\tau \ge \frac{1}{2}$ can be handled analogously. Note that the condition $\Delta_\gamma \neq 0$ depends on $\tilde{K}, V, \lambda, \lambda^*, H$, and H^*, and is an identifiability condition which guarantees that the test will be able to distinguish H_0 and H_1 for $n \to \infty$.

4.5 Test Statistic and Asymptotics under H_0 for Antisymmetric Kernels

In this section, we consider change point statistics with antisymmetric kernel $\tilde{K}$ instead of symmetric ones. An antisymmetric kernel function has the form

$$\tilde{K}(x,y) = -\tilde{K}(y,x). \tag{4.56}$$

Stute [136] suggested some estimators of the change point k^* based on U-statistics like (4.5), using antisymmetric kernels.

$$\tilde{\theta}(k) = \frac{k(n-k)}{n^2} \int_0^V \int_0^V \frac{\tilde{K}(x,y)}{\left(1-\hat{H}_n(x-)\right)\left(1-\hat{H}_n(y-)\right)} dH_k(x) dH_k^*(y).$$

Setting $r_n(\frac{k}{n}) = \tilde{\theta}_n(k)$ and interpolating to get a function on $[0,1]$, he proved the weak convergence of $n^{1/2}(r_n(nt)-\theta)$ in $\mathscr{D}[0,1]$. He also proposed $(t(1-t))^{1/2}$ as the weight function. In case of antisymmetric kernels, Horváth [62] pointed out that under H_0 estimators $\hat{\theta}(k)$ and $\tilde{\theta}(k)$ have different asymptotic distributions which, however, may be derived by essentially the same kind of argument. Therefore, we get the following analogous to Theorem 4.3.1 for antisymmetric kernels.

Let $\tilde{\theta}_3(k)$ be defined by $\hat{\theta}_3(k)$ in Section 4.3, but using an antisymmetric kernel $\tilde{K}(x,y)$. Then, we get

$$\tilde{Q}_n(k) = \frac{k(n-k)}{n^{\frac{3}{2}}} \left(\tilde{\theta}_3(k) - \theta\right).$$

Theorem 4.5.1. *Let the assumptions of Theorem 4.3.1 be fulfilled except that $\tilde{K}(x,y) = -\tilde{K}(y,x)$. Then, under the hypothesis H_0, for $n \to \infty$, $nb \to \infty$ and $b = O\left(n^{\nu-\frac{1}{2}}\right)$ for some $0 < \nu \leq \frac{1}{2}$, we have for a sequence of Brownian bridges $\{B_n(t), 0 \leq t \leq 1\}$*

$$\max_{1\leq k<n} \left|\tilde{Q}_n(k) - \sigma B_n(\frac{k}{n})\right| \left[\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-\nu} = O_p(n^{\nu-\frac{1}{2}})$$

with σ^2 as in Theorem 4.3.1.

Proof. We argue as Horváth [62] in the proof of his Theorem 3.1. First note that part a) of the proof of Theorem 4.3.1 may be used without any change as we have not exploited symmetry of $\tilde{K}$ there. At the beginning of part b) of that proof, we have introduced the kernel K^* which now is antisymmetric such that we cannot use Hall

[58]'s result. Instead we use Janson and Wichura [69] to conclude with $Z_i = (T_i, \delta_i)$ again

$$\max_{1\le k<n}\left|\sum_{i=1}^{k}\sum_{j=k+1}^{n} K^*(Z_i,Z_j) - k(n-k)\left\{\int_0^V\int_0^V \tilde{K}_1(x,y)\left(h_k(x)-\tilde{h}(x)\right)\tilde{h}(y)dxdy + \int_0^V\int_0^V \tilde{K}_1(x,y)\left(h_k^*(y)-\tilde{h}(y)\right)\tilde{h}(x)dydx\right\}\right| = O_p(n),$$

i.e., we have the same kind of approximation result as in the proof of Theorem 4.3.1. In particular, we get

$$\max_{1\le k<n}\left|\tilde{Q}_n(k)-\tilde{Q}_n^*(k)\right|\left[\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-\nu} = O_p(n^{\nu-\frac{1}{2}}),$$

where $\tilde{Q}_n^*(k)$ is defined as in the proof of Theorem 4.3.1, but now with an antisymmetric kernel.

As in the proof of Theorem 4.3.1, we can show with exactly the same arguments as in the proof of Theorem 4.3.1 that

$$\max_{1\le k<n}\left|\tilde{Q}_n^*(k)-\hat{Q}_n(k)\right|\left[\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-\nu} = O_p(n^{\nu-\frac{1}{2}}),$$

where $\hat{Q}_n(k)$ is called $Q_n^*(k)$ in the proof of Theorem 3.1 of Horváth [62]. Again, with the same arguments as in the proof of Theorem 4.3.1, we finally get

$$\max_{1\le k<n}\left|\tilde{Q}_n(k)-\hat{Q}_n(k)\right|\left[\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-\nu} = O_p(n^{\nu-\frac{1}{2}}),$$

and the rest of the result follows from Horváth [62]'s derivation of the asymptotics of $\hat{Q}_n(k)$. □

Therefore, for $\nu = \frac{1}{2}$ the asymptotic distribution of the test statistic under H_0 using Corollary 3.2 of Horváth [62] is summarized in Corollary 4.5.2.

Corollary 4.5.2. *If the conditions of Theorem 4.5.1 are satisfied, then*

$$\lim_{n\to\infty} P\left\{a(\log n)\frac{1}{\sigma}\max_{1\le k<n}\frac{Q_n(k)}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\frac{1}{2}}} \le t + d(\log n)\right\} = \exp(-e^{-t}), \qquad (4.57)$$

and

$$\lim_{n\to\infty} P\left\{ a(\log n)\frac{1}{\sigma}\max_{1\le k<n}\frac{|Q_n(k)|}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\frac{1}{2}}} \le t + d(\log n)\right\} = \exp(-2e^{-t}), \tag{4.58}$$

for all t, with

$$a(x) = (2\log x)^{\frac{1}{2}},$$

and

$$d(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log\pi.$$

We can use $\hat{\sigma}_n^2$ from (4.38) to estimate σ^2, as (4.39) holds in case of antisymmetric kernels too. Corollary 4.5.2 gives the asymptotic distribution of the test statistic under the null hypothesis of no change in the hazard functions for antisymmetric kernels. Thus we get the test statistics from (4.57) and (4.58), respectively, as

$$T_n^{(9)} = \max_{1\le k<n}\frac{1}{\hat{\sigma}_n}\frac{Q_n(k)}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\frac{1}{2}}},$$

and

$$T_n^{(10)} = \max_{1\le k<n}\frac{1}{\hat{\sigma}_n}\frac{|Q_n(k)|}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{\frac{1}{2}}}.$$

The asymptotic critical values $c_\alpha^{(1)}(n)$ and $c_\alpha^{(2)}(n)$ are now derived from Theorem 4.5.1.

4.5.1 Change Point Tests with Different Weights

As in Section 4.3.2, we can replace the standard weights ($\nu = \frac{1}{2}$) with different weights $q_\nu(t) = (t(1-t))^\nu$ for $0 < \nu < \frac{1}{2}$. As in Corollary 3.1 of Horváth [62], we have

Corollary 4.5.3. *Let the assumptions of Theorem 4.5.1 be fulfilled for some $0 < \nu < \frac{1}{2}$, $q_\nu(t) = (t(1-t))^\nu$. If H_0 holds, then*

$$\frac{1}{\sigma}\sup_{0<t<1}\frac{\tilde{Q}_n(\lceil nt\rceil)}{q_\nu(t)} \xrightarrow{\mathscr{D}} \sup_{0<t<1}\frac{B(t)}{q_\nu(t)}, \tag{4.59}$$

and

$$\frac{1}{\sigma}\sup_{0<t<1}\frac{\left|\tilde{Q}_n(\lceil nt\rceil)\right|}{q_\nu(t)} \xrightarrow{\mathscr{D}} \sup_{0<t<1}\frac{|B(t)|}{q_\nu(t)}. \tag{4.60}$$

The asymptotic distribution of the weighted test statistics under the null hypothesis of no change in hazard functions for antisymmetric kernels can be calculated from Corollary 4.5.3. Thus we obtain the test statistic from (4.59)

$$T_n^{(11)} = \sup_{0<t<1} \frac{\tilde{Q}_n(\lceil nt \rceil)}{\hat{\sigma}_n q_\nu(t)},$$

and from (4.60)

$$T_n^{(12)} = \sup_{0<t<1} \frac{|\tilde{Q}_n(\lceil nt \rceil)|}{\hat{\sigma}_n q_\nu(t)}.$$

The asymptotic critical values are $c_\alpha^{(11)}(n)$ and $c_\alpha^{(12)}(n)$ for the test statistics $T_n^{(11)}$ and $T_n^{(12)}$, respectively, where $c_\alpha^{(11)}(n)$ is the $(1-\alpha)$-quantile of $\sup_{0<t<1} \frac{B(t)}{q_\nu(t)}$ and $c_\alpha^{(12)}(n)$ is the $(1-\alpha)$-quantile of $\sup_{0<t<1} \frac{|B(t)|}{q_\nu(t)}$.

4.6 Some Remarks on a Change Point Test using $\hat{\lambda}_1(x)$

Regarding Theorem 4.2.5, we expect that the change point test of the previous sections also works with the same kind of asymptotic distribution under H_0. Let $\hat{\lambda}_{1,k}(x)$ be the estimate of $\lambda(x)$ described in Section 4.2.2, which uses only the subsample $X_1, \ldots, X_k$. Correspondingly $\hat{\lambda}^*_{1,k}(x)$ denotes the same estimate based on $X_{k+1}, \ldots, X_n$. Then, we define

$$\hat{\theta}_2(k) = \int_0^V \int_0^V \tilde{K}(x,y)\hat{\lambda}_{1,k}(x)\hat{\lambda}^*_{1,k}(y)dxdy.$$

From Theorem 4.2.8, we have under H_0

$$\sup_{0\le x\le V} \left|\hat{\lambda}_{1,k}(x) - \lambda(x)\right| = O_{a.s.}\left(\frac{1}{kb}\log\frac{1}{b}\right) + O\left(b^2\right),$$

if we assume $\alpha = 1$ in that theorem. If $\frac{k}{n} \to \gamma$, $0 < \gamma < 1$, the convergence rate is $O_{a.s.}\left(\frac{1}{nb}\log\frac{1}{b}\right) + O\left(b^2\right)$ if $b \to 0$ such that the conditions of Theorem 4.2.8 are satisfied. The same holds for $\hat{\lambda}^*_{1,k}$, and we get by the same arguments as used for Corollary 4.2.9.

Corollary 4.6.1. *Under the assumptions of Theorem 4.2.5 with $\alpha = 1$ and $\tilde{K}$ satisfying (4.10), we have under H_0*

$$\left|\hat{\theta}_2(k) - \theta\right| = O_{a.s.}\left(\left(\frac{1}{k} + \frac{1}{n-k}\right)\frac{1}{b}\log\frac{1}{b}\right) + O\left(b^2\right).$$

For instance, $k_n = \lceil \gamma n \rceil$ for some $0 < \gamma < 1$, this implies that $\hat{\theta}_2(k_n) \xrightarrow{a.s.} \theta$, more precisely,

$$\left|\hat{\theta}_2(k_n) - \theta\right| = O_{a.s.}\left(\frac{1}{nb}\log\frac{1}{b}\right) + O\left(b^2\right) = O_{a.s.}\left(\frac{1}{nb}\log\frac{1}{b}\right),$$

if $b = O\left(n^{\nu-1}\right)$ for some $0 < \nu \leq \frac{1}{2}$ as in the assumptions of Theorem 4.3.1. For getting an analogue of said theorem with $\hat{\lambda}_1$ replacing $\hat{\lambda}_2$, we would consider

$$Q_n^{'}(k) = \frac{k(n-k)}{n^{\frac{3}{2}}}\left(\hat{\theta}_2(k) - \theta\right),$$

which should be approximated by a Gaussian Process at $\frac{k}{n}$, likely the same one as in Theorem 4.3.1. This would require that for $k_n = \lceil \gamma n \rceil$, $Q_n^{'}(k_n) = O_p(1)$, which, however, does not follow from the above convergence rate which would imply only

$$\left|Q_n^{'}(k_n)\right| = O_{a.s.}\left(\frac{1}{\sqrt{nb}}\log\frac{1}{b}\right),$$

and from $b = O\left(n^{\nu-1}\right)$ for some $0 < \nu \leq \frac{1}{2}$, i.e., $b = O\left(n^{-\frac{1}{2}}\right)$, we have $\sqrt{nb} = O(1)$, and $\frac{1}{\sqrt{nb}}\log\frac{1}{b} \to \infty$. So, the rates of Diehl and Stute [41] which were meant to study uniform convergence of $\hat{\lambda}_1(x)$ to $\lambda(x)$ are not suitable for studying the asymptotics of $Q_n^{'}(k)$.

We guess that the approach of Diehl and Stute [41] to approximate $\hat{\lambda}_1(x) - \bar{\lambda}_1(x)$ by $\left(h_n(x) - Eh_n(x)\right)/(1 - H(x))$ where the latter is related closely to $\hat{\lambda}_2(x) - E\hat{\lambda}_2(x)$ may nevertheless be used for showing that $Q_n^{'}(k)$ has the same asymptotic behaviour as $Q_n(k)$ of Theorem 4.3.1, and, then, would enable us to construct change point tests using $\hat{\lambda}_1(x)$. This has to be postponed to future work.

Another approach to proving an analogue to Theorem 4.3.1 for $Q_n^{'}(k)$ would be to mimic the proof of Theorem 4.3.1, which closely follows Horváth [62]'s arguments. This is far from straightforward either as Horváth uses several deep asymptotic results for sums of i.i.d. random variables whereas

$$\hat{\lambda}_{1,k}(x) = \frac{1}{b}\sum_{i=1}^{k} K\left(\frac{x - X_{(i)}}{b}\right)\frac{\delta_{(i)}}{k - i + 1},$$

is a sum of order statistics with a rather involved kind of dependence. Note that here $X_{(i)}$, $i = 1, \ldots, k$, denote the order statistics of $X_1, \ldots, X_k$ only, and $\delta_{(1)}, \ldots, \delta_{(k)}$ are the corresponding censoring indicators.

4.7 Discussion

The main contributions of this chapter are meticulously developed nonparametric change point tests for detecting changes in the hazard distribution of survival times when the observations are subject to right censoring, by extending the U-statistic-type processes considered by Horváth [62]. In this completely nonparametric framework, we derived the asymptotic distributions of our proposed test statistics under H_0 using symmetric as well as antisymmetric kernels. Four types of change point tests were developed for two scenarios using symmetric kernels: the true parameter θ is known - the developed tests are $T_n^{(1)}$ and $T_n^{(2)}$; and when it is unknown - the developed tests are $T_n^{(3)}$ and $T_n^{(4)}$. Subsequently, we developed change point tests $T_n^{(5)}$ and $T_n^{(6)}$ using the weight function $q_\nu(t) = (t(1-t))^\nu$ for $0 < \nu < \frac{1}{2}$ when θ is known. The asymptotic behavior of our proposed test statistics under H_1 were shown in Section 4.4. In case of antisymmetric kernels, we proposed the limit behavior of tests $T_n^{(9)}$ and $T_n^{(10)}$, and the weighted asymptotic of tests $T_n^{(11)}$ and $T_n^{(12)}$ under the null hypothesis. We will investigate the finite sample properties of our developed tests via simulations and real data applications in Chapter 5. All of our developed test statistics are based on the estimator $\hat{\lambda}_2$ for symmetric and antisymmetric kernels, although we have found estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ as the equivalent estimators of Horváth's estimator. One can construct change point tests using $\hat{\lambda}_1(x)$ estimator, which can be considered in future work. The asymptotic distribution of the change point estimator in the hazard functions using $\hat{\lambda}_2$ is not explored yet and remains open for further study.

Chapter 5: Simulations and Applications

In this chapter, several simulation examples will discuss to assess the finite sample behavior with different distributions of our developed test statistics. In fact, we use the Monte Carlo simulation with exponential random variables to investigate the asymptotic properties of our proposed tests with their power performances. The proposed tests will then be applied to analyze two sets of real data collected from the cell stimulus responses observed in an animal physiology study at the University of Kaiserslautern and the breast cancer mortality rates among the recruited patients in the United States during 1973 to 2012.

The organization of the present chapter is as follows. Section 5.1 explores all the simulation results to describe the asymptotic behavior of our developed tests in Chapter 4 with corresponding power performances. Illustration of our proposed tests to detect the change point in the hazard distribution is conducted through application in two data examples and presented in Section 5.2. Finally, Section 5.3 contains some concluding remarks.

5.1 Simulations

We proposed eight different test statistics for symmetric kernel function in Section 4.3 of Chapter 4, among them $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ are unweighted test statistics whereas $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ are weighted test statistics. Another four test statistics $T_n^{(9)}$, $T_n^{(10)}$, $T_n^{(11)}$ and $T_n^{(12)}$ were also proposed in Section 4.5 for antisymmetric kernel function, where $T_n^{(11)}$ and $T_n^{(12)}$ are weighted test statistics.

All of these developed test statistics are based on U-statistic-type processes where

we use ratio-type kernel hazard estimator. Hence, we use two kinds of kernel functions: $\tilde{K}(x,y)$ for U-statistic-type processes and $K(x)$ for the ratio-type kernel hazard estimator. Typical choices of symmetric $\tilde{K}(x,y)$ kernel functions are xy, $(x-y)^2/2$ (the sample variance), $|x-y|$ (Gini's mean difference), and sign$(x+y)$ (Wilcoxon's one-sample statistic) in U-statistics, (confer Csörgö and Horváth [37]). In this simulation study, we consider $\tilde{K}(x,y) = xy$ for the symmetric kernel and $\tilde{K}(x,y) = x-y$ for the antisymmetric kernel. Usually, any symmetric kernel functions such as Uniform, Epanechnikov, Biweight and Gaussian are used for the kernel $K(x)$ in kernel hazard estimation, as explained in Section 2.2. Furthermore, for the kernel $K(x)$ Wang [143] recommended to use either the Epanechnikov kernel or the Gaussian kernel in kernel hazard estimation. Nonetheless, we are using the Epanechnikov kernel as Müller [96] found that this has certain optimal properties.

All of our developed test statistics in Sections 4.3, 4.3.2, and 4.5 of Chapter 4 contain kernel hazard estimators, which requires the optimal bandwidth selection for kernel density estimation. Good discussions of such procedures can be found in the monograph of Jones et al. [70]. The authors found that some *second generation* methods, including plug-in and smoothed bootstrap techniques, are far superior to well-known *first generation* methods, such as rules of thumb, least squares cross-validation, and biased cross-validation. They also recommend a *solve-the-equation* plug-in bandwidth selector as being most reliable in terms of overall performance. Among the various *solve-the-equation* plug-in bandwidth selectors we are interested in using the approach proposed by Sheather and Jones [130], which is known as Sheather-Jones plug-in bandwidth method.

The Sheather-Jones plug-in bandwidth method estimates the optimal bandwidth by

$$b = \left\{ \frac{R(K)}{\left(\int x^2 K(x)dx\right)^2 R(h''_{n,g(b)}))} \right\}^{1/5} n^{-1/5},$$

where $R(\varphi) = \int \varphi^2(x)dx$, K is the kernel function and $g(b)$ is the pilot bandwidth, for details confer Sheather and Jones [130]. Hence, for estimating the function $\tilde{h}(x)$ using the kernel density estimator $h_n(x) = h_{n,b}(x)$ (4.16), described in Section 4.2 of Chapter 4, we need to estimate the optimal bandwidth b for the Epanechnikov kernel using the Sheather-Jones plug-in bandwidth method. Table 5.1 represents different values of the optimal bandwidth b using the Sheather-Jones plug-in bandwidth

method.

Table 5.1: The Optimal Bandwidth for Epanechnikov kernel using the Sheather-Jones plug-in Method for Simulated Exponential Sample with Parameter μ and Sample Size n.

	μ				
n	1.0	1.5	2.0	2.5	3.0
50	0.096	0.077	0.064	0.048	0.063
100	0.068	0.081	0.054	0.031	0.022
200	0.073	0.031	0.037	0.025	0.019
500	0.058	0.032	0.023	0.022	0.017
1000	0.039	0.029	0.019	0.016	0.013

We choose the bandwidth as a global bandwidth based on the whole sample. In case, where there is a change point that may lead to somewhat sub-optimal choices as the optimal bandwidth depends on the hazard function to be estimated, and that differs between the sub-samples before and after the change point. Moreover, the optimal global bandwidth essentially depends on summary statistics of the hazard function like the total curvature (cf. Silverman [131]). So, choosing the bandwidth based on the total sample should not be a big problem if the shapes of hazard function before and after the change point are reasonably similar.

Certainly, properties and performances of our tests depend on the kernel function $\tilde{K}(x, y)$, the value of the parameter V, the kernel function $K(x)$ and the bandwidth b, and the weight $q(t)$ (for the weighted test statistics). The current section focuses on describing and discussing for all our proposed tests, organized, into two groups: unweighted and weighted statistics, their features based on the Monte Carlo simulation study.

5.1.1 Asymptotic and Simulated Critical Values

Unweighted Test Statistics

First, we calculated the asymptotic critical value $c_\alpha^{(1)}(n)$ for tests $T_n^{(1)}$, $T_n^{(3)}$ and $T_n^{(9)}$ using (4.41) for different sample sizes n at various levels of significance α and present those values in Table 5.2. Table 5.3 illustrates various asymptotic critical values

$c_\alpha^{(2)}(n)$ for tests $T_n^{(2)}$, $T_n^{(4)}$ and $T_n^{(10)}$ using (4.42). All these tests have Gumbel extreme value type distribution, where the convergences are rather slow and thus these tests are conservative for small samples. Hereafter, we did simulation to compare the simulated critical values with the approximated critical values for our proposed test statistics, and organize those values in Tables 5.4 and 5.5.

Table 5.2: Some Selected Critical Values $c_\alpha^{(1)}(n)$ at various Level of Significance α and Sample Size n for Test Statistics $T_n^{(1)}$, $T_n^{(3)}$ and $T_n^{(9)}$.

	α				
n	0.005	0.01	0.025	0.05	0.10
50	4.6054	4.1842	3.6249	3.1974	2.7616
100	4.5715	4.1735	3.6448	3.2408	2.8289
200	4.5527	4.1718	3.6658	3.2792	2.8850
300	4.5459	4.1731	3.6780	3.2996	2.9139
500	4.5402	4.1762	3.6929	3.3235	2.9469
1000	4.5361	4.1823	3.7124	3.3532	2.9871
3000	4.5350	4.1939	3.7409	3.3948	3.0419

Table 5.3: Some Selected Critical Values $c_\alpha^{(2)}(n)$ at various Level of Significance α and Sample Size n for Test Statistics $T_n^{(2)}$, $T_n^{(4)}$ and $T_n^{(10)}$.

	α				
n	0.005	0.01	0.025	0.05	0.10
50	5.0251	4.6039	4.0445	3.6171	3.1813
100	4.9681	4.5701	4.0414	3.6374	3.2256
200	4.9322	4.5513	4.0453	3.6587	3.2645
300	4.9173	4.5445	4.0494	3.6711	3.2853
500	4.9028	4.5388	4.0555	3.6862	3.3095
1000	4.8886	4.5348	4.0649	3.7057	3.3396
3000	4.8748	4.5337	4.0808	3.7346	3.3818

We performed $M = 1000$ simulations for each case. In each case, we considered a simulation where the survival time T_i and the censoring time C_i were exponentially distributed with rate μ and μ_c, respectively, under the null hypothesis (4.2). Hence, we obtained the simulated data (X_i, δ_i) by (4.1). Note that we use μ as the rate of exponential distribution instead of λ, since λ is used for hazard notation in this study.

We conducted simulations to compare the simulated critical values of unweighted test statistics $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ with the approximated critical values of those statistics for the sample sizes $n = 50, 100, 200,$ and 500 with the rate $\mu = 1.0, 1.5, 2.0, 2.5,$ and 3.0 at significance levels $\alpha = 0.01, 0.05, 0.10$. The results are reported for tests $T_n^{(1)}$ and $T_n^{(3)}$ in Table 5.4, and for tests $T_n^{(2)}$ and $T_n^{(4)}$ in Table 5.5.

Table 5.4: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(1)}(n)$ of Test Statistics $T_n^{(1)}$ and $T_n^{(3)}$ for $V = \frac{2}{\mu}$ and Optimal Bandwidth b.

		$(\alpha = 0.01)$			$(\alpha = 0.05)$			$(\alpha = 0.10)$		
n	μ	Sim.CV $T_n^{(1)}$	Sim.CV $T_n^{(3)}$	App.CV $c_\alpha^{(1)}(n)$	Sim.CV $T_n^{(1)}$	Sim.CV $T_n^{(3)}$	App.CV $c_\alpha^{(1)}(n)$	Sim.CV $T_n^{(1)}$	Sim.CV $T_n^{(3)}$	App.CV $c_\alpha^{(1)}(n)$
50	1.0	4.0393	7.5207	4.1842	2.4922	4.8602	3.1974	1.7762	3.5742	2.7616
	1.5	4.0171	6.4119	4.1842	2.5065	4.1211	3.1974	1.7874	3.0208	2.7616
	2.0	4.0468	5.6889	4.1842	2.5353	3.6552	3.1974	1.7934	2.6528	2.7616
	2.5	3.9962	4.9660	4.1842	2.5170	3.2013	3.1974	1.7919	2.3618	2.7616
	3.0	4.1284	5.0169	4.1842	2.5246	3.1318	3.1974	1.8300	2.3143	2.7616
100	1.0	3.9973	6.6437	4.1735	2.3221	4.6146	3.2408	1.6157	3.4695	2.8289
	1.5	4.0542	5.8953	4.1735	2.3482	4.0184	3.2408	1.6392	3.0604	2.8289
	2.0	4.0463	4.9734	4.1735	2.3562	3.4380	3.2408	1.6328	2.6170	2.8289
	2.5	3.9850	4.3589	4.1735	2.2972	2.9101	3.2408	1.6128	2.2330	2.8289
	3.0	4.0276	3.8296	4.1735	2.3382	2.6311	3.2408	1.5958	1.9838	2.8289
200	1.0	3.6380	7.3976	4.1718	2.3748	5.0364	3.2792	1.6192	4.0417	2.8850
	1.5	3.6786	5.7951	4.1718	2.3208	3.9692	3.2792	1.5798	3.0998	2.8850
	2.0	3.5942	5.2874	4.1718	2.3608	3.5809	3.2792	1.6412	2.8581	2.8850
	2.5	3.6385	4.6109	4.1718	2.3609	3.1661	3.2792	1.6349	2.4736	2.8850
	3.0	3.5934	4.1769	4.1718	2.3628	2.8675	3.2792	1.6013	2.2484	2.8850
500	1.0	3.7351	7.4691	4.1762	2.3875	5.3312	3.3235	1.5752	4.0003	2.9469
	1.5	3.7575	6.1999	4.1762	2.3617	4.2677	3.3235	1.6110	3.2051	2.9469
	2.0	3.7195	5.3022	4.1762	2.3522	3.6725	3.3235	1.5951	2.7689	2.9469
	2.5	3.6498	4.8213	4.1762	2.4383	3.3145	3.3235	1.5736	2.4755	2.9469
	3.0	3.8460	4.3719	4.1762	2.3679	3.0536	3.3235	1.5767	2.2619	2.9469

Table 5.4 reveals that the critical values $c_\alpha^{(1)}(n)$ obtained from the limiting distribution somehow overestimate the simulated critical values of $T_n^{(1)}$ with $V = \frac{2}{\mu}$, but is mostly close. However, the simulated critical values of $T_n^{(3)}$ overestimate the approximated critical values $c_\alpha^{(1)}(n)$ with the minimum rate of exponential variables, but these values proceeds closer as the rate is increasing and the scenario becomes clearer with larger sample sizes. Therefore, we can use the approximated critical values $c_\alpha^{(1)}(n)$ to

the proposed tests $T_n^{(1)}$ and $T_n^{(3)}$ reasonably.

Table 5.5: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(2)}(n)$ of Test Statistics $T_n^{(2)}$ and $T_n^{(4)}$ for $V = \frac{2}{\mu}$ and Optimal Bandwidth b.

		$(\alpha = 0.01)$			$(\alpha = 0.05)$			$(\alpha = 0.10)$		
n	μ	Sim.CV $T_n^{(2)}$	Sim.CV $T_n^{(4)}$	App.CV $c_\alpha^{(2)}(n)$	Sim.CV $T_n^{(2)}$	Sim.CV $T_n^{(4)}$	App.CV $c_\alpha^{(2)}(n)$	Sim.CV $T_n^{(2)}$	Sim.CV $T_n^{(4)}$	App.CV $c_\alpha^{(2)}(n)$
50	1.0	4.0393	7.5207	4.6039	2.4922	4.8602	3.6171	1.8657	3.5742	3.1813
	1.5	4.0171	6.4119	4.6039	2.5065	4.1211	3.6171	1.8646	3.0208	3.1813
	2.0	4.0468	5.6889	4.6039	2.5353	3.6552	3.6171	1.8660	2.6528	3.1813
	2.5	3.9962	4.9660	4.6039	2.5170	3.2013	3.6171	1.8644	2.3618	3.1813
	3.0	4.1284	5.0169	4.6039	2.5246	3.1318	3.6171	1.8778	2.3143	3.1813
100	1.0	3.9973	6.6437	4.5701	2.5941	4.6146	3.6374	2.5426	3.4695	3.2256
	1.5	4.0542	5.8953	4.5701	2.5848	4.0184	3.6374	2.5387	3.0604	3.2256
	2.0	4.0463	4.9734	4.5701	2.5876	3.4380	3.6374	2.5404	2.6170	3.2256
	2.5	3.9850	4.3589	4.5701	2.5939	2.9101	3.6374	2.5414	2.2330	3.2256
	3.0	4.0276	3.8296	4.5701	2.5953	2.6311	3.6374	2.5412	1.9838	3.2256
200	1.0	3.6419	7.3976	4.5513	3.5645	5.0364	3.6587	3.5311	4.0417	3.2645
	1.5	3.6786	5.7951	4.5513	3.5654	3.9692	3.6587	3.5313	3.0998	3.2645
	2.0	3.6384	5.2874	4.5513	3.5647	3.5809	3.6587	3.5308	2.8581	3.2645
	2.5	3.6423	4.6109	4.5513	3.5650	3.1661	3.6587	3.5310	2.4736	3.2645
	3.0	3.6424	4.1769	4.5513	3.5638	2.8675	3.6587	3.5314	2.2484	3.2645
500	1.0	5.5946	7.4691	4.5388	5.5525	5.3312	3.6862	5.5266	4.0003	3.3095
	1.5	5.5952	6.1999	4.5388	5.5525	4.2677	3.6862	5.5272	3.2051	3.3095
	2.0	5.5952	5.3022	4.5388	5.5523	3.6725	3.6862	5.5271	2.7689	3.3095
	2.5	5.5948	4.4813	4.5388	5.5523	3.3145	3.6862	5.5264	2.4755	3.3095
	3.0	5.5951	4.3720	4.5388	5.5525	3.0536	3.6862	5.5274	2.2619	3.3095

Table 5.5 displays the asymptotic critical values $c_\alpha^{(2)}(n)$ are close enough to the simulated critical values of $T_n^{(2)}$ with $V = \frac{2}{\mu}$ for small samples, but are overestimated for comparative larger sample size $n = 500$. However, the simulated critical values of $T_n^{(4)}$ overestimate the approximated critical values $c_\alpha^{(2)}(n)$ with the minimum rate of exponential variables, but these values proceeds closer as the rate is increasing and the scenario becomes clearer with larger sample sizes, similar to test $T_n^{(3)}$. Consequently, it is equitable to use the approximated critical values $c_\alpha^{(2)}(n)$ to the proposed tests $T_n^{(2)}$ and $T_n^{(4)}$. The parameter V has a significant impact on our developed tests, the simulated critical values are illustrating a reasonable performance with $V = \frac{2}{\mu}$, where we are assuming that $F(V) \approx 0.8646 < 1$. Whereas, the simulated critical values are not that close with $V = \frac{3}{\mu}$, where we assume $F(V) \approx 0.9502 < 1$, which

are illustrated in Tables B.1 and B.2 in Appendix B.

Weighted Test Statistics

The weight function $q(t) = (t(1-t))^{0.2}$ is considered for this simulation study. Henceforth, we used weight $q(t) = (t(1-t))^{0.2}$ in calculating the critical values for our developed tests $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$, $T_n^{(8)}$, $T_n^{(11)}$ and $T_n^{(12)}$.

The limiting Gaussian process $\{(1-t)W(t) + t\,(W(1) - W(t))\,, 0 \leq t \leq 1\}$ appeared earlier in Csörgő and Horváth [36] as the limit of U-statistic-type processes when there is no censoring. Our proposed test statistics $T_n^{(5)}$ and $T_n^{(6)}$ have an asymptotic distribution as Gaussian process divided by its weight $q(t)$, which are presented in (4.45) and (4.46), respectively. So far our knowledge, there is no known formula for the distribution function of the limiting random variable in (4.45) and (4.46). Therefore, we simulated the distribution function in (4.45) and (4.46) for $M = 20000$ repetitions and computed different percentiles to estimate the critical values at various level of significance. The obtained critical values are shown in Tables 5.6 and 5.7.

Table 5.6: Selected Critical Values $c_\alpha^{(5)}(n)$ at various Level of Significance α and n for Test Statistic $T_n^{(5)}$ with $q(t) = (t(1-t))^{0.2}$.

	α				
n	0.005	0.01	0.025	0.05	0.10
50	5.6230	5.0626	4.2644	3.5860	2.8115
100	6.4367	5.8276	4.9381	4.1241	3.2100
200	7.5343	6.7644	5.7031	4.7207	3.6609
300	8.0469	7.3130	6.2300	5.2005	4.0815
500	8.8748	8.0218	6.7716	5.7200	4.5062
1000	10.0892	9.1182	7.8216	6.5602	5.1605
3000	12.9931	11.5445	9.7071	8.1968	6.3636
10^6	41.6505	37.0325	31.2583	26.1078	20.2125

Table 5.7: Some Selected Critical Values $c_\alpha^{(6)}(n)$ at various Level of Significance α and n for Test Statistic $T_n^{(6)}$ with $q(t) = (t(1-t))^{0.2}$.

	α				
n	0.005	0.01	0.025	0.05	0.10
50	6.1179	5.6047	4.8970	4.2818	3.5980
100	7.0775	6.4923	5.6496	4.9482	4.1252
200	8.2244	7.5343	6.4932	5.6930	4.7207
300	8.7031	8.0275	7.0695	6.2155	5.1776
500	9.7780	8.9213	7.7259	6.7867	5.6984
1000	10.9802	10.1390	8.9120	7.8421	6.5995
3000	14.0463	12.9102	11.0982	9.7381	8.2104
10^6	44.3602	40.7064	35.4265	30.9392	26.0652

Our proposed tests $T_n^{(7)}$ and $T_n^{(8)}$ have an asymptotic Brownian bridge type process distribution, which are respectively $\sup\limits_{0<t<1} \frac{\{(1-2t)B(t)\}}{q(t)}$ and $\sup\limits_{0<t<1} \frac{|(1-2t)B(t)|}{q(t)}$, described in Section 4.3.2. We conducted simulation for $M = 20000$ times to compute the distribution function in $\sup\limits_{0<t<1} \frac{\{(1-2t)B(t)\}}{q(t)}$ and $\sup\limits_{0<t<1} \frac{|(1-2t)B(t)|}{q(t)}$ for different percentiles. The results for approximated critical values at various level of significance are presented in Tables 5.8 and 5.9.

Table 5.8: Some Selected Critical Values $c_\alpha^{(7)}(n)$ at various Level of Significance α and n for Test Statistic $T_n^{(7)}$.

	α				
n	0.005	0.01	0.025	0.05	0.10
50	1.2278	1.1446	1.0281	0.9321	0.8226
100	1.2490	1.1629	1.0562	0.9626	0.8591
200	1.2797	1.2088	1.0893	0.9923	0.8872
300	1.2924	1.2080	1.0968	0.9999	0.8922
500	1.3003	1.2207	1.1060	1.0074	0.8965
1000	1.3341	1.2388	1.1158	1.0227	0.9132
3000	1.3156	1.2402	1.8441	1.0364	0.9305
10^6	1.3438	1.2620	1.1440	1.0479	0.9376

Table 5.9: Some Selected Critical Values $c_{\alpha}^{(8)}(n)$ at various Level of Significance α and n for Test Statistic $T_n^{(8)}$.

	α				
n	0.005	0.01	0.025	0.05	0.10
50	1.3046	1.2191	1.1075	1.0236	0.9285
100	1.3542	1.2598	1.1470	1.0553	0.9587
200	1.3750	1.2899	1.1853	1.0893	0.9894
300	1.3678	1.2995	1.1833	1.0969	0.9970
500	1.3832	1.3069	1.1980	1.1049	1.0056
1000	1.3950	1.3175	1.2032	1.1153	1.0195
3000	1.4008	1.3201	1.2159	1.1268	1.0267
10^6	1.4203	1.3361	1.2295	1.1436	1.0442

Our developed tests $T_n^{(11)}$ and $T_n^{(12)}$ have asymptotic Brownian bridge type process distribution, i.e., $\sup\limits_{0<t<1} \frac{\{B(t)\}}{q(t)}$ and $\sup\limits_{0<t<1} \frac{|B(t)|}{q(t)}$, respectively, which are accordingly presented in (4.59) and (4.60). We conducted simulation for $M = 20000$ times to compute the distribution function in $\sup\limits_{0<t<1} \frac{\{B(t)\}}{q(t)}$ and $\sup\limits_{0<t<1} \frac{|B(t)|}{q(t)}$ for different percentiles. Results are displayed in Tables 5.10 and 5.11 for different sample sizes at various level of significance.

Table 5.10: Some Selected Critical Values $c_{\alpha}^{(11)}(n)$ at various Level of Significance α and n for Test statistic $T_n^{(11)}$.

	α				
n	0.005	0.01	0.025	0.05	0.10
50	2.0443	1.9226	1.7263	1.5553	1.3619
100	2.1360	1.9862	1.7754	1.5947	1.3947
200	2.1255	1.9882	1.7915	1.6190	1.4214
300	2.1436	2.0166	1.8201	1.6395	1.4398
500	2.1757	2.0254	1.7915	1.6300	1.4378
1000	2.1804	2.0376	1.8426	1.6539	1.4614
3000	2.2111	2.0629	1.8441	1.6653	1.4762
10^6	2.1830	2.0429	1.8501	1.6764	1.4805

Table 5.11: Some Selected Critical Values $c_{\alpha}^{(12)}(n)$ at various Level of Significance α and n for Test statistic $T_n^{(12)}$.

	α				
n	0.005	0.01	0.025	0.05	0.10
50	2.1953	2.0699	1.8675	1.7210	1.5505
100	2.2592	2.1253	1.9262	1.7669	1.5865
200	2.2768	2.1381	1.9551	1.8016	1.6254
300	2.3077	2.1386	1.9677	1.8070	1.6317
500	2.2673	2.1608	1.9682	1.8072	1.6353
1000	2.3349	2.1973	1.9928	1.8342	1.6531
3000	2.3289	2.2016	2.0031	1.8391	1.6608
10^6	2.3076	2.1769	1.9998	1.8501	1.6764

Table 5.12: Comparison of Simulated and Approximated Critical Values $c_{\alpha}^{(5)}(n)$ of Test Statistic $T_n^{(5)}$ with $q(t) = (t(1-t))^{0.2}$ and $V = \frac{2}{\mu}$ and Optimal Bandwidth b.

		($\alpha = 0.01$)		($\alpha = 0.05$)		($\alpha = 0.10$)	
n	μ	Sim.CV	App.CV	Sim.CV	App.CV	Sim.CV	App.CV
50	1.0	1.2416	5.0626	0.8101	3.5860	0.5885	2.8115
	1.5	1.2347	5.0626	0.8270	3.5860	0.5890	2.8115
	2.0	1.2439	5.0626	0.8316	3.5860	0.5844	2.8115
	2.5	1.2283	5.0626	0.8258	3.5860	0.5891	2.8115
	3.0	1.2689	5.0626	0.8480	3.5860	0.6070	2.8115
100	1.0	1.0014	5.8276	0.6350	4.1241	0.4327	3.2100
	1.5	1.0167	5.8276	0.6044	4.1241	0.4402	3.2100
	2.0	1.0158	5.8276	0.6041	4.1241	0.4429	3.2100
	2.5	0.9979	5.8276	0.6148	4.1241	0.4413	3.2100
	3.0	1.0195	5.8276	0.6247	4.1241	0.4294	3.2100
200	1.0	0.5754	6.7644	0.3134	4.7207	0.1386	3.6609
	1.5	0.5723	6.7644	0.3076	4.7207	0.1300	3.6609
	2.0	0.5684	6.7644	0.3129	4.7207	0.1383	3.6609
	2.5	0.5765	6.7644	0.3041	4.7207	0.1312	3.6609
	3.0	0.5820	6.7644	0.2997	4.7207	0.1381	3.6609
500	1.0	0.1224	8.0218	0.0000	5.7200	0.0000	4.5062
	1.5	0.1181	8.0218	0.0000	5.7200	0.0000	4.5062
	2.0	0.1147	8.0218	0.0000	5.7200	0.0000	4.5062
	2.5	0.1251	8.0218	0.0000	5.7200	0.0000	4.5062
	3.0	0.1113	8.0218	0.0000	5.7200	0.0000	4.5062

We performed $M = 1000$ simulations to compare the simulated critical values with the approximated critical values for our proposed weighted test statistics $T_n^{(5)}$, $T_n^{(6)}$,

$T_n^{(7)}$, $T_n^{(8)}$, $T_n^{(11)}$ and $T_n^{(12)}$ for the sample sizes $n = 50, 100, 200$, and 500 with the rate $\mu = 1.0, 1.5, 2.0, 2.5$, and 3.0 at significance levels $\alpha = 0.01, 0.05, 0.10$, and we organized those values in Tables 5.12, 5.13, 5.14 and 5.15. In each case, we considered the data generation process similar to that of our developed unweighted tests, i.e., the survival time T_i and the censoring time C_i were exponentially distributed with rate μ and μ_c, respectively, under the null hypothesis (4.2). Hence, we obtained the simulated data (X_i, δ_i) by (4.1).

Table 5.13: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(6)}(n)$ of Test Statistic $T_n^{(6)}$ with $q(t) = (t(1-t))^{0.2}$ and $V = \frac{2}{\mu}$ and Optimal Bandwidth b.

		$(\alpha = 0.01)$		$(\alpha = 0.05)$		$(\alpha = 0.10)$	
n	μ	Sim.CV	App.CV	Sim.CV	App.CV	Sim.CV	App.CV
50	1.0	1.2440	5.6047	1.2122	4.2818	1.1980	3.5980
	1.5	1.2469	5.6047	1.2118	4.2818	1.1971	3.5980
	2.0	1.2483	5.6047	1.2116	4.2818	1.1968	3.5980
	2.5	1.2434	5.6047	1.2117	4.2818	1.1970	3.5980
	3.0	1.2689	5.6047	1.2101	4.2818	1.1948	3.5980
100	1.0	1.7051	6.4923	1.6792	4.9482	1.6640	4.1252
	1.5	1.7007	6.4923	1.6786	4.9482	1.6626	4.1252
	2.0	1.7017	6.4923	1.6787	4.9482	1.6632	4.1252
	2.5	1.7043	6.4923	1.6794	4.9482	1.6639	4.1252
	3.0	1.7053	6.4923	1.6793	4.9482	1.6642	4.1252
200	1.0	2.3756	7.5343	2.3437	5.6930	2.3255	4.7207
	1.5	2.3766	7.5343	2.3447	5.6930	2.3263	4.7207
	2.0	2.3754	7.5343	2.3437	5.6930	2.3254	4.7207
	2.5	2.3762	7.5343	2.3441	5.6930	2.3262	4.7207
	3.0	2.3764	7.5343	2.3444	5.6930	2.3264	4.7207
500	1.0	3.6901	8.9213	3.6632	6.7867	3.6454	5.6984
	1.5	3.6904	8.9213	3.6630	6.7867	3.6461	5.6984
	2.0	3.6905	8.9213	3.6629	6.7867	3.6460	5.6984
	2.5	3.6902	8.9213	3.6630	6.7867	3.6456	5.6984
	3.0	3.6904	8.9213	3.6632	6.7867	3.6462	5.6984

Using the Sheather-Jones plug-in optimal bandwidth b from Table 5.1, $V = \frac{2}{\mu}$ and the weight $q(t) = (t(1-t))^{0.2}$, we found that the approximated critical values $c_\alpha^{(5)}(n)$ of test statistic $T_n^{(5)}$ are highly overestimating the simulated critical values of $T_n^{(5)}$ for all cases. So, the tests using the approximated values are quite conservative. The simulated critical values are proceeding to zero as sample size increases for test $T_n^{(5)}$, which reveals in Table 5.12. Oppositely, the simulated critical values for test $T_n^{(6)}$ are

proceeding to the approximated critical values $c_\alpha^{(6)}(n)$ as sample size increases, those are illustrated in Table 5.13.

Table 5.14: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(7)}(n)$ of Test Statistic $T_n^{(7)}$ with $q(t) = (t(1-t))^{0.2}$, $V = \frac{2}{\mu}$ and Optimal Bandwidth b.

		$(\alpha = 0.01)$		$(\alpha = 0.05)$		$(\alpha = 0.10)$	
n	μ	Sim.CV	App.CV	Sim.CV	App.CV	Sim.CV	App.CV
50	1.0	2.5172	1.2620	1.6130	1.0479	1.2034	0.9376
	1.5	2.0834	1.2620	1.3499	1.0479	1.0220	0.9376
	2.0	1.8346	1.2620	1.1810	1.0479	0.9077	0.9376
	2.5	1.6160	1.2620	1.0487	1.0479	0.7957	0.9376
	3.0	1.5902	1.2620	1.0086	1.0479	0.8026	0.9376
100	1.0	1.7715	1.2620	1.2344	1.0479	0.9720	0.9376
	1.5	1.5667	1.2620	1.1097	1.0479	0.8741	0.9376
	2.0	1.3376	1.2620	0.9383	1.0479	0.7443	0.9376
	2.5	1.1382	1.2620	0.8031	1.0479	0.6305	0.9376
	3.0	1.0272	1.2620	0.7056	1.0479	0.5618	0.9376
200	1.0	1.3507	1.2620	0.9008	1.0479	0.5911	0.9376
	1.5	1.0976	1.2620	0.6887	1.0479	0.4534	0.9376
	2.0	0.9598	1.2620	0.6343	1.0479	0.4192	0.9376
	2.5	0.8636	1.2620	0.5564	1.0479	0.3660	0.9376
	3.0	0.7946	1.2620	0.5051	1.0479	0.3274	0.9376
500	1.0	0.6971	1.2620	0.2389	1.0479	0.1891	0.9376
	1.5	0.5280	1.2620	0.1928	1.0479	0.1533	0.9376
	2.0	0.4648	1.2620	0.1651	1.0479	0.1324	0.9376
	2.5	0.4267	1.2620	0.1503	1.0479	0.1192	0.9376
	3.0	0.3730	1.2620	0.1365	1.0479	0.1088	0.9376

Table 5.14 illustrates the approximated critical values $c_\alpha^{(7)}(n)$ and the simulated critical values of $T_n^{(7)}$ for the sample sizes $n = 50, 100, 200$, and 500 with the rate $\mu = 1.0, 1.5, 2.0, 2.5$, and 3.0 at significance levels $\alpha = 0.01, 0.05, 0.10$ using the Sheather-Jones plug-in optimal bandwidth b from Table 5.1, $V = \frac{2}{\mu}$ and the weight $q(t) = (t(1-t))^{0.2}$, we found that the simulated critical values of $T_n^{(7)}$ slightly overestimate $c_\alpha^{(7)}(n)$ for the small size, e.g., $n = 50, 100$, but underestimated for larger sample size. In comparing the results for given n and various μ, we have to recall that the test procedure parameter V depends on μ in our setup, and this may have an influence.

Table 5.15: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(8)}(n)$ of Test Statistic $T_n^{(8)}$ with $q(t) = (t(1-t))^{0.2}$, $V = \frac{2}{\mu}$ and Optimal Bandwidth b.

		$(\alpha = 0.01)$		$(\alpha = 0.05)$		$(\alpha = 0.10)$	
n	μ	Sim.CV	App.CV	Sim.CV	App.CV	Sim.CV	App.CV
50	1.0	2.5172	1.3361	1.6130	1.1436	1.2034	1.0442
	1.5	2.0834	1.3361	1.3499	1.1436	1.0220	1.0442
	2.0	1.8346	1.3361	1.1810	1.1436	0.9077	1.0442
	2.5	1.6160	1.3361	1.0487	1.1436	0.7957	1.0442
	3.0	1.5902	1.3361	1.0086	1.1436	0.8026	1.0442
100	1.0	1.7715	1.3361	1.2344	1.1436	0.9720	1.0442
	1.5	1.5667	1.3361	1.1097	1.1436	0.8741	1.0442
	2.0	1.3376	1.3361	0.9383	1.1436	0.7443	1.0442
	2.5	1.1382	1.3361	0.8031	1.1436	0.6305	1.0442
	3.0	1.0272	1.3361	0.7056	1.1436	0.5618	1.0442
200	1.0	1.3507	1.3361	0.9008	1.1436	0.5911	1.0442
	1.5	1.0976	1.3361	0.6887	1.1436	0.4534	1.0442
	2.0	0.9598	1.3361	0.6343	1.1436	0.4192	1.0442
	2.5	0.8636	1.3361	0.5564	1.1436	0.3660	1.0442
	3.0	0.7946	1.3361	0.5051	1.1436	0.3274	1.0442
500	1.0	0.6971	1.3361	0.2389	1.1436	0.1891	1.0442
	1.5	0.5280	1.3361	0.1928	1.1436	0.1533	1.0442
	2.0	0.4648	1.3361	0.1651	1.1436	0.1324	1.0442
	2.5	0.4267	1.3361	0.1503	1.1436	0.1192	1.0442
	3.0	0.3730	1.3361	0.1365	1.1436	0.1088	1.0442

Table 5.15 also reports that the test statistic $T_n^{(8)}$ slightly overestimates its approximated critical values $c_\alpha^{(8)}(n)$ for the small size, e.g., $n = 50, 100$ with the rate $\mu = 1.0, 1.5, 2.0, 2.5$, and 3.0 at significance levels $\alpha = 0.01, 0.05, 0.10$, but underestimates for larger sample size. Therefore, it turned out that the critical values obtained from the limiting distribution for tests $T_n^{(7)}$ and $T_n^{(8)}$ are roughly close, hence we can use the approximated critical values $c_\alpha^{(7)}(n)$ and $c_\alpha^{(8)}(n)$, respectively, to the proposed tests.

From Tables 5.4, 5.5, 5.12-5.15, the asymptotic critical values obtained from the limiting distributions for all proposed tests (except $T_n^{(5)}$) are somehow roughly close to the true one. Though, the parameter V has a significant impact on our developed tests, the simulated critical values are illustrating a reasonable performance with $V = \frac{2}{\mu}$, where we are assuming that $F(V) \approx 0.8646 < 1$. Whereas, the simulated critical values are not that close with $V = \frac{3}{\mu}$, where we assume $F(V) \approx 0.9502 < 1$, which are illustrated in Appendix B. The standard errors for $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$, $T_n^{(4)}$, $T_n^{(5)}$, $T_n^{(6)}$,

$T_n^{(7)}$ and $T_n^{(8)}$ are 0.9117, 0.4151, 1.1203, 1.1137, 0.0280, 0.0132, 0.1574 and 0.1565 in case of $n = 50$, $\mu = 3$ and $V = \frac{2}{\mu}$; and 0.8341, 0.1278, 0.9842, 0.9777, 0.0038, 0.0121, 0.0603 and 0.0592 for $n = 200$, $\mu = 3$ and $V = \frac{2}{\mu}$; respectively, which indicates that the standard errors decrease for larger sample size. Smaller standard errors imply that simulations $M = 1000$ is quite fair, but we will get even improved results with the larger simulation repetitions and the larger sample sizes. Because of the computational time restriction, we limit our simulation with $M = 1000$ repetitions. We can reasonably, therefore, use the approximated critical values to the proposed tests and to estimate the power of the corresponding tests.

5.1.2 Simulated Powers

We conducted the Monte Carlo simulation with $M = 1000$ repetitions to investigate the power performance of our proposed test statistics $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$, $T_n^{(4)}$, $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ for $n = 50, 100, 200, 500$, $V = \frac{2}{\mu_1}$, the Sheather-Jones plug-in optimal bandwidth b from Table 5.1, $\mu_1 = 1.0, 1.5, 2.0$, $\mu_2 = 1.0, 1.5, 2.0$ and $\mu_c = 1.0, 1.5, 2.0$. Here, μ_1 and μ_2 are the rate of exponential survival time distributions before and after the change point, respectively, and μ_c is the rate of exponential censored distribution. In each case, we assume under the alternative hypothesis (4.2) that

$$F_1 = \text{Exp}(\mu_1), \quad F_2 = \text{Exp}(\mu_2), \quad \text{and} \quad G = \text{Exp}(\mu_c).$$

We ran simulation to investigate the power of all proposed tests by dealing with different sample sizes but holding the same location of change point, i.e., $k^* = 0.50$, which means 50% observations are from the before-change sample and 50% observations are in the after-change sample. Tables 5.16-5.17 display the results of the power analysis for all the developed tests with samples of size $n = 50, 100, 200, 500$.

Table 5.16 shows that our proposed tests $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ have almost the size corrected powers under the null hypothesis of no change in the hazard functions of survival times at the level $\alpha = 0.01$ and $\alpha = 0.05$. However, all these powers are not strong enough for $V = \frac{2}{\mu_1}$ with the Sheather-Jones plug-in optimal bandwidth b. Clearly, the power for the choice $\mu_1 = 1.5$, $\mu_2 = 1.0$, and $\mu_c = 1.0$ is the best among all other combinations for all sample sizes, e.g., the power of tests $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ are 0.267, 0.248, 0.178, and 0.178, respectively, at the level $\alpha = 0.05$ for $n = 100$.

Table 5.16: Simulated Powers of Test Statistics $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ for $V = \frac{2}{\mu_1}$, and Optimal Bandwidth b, using the Simulated Critical Value for $\mu_1 = \mu_2 = 1$.

					($\alpha = 0.01$)				($\alpha = 0.05$)			
n	Censor	μ_1	μ_2	μ_c	$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$	$T_n^{(4)}$	$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$	$T_n^{(4)}$
50	50.2%	1.0	1.0	1.0	0.006	0.006	0.015	0.015	0.046	0.046	0.050	0.050
	40.2%	1.0	1.5	1.0	0.048	0.048	0.018	0.018	0.149	0.149	0.065	0.065
	45.1%	1.5	1.0	1.0	0.139	0.139	0.106	0.106	0.229	0.229	0.136	0.136
	55.2%	1.5	1.0	1.5	0.050	0.050	0.034	0.034	0.119	0.119	0.079	0.079
	58.5%	2.0	1.0	2.0	0.101	0.101	0.049	0.049	0.155	0.155	0.111	0.111
	53.8%	2.0	1.5	2.0	0.037	0.037	0.013	0.013	0.103	0.103	0.054	0.054
100	50.2%	1.0	1.0	1.0	0.009	0.009	0.019	0.019	0.056	0.046	0.058	0.058
	40.3%	1.0	1.5	1.0	0.043	0.043	0.023	0.023	0.148	0.128	0.092	0.092
	45.4%	1.5	1.0	1.0	0.158	0.158	0.103	0.103	0.267	0.248	0.178	0.178
	55.3%	1.5	1.0	1.5	0.061	0.061	0.049	0.049	0.146	0.166	0.101	0.101
	58.6%	2.0	1.0	2.0	0.093	0.093	0.062	0.062	0.159	0.149	0.118	0.118
	53.9%	2.0	1.5	2.0	0.049	0.049	0.026	0.026	0.122	0.181	0.070	0.070
200	50.0%	1.0	1.0	1.0	0.006	0.006	0.010	0.010	0.033	0.043	0.031	0.031
	40.0%	1.0	1.5	1.0	0.055	0.054	0.013	0.013	0.132	0.058	0.057	0.057
	45.0%	1.5	1.0	1.0	0.147	0.147	0.107	0.107	0.235	0.149	0.152	0.152
	55.0%	1.5	1.0	1.5	0.059	0.074	0.039	0.039	0.111	0.142	0.091	0.091
	58.4%	2.0	1.0	2.0	0.100	0.100	0.049	0.049	0.148	0.105	0.105	0.105
	53.6%	2.0	1.5	2.0	0.043	0.079	0.018	0.018	0.100	0.205	0.057	0.057
500	49.9%	1.0	1.0	1.0	0.011	0.014	0.010	0.010	0.054	0.050	0.049	0.049
	39.9%	1.0	1.5	1.0	0.057	0.024	0.029	0.029	0.127	0.054	0.054	0.054
	44.9%	1.5	1.0	1.0	0.161	0.119	0.101	0.101	0.233	0.121	0.159	0.159
	54.9%	1.5	1.0	1.5	0.060	0.042	0.029	0.029	0.123	0.079	0.083	0.083
	58.3%	2.0	1.0	2.0	0.087	0.040	0.046	0.046	0.145	0.060	0.092	0.092
	53.5%	2.0	1.5	2.0	0.040	0.121	0.012	0.012	0.108	0.216	0.046	0.046

Similar power performances are also observed in Table 5.17 for our developed weighted tests $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ with the weight $q(t) = (t(1-t))^{0.2}$, $V = \frac{2}{\mu_1}$, and the Sheather-Jones plug-in optimal bandwidth b. A reasonable power performance is found in the case of parameter choice $\mu_1 = 1.5$, $\mu_2 = 1.0$, and $\mu_c = 1.0$ with $n = 500$ at the level $\alpha = 0.05$, for instance, the observed powers for the tests $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ are 0.178, 0.120, 0.253, and 0.306, respectively.

Table 5.17: Simulated Powers of Test Statistics $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ for $q(t) = (t(1-t))^{0.2}$, $V = \frac{2}{\mu_1}$, and Optimal Bandwidth b, using the Simulated Critical Value for $\mu_1 = \mu_2 = 1$.

					($\alpha = 0.01$)				($\alpha = 0.05$)			
n	Censor	μ_1	μ_2	μ_c	$T_n^{(5)}$	$T_n^{(6)}$	$T_n^{(7)}$	$T_n^{(8)}$	$T_n^{(5)}$	$T_n^{(6)}$	$T_n^{(7)}$	$T_n^{(8)}$
50	50.2%	1.0	1.0	1.0	0.006	0.006	0.015	0.015	0.050	0.046	0.052	0.052
	40.2%	1.0	1.5	1.0	0.059	0.058	0.014	0.014	0.158	0.063	0.076	0.076
	45.1%	1.5	1.0	1.0	0.157	0.157	0.062	0.062	0.233	0.168	0.143	0.143
	55.2%	1.5	1.0	1.5	0.061	0.097	0.035	0.035	0.117	0.188	0.078	0.078
	58.5%	2.0	1.0	2.0	0.109	0.110	0.044	0.044	0.153	0.122	0.111	0.111
	53.8%	2.0	1.5	2.0	0.041	0.086	0.011	0.011	0.108	0.189	0.052	0.052
100	50.2%	1.0	1.0	1.0	0.011	0.010	0.013	0.013	0.054	0.053	0.059	0.059
	40.3%	1.0	1.5	1.0	0.154	0.001	0.019	0.019	0.054	0.051	0.104	0.104
	45.4%	1.5	1.0	1.0	0.266	0.161	0.102	0.102	0.180	0.164	0.185	0.185
	55.3%	1.5	1.0	1.5	0.135	0.066	0.050	0.050	0.067	0.140	0.102	0.102
	58.6%	2.0	1.0	2.0	0.159	0.031	0.061	0.061	0.101	0.036	0.117	0.117
	53.9%	2.0	1.5	2.0	0.118	0.100	0.025	0.025	0.057	0.202	0.080	0.080
200	50.0%	1.0	1.0	1.0	0.007	0.013	0.006	0.006	0.046	0.055	0.041	0.041
	40.0%	1.0	1.5	1.0	0.058	0.020	0.015	0.015	0.137	0.060	0.065	0.065
	45.0%	1.5	1.0	1.0	0.114	0.121	0.161	0.161	0.156	0.151	0.116	0.116
	55.0%	1.5	1.0	1.5	0.044	0.038	0.029	0.029	0.082	0.109	0.065	0.065
	58.4%	2.0	1.0	2.0	0.045	0.020	0.028	0.028	0.070	0.060	0.068	0.068
	53.6%	2.0	1.5	2.0	0.037	0.079	0.020	0.020	0.083	0.184	0.063	0.063
500	49.9%	1.0	1.0	1.0	0.006	0.013	0.008	0.008	0.047	0.046	0.056	0.057
	39.9%	1.0	1.5	1.0	0.055	0.020	0.016	0.016	0.088	0.050	0.136	0.137
	44.9%	1.5	1.0	1.0	0.138	0.120	0.122	0.122	0.178	0.120	0.253	0.306
	54.9%	1.5	1.0	1.5	0.027	0.028	0.023	0.023	0.054	0.060	0.075	0.083
	58.3%	2.0	1.0	2.0	0.027	0.020	0.020	0.020	0.051	0.050	0.108	0.127
	53.5%	2.0	1.5	2.0	0.013	0.103	0.026	0.026	0.059	0.184	0.081	0.081

For investigating the improved power performance for all of our developed tests, we also conducted Monte Carlo simulations with $V = \frac{3}{\mu_1}$ and keeping remaining parameters unchanged. Tables 5.18 and 5.19 illustrate that results. Note that we have used the simulated critical values for $\mu_1 = \mu_2 = 1$ without censoring to calculate powers for each of our proposed tests. Therefore, due to censoring, we have found not exactly the size corrected powers (but almost close to) for our proposed tests under the null hypothesis of no change in the hazard functions of survival times at the levels $\alpha = 0.01$ and $\alpha = 0.05$. Tables 5.18 and 5.19 reveal that results.

Table 5.18: Simulated Powers of Test Statistics $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ for $V = \frac{3}{\mu_1}$, and Optimal Bandwidth b, using the Simulated Critical Value for $\mu_1 = \mu_2 = 1$.

					($\alpha = 0.01$)				($\alpha = 0.05$)			
n	Censor	μ_1	μ_2	μ_c	$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$	$T_n^{(4)}$	$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$	$T_n^{(4)}$
50	50.2%	1.0	1.0	1.0	0.009	0.009	0.010	0.010	0.046	0.048	0.050	0.050
	45.3%	1.0	1.5	1.0	0.263	0.951	0.113	0.113	0.335	0.954	0.234	0.234
	45.1%	1.5	1.0	1.0	0.383	0.609	0.234	0.234	0.424	0.618	0.313	0.313
	36.9%	1.5	2.0	1.0	0.374	0.166	0.101	0.101	0.502	0.184	0.266	0.266
	55.2%	1.5	1.0	1.5	0.269	0.889	0.147	0.147	0.312	0.894	0.224	0.224
	46.7%	1.5	2.0	1.5	0.199	0.471	0.059	0.059	0.332	0.495	0.173	0.173
	58.5%	2.0	1.0	2.0	0.269	0.626	0.160	0.160	0.296	0.642	0.215	0.215
	53.8%	2.0	1.5	2.0	0.279	0.928	0.118	0.118	0.331	0.930	0.208	0.208
100	50.2%	1.0	1.0	1.0	0.010	0.010	0.011	0.011	0.050	0.050	0.053	0.053
	45.2%	1.0	1.5	1.0	0.442	0.631	0.196	0.196	0.499	0.651	0.343	0.343
	45.4%	1.5	1.0	1.0	0.403	0.523	0.284	0.284	0.433	0.531	0.359	0.359
	36.9%	1.5	2.0	1.0	0.357	0.082	0.119	0.119	0.490	0.089	0.272	0.272
	55.3%	1.5	1.0	1.5	0.290	0.944	0.184	0.184	0.319	0.951	0.256	0.256
	46.7%	1.5	2.0	1.5	0.208	0.461	0.081	0.081	0.341	0.514	0.201	0.201
	58.6%	2.0	1.0	2.0	0.249	0.625	0.175	0.175	0.281	0.649	0.229	0.229
	53.9%	2.0	1.5	2.0	0.305	0.973	0.157	0.157	0.346	0.976	0.256	0.256
200	50.0%	1.0	1.0	1.0	0.006	0.006	0.010	0.010	0.053	0.051	0.043	0.043
	44.9%	1.0	1.5	1.0	0.250	1.000	0.148	0.148	0.322	1.000	0.237	0.237
	45.0%	1.5	1.0	1.0	0.384	0.393	0.266	0.266	0.422	0.406	0.333	0.333
	36.7%	1.5	2.0	1.0	0.355	0.333	0.135	0.135	0.479	0.354	0.271	0.271
	55.0%	1.5	1.0	1.5	0.276	0.978	0.183	0.183	0.315	0.979	0.238	0.238
	46.2%	1.5	2.0	1.5	0.196	0.461	0.069	0.069	0.313	0.510	0.171	0.171
	58.4%	2.0	1.0	2.0	0.272	0.560	0.174	0.174	0.299	0.582	0.226	0.226
	53.6%	2.0	1.5	2.0	0.280	0.994	0.147	0.147	0.330	0.996	0.218	0.218
500	49.9%	1.0	1.0	1.0	0.006	0.011	0.008	0.008	0.044	0.049	0.051	0.051
	44.9%	1.0	1.5	1.0	0.015	0.895	0.017	0.017	0.018	0.963	0.077	0.077
	44.9%	1.5	1.0	1.0	0.081	0.143	0.086	0.086	0.084	0.188	0.142	0.142
	36.6%	1.5	2.0	1.0	0.031	0.073	0.042	0.042	0.068	0.094	0.063	0.063
	54.9%	1.5	1.0	1.5	0.049	0.693	0.041	0.041	0.067	0.817	0.070	0.070
	46.3%	1.5	2.0	1.5	0.023	0.096	0.037	0.037	0.031	0.102	0.064	0.064
	58.3%	2.0	1.0	2.0	0.080	0.048	0.050	0.050	0.100	0.055	0.078	0.078
	53.5%	2.0	1.5	2.0	0.032	0.900	0.015	0.015	0.051	0.951	0.065	0.065

Comparing the simulated powers of all the unweighted tests, we observe in Table 5.18 that $T_n^{(2)}$ test performs better than the tests $T_n^{(1)}$, $T_n^{(3)}$ and $T_n^{(4)}$ in all cases considered here. Nevertheless, $T_n^{(1)}$ and $T_n^{(2)}$ tests are based on the same functional with known θ.

Table 5.19: Simulated Powers of Test Statistics $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ for $V = \frac{3}{\mu_1}$, and Optimal Bandwidth b, using the Simulated Critical Value for $\mu_1 = \mu_2 = 1$.

					($\alpha = 0.01$)				($\alpha = 0.05$)			
n	Censor	μ_1	μ_2	μ_c	$T_n^{(5)}$	$T_n^{(6)}$	$T_n^{(7)}$	$T_n^{(8)}$	$T_n^{(5)}$	$T_n^{(6)}$	$T_n^{(7)}$	$T_n^{(8)}$
50	50.2%	1.0	1.0	1.0	0.010	0.010	0.013	0.013	0.050	0.051	0.051	0.051
	45.3%	1.0	1.5	1.0	0.267	0.939	0.122	0.122	0.334	0.942	0.223	0.223
	45.1%	1.5	1.0	1.0	0.385	0.512	0.232	0.232	0.425	0.526	0.322	0.322
	36.9%	1.5	2.0	1.0	0.386	0.064	0.120	0.120	0.505	0.078	0.264	0.264
	55.2%	1.5	1.0	1.5	0.270	0.866	0.147	0.147	0.311	0.874	0.222	0.222
	46.7%	1.5	2.0	1.5	0.207	0.430	0.068	0.068	0.333	0.453	0.172	0.172
	58.5%	2.0	1.0	2.0	0.270	0.581	0.163	0.163	0.296	0.601	0.219	0.219
	53.8%	2.0	1.5	2.0	0.283	0.917	0.121	0.121	0.331	0.918	0.194	0.194
100	50.2%	1.0	1.0	1.0	0.010	0.010	0.011	0.011	0.050	0.050	0.049	0.049
	45.2%	1.0	1.5	1.0	0.288	0.982	0.184	0.184	0.354	0.984	0.268	0.268
	45.4%	1.5	1.0	1.0	0.406	0.394	0.300	0.300	0.433	0.407	0.378	0.378
	36.9%	1.5	2.0	1.0	0.374	0.020	0.158	0.158	0.491	0.078	0.296	0.296
	55.3%	1.5	1.0	1.5	0.293	0.928	0.198	0.198	0.319	0.931	0.257	0.257
	46.7%	1.5	2.0	1.5	0.221	0.441	0.103	0.103	0.348	0.497	0.208	0.208
	58.6%	2.0	1.0	2.0	0.249	0.545	0.185	0.185	0.282	0.569	0.230	0.230
	53.9%	2.0	1.5	2.0	0.310	0.963	0.175	0.175	0.346	0.969	0.252	0.252
200	50.0%	1.0	1.0	1.0	0.007	0.011	0.006	0.006	0.046	0.050	0.044	0.044
	44.9%	1.0	1.5	1.0	0.161	1.000	0.096	0.096	0.161	1.000	0.179	0.179
	45.0%	1.5	1.0	1.0	0.210	0.237	0.170	0.170	0.210	0.250	0.297	0.304
	36.7%	1.5	2.0	1.0	0.313	0.013	0.119	0.119	0.313	0.054	0.276	0.276
	55.0%	1.5	1.0	1.5	0.134	0.974	0.097	0.097	0.134	0.977	0.170	0.170
	46.2%	1.5	2.0	1.5	0.170	0.448	0.059	0.059	0.170	0.497	0.171	0.171
	58.4%	2.0	1.0	2.0	0.099	0.476	0.074	0.074	0.099	0.498	0.151	0.151
	53.6%	2.0	1.5	2.0	0.156	0.992	0.092	0.092	0.156	0.993	0.149	0.149
500	49.9%	1.0	1.0	1.0	0.006	0.011	0.008	0.008	0.048	0.050	0.052	0.053
	44.9%	1.0	1.5	1.0	0.017	1.000	0.051	0.051	0.017	1.000	0.107	0.107
	44.9%	1.5	1.0	1.0	0.038	0.105	0.207	0.240	0.038	0.117	0.323	0.324
	36.6%	1.5	2.0	1.0	0.060	0.011	0.114	0.114	0.069	0.051	0.260	0.261
	54.9%	1.5	1.0	1.5	0.033	1.000	0.054	0.057	0.071	1.000	0.159	0.162
	46.3%	1.5	2.0	1.5	0.011	0.450	0.032	0.032	0.011	0.522	0.274	0.274
	58.3%	2.0	1.0	2.0	0.027	0.405	0.086	0.096	0.067	0.428	0.108	0.134
	53.5%	2.0	1.5	2.0	0.013	1.000	0.023	0.023	0.013	1.000	0.476	0.476

Table 5.19 reports that the test $T_n^{(6)}$ has better powers than other weighted tests $T_n^{(5)}$, $T_n^{(7)}$ and $T_n^{(8)}$ under the parameter combinations considered in this experiment with different censoring percentages.

5.1.3 Evaluation of Change Point Estimators

Let the estimator of the change point k^* be the point k where the test statistic takes its maximum for the unweighted test $T_n^{(1)}$, that is,

$$\hat{k^*} = \arg \max_{1 \leq k < n} T_n^{(1)}.$$

This estimator also holds for the unweighted tests $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$. However, for the weighted test $T_n^{(5)}$, the change point estimator is given by

$$\frac{\hat{k^*}}{n} = \arg \max_{0 < t < 1} T_n^{(5)},$$

which is also true for the tests $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$.

Table 5.20: Simulated Estimates of Change Point in the Hazard Distribution using Test Statistics $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ for $V = \frac{3}{\mu_1}$, and Optimal Bandwidth b.

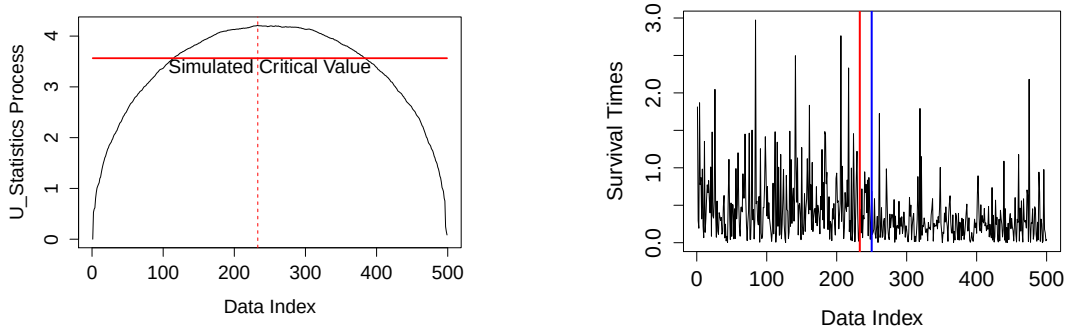
						Estimate of k^*				Std. Error of k^*			
n	k^*	Censor	μ_1	μ_2	μ_c	$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$	$T_n^{(4)}$	$T_n^{(1)}$	$T_n^{(2)}$	$T_n^{(3)}$	$T_n^{(4)}$
50	25	55.2%	1.5	1.0	1.5	16.65	**29.00**	32.08	31.79	22.42	**9.34**	18.88	19.81
100	50	55.3%	1.5	1.0	1.5	33.25	**56.88**	63.03	62.77	45.98	**17.58**	41.03	42.02
200	100	55.0%	1.5	1.0	1.5	66.06	**106.32**	123.05	121.76	92.97	**25.38**	86.85	89.21
500	250	54.9%	1.5	1.0	1.5	160.27	**253.63**	313.50	302.55	232.30	**33.97**	222.80	228.70

Table 5.21: Simulated Estimates of Change Point in the Hazard Distribution using Test Statistics $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ for $V = \frac{3}{\mu_1}$, and Optimal Bandwidth b.

						Estimate of k^*				Std. Error of k^*			
n	k^*	Censor	μ_1	μ_2	μ_c	$T_n^{(5)}$	$T_n^{(6)}$	$T_n^{(7)}$	$T_n^{(8)}$	$T_n^{(5)}$	$T_n^{(6)}$	$T_n^{(7)}$	$T_n^{(8)}$
50	25	55.2%	1.5	1.0	1.5	15.97	**26.44**	31.32	30.78	22.87	**6.78**	18.69	19.11
100	50	55.3%	1.5	1.0	1.5	32.56	**50.31**	62.01	61.74	46.42	**6.39**	39.21	39.48
200	100	55.0%	1.5	1.0	1.5	36.50	**99.83**	101.80	102.01	67.42	**5.01**	78.98	78.95
500	250	54.9%	1.5	1.0	1.5	39.46	**249.84**	237.52	226.39	41.29	**9.68**	190.13	189.55

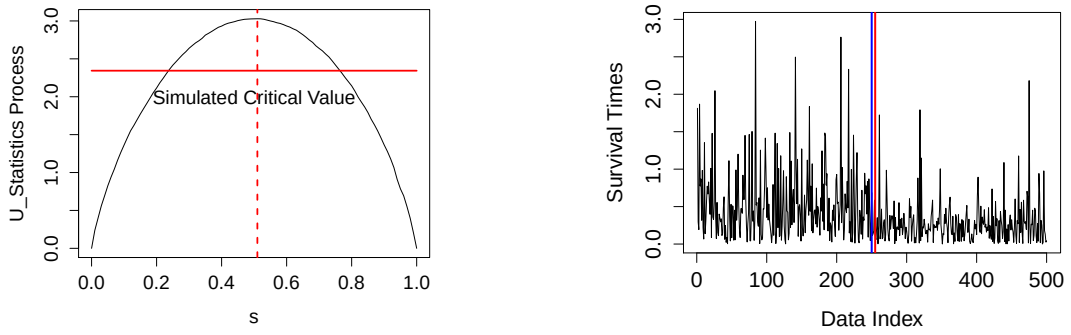
It is evident from Tables 5.20 and 5.21 that change point estimators are quite good for the two-sided tests $T_n^{(2)}$ and $T_n^{(6)}$. Tests $T_n^{(2)}$ and $T_n^{(6)}$ use knowledge of θ, which is too good for practice. The unweighted tests $T_n^{(3)}$ and $T_n^{(4)}$ for unknown θ overestimate

k^* due to particular λ, λ^*. On the other hand, the weighted tests $T_n^{(7)}$ and $T_n^{(8)}$ for unknown θ perform better, as other weight $q(t)$ may be helpful.



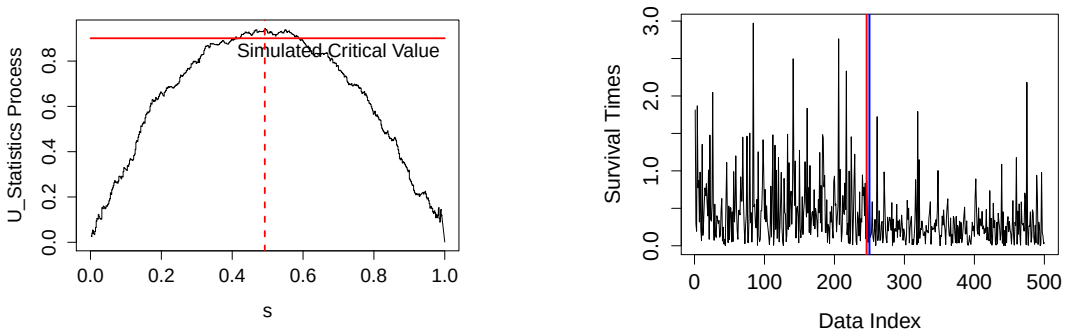
(a) (b)

Figure 5.1: Detection of change point in the hazard distribution using the test $T_n^{(2)}$. (a) U-statistics process for $T_n^{(2)}$, simulated critical value, and the location of significant change point are presented; and (b) True (blue) and estimated (red) change points are located in the simulated survival data.



(a) (b)

Figure 5.2: Detection of change point in the hazard distribution using the test $T_n^{(6)}$. (a) U-statistics process for $T_n^{(6)}$, simulated critical value, and the location of significant change point are presented; and (b) True (blue) and estimated (red) change points are located in the simulated survival data.



(a) (b)

Figure 5.3: Detection of change point in the hazard distribution using the test $T_n^{(8)}$. (a) U-statistics process for $T_n^{(8)}$, simulated critical value, and the location of significant change point are presented; and (b) True (blue) and estimated (red) change points are located in the simulated survival data.

Figures 5.1, 5.2, and 5.3 present the potential impact of our proposed tests to the simulated data with $\mu_1 = 1.0$, $\mu_2 = 1.5$, and $\mu_c = 1.0$ for $n = 500$ at the level $\alpha = 0.05$, we observed that our tests significantly identified the change point in the hazard distribution with 56.3% censoring.

5.2 Real Data Applications

We apply our developed tests for detecting change points in the hazard functions over the course of time to two real data examples; breast cancer mortality, and cell stimulus responses. The first data set is well-known survival data and the later one is a cognitive data. Breast cancer mortality data contains 78.9% of censored observations, whereas the cell stimulus response data contains only 0.5% of censored observations, hence we are interested in investigating our developed methodologies with very high as well as with very low percentages of censored scenarios. As the value of the true parameter θ is unknown for these two data examples, we applied the unweighted tests $T_n^{(3)}$ and $T_n^{(4)}$ along with the weighted tests $T_n^{(7)}$ and $T_n^{(8)}$ to investigate the change point in the hazard distributions.

5.2.1 Cell Stimulus Response Data

We consider the cell stimulus response data again, which we have studied intensively in Chapter 3. Let us first note that the data contain delays, i.e., waiting times between the stimulus and the onset of the response, such that we may interpret them as survival times which are also waiting times between entering the study (time of origin) and the occurrence of the event of interest (time of failure).

Applying the change point tests developed in Chapter 4 to this data is not completely justified as the sequence of delay data are coming from the same cell and are likely not a sequence of independent data ordered in time. Nevertheless, we apply hazard function change point tests to these data and look what happens.

The cell stimulus response data contains only 15 censored observations out of 3000, so it has very low censoring percentage 0.5%. To detect a change in the hazard distribution over the times in the cell stimulus responses, we used unweighted tests $T_n^{(3)}$ and $T_n^{(4)}$, and weighted tests $T_n^{(7)}$ and $T_n^{(8)}$, where we estimated the true parameter θ by $\hat{\theta}^{(3)}(\lfloor \frac{n}{2} \rfloor)$ and the variance by $\hat{\sigma}_n^2$. We considered the symmetric kernel function $\tilde{K}(x, y) = xy$; $V = 3.1$ to fulfill Assumption (D) and (4.4), since $\max(T_i) = 3.9$ and $\max(C_i) = 3.7$; and the optimal bandwidth $b = 3e^{-11}$ for $n = 3000$ as per the condition (4.17).

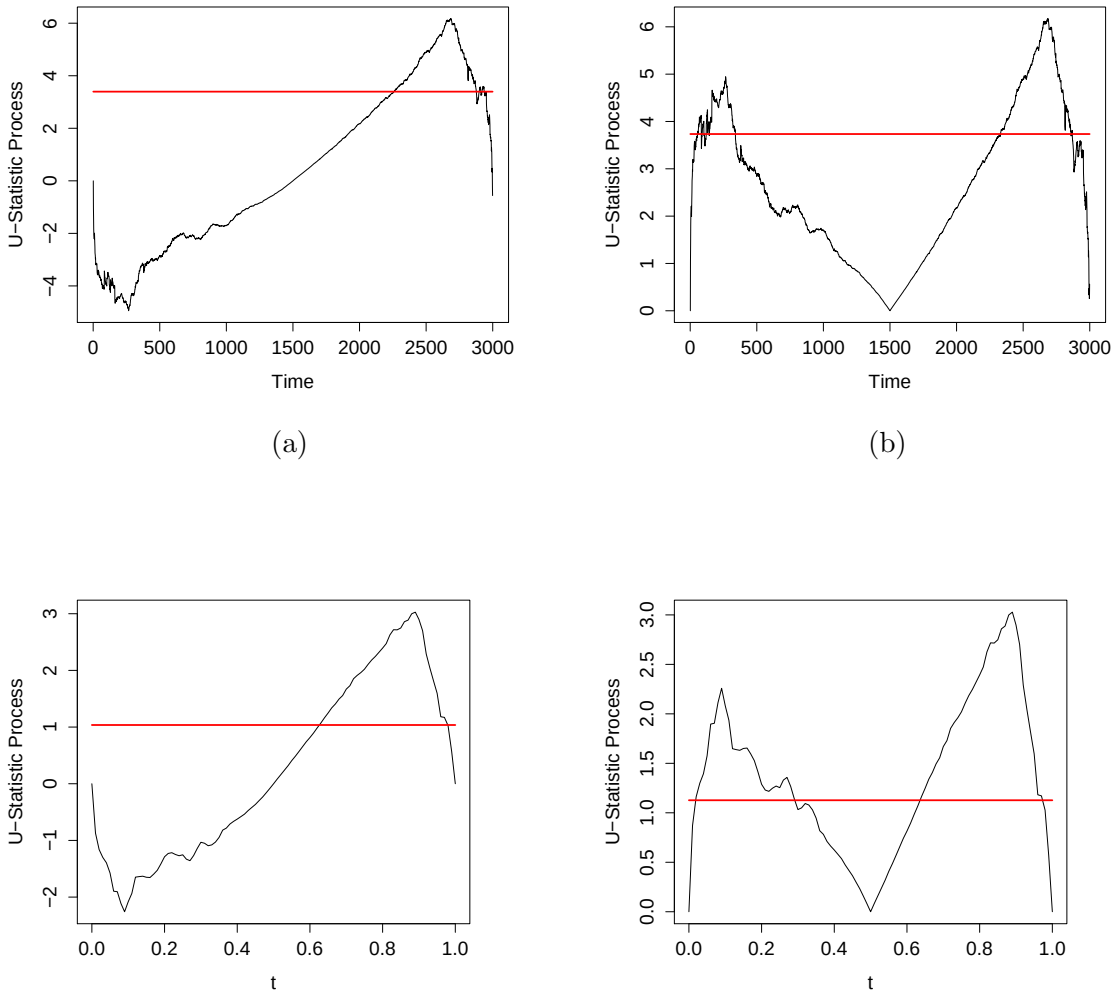


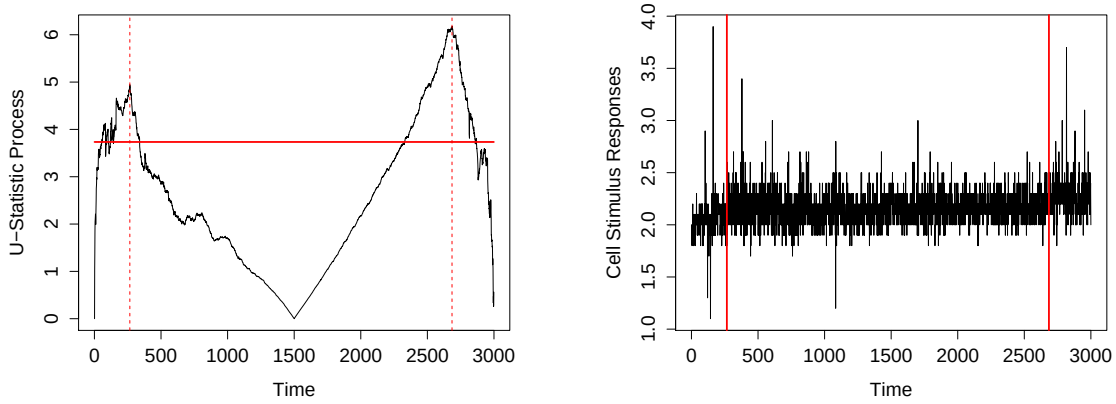
Figure 5.4: Detection of change point in the hazard distribution using the test statistic $T_n^{(3)}$ in (a), $T_n^{(4)}$ in (b), $T_n^{(7)}$ in (c), and $T_n^{(8)}$ in (d) with the optimal bandwidth $b = 3e^{-11}$, and the asymptotic critical values are presented by the red horizontal lines.

As we conducted our test at 5% level of significance, hence the critical value for the test $T_n^{(3)}$ is $c_\alpha^{(1)}(n) = 3.3948$, and and for the test $T_n^{(4)}$ is $c_\alpha^{(2)}(n) = 3.7346$ with $n = 3000$ from Tables 5.2, and 5.3, respectively. The critical values for the weighted tests $T_n^{(7)}$ and $T_n^{(8)}$ are observed as $c_\alpha^{(7)}(n) = 1.0364$ and $c_\alpha^{(8)}(n) = 1.1268$, respectively, from Tables 5.8, and 5.9 at $\alpha = 0.05$ with $n = 3000$. There is an evidence of significant change point in the hazard distribution of cell stimulus responses at 5% level of significance in Figure 5.4.

Multiple change points can also be tested by using the binary segmentation procedure described in Algorithm 3, Chapter 3. For our developed tests we are rewriting that in the following algorithm for the test $T_n^{(4)}$.

Algorithm 5 The generic Binary Segmentation algorithm to detect multiple change points using the test $T_n^{(4)}$.

1: Set the data for testing $X_1, \ldots, X_n$ and the optimal bandwidth b.
2: Calculate the test statistic $T_n^{(4)}$ and the critical value $c_\alpha^{(2)}(n)$ at α level of significance. if $T_n^{(4)} > c_\alpha^{(2)}(n)$ then select $\tilde{m}_1 := \arg\max T_n^{(4)}$.
3: Split the data in to two segments, i.e., $X_1, \ldots, X_{\tilde{m}_1}$ and $X_{\tilde{m}_1+1}, \ldots, X_n$, calculate the critical value $c_\alpha^{(2)}(n)$ and redefine the bandwidth b for each segment with different sample sizes n. Repeat step 1 and 2 for each segment until no significant change points are detected.
4: Obtain $\tilde{m}_1, \ldots, \tilde{m}_q$.

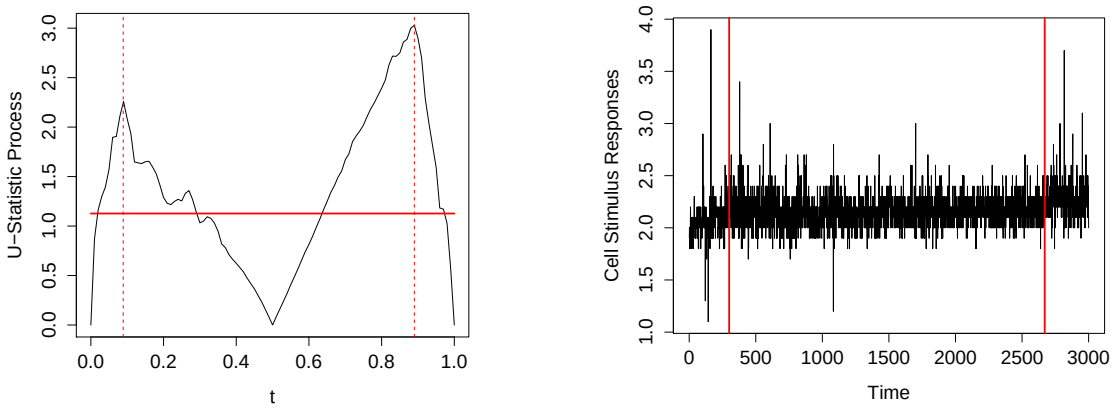


(a) (b)

Figure 5.5: U-statistic process for the test statistic $T_n^{(4)}$ to detect multiple change points in the hazard distributions of cell stimulus responses with the asymptotic critical value is presented by the red horizontal line and the location of change points are indicated by a dotted vertical red line in (a); and the estimated multiple change points are indicated by the red vertical line in (b).

Using the binary segmentation procedure for test $T_n^{(4)}$ with $c_\alpha^{(2)}(n) = 3.7346$ for the entire sample $n = 3000$ and the bandwidth $b = 3e^{-11}$, we found that the change point in the hazard distribution of cell stimulus responses at $\tilde{m}_1 = 2685$ using

$\tilde{m}_1 := \arg\max T_n^{(4)}$. At the step 2 of Algorithm 5, we split the entire sample into two parts: $X_1, \ldots, X_{2685}$ and $X_{2686}, \ldots, X_{3000}$. Hereafter, we recalculated the critical values for each segments and found the significant change point at $\tilde{m}_2 = 266$ for the first segment of data $X_1, \ldots, X_{2685}$ with $c_\alpha^{(2)}(n) = 3.7318$ for the sample size $n = 2685$ and the bandwidth $b = 3e^{-13}$, which is presented in Figure 5.5.



(a) (b)

Figure 5.6: U-statistic process for the test statistic $T_n^{(8)}$ to detect multiple change points in the hazard distributions of cell stimulus responses with the asymptotic critical value is presented by the red horizontal line and the location of change points are indicated by a dotted vertical red line in (a); and the estimated multiple change points are indicated by the red vertical line in (b).

Again, using the binary segmentation procedure of Algorithm 5 for the test $T_n^{(8)}$ with $c_\alpha^{(8)}(n) = 1.1268$ for the entire sample $n = 3000$ with the bandwidth $b = 3e^{-11}$, we found the change point in hazard function at $t = 0.89$ so in the sample at $\tilde{m}_1 = 2670$ using $\tilde{m}_1 := \arg\max T_n^{(8)}$. At the step 2 of Algorithm 5, we split the entire sample into two parts: $X_1, \ldots, X_{2670}$ and $X_{2671}, \ldots, X_{3000}$. Hereafter, we recalculated the critical values for each segments and found the significant change point at $t = 0.10$ which turns on the sample at $\tilde{m}_2 = 300$ for the first segment with $c_\alpha^{(8)}(n) = 1.1252$ for the sample size $n = 2670$ and the bandwidth $b = 3e^{-13}$, which is reported in Figure 5.6.

Figure 5.5 illustrates the two significant change points in the hazard distribution

at 2685 and 266 using the test $T_n^{(4)}$, we also estimated two change points in the hazard distribution of cell stimulus responses at 2670 and 300 using the test $T_n^{(8)}$ which are shown in Figure 5.6. Both the significant change points, which are detected by $T_n^{(4)}$ and $T_n^{(8)}$, are quite close.

The change points identified by these tests are close to the change points which are found in Chapter 3 by various other methods. So, it seems to be confirmed that there are changes happening around observations 265 and 2680, i.e., roughly after 9% and 89% of the sample, respectively. If this is rather a change of hazard, a change of the mean, or a change in regression has to be investigated further, using data from other cells too.

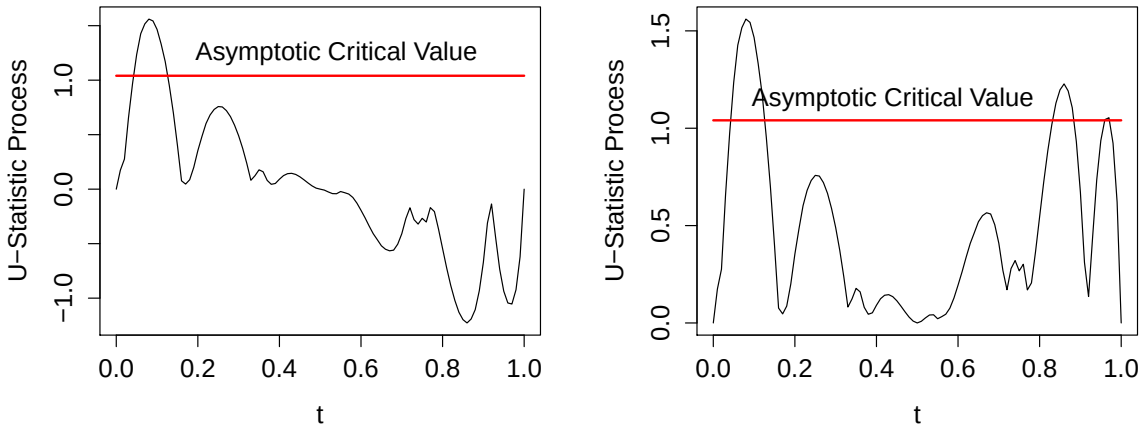
5.2.2 Breast Cancer Mortality Data

We use the Surveillance, Epidemiology, and End Results (SEER) Program Public-Use Data (1973-2010), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2013, based on the November 2012 submission. This data contains cancer incidence and survival for cases diagnosed from 1973 to 2010, follow-up continued until December 31, 2012. Our focus is on the breast cancer mortality data only. Calculation of hazard functions will be based on the breast cancer patient's risk of death, which is described in Section 2.4.2.

For the purpose of analysis we excluded patients with unknown follow-up time and not a case of first tumor. We define an event as death from breast cancer. If a patient dies from another cause they are censored at the time of their death. The breast cancer mortality data contains 78.9% of censored observations, where there are 576250 observations among them 455196 are being censored. For such big data set, we used only the weighted tests $T_n^{(7)}$ and $T_n^{(8)}$ to find out the change point in the hazard functions, since the convergences of the unweighted tests $T_n^{(3)}$ and $T_n^{(4)}$ are rather slow and thus these tests are conservative only for small samples. To calculate the hazard rates, we converted monthly survival times (*SurvM*) into yearly survival times and used *DeathCause* variable as the censoring indicator variable.

Before using tests $T_n^{(7)}$ and $T_n^{(8)}$ we considered some values for our parameters of interest to fulfill all the assumptions in Theorem 4.3.4 of Chapter 4. We considered the symmetric kernel function $\tilde{K}(x, y) = xy$; $V = 30$ for fulfilling Assumption (D)

and (4.4), since $\max(T_i) = 37.92$ and $\max(C_i) = 37.41$; and the bandwidth $b = 0.05$ for $n = 576250$ as per the condition (4.17). We conducted tests $T_n^{(7)}$ and $T_n^{(8)}$ to detect change points in the hazard distribution at 5% level of significance with the asymptotic critical values $c_\alpha^{(7)}(n) = 1.0479$ and $c_\alpha^{(8)}(n) = 1.1436$ from Table 5.8 and Table 5.9, respectively, corresponding U-statistic processes are illustrated in Figure 5.7.



(a) (b)

Figure 5.7: Detection of change point in the hazard distribution using test $T_n^{(7)}$ in (a), and test $T_n^{(8)}$ in (b) with the Sheather-Jones plug-in bandwidth $b = 0.05$, and the asymptotic critical values are presented by the red horizontal lines.

Figure 5.7 implies a significant change point in the hazard distribution at 46100 data position. Multiple change points in the hazard distribution can also be tested and estimated by using the binary segmentation procedure described in Algorithm 3 of Chapter 3. We rewrite that algorithm for test $T_n^{(8)}$ as Algorithm 6 and the necessary asymptotic critical values for tests $T_n^{(7)}$ and $T_n^{(8)}$ are reported in Table 5.22.

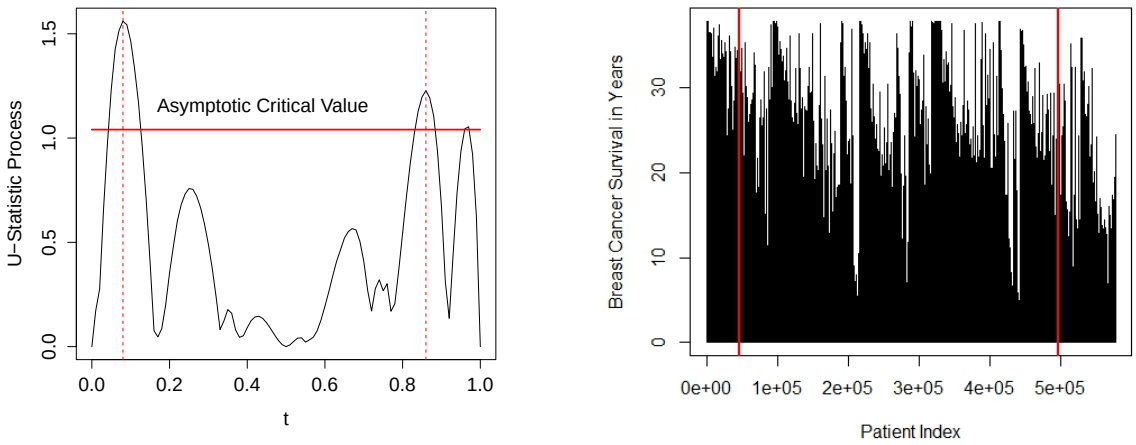
Algorithm 6 The generic Binary Segmentation algorithm to detect multiple change points using the test $T_n^{(8)}$.

1: Set the data for testing $X_1, \ldots, X_n$ and define the optimal bandwidth b and parameter V.
2: Calculate the test statistic $T_n^{(8)}$ and the critical value $c_\alpha^{(8)}(n)$ at α level of significance. if $T_n^{(8)} > c_\alpha^{(8)}(n)$ then select $\tilde{m}_1 := \arg\max T_n^{(8)}$.
3: Split the data in to two segments, i.e., $X_1, \ldots, X_{\tilde{m}_1}$ and $X_{\tilde{m}_1+1}, \ldots, X_n$, calculate the critical value $c_\alpha^{(8)}(n)$ and redefine the optimal bandwidth b and parameter V for each segment with different sample sizes n. Repeat step 1 and 2 for each segment until no significant change points are detected.
4: Obtain $\tilde{m}_1, \ldots, \tilde{m}_q$.

Table 5.22: Some Selected Critical Values of $T_n^{(7)}$ and $T_n^{(8)}$ Test Statistics at various α and n.

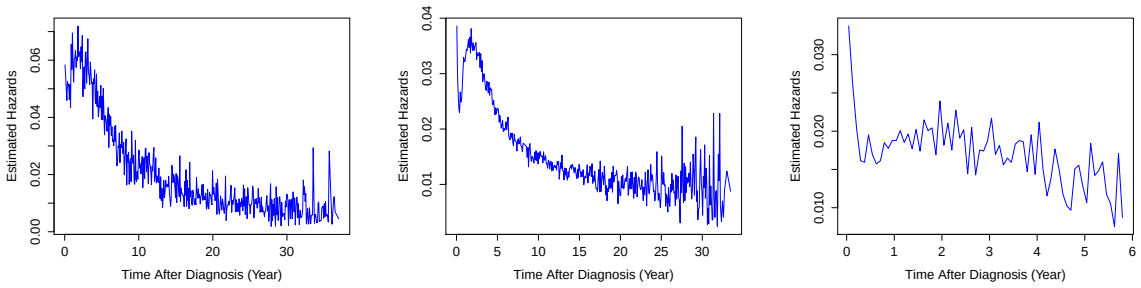
		α			
$c_\alpha^{(i)}(n)$	n	0.005	0.01	0.025	0.05
$c_\alpha^{(7)}(n)$	576250	1.3318	1.2579	1.1400	1.0406
	530150	1.3136	1.2385	1.1351	1.0398
$c_\alpha^{(8)}(n)$	576250	1.4056	1.3335	1.2242	1.1375
	530150	1.3932	1.3198	1.2161	1.1333

Using the binary segmentation procedure of Algorithm 6 for the test $T_n^{(8)}$ with $c_\alpha^{(8)}(n) = 1.1375$ for the entire sample $n = 576250$ and the Sheather-Jones plug-in optimal bandwidth $b = 0.05$, we found that the change point in the hazard distribution at $t = 0.08$ so in the sample at $\tilde{m}_1 = 46100$ by $\tilde{m}_1 := \arg\max T_n^{(8)}$. At the step 2 of Algorithm 5, we split the entire sample into two parts: $X_1, \ldots, X_{46100}$ and $X_{46101}, \ldots, X_{576250}$. Hereafter, we recalculated the critical values for each segments and found the significant change point at $t = 0.86$ which turns on the sample at $\tilde{m}_2 = 495575$. Figure 5.8 illustrates those significant change points in the hazard distribution for breast cancer mortality data.



(a) (b)

Figure 5.8: U-statistic process for the test statistic $T_n^{(8)}$ to detect multiple change points in the hazard distributions of breast cancer mortality with the asymptotic critical value is presented by the red horizontal line and the location of change points are indicated by a dotted vertical red line in (a); and the estimated multiple change points are indicated by the red vertical line in (b).



(a) (b) (c)

Figure 5.9: Estimated hazard functions using Nelson-Aalen estimator for the breast cancer mortality data during the time period 1973 to 1976 in (a); 1977 to 2004 in (b); and 2005 to 2010 in (c).

Figure 5.8 reveals the two significant change points at years $\tilde{m}_1 = 2.79$ and $\tilde{m}_2 = 31.04$ in the hazard distributions of the breast cancer mortality data using our developed test $T_n^{(8)}$ at the level $\alpha = 0.05$. Initially, using test $T_n^{(8)}$ we observed the significant change point at $t = 0.08$ and $t = 0.86$. The total observation period consists of 434

months. Hence, for the first change point $t = 0.08$, we found $nt = 434 \times 0.08 = 34.72$, which in years is $\tilde{m}_1 = 2.79$. Similarly, for the second change point $t = 0.86$, we found $nt = 434 \times 0.86 = 373.24$, which is in years $\tilde{m}_2 = 31.04$. Hence, the changes in hazard happened around the years 1976 and 2004.

Estimated hazard functions using the Nelson-Aalen estimator for different time periods of the breast cancer mortality data show three different patterns of hazard functions in Figure 5.9. During the time period from 1973 to 1976, breast cancer patients in the United States experienced higher hazard rate, e.g., 0.06, of mortality, than other two time periods. Starting from the higher hazard rate, breast cancer patients attended 0.02 hazard rate of mortality after 8 years during the time period 1973 to 1976, after 6 years during the time period 1977 to 2004, and after only half a year during the time period 2005 to 2010. These changes in hazard functions happen due to the improvement in medical breakthroughs, treatments or diagnosis at that time.

5.3 Discussion

This chapter is devoted to investigating the theoretical findings of our proposed tests in Chapter 4 by simulations, and hence, the practical impact of the real data analysis is demonstrated by successfully detecting the significance change points in the hazard functions to the breast cancer mortality data and the cell stimulus response data.

Chapter 6: Summary and Conclusions

Motivated by the long-standing demands in survival analysis to have a deeper understanding of the overall survival trends with its changes for an entire population, we rigorously developed nonparametric methods for detecting change points in the hazard distribution under the right random censorship, since nearly a universal feature of survival data is censoring, the most common form of which is right-censoring. Our proposed tests, inspired by Horváth [62], are based on U-statistics to detect change in the hazard distribution using symmetric and antisymmetric kernels. Our developed methods do not require any model assumptions to find out significant changes in the hazard distribution, hence, are not affected by model misspecification errors. Instead, following a model implicitly and concerning about corresponding model misspecification errors which is done by all the existing methods, our approach does not need to take into account model misspecification errors.

We extended the U-statistic-type nonparametric change point statistics to detect the change in the distribution of the observations under random censorship considered by Horváth [62] to the hazard function case. Using symmetric and antisymmetric kernel functions we developed different types of test statistics, for instance, we proposed eight different test statistics for symmetric kernel function in Sections 4.3 and 4.3.2 of Chapter 4, among them $T_n^{(1)}$, $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$ are unweighted test statistics whereas $T_n^{(5)}$, $T_n^{(6)}$, $T_n^{(7)}$ and $T_n^{(8)}$ are weighted test statistics. Another four test statistics $T_n^{(9)}$, $T_n^{(10)}$, $T_n^{(11)}$ and $T_n^{(12)}$ also proposed in Section 4.5 for antisymmetric kernel function, where $T_n^{(11)}$ and $T_n^{(12)}$ are weighted test statistics. We derived the asymptotic distributions of our proposed test statistics; where the the unweighted statistics $T_n^{(1)}$, $T_n^{(2)}$ $T_n^{(3)}$, $T_n^{(4)}$, $T_n^{(9)}$ and $T_n^{(10)}$ have a Gumbel extreme value type distribution; the weighted tests $T_n^{(5)}$ and $T_n^{(6)}$ have the limiting Gaussian process distribution; $T_n^{(7)}$ and $T_n^{(8)}$ have an asymptotic Brownian bridge type process distribution; and $T_n^{(11)}$ and

$T_n^{(12)}$ also have asymptotic Brownian bridge type process distribution, i.e., $\sup_{0<t<1} \frac{\{B(t)\}}{q(t)}$ and $\sup_{0<t<1} \frac{|B(t)|}{q(t)}$, respectively. As we already know that the convergences of Gumbel extreme value type distribution are rather slow and thus these tests are conservative for small samples, the asymptotic critical values are reported in Tables 5.2-5.3. We ran Monte Carlo simulation with 20,000 repetitions to simulate the approximated critical values for all the weighted tests and presented in Tables 5.6-5.11. We also showed that our proposed tests have an asymptotic power of 1 under the alternative hypothesis in Section 4.4.

We conducted different simulations to investigate the theoretical finding of our proposed tests in Chapter 5. In the simulation study, we used the kernel function $\tilde{K}(x, y) = xy$; the value of the parameter either $V = \frac{2}{\mu}$ having assumption $F(V) \approx 0.8646 < 1$ or $V = \frac{3}{\mu}$ having $F(V) \approx 0.9502 < 1$; the Epanechnikov kernel function $K(x)$ with the Sheather-Jones plug-in optimal bandwidth b were used to estimate the kernel density estimator $h_n(x)$ (4.16) in the ratio-type kernel hazard estimator; and the weight function $q(t) = (t(1-t))^{0.2}$ was considered for comparing with the standard weight $\sqrt{t(1-t)}$. We did 1000 simulations to investigate the simulated and approximated critical values in Section 5.1.1, where it turned out that the critical values obtained from the limiting distribution for all these tests are roughly close. Hence, we can use the approximated critical values from Tables 5.2-5.3 and 5.6-5.11 for the corresponding proposed tests to detect the change point in the hazard distribution.

We found the better power performance of all our proposed tests with $V = \frac{3}{\mu_1}$. This is due to the fact that the critical values of all these tests with $V = \frac{3}{\mu_1}$ slightly smaller than that of all these tests with $V = \frac{2}{\mu_1}$. Hence, it is apparently very important to choose an appropriate V for all these proposed tests.

We used two actual data sets to illustrate the potential impact of our developed tests for identifying the significant change point in the hazard distribution. In the survival study, we examined that the breast cancer mortality has two significant change points around 1974 and 2004 in mortality hazard distribution. In the cognitive study, we realized two significant change points in the cell stimulus response's hazard distribution at about 9% and 89% of the whole sample. This is in line with the change points found by the majority of more common change point tests, which we have studied in

Chapter 3. For detecting multiple change points in the hazard distribution, we used the binary segmentation procedure described in Algorithm 3 of Chapter 3.

Chapter 3 attempts to raise the awareness of the missing data problem in change point analysis by illustrating different classical nonparametric change point procedures. Additionally, we also demonstrated the change point methods for detecting a change in the distribution of the observations under either missingness or random censorship in Section 3.4.3. Hence, we experienced that it is important to know and use the right data feature for investigating change point correctly.

In Chapter 2, we also developed the Cross-Entropy algorithm to estimate multiple change points for the hazard functions in the *piecewise constant model* (2.25) as well as the *piecewise linear model* (2.29), and observed that the CE method performed better than the existing counterpart Nelder-Mead Simplex algorithm through simulations and real data applications. This method can also be used as a detection method for identifying multiple change points in the hazard functions.

Finally, we wish to conclude by mentioning some related areas as further research. The estimation of the change point in the hazard functions is insufficient without previous testing of the existence of this change along with inevitable censoring. Hence, this thesis contains useful, not affected by the model misspecification errors and easily manageable tests for detecting changes in the hazard distribution under the right random censoring. We have demonstrated their practical applications through simulations and real data examples. However, we did not establish the consistency and the asymptotic distribution of the change point estimator in this context, which can be considered as a further research topic.

Another issue, not studied in this thesis, is the choice of the parameter V. Certainly, the performances of our developed tests depend on the parameter V and the bandwidth b. Although we have shown the satisfactory results in the simulation study, the question of finding adaptive methods for the choice of the parameter V is not explored yet and remains open.

A fascinating future research area is to develop a nonparametric change point test of the hazard functions for the functional data feature in the case of i.i.d. as well as

censoring. Furthermore, our proposed methods can be extended to the right censoring and left truncation scenario, which seems more rational in some survival studies.

Another interesting field of further research is an implementation of the sequential (online) change point method in the hazard functions, which allows detecting change as soon as possible while the data are collected. As in survival analysis, the sequential method is more reasonable.

Appendix A: Breast Cancer Dataset

We use the Surveillance, Epidemiology, and End Results (SEER) Program Public-Use Data (1973-2010), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2013, based on the November 2012 submission. This data contains cancer incidence and survival for cases diagnosed from 1973 to 2010, follow-up continued until December 31, 2012. Our focus is on the breast cancer data only.

Table A.1: Variables descriptions for the Breast Cancer Patients using SEER Data from 1973-2010.

Variables Name	**Description and Values**
Age	Age at diagnosis. Actual age in years. 999 for unknown age.
Race	Race of the Patients: 1- White, 2-Black, 3- Other(American Indian/AK Native, Asian/Pacific Islander), 7- other unspecified, 9- unknown.
DateYr	The year of diagnosis of the tumor
DateMo	The month of diagnosis of the tumor
DeathCause	Indicates cause-specific survival: 0- Alive or dead of other cause,1- Dead, 9- N/A not first tumor.
SurvM	Survival Months of the patients. 9999 for unknown.

Appendix B: Simulated Critical Values

Table B.1: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(1)}(n)$ of Test Statistics $T_n^{(1)}$ and $T_n^{(3)}$ for $V = \frac{3}{\mu}$ and Optimal Bandwidth b.

		$(\alpha = 0.01)$			$(\alpha = 0.05)$			$(\alpha = 0.10)$		
n	μ	Sim.CV $T_n^{(1)}$	Sim.CV $T_n^{(3)}$	App.CV $c_\alpha^{(1)}(n)$	Sim.CV $T_n^{(1)}$	Sim.CV $T_n^{(3)}$	App.CV $c_\alpha^{(1)}(n)$	Sim.CV $T_n^{(1)}$	Sim.CV $T_n^{(3)}$	App.CV $c_\alpha^{(1)}(n)$
50	1.0	0.4834	3.1680	4.1842	0.1073	1.8071	3.1974	0.0000	1.3146	2.7616
	1.5	0.4901	2.7742	4.1842	0.1129	1.5525	3.1974	0.0000	1.1074	2.7616
	2.0	0.4847	2.4753	4.1842	0.1164	1.3998	3.1974	0.0000	0.9842	2.7616
	2.5	0.4741	2.1789	4.1842	0.1159	1.2228	3.1974	0.0000	0.8617	2.7616
	3.0	0.4735	2.2476	4.1842	0.1292	1.2733	3.1974	0.0000	0.8895	2.7616
100	1.0	0.4472	2.9021	4.1735	0.0882	1.6839	3.2408	0.0000	1.2777	2.8289
	1.5	0.4605	2.7801	4.1735	0.0979	1.6141	3.2408	0.0000	1.2073	2.8289
	2.0	0.4609	2.3635	4.1735	0.0974	1.3648	3.2408	0.0000	1.0205	2.8289
	2.5	0.4448	1.9063	4.1735	0.0974	1.0890	3.2408	0.0000	0.8401	2.8289
	3.0	0.4661	1.6888	4.1735	0.0975	0.9705	3.2408	0.0000	0.7244	2.8289
200	1.0	0.4299	2.8189	4.1718	0.0818	1.8477	3.2792	0.0000	1.3958	2.8850
	1.5	0.4300	2.1103	4.1718	0.0831	1.3867	3.2792	0.0000	1.0358	2.8850
	2.0	0.4216	2.0166	4.1718	0.0812	1.3049	3.2792	0.0000	0.9851	2.8850
	2.5	0.4331	1.7357	4.1718	0.0821	1.1503	3.2792	0.0000	0.8462	2.8850
	3.0	0.4183	1.5737	4.1718	0.0814	1.0307	3.2792	0.0000	0.7723	2.8850
500	1.0	0.5746	3.3709	4.1762	0.1491	2.0741	3.3235	0.0000	1.3620	2.9469
	1.5	0.5602	2.6758	4.1762	0.1368	1.6466	3.3235	0.0000	1.0857	2.9469
	2.0	0.5846	2.3250	4.1762	0.1467	1.4291	3.3235	0.0000	0.9298	2.9469
	2.5	0.5866	2.1308	4.1762	0.1385	1.3173	3.3235	0.0000	0.8444	2.9469
	3.0	0.5746	1.9264	4.1762	0.1313	1.1741	3.3235	0.0000	0.7667	2.9469

Table B.2: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(2)}(n)$ of Test Statistics $T_n^{(2)}$ and $T_n^{(4)}$ for $V = \frac{3}{\mu}$ and Optimal Bandwidth b.

		($\alpha = 0.01$)			($\alpha = 0.05$)			($\alpha = 0.10$)		
n	μ	Sim.CV $T_n^{(2)}$	Sim.CV $T_n^{(4)}$	App.CV $c_\alpha^{(2)}(n)$	Sim.CV $T_n^{(2)}$	Sim.CV $T_n^{(4)}$	App.CV $c_\alpha^{(2)}(n)$	Sim.CV $T_n^{(2)}$	Sim.CV $T_n^{(4)}$	App.CV $c_\alpha^{(2)}(n)$
50	1.0	1.5811	3.1680	4.6039	1.5748	1.8071	3.6171	1.5715	1.3146	3.1813
	1.5	1.5810	2.7742	4.6039	1.5747	1.5525	3.6171	1.5715	1.1074	3.1813
	2.0	1.5809	2.4753	4.6039	1.5746	1.3998	3.6171	1.5714	0.9842	3.1813
	2.5	1.5809	2.1789	4.6039	1.5747	1.2228	3.6171	1.5714	0.8617	3.1813
	3.0	1.5806	2.2476	4.6039	1.5741	1.2733	3.6171	1.5709	0.8895	3.1813
100	1.0	2.2263	2.9021	4.5701	2.2194	1.6839	3.6374	2.2160	1.2777	3.2256
	1.5	2.2260	2.7801	4.5701	2.2191	1.6141	3.6374	2.2155	1.2073	3.2256
	2.0	2.2261	2.3635	4.5701	2.2192	1.3648	3.6374	2.2156	1.0205	3.2256
	2.5	2.2263	1.9063	4.5701	2.2194	1.0890	3.6374	2.2159	0.8401	3.2256
	3.0	2.2263	1.6888	4.5701	2.2194	0.9705	3.6374	2.2160	0.7244	3.2256
200	1.0	3.1369	2.8189	4.5513	3.1293	1.8477	3.6587	3.1242	1.3958	3.2645
	1.5	3.1371	2.1103	4.5513	3.1295	1.3867	3.6587	3.1244	1.0358	3.2645
	2.0	3.1369	2.0166	4.5513	3.1293	1.3049	3.6587	3.1242	0.9851	3.2645
	2.5	3.1370	1.7357	4.5513	3.1294	1.1503	3.6587	3.1242	0.8462	3.2645
	3.0	3.1370	1.5737	4.5513	3.1295	1.0307	3.6587	3.1243	0.7723	3.2645
500	1.0	4.9451	3.3709	4.5388	4.9360	2.0741	3.6862	4.9309	1.3620	3.3095
	1.5	4.9452	2.6758	4.5388	4.9361	1.6466	3.6862	4.9310	1.0857	3.3095
	2.0	4.9452	2.3250	4.5388	4.9361	1.4291	3.6862	4.9310	0.9298	3.3095
	2.5	4.9451	2.1308	4.5388	4.9360	1.3173	3.6862	4.9309	0.8444	3.3095
	3.0	4.9452	1.9264	4.5388	4.9361	1.1741	3.6862	4.9310	0.7667	3.3095

Table B.3: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(5)}(n)$ of Test statistic $T_n^{(5)}$ with $q(t) = (t(1-t))^{0.2}$ and $V = \frac{3}{\mu}$ and Optimal Bandwidth b.

		$(\alpha = 0.01)$		$(\alpha = 0.05)$		$(\alpha = 0.10)$	
n	μ	Sim.CV	App.CV	Sim.CV	App.CV	Sim.CV	App.CV
50	1.0	0.1486	5.0626	0.0347	3.5860	0.0000	2.8115
	1.5	0.1506	5.0626	0.0352	3.5860	0.0000	2.8115
	2.0	0.1489	5.0626	0.0358	3.5860	0.0000	2.8115
	2.5	0.1457	5.0626	0.0360	3.5860	0.0000	2.8115
	3.0	0.1455	5.0626	0.0408	3.5860	0.0000	2.8115
100	1.0	0.1120	5.8276	0.0221	4.1241	0.0000	3.2100
	1.5	0.1153	5.8276	0.0245	4.1241	0.0000	3.2100
	2.0	0.1154	5.8276	0.0243	4.1241	0.0000	3.2100
	2.5	0.1113	5.8276	0.0244	4.1241	0.0000	3.2100
	3.0	0.1167	5.8276	0.0244	4.1241	0.0000	3.2100
200	1.0	0.0000	6.7644	0.0000	4.7207	0.0000	3.6609
	1.5	0.0000	6.7644	0.0000	4.7207	0.0000	3.6609
	2.0	0.0000	6.7644	0.0000	4.7207	0.0000	3.6609
	2.5	0.0000	6.7644	0.0000	4.7207	0.0000	3.6609
	3.0	0.0000	6.7644	0.0000	4.7207	0.0000	3.6609
500	1.0	0.0000	8.0218	0.0000	5.7200	0.0000	4.5062
	1.5	0.0000	8.0218	0.0000	5.7200	0.0000	4.5062
	2.0	0.0000	8.0218	0.0000	5.7200	0.0000	4.5062
	2.5	0.0000	8.0218	0.0000	5.7200	0.0000	4.5062
	3.0	0.0000	8.0218	0.0000	5.7200	0.0000	4.5062

Table B.4: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(6)}(n)$ of Test statistic $T_n^{(6)}$ with $q(t) = (t(1-t))^{0.2}$ and $V = \frac{3}{\mu}$ and Optimal Bandwidth b.

		$(\alpha = 0.01)$		$(\alpha = 0.05)$		$(\alpha = 0.10)$	
n	μ	Sim.CV	App.CV	Sim.CV	App.CV	Sim.CV	App.CV
50	1.0	1.0432	5.6047	1.0390	4.2818	1.0369	3.5980
	1.5	1.0431	5.6047	1.0390	4.2818	1.0368	3.5980
	2.0	1.0431	5.6047	1.0389	4.2818	1.0368	3.5980
	2.5	1.0431	5.6047	1.0389	4.2818	1.0368	3.5980
	3.0	1.0429	5.6047	1.0386	4.2818	1.0365	3.5980
100	1.0	1.4688	6.4923	1.4642	4.9482	1.4619	4.1252
	1.5	1.4686	6.4923	1.4640	4.9482	1.4616	4.1252
	2.0	1.4686	6.4923	1.4640	4.9482	1.4617	4.1252
	2.5	1.4688	6.4923	1.4642	4.9482	1.4619	4.1252
	3.0	1.4688	6.4923	1.4642	4.9482	1.4619	4.1252
200	1.0	2.0696	7.5343	2.0645	5.6930	2.0611	4.7207
	1.5	2.0697	7.5343	2.0646	5.6930	2.0613	4.7207
	2.0	2.0696	7.5343	2.0645	5.6930	2.0611	4.7207
	2.5	2.0696	7.5343	2.0646	5.6930	2.0612	4.7207
	3.0	2.0697	7.5343	2.0646	5.6930	2.0612	4.7207
500	1.0	3.2565	8.9213	3.2531	6.7867	3.2625	5.6984
	1.5	3.2566	8.9213	3.2532	6.7867	3.2626	5.6984
	2.0	3.2566	8.9213	3.2532	6.7867	3.2626	5.6984
	2.5	3.2565	8.9213	3.2531	6.7867	3.2626	5.6984
	3.0	3.2566	8.9213	3.2532	6.7867	3.2626	5.6984

Table B.5: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(7)}(n)$ of Test statistic $T_n^{(7)}$ with $q(t) = (t(1-t))^{0.2}$, $V = \frac{3}{\mu}$ and Optimal Bandwidth b.

		$(\alpha = 0.01)$		$(\alpha = 0.05)$		$(\alpha = 0.10)$	
n	μ	Sim.CV	App.CV	Sim.CV	App.CV	Sim.CV	App.CV
50	1.0	1.0431	1.2620	0.6384	1.0479	0.4373	0.9376
	1.5	0.9053	1.2620	0.5525	1.0479	0.3692	0.9376
	2.0	0.7960	1.2620	0.4871	1.0479	0.3319	0.9376
	2.5	0.7081	1.2620	0.4326	1.0479	0.2907	0.9376
	3.0	0.6980	1.2620	0.4247	1.0479	0.2971	0.9376
100	1.0	0.7267	1.2620	0.4846	1.0479	0.3523	0.9376
	1.5	0.6962	1.2620	0.4533	1.0479	0.3356	0.9376
	2.0	0.5919	1.2620	0.3820	1.0479	0.2795	0.9376
	2.5	0.4774	1.2620	0.3178	1.0479	0.2319	0.9376
	3.0	0.4229	1.2620	0.2731	1.0479	0.1998	0.9376
200	1.0	0.5720	1.2620	0.3087	1.0479	0.2001	0.9376
	1.5	0.4329	1.2620	0.2293	1.0479	0.1439	0.9376
	2.0	0.4011	1.2620	0.2176	1.0479	0.1409	0.9376
	2.5	0.3554	1.2620	0.1895	1.0479	0.1204	0.9376
	3.0	0.3112	1.2620	0.1714	1.0479	0.1073	0.9376
500	1.0	0.2559	1.2620	0.0973	1.0479	0.0724	0.9376
	1.5	0.2060	1.2620	0.0781	1.0479	0.0578	0.9376
	2.0	0.1743	1.2620	0.0671	1.0479	0.0498	0.9376
	2.5	0.1631	1.2620	0.0611	1.0479	0.0455	0.9376
	3.0	0.1484	1.2620	0.0553	1.0479	0.0410	0.9376

Table B.6: Comparison of Simulated and Approximated Critical Values $c_\alpha^{(8)}(n)$ of Test statistic $T_n^{(8)}$ with $q(t) = (t(1-t))^{0.2}$, $V = \frac{3}{\mu}$ and Optimal Bandwidth b.

		$(\alpha = 0.01)$		$(\alpha = 0.05)$		$(\alpha = 0.10)$	
n	μ	Sim.CV	App.CV	Sim.CV	App.CV	Sim.CV	App.CV
50	1.0	1.0431	1.3361	0.6384	1.1436	0.4373	1.0442
	1.5	0.9053	1.3361	0.5525	1.1436	0.3692	1.0442
	2.0	0.7960	1.3361	0.4871	1.1436	0.3319	1.0442
	2.5	0.7081	1.3361	0.4326	1.1436	0.2907	1.0442
	3.0	0.6980	1.3361	0.4247	1.1436	0.2971	1.0442
100	1.0	0.7267	1.3361	0.4846	1.1436	0.3523	1.0442
	1.5	0.6962	1.3361	0.4533	1.1436	0.3356	1.0442
	2.0	0.5919	1.3361	0.3820	1.1436	0.2795	1.0442
	2.5	0.4774	1.3361	0.3178	1.1436	0.2319	1.0442
	3.0	0.4229	1.3361	0.2731	1.1436	0.1998	1.0442
200	1.0	0.5720	1.3361	0.3087	1.1436	0.2001	1.0442
	1.5	0.4329	1.3361	0.2293	1.1436	0.1439	1.0442
	2.0	0.4011	1.3361	0.2176	1.1436	0.1409	1.0442
	2.5	0.3554	1.3361	0.1895	1.1436	0.1204	1.0442
	3.0	0.3112	1.3361	0.1714	1.1436	0.1073	1.0442
500	1.0	0.2559	1.3361	0.0973	1.1436	0.0724	1.0442
	1.5	0.2060	1.3361	0.0781	1.1436	0.0578	1.0442
	2.0	0.1743	1.3361	0.0671	1.1436	0.0498	1.0442
	2.5	0.1631	1.3361	0.0611	1.1436	0.0455	1.0442
	3.0	0.1484	1.3361	0.0553	1.1436	0.0410	1.0442

Bibliography

[1] Albers, W., Akritas, M.G.: Combined rank tests for the two-sample problem with randomly censored data. Journal of the American Statistical Association 82(398), 648–655 (1987)

[2] Aly, E.E.: Change point tests for randomly censored data. In: Szyszkowicz, B. (ed.) Asymptotic Methods in Probability and Statistics, pp. 503–513. North-Holland Amsterdam (1998)

[3] Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N.: Statistical models based on counting processes. Springer Science & Business Media, New York (1993)

[4] Andreou, E., Ghysels, E.: Structural breaks in financial time series. In: Handbook of financial time series, pp. 839–870. Springer (2009)

[5] Andrews, D.W., Ploberger, W.: Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62(6), 1383–1414 (1994)

[6] Anis, M.Z.: Inference on a sharp jump in hazard rate: a review. Economic Quality Control 24(2), 213–229 (2009)

[7] Antoch, J., Hušková, M., Jarušková, D.: Change point detection. In: Lecture Notes of the 5th IASC Summer School. pp. 1–76. ISI, Voorburg (2000)

[8] Antoch, J., Hušková, M., Prášková, Z.: Effect of dependence on statistics for determination of change. Journal of Statistical Planning and Inference 60(2), 291–310 (1997)

[9] Antoch, J., Hušková, M.: Estimators of changes. In: Nonparametrics, Asymptoties and Time Series, pp. 533–578. Marcel Dekker, New York (1999)

[10] Antoch, J., Hušková, M.: Permutation tests in change point analysis. Statistics & probability letters 53(1), 37–46 (2001)

[11] Antoch, J., Hušková, M., Jarušková, D.: Off-line statistical process control. In: Multivariate Total Quality Control, pp. 1–86. Springer (2002)

[12] Antoniadis, A., Gijbels, I., MacGibbon, B.: Non-parametric estimation for the location of a change-point in an otherwise smooth hazard function under random censoring. Scandinavian journal of statistics 27(3), 501–519 (2000)

[13] Bacon, D.W., Watts, D.G.: Estimating the transition between two intersecting straight lines. Biometrika 58(3), 525–534 (1971)

[14] Bai, J., Perron, P.: Estimating and testing linear models with multiple structural changes. Econometrica 66(1), 47–78 (1998)

[15] Basseville, M., Nikiforov, I.V.: Detection of abrupt changes: theory and application. Prentice Hall, New York (1993)

[16] Bauer, P., Hack, P.: An extension of the mosum technique for quality control. Technometrics 22(1), 1–7 (1980)

[17] Bauer, P., Hackl, P.: The use of mosums for quality control. Technometrics 20(4), 431–436 (1978)

[18] Benham, T., Duan, Q., Kroese, D.P., Liquet, B.: Ceoptim: Cross-entropy r package for optimization. arXiv preprint arXiv:1503.01842 (2015)

[19] Bhore, R., Huque, M.: Estimation and testing for change in hazard for time-to-event endpoints. In: Design and Analysis of Clinical Trials with Time-to-Event Endpoints. pp. 195–226. CRC Press (2008)

[20] Billingsley, P.: Probability and measure. John Wiley & Sons, New York, 3rd edn. (1995)

[21] Breiman, L.: Probability. Addison-Wesley, Massachusetts (1968)

[22] Brockwell, P.J., Davis, R.A.: Introduction to time series and forecasting. Springer, New York, 2nd edn. (2002)

[23] Brodsky, E., Darkhovsky, B.: Nonparametric Methods in Change Point Problems. Kluwer Academic, Dordrecht (1993)

[24] Brown, R.L., Durbin, J., Evans, J.M.: Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society 37(2), 149–192 (1975)

[25] Buhamra, S., Aly, E.E., Al-Awadhi, S.: Nonparametric change point tests for randomly censored data. Advances & Application in Statistics 2(2), 199–216 (2002)

[26] Burke, M., Horváth, L.: Density and failure rate estimation in a competing risks model. Sankhyā: The Indian Journal of Statistics, Series A 46(1), 135–154 (1984)

[27] Cai, T., Hyndman, R.J., Wand, M.: Mixed model-based hazard estimation. Journal of Computational and Graphical Statistics 11(4), 784–798 (2002)

[28] Carlstein, E., Müller, H.G., Siegmund, D.: Change-point problems. In IMS Lecture Notes–Monograph Series 23 (1994)

[29] Chang, I.S., Chen, C.H., Hsiung, C.A.: Estimation in change-point hazard rate models with random censorship. Lecture Notes-Monograph Series 23, 78–92 (1994)

[30] Chen, J., Gupta, A.K.: Parametric Statistical Change Point Analysis. Birkhäuse, Boston (2000)

[31] Chen, J., Gupta, A.K.: Parametric statistical change point analysis: With applications to genetics, medicine, and finance. Birkhäuse, Boston (2012)

[32] Chernoff, H., Zacks, S.: Estimating the current mean of a normal distribution which is subjected to changes in time. The Annals of Mathematical Statistics 35(3), 999–1018 (1964)

[33] Chu, C.S.J., Hornik, K., Kaun, C.M.: Mosum tests for parameter constancy. Biometrika 82(3), 603–617 (1995)

[34] Chu, C.S.J., Hornik, K., Kuan, C.M.: The moving-estimates test for parameter stability. Econometric Theory 11(04), 699–720 (1995)

[35] Csörgo, M., Horváth, L.: Nonparametric tests for the changepoint problem. Journal of Statistical Planning and Inference 17, 1–9 (1987)

[36] Csörgő, M., Horváth, L.: Invariance principles for changepoint problems. Journal of Multivariate Analysis 27(1), 151–168 (1988)

[37] Csörgö, M., Horváth, L.: Limit theorems in change-point analysis. John Wiley & Sons Inc, New York (1997)

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

[39] Davison, A.C., Hinkley, D.V.: Bootstrap methods and their application. Cambridge University Press, Cambridge MA (1997)

[40] De Boer, P.T., Kroese, D.P., Mannor, S., Rubinstein, R.Y.: A tutorial on the cross-entropy method. Annals of operations research 134, 19–67 (2005)

[41] Diehl, S., Stute, W.: Kernel density and hazard function estimation in the presence of censoring. Journal of Multivariate Analysis 25, 299–310 (1988)

[42] Dupuy, J.F.: Detecting change in a hazard regression model with right-censoring. Journal of Statistical Planning and Inference 139(5), 1578–1586 (2009)

[43] Efron, B., Tibshirani, R.J.: An introduction to the bootstrap. Chapman & Hall, New York (1993)

[44] Evans, G.E., Sofronov, G.Y., Keith, J.M., Kroese, D.P.: Estimating change-points in biological sequences via the cross-entropy method. Annals of Operations Research 189(1), 155–165 (2011)

[45] Ferger, D.: Change-point analysis of censored data. Economic Quality Control 13(3), 167–177 (1998)

[46] Fleming, T.R., Harrington, D.P.: Counting processes and survival analysis. John Wiley & Sons, New York (1991)

[47] Franke, J., Kirch, C., Kamgaing, J.T.: Changepoints in times series of counts. Journal of Time Series Analysis 33(5), 757–770 (2012)

[48] Fryzlewicz, P., et al.: Wild binary segmentation for multiple change-point detection. The Annals of Statistics 42(6), 2243–2281 (2014)

[49] Gehan, E.A.: A generalized wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52(1-2), 203–223 (1965)

[50] Gijbels, I., Gürler, Ü.: Estimation of a change point in a hazard function based on censored data. Lifetime Data Analysis 9(4), 395–411 (2003)

[51] Gombay, E.: Change detection in linear regression with time series errors. Canadian Journal of Statistics 38(1), 65–79 (2010)

[52] Gombay, E., Liu, S.: A nonparametric test for change in randomly censored data. The Canadian Journal of Statistics/La Revue Canadienne de Statistique 28(1), 113–121 (2000)

[53] Good, P.: Permutation, Parametric, and Bootstrap Tests of Hypotheses. Springer, New York, 3rd edn. (2005)

[54] Goodman, M.S., Li, Y., Tiwari, R.C.: Survival analysis with change point hazard functions. Tech. rep., Harvard University Biostatistics Working Paper Series, Working Paper 40 (2006), http://biostats.bepress.com/cgi/viewcontent.cgi?article=1043&context=harvardbiostat

[55] Goodman, M.S., Li, Y., Tiwari, R.C.: Detecting multiple change points in piecewise constant hazard functions. Journal of applied statistics 38(11), 2523–2532 (2011)

[56] Green, P.J.: Reversible jump markov chain monte carlo computation and bayesian model determination. Biometrika 82(4), 711–732 (1995)

[57] Györfi, L., Härdle, W., Sarda, P., Vieu, P.: Nonparametric curve estimation from time series, vol. Lecture notes in Statistics 60. Springer-Verlag, Heidelberg (1989)

[58] Hall, P.: On the invariance principle for u-statistics. Stochastic Processes and their Applications 9, 163–174 (1979)

[59] Henderson, R.: A problem with the likelihood ratio test for a change-point hazard rate model. Biometrika 77(4), 835–843 (1990)

[60] Hobbs, B.: On nonparametric hazard estimation. J Biom Biostat 6(232), 2 (2015)

[61] Horváth, L.: Detection of changes in linear sequences. Annals of the Institute of Statistical Mathematics 49(2), 271–283 (1997)

[62] Horváth, L.: Tests for changes under random censorship. Journal of statistical planning and inference 69, 229–243 (1998)

[63] Hušková, M.: Asymptotics for robust mosum. Commentationes Mathematicae Universitatis Carolinae 31(2), 345–356 (1990)

[64] Hušková, M.: Permutation principle and bootstrap in change point analysis. Fields Inst. Commun 44, 273–291 (2004)

[65] Hušková, M., Neuhaus, G.: Change point analysis for censored data. Journal of statistical planning and inference 126, 207–223 (2004)

[66] Hušková, M., Picek, J.: Bootstrap in detection of changes in linear regression. Sankhyā: The Indian Journal of Statistics 67(2), 200–226 (2005)

[67] Hušková, M., Slabỳ, A.: Permutation tests for multiple changes. Kybernetika 37(5), 605–622 (2001)

[68] James, B., James, K.L., Siegmund, D.: Tests for a change-point. Biometrika 74(1), 71–83 (1987)

[69] Janson, S., Wichura, M.J.: Invariance principles for stochastic area and related stochastic integrals. Stochastic processes and their applications 16(1), 71–84 (1984)

[70] Jones, M.C., Marron, J.S., Sheather, S.J.: A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association 91(433), 401–407 (1996)

[71] Kalbfleisch, J.D., Prentice, R.L.: The statistical analysis of failure time data. John Wiley & Sons, New Jersey, 2nd edn. (2002)

[72] Kander, Z., Zacks, S.: Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. The Annals of Mathematical Statistics 37(5), 1196–1210 (1966)

[73] Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. Journal of the American statistical association 53(282), 457–481 (1958)

[74] Kirch, C.: Resampling methods for the change analysis of dependent data. Ph.D. thesis, Universität zu Köln (2006)

[75] Kirch, C., Muhsal, B.: A mosum procedure for the estimation of multiple random change points. Preprint (2014)

[76] Kirch, C., Steinebach, J.: Permutation principles for the change analysis of stochastic processes under strong invariance. Journal of computational and applied mathematics 186(1), 64–88 (2006)

[77] Komárková, L.: Change point problem for censored data. Ph.D. thesis, Department of Statistics, Charles University, Prague (2004)

[78] Kosorok, M.R., Song, R.: Inference under right censoring for transformation models with a change-point based on a covariate threshold. The Annals of Statistics 35(3), 957–989 (2007)

[79] Krämer, W., Ploberger, W., Alt, R.: Testing for structural change in dynamic models. Econometrica 56(6), 1355–1369 (1988)

[80] Kuan, C.M., Hornik, K.: The generalized fluctuation test: A unifying view. Econometric Reviews 14(2), 135–161 (1995)

[81] Langlands, A.O., Pocock, S.J., Kerr, G.R., Gore, S.M.: Long-term survival of patients with breast cancer: a study of the curability of the disease. British Medical Journal 2, 1247–1251 (1979)

[82] Lerman, P.M.: Fitting segmented regression models by grid search. Applied Statistics 29(1), 77–84 (1980)

[83] Leung, K.M., Elashoff, R.M., Afifi, A.A.: Censoring issues in survival analysis. Annual review of public health 18, 83–104 (1997)

[84] Liang, K.Y., Self, S.G., Liu, X.: The cox proportional hazards model with change point: An epidemiologic application. Biometrics 46(3), 783–793 (1990)

[85] Little, R.J.A., Rubin, D.B.: Statistical analysis with missing data. John Wiley & Sons, New York (1987)

[86] Little, R.J.A., Rubin, D.B.: Statistical analysis with missing data. John Wiley & Sons, New Jersey, 2nd edn. (2002)

[87] Liu, S.: Nonparametric Test for Change Point Problems with Random Censorship. Ph.D. thesis, Department of Mathematical Sciences, University of Alberta, Edmonton (1998)

[88] Lo, S., Mack, Y., Wang, J.: Density and hazard rate estimation for censored data via strong representation of the kaplan-meier estimator. Probability theory and related fields 80(3), 461–473 (1989)

[89] Loader, C.R.: Inference for a hazard rate change point. Biometrika 78(4), 749–757 (1991)

[90] Mantel, N.: Ranking procedures for arbitrarily restricted observation. Biometrics 23(1), 65–78 (1967)

[91] Marmer, V.: Linear Processes II: Spectral Density, Lag Operator, ARMA. Tech. rep., University of British Columbia (11 2016), http://faculty.arts.ubc.ca/vmarmer/econ627/627_10_02.pdf

[92] Marušiaková, M., Hušková, M.: Tests for multiple changes in linear regression models. Ph.D. thesis, Charles University in Prague (2009)

[93] Matthews, D.E., Farewell, V.T.: On testing for a constant hazard against a change-point alternative. Biometrics 38(2), 463–468 (1982)

[94] Matthews, D., Farewell, V., Pyke, R., et al.: Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. The Annals of Statistics 13(2), 583–591 (1985)

[95] McKnight, P.E., McKnight, K.M., Sidani, S., Figueredo, A.J.: Missing data: A gentle introduction. Guilford Press (2007)

[96] Müller, H.G.: Nonparametric regression analysis of longitudinal data. Springer, New York (1988)

[97] Müller, H.G., Wang, J.L.: Change-point models for hazard functions. Lecture Notes-Monograph Series 23, 224–241 (1994)

[98] Müller, H.G., Wang, J.L.: Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models. Biometrika 77(2), 305–314 (1990)

[99] Müller, H.G., Wang, J.L.: Locally adaptive hazard smoothing. Probability Theory and Related Fields 85(4), 523–538 (1990)

[100] Müller, H.G., Wang, J.L.: An invariance principle for discontinuity estimation in smooth hazard functions under random censoring. Sankhyā: The Indian Journal of Statistics, Series A 58(3), 392–402 (1996)

[101] Murthy, V.: Estimation of jumps, reliability and hazard rate. The Annals of Mathematical Statistics 36(3), 1032–1040 (1965)

[102] Nelder, J.A., Mead, R.: A simplex method for function minimization. The computer journal 7(4), 308–313 (1965)

[103] Nelson, W.: Theory and applications of hazard plotting for censored failure data. Technometrics 14, 945–966 (1972)

[104] Nguyen, H., Rogers, G., Walker, E.: Estimation in change-point hazard rate models. Biometrika 71(2), 299–304 (1984)

[105] Nyarige, E.G.: The bootstrap for the functional autoregressive model FAR(1). Ph.D. thesis, Department of Mathematics, University of Kaiserslautern, Germany (2016)

[106] Page, E.: Continuous inspection schemes. Biometrika 41(1/2), 100–115 (1954)

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

[108] Page, E.: On problems in which a change in a parameter occurs at an unknown point. Biometrika 44(1/2), 248–252 (1957)

[109] Patil, P.: Nonparametric hazard rate estimation by orthogonal wavelet methods. Journal of Statistical Planning and Inference 60(1), 153–168 (1997)

[110] Perron, P.: Dealing with structural breaks. Palgrave handbook of econometrics 1, 278–352 (2006)

[111] Ploberger, W., Krämer, W.: The cusum test with ols residuals. Econometrica 60(2), 271–285 (1992)

[112] Ploberger, W., Krämer, W., Kontrus, K.: A new test for structural stability in the linear regression model. Journal of Econometrics 40(2), 307–318 (1989)

[113] Polushina, T., Sofronov, G.: Change-point detection in binary markov dna sequences by the cross-entropy method. In: Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on. pp. 471–478. IEEE (2014)

[114] Pons, O.: Estimation in a cox regression model with a change-point at an unknown time. Statistics: A Journal of Theoretical and Applied Statistics 36(2), 101–124 (2002)

[115] Preuss, P., Puchstein, R., Dette, H.: Detection of multiple structural breaks in multivariate time series. Journal of the American Statistical Association 110(510), 654–668 (2015)

[116] Priyadarshana, M., Polushina, T., Sofronov, G.: Hybrid algorithms for multiple change-point detection in biological sequences. In: Signal and Image Analysis for Biomedical and Life Sciences, pp. 41–61. Springer (2015)

[117] Qian, L., Zhang, W.: Multiple change-point detection in piecewise exponential hazard regression models with long-term survivors and right censoring. In: Contemporary Developments in Statistical Theory, pp. 289–304. Springer (2014)

[118] Ramlau-Hansen, H.: Smoothing counting process intensities by means of kernel functions. The Annals of Statistics 11(2), 453–466 (1983)

[119] Reeves, J., Chen, J., Wang, X.L., Lund, R., Lu, Q.Q.: A review and comparison of changepoint detection techniques for climate data. Journal of Applied Meteorology and Climatology 46(6), 900–915 (2007)

[120] Rice, J., Rosenblatt, M.: Estimation of the log survivor function and hazard function. Sankhyā: The Indian Journal of Statistics, Series A 38(1), 60–78 (1976)

[121] Rubin, D.B.: Inference and missing data. Biometrika 63(3), 581–592 (1976)

[122] Rubinstein, R.Y.: Optimization of computer simulation models with rare events. European Journal of Operational Research 99(1), 89–112 (1997)

[123] Rubinstein, R.Y., Kroese, D.P.: The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning. Springer-Verlag (2004)

[124] Rubinstein, R.Y., Kroese, D.P.: Simulation and the Monte Carlo method. John Wiley & Sons, New York, 2nd edn. (2008)

[125] Schafer, H.: A note on data-adaptive kernel estimation of the hazard and density function in the random censorship situation. The Annals of Statistics 13(2), 818–820 (1985)

[126] Schafer, J.L.: Analysis of incomplete multivariate data. Chapman & Hall, London (1997)

[127] Schafer, J.L., Graham, J.W.: Missing data: our view of the state of the art. Psychological methods 7(2), 147–177 (2002)

[128] Serfling, R.J.: Approximation theorems of mathematical statistics. John Wiley & Sons, New York (1980)

[129] Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). Biometrika 52(3/4), 591–611 (1965)

[130] Sheather, S.J., Jones, M.C.: A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society. Series B (Methodological) 53(3), 683–690 (1991)

[131] Silverman, B.W.: Density estimation for statistics and data analysis. CRC press, Florida (1986)

[132] Singpurwalla, N.D., Wong, M.Y.: Kernel estimators of the failure-rate function and density estimation: an analogy. Journal of the american Statistical association 78(382), 478–481 (1983)

[133] Sonesson, C., Bock, D.: A review and discussion of prospective statistical surveillance in public health. Journal of the Royal Statistical Society: Series A (Statistics in Society) 166(1), 5–21 (2003)

[134] Stephens, D.: Bayesian retrospective multiple-changepoint identification. Applied Statistics 43(1), 159–178 (1994)

[135] Stoumbos, Z.G., Reynolds Jr, M.R., Ryan, T.P., Woodall, W.H.: The state of statistical process control as we proceed into the 21st century. Journal of the American Statistical Association 95(451), 992–998 (2000)

[136] Stute, W.: Changepoint problems under random censorship. Statistics: A jounal of theoretical and applied statistics 27(3-4), 255–266 (1996)

[137] Tanner, M.A., Wong, W.H.: The estimation of the hazard function from randomly censored data by the kernel method. The Annals of Statistics 11(3), 989–993 (1983)

[138] Ulm, K.: A statistical method for assessing a threshold in epidemiological studies. Statistics in medicine 10(3), 341–349 (1991)

[139] Van der Vaart, A.W.: Asymptotic statistics. Cambridge UniversityPpress, Cambridge (1998)

[140] Venkatraman, E.S.: Consistency results in multiple change-point problems. Ph.D. thesis, to the Department of Statistics.Stanford University (1992)

[141] Vostrikova, L.J.: Detecting disorder in multidimensional random-processes. Soviet Mathematics Doklady 24, 55–59 (1981)

[142] Wang, J.L.: Smoothing hazard rates. Encyclopedia of biostatistics 2, 4986–4997 (2005)

[143] Wang, J.L.: Smoothing hazard rates. In: Wiley StatsRef: statistics reference online. Wiley (2014)

[144] Watson, G., Leadbetter, M.: Hazard analysis i. Biometrika 51(1/2), 175–184 (1964)

[145] Watson, G., Leadbetter, M.: Hazard analysis ii. Sankhyā: The Indian Journal of Statistics 26(1), 101–116 (1964)

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

[147] Worsley, K.J.: An improved bonferroni inequality and applications. Biometrika 69(2), 297–302 (1982)

[148] Worsley, K.J.: Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. Biometrics 44(1), 259–263 (1988)

[149] Yandell, B.S.: Nonparametric inference for rates with censored survival data. The Annals of Statistics 11(4), 1119–1135 (1983)

[150] Yao, Y.C.: Maximum likelihood estimation in hazard rate models with a change-point. Communications in Statistics-Theory and Methods 15(8), 2455–2466 (1986)

[151] Zeileis, A., Leisch, F., Hornik, K., Kleiber, C.: strucchange. an r package for testing for structural change in linear regression models. Journal of Statistical Software 7(2), 1–38 (2002)

[152] Zeileis, A., Shah, A., Patnaik, I.: Testing, monitoring, and dating structural changes in exchange rate regimes. Computational Statistics & Data Analysis 54(6), 1696–1706 (2010)

[153] Zhao, X., Wu, X., Zhou, X.: A change-point model for survival data with long-term survivors. Statistica Sinica 19(1), 377–390 (2009)

Scientific Career

2000	Bachelor of Science in Statistics (held in 2002), Jahangirnagar University, Bangladesh
2001	Master of Science in Statistics (held in 2004), Jahangirnagar University, Bangladesh
2006	Master of Philosophy (Coursework completed) in Statistics (held in 2008), Jahangirnagar University, Bangladesh
01/2005 - 12/2005	Teaching Assistant, Department of Applied Physics and Communication Engineering, East West University, Bangladesh
01/2006 - 05/2006	Adjunct Faculty, Department of Applied Physics and Communication Engineering, East West University, Bangladesh
06/2006 - 12/2009	Lecturer, Department of Statistics, Jahangirnagar University, Bangladesh
12/2009 - till date	Assistant Professor, Department of Statistics, Jahangirnagar University, Bangladesh
08/2013 - 2017	PhD in Mathematics, University of Kaiserslautern, Germany

Wissenschaftlicher Werdegang

2000	Bachelor of Science in Statistik (erhalten 2002), Jahangirnagar University, Bangladesch
2001	Master of Science in Statistik (erhalten 2004), Jahangirnagar University, Bangladesch
2006	Master of Philosophy (Kurse abgeschlossan) in Statistik (erhalten 2008), Jahangirnagar University, Bangladesch
01/2005 - 12/2005	Teaching Assistant, Department of Applied Physics and Communication Engineering, East West University, Bangladesch
01/2006 - 05/2006	Adjunct Faculty, Department of Applied Physics and Communication Engineering, East West University, Bangladesch
06/2006 - 12/2009	Lecturer, Department of Statistics, Jahangirnagar University, Bangladesch
12/2009 - jetzt	Assistant Professor, Department of Statistics, Jahangirnagar University, Bangladesch
08/2013 - 2017	Doktorentin im Fach Mathematik, Technische Universität Kaiserslautern, Deutschland