Fall 12-2020

Inference and Estimation in Change Point Models for Censored Data

Kristine Gierz
Old Dominion University, kgierz@icloud.com

INFERENCE AND ESTIMATION IN CHANGE POINT MODELS FOR CENSORED DATA

by

Kristine Gierz
B.S. June 2013, University of California, Los Angeles
M.S. December 2016, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

MATHEMATICS AND STATISTICS

OLD DOMINION UNIVERSITY
December 2020

Approved by:

Kayoung Park (Director)

Lucia Tabacu (Member)

N. Rao Chaganty (Member)

Abdullah M. Al-Taiar (Member)

ABSTRACT

INFERENCE AND ESTIMATION IN CHANGE POINT MODELS FOR CENSORED DATA

Kristine Gierz
Old Dominion University, 2020
Director: Dr. Kayoung Park

In general, the change point problem considers inference of a change in distribution for a set of time-ordered observations. This has applications in a large variety of fields, and can also apply to survival data. With improvements to medical diagnoses and treatments, incidences and mortality rates have changed. However, the most commonly used analysis methods do not account for such distributional changes. In survival analysis, change point problems can concern a shift in a distribution for a set of time-ordered observations, potentially under censoring or truncation.

In this dissertation, we first propose a sequential testing approach for detecting multiple change points in the Weibull accelerated failure time model, since this is sufficiently flexible to accommodate increasing, decreasing, or constant hazard rates and is also the only continuous distribution for which the accelerated failure time model can be reparametrized as a proportional hazards model. Our sequential testing procedure does not require the number of change points to be known; this information is instead inferred from the data. We conduct a simulation study to show that the method accurately detects change points and estimates the model. The numerical results along with a real data application demonstrate that our proposed method can detect change points in the hazard rate.

In survival analysis, most existing methods compare two treatment groups for the entirety of the study period. Some treatments may take a length of time to show effects in subjects. This has been called the time-lag effect in the literature, and in cases where time-lag effect is considerable, such methods may not be appropriate to detect significant differences between two groups. In the second part of this dissertation, we propose a novel non-parametric approach for estimating the point of treatment time-lag effect by using an empirical divergence measure. Theoretical properties of the estimator are studied. The results from the simulated data and real data example support our proposed method.

Dedicated to my parents for all of their love and support, and to my husband Frank for his patience and encouragement.

ACKNOWLEDGEMENTS

Firstly, I thank my advisor, Dr. Kayoung Park, for recommending this topic of research to me. Her advice and guidance have been invaluable, and it would not be possible to complete this research without her mentorship. I am incredibly grateful for all of her encouragement, which has helped further this work beyond measure.

I would additionally like to thank my advisory committee, Dr. N Rao Chaganty, Dr. Lucia Tabacu, and Dr. Abdullah Al-Taiar. Their insightful suggestions and expertise truly helped progress this research and greatly improved the quality of this dissertation. I am extremely appreciative of their support for the duration of this work.

Additionally, I would like to thank the Department of Mathematics and Statistics for providing the opportunity to work as a Graduate Teaching Assistant and providing the majority of the funding to complete my graduate studies. Finally, I would like to acknowledge the award from the Department of Defense's SMART program that funded the final portion of my dissertation research.

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES xi

Chapter

1. INTRODUCTION AND MOTIVATION 1

2. BACKGROUND 5
   - 2.1 REGRESSION MODELS FOR SURVIVAL ANALYSIS 5
   - 2.2 EMPIRICAL DIVERGENCE BASED ON U-STATISTICS 9

3. DETECTION OF MULTIPLE CHANGE POINTS IN AN ACCELERATED FAILURE TIME MODEL USING SEQUENTIAL TESTING 11
   - 3.1 INTRODUCTION 11
   - 3.2 WEIBULL ACCELERATED FAILURE TIME MODEL 13
   - 3.3 MAXIMUM LIKELIHOOD ESTIMATION FOR CHANGE POINT MODEL 15
   - 3.4 HYPOTHESIS TESTING 17
   - 3.5 SIMULATION STUDY 22
   - 3.6 REAL DATA APPLICATIONS 31
   - 3.7 CONCLUSION AND FURTHER RESEARCH 40

4. A NON-PARAMETRIC APPROACH TO EVALUATION THE POINT OF TREATMENT TIME-LAG EFFECT 42
   - 4.1 INTRODUCTION 42
   - 4.2 PROPOSED EMPIRICAL DIVERGENCE METHOD 44
   - 4.3 SIMULATION STUDY 49
   - 4.4 REAL DATA APPLICATIONS 57
   - 4.5 CONCLUSIONS AND FURTHER RESEARCH 62

5. SUMMARY AND FUTURE RESEARCH 64
   - 5.1 SUMMARY 64
   - 5.2 FUTURE RESEARCH 65

REFERENCES 67

APPENDICES
   - A. SELECTED R CODES 77
   - B. ACCELERATED FAILURE TIME DERIVATIONS 82
   - B.1 AFT CUMULATIVE HAZARD 82

B.2 MAXIMUM LIKELIHOOD ESTIMATION DETAILS 84
B.3 WALD TEST VARIANCE 87
B.4 ADDITIONAL SIMULATION RESULTS 88
B.5 SEER LOCALIZED/REGIONAL PROSTATE CANCER CASE ANALYSIS 88
C. DETAILED PROOFS FOR CONVERGENCE AND CONSISTENCY OF EMPIRICAL DIVERGENCE MEASURE 92

VITA 96

LIST OF TABLES

Table Page

1 Averaged point estimates (MEAN) and MSE values of the estimated parameters of the proposed model and piecewise constant hazard model based on 1000 replicated simulations for (3.4). 23

2 Averaged point estimates (MEAN) and MSE values of the estimated parameters of the proposed model and piecewise constant hazard model based on 1000 replicated simulations for (3.5). 25

3 Averaged point estimates (MEAN) and MSE values of the estimated parameters of the proposed model and piecewise constant hazard model based on 1000 replicated simulations for (3.6). 26

4 Median of point estimates (MEDIAN) and mean absolute deviation (MAD) values of the estimated parameters of the proposed method and piecewise constant hazard method based on 1000 replicated simulations for (3.6). 28

5 Averaged point estimates (MEAN) and MSE values of the estimated parameters of the proposed model and piecewise constant hazard model based on 1000 replicated simulations for (3.7). 29

6 Power and Type I error results for the sequential hypothesis testing. 32

7 Point estimates and 95% confidence intervals in parentheses for SEER prostate cancer data considering only distant cases by the proposed model and the constant hazard model by Goodman et al. (2011). 37

8 Mean point estimate (PE), variance, and mean square error (MSE) of the estimate for τ in model (4.3). In the table, CR denotes the censoring rate. 50

9 Mean point estimate (PE), variance, and mean square error (MSE) of the estimate for τ in model (4.4). In the table, CR denotes the censoring rate. 53

10 Percent coverage and average length of 95% BCa confidence intervals for Weibull distribution simulations following (4.3). CR denotes the censoring rate. 56

11 Number of significant test statistics out of 1000 simulations with $B = 1000$ permutations for each at $\alpha = 0.05$ level of significance. CR denotes censoring rate. 57

12 Averaged point estimates (MEAN) and MSE values of the estimated parameters of the proposed model and piecewise constant hazard model based on 1000 replicated simulations for B.1. 89

13 Point estimates and 95% confidence intervals in parentheses for SEER prostate cancer data considering only localized/regional cases by the proposed model and the constant hazard model by Goodman et al. (2011). 90

LIST OF FIGURES

Figure Page

1 Estimation from one simulation with settings for 40% censoring and sample size 500 as in Table 2. The solid lines are the Nelson-Aalen estimate of cumulative hazard, and the vertical dash-dotted line shows the value of the change point. . . 20

2 Estimation from one simulation with settings for 0% censoring and sample size 500 as in Table 3. The solid lines are the Nelson-Aalen estimate of cumulative hazard, and the vertical dash-dotted lines show the values of the change points. . 21

3 Vertical lines represent change point estimates from the proposed method. Left: Represents estimation without the consideration of any covariates. Right: Represents the estimation when considering the treatment covariate. 33

4 Vertical lines represent change point estimates from the proposed method. Left: Represents estimation without the consideration of any covariates. Right: Represents the estimation when considering the treatment covariate. 34

5 Nelson-Aalen cumulative hazard estimates of distant and localized/regional stages for SEER prostate cancer incidence data. 36

6 The vertical dotted lines represent the location of the change point estimates for the proposed method. Left: The figure represents estimation with no covariates included. Right: The figure represents estimation with the race covariate included. 38

7 The vertical dotted lines represent the location of the change point estimates for the proposed method. Left: The figure represents estimation with no covariates included. Right: The figure represents estimation with the race covariate included. 39

8 Kaplan-Meier curves with the black dotted line showing the control group and the grey dotted line showing the treatment group. Data was simulated from a Weibull distribution with $N = 100, \beta = 1, \lambda = 0.5, \nu = 1.5,$ and $\tau = 1$. 45

9 Density curves of the estimated time-lag points for 1000 replications, with the sample size of 500 and the censoring rate of 20% in model (4.3) with $\beta = 1$. The black dotted line represents the true value for $\tau = 1$. 52

10 Density curves of the estimated time-lag points for 1000 replications, with the sample size of 500 and the censoring rate of 20% in model (4.4) with $\beta = 1.25$. The black dotted line represents the true value for $\tau = 3$. 54

11 Kaplan-Meier curves for the distribution of event time in lung cancer according to two groups; the black dotted line denotes the standard treatment group and the grey dotted line denotes the test treatment group. 58

12 Kaplan-Meier curves for the distribution of time to death for breast cancer; the black dotted line denotes the standard treatment group without hormone replacement therapy and the grey dotted line denotes the hormone therapy group. 60

13 Left: Estimation without covariates. Right: This figure includes estimation with race and averaged age covariates. 91

CHAPTER 1

INTRODUCTION AND MOTIVATION

Change point detection seeks to identify times when the probability distribution of a stochastic process or time series changes. Generally speaking, the change point problem can concern both whether or not a change (or multiple changes) has occurred, and identifying the time point of any changes. There is vast literature on differing change point analysis methods, with applications in a wide variety of fields, including financial modeling (Talih and Hengartner, 2005; Zeileis et al., 2010), bioinformatics (Erdman and Emerson, 2008; Muggeo and Adelfio, 2011), signal processing (Kim et al., 2009), climatology (Reeves et al., 2007), oceanography (Killick et al., 2010), and medical imaging (Nam et al., 2012). These specific applications may be concerned with changes in the mean, variance, correlation, or other measure.

One major area of research used in many of these application is statistical process control, or statistical quality control. Statistical process control aims to employ statistical methods to monitor and control a process. This ensures that the process runs efficiently and allows for adjustments in the process to be made, with potentially life-saving effects. Statistical process control was introduced by Walter A Shewhart in the early 1920s in the form of Shewhart charts. These charts have upper control limits and lower control limits that indicate a threshold at which process output is considered unusual. An alternative for Shewhart charts called cumulative sum (or CUSUM) charts was proposed by Page (1954), which is more efficient in detecting small shifts in the process. This was followed by the exponentially weighted moving average (EWMA) chart proposed by Roberts (1959). The EWMA chart is sensitive to small shifts, but does not match Shewhart charts in detecting larger changes. These charts have more recently been adapted to monitor censored data (see, for example, Sego et al. (2009)).

There has been an increasing need for methods involving change point methods when dealing with censored data, especially when considering changes in distribution of survival probabilities or hazard rates (Jemal et al., 2017). Common methods of analysis for censored data can be non-parametric (the Kaplan-Meier estimator and log-rank type tests), semi-parametric (proportional hazards models), or parametric (parametric proportional hazards models and accelerated failure time models). Many of these existing standard models and

methods of analysis have certain assumptions that would be violated if a change point exists. The proportional hazards model proposed by Cox (1972), for example, assumes that covariates are multiplicatively related to the hazard. As noted by He et al. (2013), the inferences and conclusions made from the proportional hazards model may be invalid and misleading if there is a change point. Due to the prominence of the proportional hazards model in survival analysis, a number of approaches have been suggested that involve change points, including those proposed by Chen and Baron (2014), He et al. (2013), Liu et al. (2008), Lee et al. (2020), and Park and Qiu (2018). Another common distributional assumption in survival analysis is that of a constant hazard rate, leading to work on constant piecewise hazard models proposed first by Matthews and Farewell (1982) and expanded much later by Goodman et al. (2011). Many other methods have put forward, including non-parametric methods such as the weighted log-rank statistic approach of Zucker and Lakatos (1990) and Dinse et al. (1993), and the stump regression of Brazzale et al. (2019). The simplest of these methods involve a single change point in survival data without covariates, while the most complex is adaptable to multiple change points in survival data that includes explanatory covariates.

Asymptotic theory of maximum likelihood estimates under a change point is far from trivial, and in some cases intractable. Matthews and Farewell (1982) first noted problems with boundedness, but mentioned that the distribution of a likelihood ratio test to determine the existence of a change point is very close to a Chi-squared distribution with two degrees of freedom. Yao (1986) provided some constraints and asymptotic theory for a maximum likelihood estimator of a single change point in a hazard rate, which Goodman et al. (2011) then used in order to construct a Wald-type test statistic, thereby avoiding distributional issues with a likelihood ratio test. Most other methods use Monte Carlo methods in order to find the empirical distribution of the test statistic and overcome difficulties in asymptotic theory. Some methods involve the necessity to assume the number of change points present in the model in order to carry out estimation and hypothesis testing. A newer trend in literature is sequential hypothesis testing using increasingly conservative significance levels in order to fit the most parsimonious model (He et al., 2013; Liu et al., 2007; Lee et al., 2020).

Since the assumptions of the proportional hazards model of Cox (1972) are often violated in real data applications, the accelerated failure time (AFT) model has been used as an alternative. Where the proportional hazards model assumes that the effect of a covariate is to multiply the hazard by some constant, the AFT model instead has the assumption that

the effect of a covariate is to accelerate (or decelerate) the life course of a disease by some constant. However, the AFT model is typically fully parametric and therefore requires the stricter distributional assumptions. A semi-parametric alternative to the AFT model was created by Buckley and James (1979), but was argued by Wei (1992) to lack theoretical justification. The log-logistic distribution provides the most commonly used AFT model, as it can have a non-monotonic hazard function which increases at early times and decreases at later times. It is somewhat similar in shape to the log-normal distribution with the exception that it has heavier tails. The Weibull distribution is another commonly used distribution, because it includes the exponential distribution as a special case, and can be re-parameterized as a proportional hazards model. It is the only family of distributions to have this property, which gives it more flexibility. However, the applicability of this model could be limited by the fact that the hazard function is monotonic.

A popular theory in estimation is U-statistics, where the "U" stands for "unbiased." These statistics arise naturally from minimum-variance unbiased estimators (MVUEs). In non-parametric statistics, the theory of U-statistics can be used for to find estimators and construct hypothesis tests. The term originally comes from Hoeffding (1948), and it is particularly helpful in the context of independent and identically-distributed random variables. There has also been work into U-statistics for random variables that are not independent and identically-distributed (for example, see Denker and Keller (1983)). U-statistics have also been used for change point analysis. Székely et al. (2005) proposed a multivariate divergence measure that is an extension of Ward's minimum variance method, which is a method from cluster analysis. Matteson and James (2014) adapted this to an empirical divergence measure based on U-statistics to find change points in multivariate time series. These methods can be computationally expensive, especially when considering confidence intervals or hypothesis testing (which can be carried out via Monte Carlo methods such as bootstrapping or permutation tests).

In this dissertation, we propose and develop two different methods of change point estimation and inference for censored data. In Chapter 2 we introduce the background information necessary to construct these estimates. Specifically, we first review the literature and theory required to use sequential testing and maximum likelihood estimation in order to create an AFT model with multiple change points based on the Weibull distribution. Next, we give an overview on the theory of U-statistics and non-parametric multivariate time series methodology that we will adapt to be used for survival data.

In Chapter 3, we discuss the detection of multiple change points in an accelerated failure

time model using sequential testing. We use maximum likelihood estimation, as He et al. (2013) used for similar methodology concerning the Cox proportional hazards model. Due to known issues with the distribution of a likelihood ratio test statistic in order to determine the existence of a change point or change points, we use a model-based bootstrap procedure to find the empirical distribution of the test statistic. We use an alpha-spending function proposed by Goodman et al. (2011) to ensure increasingly conservative significance levels to test for each consecutive change point. Simulations studies are performed to evaluate the accuracy of estimation as well as the power and Type I error of the sequential hypothesis testing procedure. Finally, the method is applied to a prostate cancer data set from the Surveillance, Epidemiology, and End Results Program as well as a bladder tumor recurrence data from the Veteran's Administration in order to show the flexibility of the model compared to other models.

In Chapter 4, we propose a non-parametric method of finding the point of treatment time-lag effect based on the previous method of Matteson and James (2014) for multivariate time series. In this method, we compare a "treatment group" to a "control group." Specifically, the two groups are quite similar up until a certain time point, and then differ after the treatment takes effect. We give some theoretical properties of the empirical divergence measure based on the Kaplan-Meier survival estimates, including strong consistency of the estimator. In order to show the advantage of the estimation technique in this chapter, we provide simulation results for accuracy, empirical coverage probability of bootstrap confidence intervals, and reliability of a permutation test to test the null hypothesis of no change point. The method is then applied to a Veteran's Administration Lung Cancer Trial data set and a German breast cancer data set.

In Chapter 5 we give a summary and conclusion of the methods and discuss avenues for future research. Some `R` code for reproducibility, more detailed proofs, and additional simulations can be found in the Appendix.

CHAPTER 2

BACKGROUND

2.1 REGRESSION MODELS FOR SURVIVAL ANALYSIS

The analysis of time-to-event data is most commonly called survival analysis. It may be referred to as reliability theory within engineering, duration analysis within economics, and event history analysis within sociology. If we consider a nonnegative random variable representing the failure time of an individual, the probability distribution can be described in many ways.

The standard estimator for the survival function of censored data was first proposed by Kaplan and Meier (1958), and is often called the product-limit estimator. As proposed by Kaplan and Meier (1958) and Greenwood (1926):

Definition 1. *Suppose that events occur at T distinct times $t_1 < t_2 < \cdots < t_T$, and that there are d_i events with y_i individuals at risk at time t_i. Then, the product-limit estimator is defined as:*

$$\widehat{S}(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_i \leq t} \left[1 - \frac{d_i}{y_i}\right] & \text{if } t_1 \leq t. \end{cases}$$

with variance given by Greenwood's formula:

$$\widehat{V}\left[\widehat{S}(t)\right] = \widehat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{y_i(y_i - d_i)}.$$

It is clear that $\widehat{S}(t)$ is a step function that jumps at the observed event times. The size of jumps depends on both the number of events observed at time t_i and on the pattern of censored observations before t_i.

The product-limit estimator suggested by Kaplan and Meier (1958) is an efficient way to empirically estimate the survival function for right-censored data. It can also be used to estimate the cumulative hazard function as $\widehat{\Lambda}(t) = -\log\left[\widehat{S}(t)\right]$. A popularly-used alternative to estimating the cumulative hazard function is known as the Nelson-Aalen estimate. It was first proposed by Nelson (1972) in the context of reliability. It was later rediscovered by Aalen (1978) using counting process techniques.

Definition 2. *The cumulative hazard estimator defined by Nelson (1972) and Aalen (1978) is*

$$\tilde{\Lambda}(t) = \begin{cases} 0 & \text{if } t \leq t_1 \\ \sum_{t_i \leq t} \frac{d_i}{y_i} & \text{if } t_1 \leq t. \end{cases}$$

Then an alternative estimator of the survival function is $\tilde{S}(t) = \exp\left[-\tilde{\Lambda}(t)\right]$. One of the primary uses of the Nelson-Aalen estimator is in selecting between parametric models.

2.1.1 ACCELERATED FAILURE TIME MODEL

One model that has been proposed to analyze censored data (potentially with explanatory covariates) is the accelerated failure time model.

Definition 3. *The accelerated failure time (AFT) model is defined by the relationship*

$$S(t; \boldsymbol{\zeta}) = S_0\left[\exp(t\boldsymbol{\theta}'\boldsymbol{\zeta})\right], \quad \forall t \tag{2.1}$$

where $\boldsymbol{\zeta}$ is a vector of explanatory covariates and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ is a corresponding p-dimensional vector of regression coefficients. The factor $\exp(\boldsymbol{\theta}'\boldsymbol{\zeta})$ is called the acceleration factor. An implication of the model is that the hazard rate is related to the baseline hazard rate $h_0(t)$ by

$$h(t; \boldsymbol{\zeta}) = \exp(\boldsymbol{\theta}'\boldsymbol{\zeta}) h_0(\exp(t\boldsymbol{\theta}'\boldsymbol{\zeta})), \quad \forall t. \tag{2.2}$$

The AFT model is parametric, with the log-logistic and Weibull distributions being two underlying distributions that are commonly used. The log-logistic distribution has a hazard rate that is hump-shaped, and is the only distribution for which the AFT model also has a representation as a proportional odds model. The Weibull distribution, on the other hand, is very flexible because it has a hazard rate that can be either increasing, decreasing, or constant. It is the only family of distributions for which the AFT model also gives rise to a proportional hazards model as described by Cox (1972). In this dissertation, we will focus on an AFT model based on the Weibull distribution due to these desirable properties.

There are several different parameterizations which can be used for the two-parameter Weibull distribution. For the rest of the dissertation, we will use the one as outlined here.

Definition 4. *Assume the shape parameter $\nu > 0$ and the scale parameter $\lambda > 0$. The survival function is given by*

$$S(t) = \exp(-\lambda t^{\nu}),$$

and the hazard rate is therefore

$$h(t) = \lambda\nu t^{\nu-1}.$$

If we take the transform $V = \log(T)$*, then the survival function is* $S_V(v) = \exp(-\lambda e^{\nu v})$*. This also leads to a proportional hazards model for* T *with a Weibull baseline hazard* $h(t;\boldsymbol{\zeta}) = (\nu\lambda t^{\nu-1})\exp(\boldsymbol{\beta}'\boldsymbol{\zeta})$*. Letting* $\lambda = \exp(-\mu/\sigma)$*,* $\sigma = 1/\nu$*, and* $\boldsymbol{\beta} = \boldsymbol{\gamma}/\sigma$*,* V *follows a log-linear model with*

$$V = \log(T) = \mu + \boldsymbol{\gamma}'\boldsymbol{\zeta} + \sigma W \tag{2.3}$$

where W *follows the extreme value distribution.*

Then we know that W has the probability function

$$f_W(w) = \exp(w - e^w)$$

and survival function

$$S_W(w) = \exp(-e^w)$$

for $-\infty < w < \infty$. For the AFT representation, it is true that

$$h(t;\boldsymbol{\zeta}) = \exp(\boldsymbol{\theta}'\boldsymbol{\zeta})h_0\left[t\exp(\boldsymbol{\theta}'\boldsymbol{\zeta})\right].$$

Comparing the hazard rates for the proportional hazards representation versus the accelerated failure times model, we can see that $\boldsymbol{\theta} = \boldsymbol{\beta}/\nu$, or $\boldsymbol{\theta} = -\boldsymbol{\gamma}$. It is important to note that when $\nu = 1$ (or, equivalently, $\sigma = 1$) the hazard rate is constant, and the Weibull distribution reduces to the exponential distribution. When $\nu > 1$, the hazard rate is increasing, and when $\nu < 1$, the hazard rate is decreasing.

The likelihood function for right-censored data is given by

$$\begin{aligned} L &= \prod_{i=1}^{n} \left[f_V(v_i)\right]^{\delta_i} \left[S_V(v_i)\right]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[\frac{1}{\sigma} f_W\left(\frac{y_i - \mu - \boldsymbol{\gamma}'\boldsymbol{\zeta}}{\sigma}\right)\right]^{\delta_i} \left[S_W\left(\frac{y_i - \mu - \boldsymbol{\gamma}'\boldsymbol{\zeta}}{\sigma}\right)\right]^{1-\delta_i}, \end{aligned}$$

where δ_i is a censoring indicator taking value 1 if the time point is not censored, and value 0 if it is censored.

From this, estimates of μ, σ, and $\boldsymbol{\gamma}$ can be found numerically by maximizing the likelihood function. As for the proportional hazards model suggested by Cox (1972), commonly used statistical software includes routines to carry out this maximization. The variance-covariance matrix of the log-linear parameters is also available in such routines. Let $\widehat{\sigma}$ and

$\widehat{\mu}$ be the maximum likelihood estimators (MLEs) of σ and μ, respectively. Then, by the invariance property of the maximum likelihood, the MLEs of λ and μ are given by

$$\widehat{\lambda} = \exp(-\widehat{\mu}/\widehat{\sigma}), \quad \widehat{\nu} = 1/\widehat{\sigma}, \text{ and } \quad \widehat{\boldsymbol{\theta}} = -\widehat{\boldsymbol{\gamma}}.$$

Definition 5. *Applying the delta method (or method of statistical differences) as seen in Elandt-Johnson and Johnson (1980, pp 69-72), for $j, k = 1, \ldots, p$*

$$\begin{aligned}
Var(\widehat{\lambda}) &= \exp\left(\frac{-2\widehat{\mu}}{\widehat{\sigma}}\right)\left[\frac{Var(\widehat{\mu})}{\widehat{\sigma}^2} - 2\frac{\widehat{\mu}\,Cov(\widehat{\mu},\widehat{\sigma})}{\widehat{\sigma}^3} + \frac{\widehat{\mu}^2\,Var(\widehat{\sigma})}{\widehat{\sigma}^4}\right] \\
Var(\widehat{\nu}) &= \frac{Var(\widehat{\sigma})}{\widehat{\sigma}^4} \\
Cov(\widehat{\lambda},\widehat{\nu}) &= \exp\left(\frac{-\widehat{\mu}}{\widehat{\sigma}}\right)\left[\frac{Cov(\widehat{\mu},\widehat{\sigma})}{\widehat{\sigma}^3} - \frac{\widehat{\mu}\,Var(\widehat{\sigma})}{\widehat{\sigma}^4}\right] \\
Cov(\widehat{\beta}_j,\widehat{\beta}_k) &= \frac{Cov(\widehat{\gamma}_j,\widehat{\gamma}_k)}{\widehat{\sigma}^2} - \frac{\widehat{\gamma}_j\,Cov(\widehat{\gamma}_j,\widehat{\sigma})}{\widehat{\sigma}^3} - \frac{\widehat{\gamma}_k\,Cov(\widehat{\gamma}_k,\widehat{\sigma})}{\widehat{\sigma}^3} \\
Cov(\widehat{\lambda},\widehat{\beta}_j) &= \exp\left(-\frac{\widehat{\mu}}{\widehat{\sigma}}\right)\left[\frac{Cov(\widehat{\gamma}_j),\widehat{\mu})}{\widehat{\sigma}^2} - \frac{\widehat{\gamma}_j\,Cov(\widehat{\gamma}_j,\widehat{\sigma})}{\widehat{\sigma^3}} - \frac{\widehat{\mu}\,Cov(\widehat{\mu},\widehat{\sigma})}{\widehat{\sigma}^3} + \frac{\widehat{\gamma}_j\widehat{\mu}\,Var(\widehat{\sigma})}{\widehat{\sigma}^4}\right] \\
Cov(\widehat{\nu},\widehat{\beta}_j) &= \frac{Cov(\widehat{\gamma}_j,\widehat{\sigma})}{\widehat{\sigma}^3} - \frac{\widehat{\gamma}_j\,Var(\widehat{\sigma})}{\widehat{\sigma}^4}.
\end{aligned}$$

Proof. The delta method is based upon a Taylor series expansion of a continuous function $g(\cdot)$ of the maximum likelihood estimators of a vector of parameters. Let ψ_1 and ψ_2 be the parameters of interest with $\widehat{\psi}_1$ and $\widehat{\psi}_2$ being the maximum likelihood estimates. Let $\theta_1 = g_1(\psi_1, \psi_2)$ and $\theta_2 = g_2(\psi_1, \psi_2)$. By the invariance principle of the maximum likelihood, the MLEs of θ_1 and θ_2 are $g_k(\widehat{\psi}_1, \widehat{\psi}_2)$ for $k = 1, 2$.

In order to apply the delta method, we expand $g_k(\widehat{\psi}_1, \widehat{\psi}_2)$ in a first-order Taylor series about the true values of ψ_1 and ψ_2. So,

$$g_k(\widehat{\psi}_1, \widehat{\psi}_2) - g_k(\psi_1, \psi_2) = (\widehat{\psi}_1 - \psi_1)\frac{\partial g_k(\widehat{\psi}_1, \widehat{\psi}_2)}{\partial \widehat{\psi}_1} + (\widehat{\psi}_2 - \psi_2)\frac{\partial g_k(\widehat{\psi}_1, \widehat{\psi}_2)}{\partial \widehat{\psi}_2}$$

Then for large samples, if we let $g_k^h = \frac{\partial g_k(\widehat{\psi}_1, \widehat{\psi}_2)}{\partial \widehat{\psi}_h}$ with $h = 1, 2$, we have for $k, m = 1, 2$

$$\begin{aligned}
\text{Cov}\left[g_k(\widehat{\psi}_1, \widehat{\psi}_2), g_m(\widehat{\psi}_1, \widehat{\psi}_2)\right] = \text{E}\Big\{ & g_k^1 g_m^1(\widehat{\psi}_1 - \psi_1)^2 + \left(g_k^1 g_m^2 + g_k^2 g_m^1\right)\left(\widehat{\psi}_1 - \psi_1\right)\left(\widehat{\psi}_2 - \psi_2\right) \\
& + g_k^2 g_m^2(\widehat{\psi}_2 - \psi_2)^2\Big\}
\end{aligned}$$

or, for computational purposes

$$\text{Cov}\left[g_k(\widehat{\psi}_1, \widehat{\psi}_2), g_m(\widehat{\psi}_1, \widehat{\psi}_2)\right] = g_k^1 g_m^1 \text{Var}(\widehat{\psi}_1) + (g_k^1 g_m^2 + g_k^2 g_m^1)\text{Cov}(\widehat{\psi}_1, \widehat{\psi}_2) + g_k^2 g_m^2 \text{Var}(\widehat{\psi}_2).$$

This can be expanded beyond two variables in order to get the results for the AFT model. □

The results above are known for the accelerated failure time model with no change points. In Chapter 3 we will consider a model that involves change points in the scale parameter.

2.2 EMPIRICAL DIVERGENCE BASED ON U-STATISTICS

The term "U-statistic" comes from Hoeffding (1948). It stands for unbiased statistic, and is especially important in estimation theory. In elementary statistics, U-statistics arise naturally from minimum-variance unbiased estimators. They are defined as follows.

Definition 6. *Let $f : \mathbb{R}^r \longrightarrow \mathbb{R}$ be a real-valued or complex-valued function of r variables. For each $n \geq r$ the associated U-statistic $f_n : \mathbb{R}^n \longrightarrow \mathbb{R}$ is equal to the average over ordered samples $\varphi(1), \ldots, \varphi(r)$ of size r of the sample values $f(x_\varphi)$. In other words, $f_n(x_1, \ldots, x_n) = ave\ f(x_{\varphi(1)}, \ldots, x_{\varphi(r)})$, the average being taken over distinct ordered samples of size r taken from $\{1, \ldots, n\}$. Each U-statistic $f_n(x_1, \ldots, x_n)$ is necessarily a symmetric function.*

Within the realm of non-parametric statistics, the theory of U-statistics is used to establish statistical procedures such as estimators or tests. The theory has also been used to study more general statistics, as well as stochastic processes. As in the initial work of Hoeffding (1948), much work relating to U-statistics is in the context of independent and identically-distributed variables (for early examples, Fisher (1929) and Tukey (1950)). Later, Hoeffding (1961) proved some asymptotic results of U-statistics using the Strong Law of Large Numbers.

While most work with U-statistics is for independent and identically-distributed variables, there has been some research into U-statistics for variables that do not fit these criteria, such as the work of Denker and Keller (1983) into U-statistics for weakly dependent processes and Dehling (2006) on limit theorems for dependent U-statistics.

In Chapter 4 we will be introducing a method to estimate the point of treatment time-lag effect by using an empirical divergence measure based on U-statistics. The treatment time-lag effect has been referred to in the literature as the time that it takes a treatment to take effect. This means that the survival of a control group would have distribution F_1 throughout the study, while a treatment group has distribution F_1 until the change point τ, and a different distribution F_2 afterwards (Dinse et al., 1993; Park and Qiu, 2018; Zucker and Lakatos, 1990). Let $Z_1, \ldots, Z_T \in \mathbb{R}$ be an independent sequence of time-ordered observations. Specifically, let $Z_l, l = 1, \ldots, T$ be the difference of the survival probabilities of a treatment group and a control group at event time t_l. In a naive approach to a univariate case, the well-known Kolmogorov-Smirnoff test could be used to test for homogeneity of

distribution.

For two random variables $X, Y \in \mathbb{R}$, let ϕ_x and ϕ_y denote the characteristic functions of X and Y, respectively. A divergence measure between distributions may be defined as

$$\int_{\mathbb{R}} |\phi_x(t) - \phi_y(t)|^2 \, w(t) dt \tag{2.4}$$

where $w(t)$ denotes an arbitrary positive weight function for which the integral exists. Some weight functions have been proposed by Székely et al. (2005) in order to create a divergence measure that is based on the Euclidean distances. We will adapt this clustering approach and apply it to censored data in order to find an estimate for the point of treatment time-lag effect as described previously.

As this approach is non-parametric, it avoids the difficulties of the distributional assumptions for the parametric accelerated failure time model described in the previous section. However, we only find one change point with this method, while the AFT method uses sequential testing for multiple change points and can be more flexible in that aspect.

CHAPTER 3

DETECTION OF MULTIPLE CHANGE POINTS IN AN ACCELERATED FAILURE TIME MODEL USING SEQUENTIAL TESTING

3.1 INTRODUCTION

There has been a wide variety of research into change-point analysis, which can be broadly interpreted as the point at which distribution changes. Some different applications include financial modelling (Talih and Hengartner, 2005), bioinformatics (Muggeo and Adelfio, 2011), signal processing (Kim et al., 2009), and control charts (Sego et al., 2009). Within the scope of survival analysis, the change-point problem can concern a shift in distribution for a set of time-ordered observations. One of the most widely adopted methods of survival analysis utilizes the semi-parametric proportional hazards model (Cox, 1972). The model, although often used, requires the key assumption that the hazard ratio function is constant over time. This is frequently violated in survival data as there is likely to be a lag period before an experimental treatment takes effect. A parametric alternative to this is the accelerated failure time (AFT) model, which requires distributional assumptions and instead of assuming that a covariate has a multiplicative effect on the hazard, assumes that a covariate accelerates the hazard by a constant. A semi-parametric version of the AFT model that does not have the restrictive distributional assumption has been proposed by Buckley and James (1979) but was argued by Wei (1992) to lack theoretical justification.

Prostate cancer is the most commonly diagnosed non-skin cancer and the second leading cause of cancer death in men, behind only lung cancer (Brawley, 2012). Prostate cancer demographics have changed dramatically in the past 30 years. According to the National Cancer Institute, there have been reductions in mortality rates of cancer over time, which has been attributed to successful treatments, improved methods of diagnosis, and other public health programs (Jemal et al., 2017). This has created an interest to better understand the impacts of scientific breakthroughs (along with increasingly effective treatments) on the survival for a population. Although many methods include a single change point, it is more

probable that a model with two or more change points may be appropriate, which motivates further research into the area.

In existing work, the focus has largely been on one change point. Several approaches estimate a change in the hazard rate when there is one group that receives standard of care and another that receives experimental treatment. In this case, the change points are restricted to only occur in the experimental treatment group. Park and Qiu (2018) is a semi-parametric model, while the earlier method by Dinse et al. (1993) uses a test statistic based on a difference on survival probabilities between groups, and Zucker and Lakatos (1990) investigate the log-rank test statistic. He et al. (2013) consider a MLE approach to a multiple change point model with sequential testing based on the proportional hazards model that has change points in the hazard ratio function. In the case of one group, Matthews and Farewell (1982) first suggested a likelihood ratio test for a constant hazard against a single change-point alternative, with Yao (1986) proving asymptotic distribution with some restrictions, and Goodman et al. (2011) extending to multiple change points in the model with sequential testing utilizing a Wald-type test statistic.

In this chapter, we will propose a MLE approach for finding multiple change points, since the likelihood function of survival distributions is relatively easily found. The AFT model with a Weibull distribution is more flexible than some previously suggested models, as it is the only family of distributions for which it is true that it can (as a special case) be the exponential distribution and is the only continuous distribution that gives rise to both a proportional hazard and an accelerated failure time (AFT) model (Klein and Moeschberger, 2003). This model allows for increasing or decreasing hazard rate, or can be simplified to the piecewise constant hazard model of Goodman et al. (2011), and can be re-parameterized to the proportional hazards change point model similar to the one proposed by He et al. (2013). Contrary to He et al. (2013), although we will consider covariates, we do not consider the case where a change point occurs only in one group of a binary covariate.

Since there are known complexities on the asymptotic distributions of the likelihood ratio statistics in the change point setting (see, for example, Henderson (1990) and Liu et al. (2008)), Goodman et al. (2011) used a Wald-type test statistic based on the restricted MLE that assumes change points are larger than the first non-censored survival time and smaller than the second to last non-censored time, which has asymptotic properties proven by Yao (1986). Although this circumvents problems with the distribution of the likelihood ratio test statistic, it creates the necessity to derive the variance of the estimates, which is sometimes not possible and requires numerical approximations of the hessian matrix (Gill

and King, 2003). This is often not trivial, and we did not find this method to be effective for our model. He et al. (2013) and Lee et al. (2020) (among others) deal with the issue by using Monte Carlo methods to determine the distributions of the Wald or likelihood ratio test statistics, which is computationally expensive. We will describe a sequential testing approach using Monte Carlo methods based on the likelihood ratio test statistic, using a decreasing alpha spending function first suggested by Goodman et al. (2011). This does not require for the number of change points to be known ahead of time, and uses increasingly conservative confidence levels in order to find a parsimonious model.

The rest of this chapter is structured as follows. We will define the Weibull AFT model and its reparameterization to a proportional hazards model in Section 3.2, and then define the change point scale parameter function in Section 3.3. In Section 3.4, we describe the sequential testing and model selection method. Sections 3.5 and 3.6 include the simulation study and real data application, which show that Type I error, power, and estimation of parameters are robust for our method. We conclude with some discussion in Section 3.7.

3.2 WEIBULL ACCELERATED FAILURE TIME MODEL

We can begin by considering the likelihood function for a right-censored, continuous time-to-event variable. Here, let U_i represent the iid survival time for patient i for $i = 1, \ldots, N$. The data are observed in pairs (t_i, δ_i) where

$$t_i = \min(U_i, C_i) \quad \text{and} \quad \delta_i = \begin{cases} 1 & \text{if } U_i \leq C_i \\ 0 & \text{if } U_i > C_i. \end{cases}$$

Here, $C_1, \ldots, C_N$ are the random censoring times which are assumed to be independent of the survival times, and given information from covariates we assume that the C_i's are stochastically independent of each other. The survival function for the underlying uncensored failure time variable U_i is $P(U_i > u) = S(u; \boldsymbol{\theta}, \boldsymbol{\zeta_i})$ with corresponding density $f(u; \boldsymbol{\theta}, \boldsymbol{\zeta_i})$, where $\boldsymbol{\theta}$ is a vector of parameters for the failure time model and $\boldsymbol{\zeta_i}$ is a vector of covariates associated with the ith individual. Note that the random censorship model includes the special case of type I censoring, where the censoring time of each individual is fixed in advance, as well as the case where individuals enter the study at random over time and the analysis is carried out at a prespecified time.

The accelerated failure time (AFT) model is based on the assumption that the survival of patient i, with observed data $(t_i, \delta_i, \boldsymbol{\zeta_i})$ is the same as the baseline survival function with the time scale changed by a factor known as the acceleration factor. Depending on the sign

of the acceleration factor, the time is either accelerated by a constant or degraded by a constant (Kalbfleisch and Prentice, 2011).

For the Weibull distribution, we will use the following parameterization:

$$f_0(t_i) = \lambda\nu t_i^{\nu-1}\exp\left(-\lambda t_i^{\nu}\right), \quad S_0(t_i) = \exp\left(-\lambda t_i^{\nu}\right), \quad h_0(t_i) = \nu\lambda t_i^{\nu-1} \tag{3.1}$$

where $f_0(t_i)$ is the baseline density, $S_0(t_i)$ is the baseline survival function, $h_0(t_i)$ is the baseline hazard function, $\nu > 0$ is a shape parameter and $\lambda > 0$ is a scale parameter. The Weibull distribution is flexible enough to accommodate increasing ($\nu > 1$), decreasing ($\nu < 1$), or constant hazard rates ($\nu = 1$). The Weibull distribution is also the only continuous distribution that yields both a proportional hazards and an accelerated failure time model. A proportional hazards model with a Weibull baseline hazard rate can be written as

$$h(t_i;\boldsymbol{\zeta_i}) = (\nu\lambda t_i^{\nu-1})\exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \tag{3.2}$$

which is equivalent to the AFT model we will describe. If we redefine the parameters as $\lambda = \exp(-\mu/\sigma)$, $\sigma = 1/\nu$, and $\boldsymbol{\beta} = -\boldsymbol{\gamma}/\sigma$ then the log transform of time $V_i = \log(T_i)$ gives the log linear model

$$V_i = \mu + \boldsymbol{\gamma}'\boldsymbol{\zeta_i} + \sigma W_i$$

where W_i has the extreme value distribution with probability density function $f_W(w_i) = \exp(w_i - e^{w_i})$ and survival function $S_W(w_i) = \exp(-e^{w_i}), -\infty < w_i < \infty$. The underlying probability density and survival functions, respectively, for V_i are

$$f_V(V_i;\boldsymbol{\zeta_i}) = \frac{1}{\sigma}\exp\left[\frac{V_i - \mu - \boldsymbol{\gamma}'\boldsymbol{\zeta_i}}{\sigma} - \exp\left(\frac{V_i - \mu - \boldsymbol{\gamma}'\boldsymbol{\zeta_i}}{\sigma}\right)\right]$$

and

$$S_V(V_i;\boldsymbol{\zeta_i}) = \exp\left[-\exp\left(\frac{V_i - \mu - \boldsymbol{\gamma}'\boldsymbol{\zeta_i}}{\sigma}\right)\right]$$

where σ is a scale parameter and μ is a shape parameter.

In order to see why this is an accelerated failure time model, let $S_0(t_i)$ denote the baseline survival function of $T_i = \exp(V_i)$ when $\boldsymbol{\zeta_i}$ is zero (i.e. without the presence of explanatory covariates). In other words, $S_0(t_i)$ is the baseline survival function of $\exp(\mu + \sigma W_i)$. Then, it follows that

$$\begin{aligned} S(t_i;\boldsymbol{\zeta_i}) &= P(T_i > t_i|\boldsymbol{\zeta_i}) = P(V_i > \log(t_i)|\boldsymbol{\zeta_i}) \\ &= P(\mu + \sigma W_i > \log(t_i) - \boldsymbol{\gamma}'\boldsymbol{\zeta_i}|\boldsymbol{\zeta_i}) \\ &= P(e^{\mu+\sigma W_i} > t_i\exp(-\boldsymbol{\gamma}'\boldsymbol{\zeta_i})|\boldsymbol{\zeta_i}) \\ &= S_0(t_i\exp(-\boldsymbol{\gamma}\boldsymbol{\zeta_i})). \end{aligned}$$

Note that the effect of the covariates in the original time scale is to change the time scale by the acceleration factor $\exp(-\boldsymbol{\gamma}\boldsymbol{\zeta_i})$. Also notice that the hazard rate of an individual with a covariate $\boldsymbol{\zeta_i}$ for this class of models is related to the baseline hazard rate in (3.1) by

$$\begin{aligned} h(t_i;\boldsymbol{\zeta_i}) &= h_0\left[t_i\exp(-\boldsymbol{\gamma'}\boldsymbol{\zeta_i})\right]\exp(-\boldsymbol{\gamma'}\boldsymbol{\zeta_i}) \\ &= \nu\lambda(t_i\exp(-\boldsymbol{\gamma'}\boldsymbol{\zeta_i}))^{\nu-1}\exp(-\boldsymbol{\gamma'}\boldsymbol{\zeta_i}) \\ &= (\nu\lambda t_i^{\nu-1})\exp(\sigma\boldsymbol{\beta'}\boldsymbol{\zeta_i})^{1/\sigma} \\ &= (\nu\lambda t_i^{\nu-1})\exp(\boldsymbol{\beta'}\boldsymbol{\zeta_i}) \end{aligned}$$

which is equivalent to the proportional hazards model described in (3.2).

3.3 MAXIMUM LIKELIHOOD ESTIMATION FOR CHANGE POINT MODEL

In the change point model we propose for the Weibull AFT, ν and $\boldsymbol{\beta}$ will remain unchanged, although we could also detect changes in the shape parameter ν, or even in covariate effects $\boldsymbol{\beta}$. Using the accelerated failure time model for a Weibull distribution as described in Section 3.2, we suggest a model that has a change point in the scale parameter. For a log-transform of time (as used in most statistical software), changes in the scale parameter λ would correspond to changes in μ since ν (and therefore σ) are held constant.

We propose the change point model:

$$\lambda(t_i) = \begin{cases} \lambda_1 & \text{if } 0 < t_i \leq \tau_1 \\ \lambda_2 & \text{if } \tau_1 < t_i \leq \tau_2 \\ \quad\vdots & \\ \lambda_{k+1} & \text{if } t_i > \tau_k \end{cases}$$

where $\lambda(t_i)$ is the scale parameter function for the model, $0 = \tau_0 < \tau_1 < \cdots < \tau_k < \tau_{k+1} = \infty$ are the change points, k is the number of change points in the model, and λ_j with $j = 1, \ldots, k$ is the value of the scale parameter between time points τ_{j-1} and τ_j.

As in Kalbfleisch and Prentice (2011), if t_i is the observed (possibly censored) failure time, for noninformative random censorship (meaning the censoring density function does not depend on the covariates) we have the well-known likelihood function on the data $(t_i, \delta_i, \boldsymbol{\zeta_i})$. Conditional on the parameter vector $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \nu, \lambda_1, \ldots, \lambda_{k+1}, \tau_1, \ldots, \tau_k\}$, the likelihood function is given by

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^{N} f(t_i;\boldsymbol{\zeta_i})^{\delta_i} S(t_i;\boldsymbol{\zeta_i})^{1-\delta_i}.$$

This likelihood is of the form $L(\boldsymbol{\theta}) = \prod L_i(\boldsymbol{\theta})$, where $L_i(\boldsymbol{\theta})$ is $f(t_i; \boldsymbol{\zeta_i})$ for a failure time and $S(t_i; \boldsymbol{\zeta_i})$ for a censored time. The hazard function for the change point model can be written using an indicator function, $I(\cdot)$, so that $h(t_i; \boldsymbol{\zeta_i}) = \nu\lambda_j t_i^{\nu-1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) I(\tau_{j-1} < t_i \leq \tau_j)$. Then, the cumulative hazard function is $H(t_i; \boldsymbol{\zeta_i}) = \int_0^{t_i} h(u; \boldsymbol{\zeta_i}) du$. We find that the cumulative hazard function (see derivation in Appendix B.1.1) can be described by $H(t_i; \boldsymbol{\zeta_i}) = \sum_{j=1}^{k+1} \lambda_j \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - (t_i \wedge \tau_{j-1})^\nu\right]$ since for $j > 1$, if $t_i < \tau_{j-1}$ then $(t_i \wedge \tau_j)^\nu - (t_i \wedge \tau_{j-1})^\nu = 0$. Here, $(t_i \wedge \tau_j) = \min(t_i, \tau_j)$.

Now, the log likelihood function $l(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta}))$ is

$$\sum_{i=1}^{N} \sum_{j=1}^{k+1} \left\{ [X(\tau_j) - X(\tau_{j-1})] \log\left[\nu\lambda_j t_i^{\nu-1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\right] - \lambda_j \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - (t_i \wedge \tau_{j-1})^\nu\right] \right\}$$

where $X(t) = \delta_i I(t_i \leq t)$, and $\sum_{i=1}^{N} X(t)$ can be interpreted as the number of deaths up until time t. If we fix $\tau_j, j = 1, \ldots, k$, the estimate $\widehat{\lambda}_j$ for $\lambda_j, j = 1, \ldots, k+1$ maximizes $l(\boldsymbol{\theta}; \tau_j, \ldots, \tau_k)$:

$$\sum_{i=1}^{N} \left\{ [X(\tau_j) - X(\tau_{j-1})] \log\left[\nu\lambda_j t_i^{\nu-1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\right] - \lambda_j \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - \tau_{j-1}^\nu\right] I(t_i > \tau_{j-1}) \right\}.$$

Some work (details in Appendix B.2) shows that the ML estimate is

$$\widehat{\lambda}_j = \frac{\sum_{i=1}^{N} [X(\tau_j) - X(\tau_{j-1})]}{\sum_{i=1}^{N} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - \tau_{j-1}^\nu\right] I(t_i > \tau_{j-1})}.$$

If we substitute this in to $l(\boldsymbol{\theta})$ we can find the profile likelihood of ν and τ_j:

$$l(\boldsymbol{\theta}) = \sum_i \sum_j [X(\tau_j) - X(\tau_{j-1})] \left\{ \log(\nu t_i^{\nu-1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})) + \log\left[\frac{\sum_i [X(\tau_j) - X(\tau_{j-1})]}{\sum_i \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - \tau_{j-1}^\nu\right] I(t_i > \tau_{j-1})} \right] \right\}. \tag{3.3}$$

The maximum likelihood estimates of ν and $\tau_j (j = 1, ..., k)$ are found numerically using optimization routines, such as the Nelder-Mead simplex algorithm present in the `opm` function found in `R` (Nash, 2014; R Core Team, 2019). Then, these can be substituted back to find the ML estimates for $\lambda_j (j = 1, ..., k+1)$. As noted by Goodman et al. (2011), the Nelder-Mead simplex requires user-supplied starting values, which they found needed to be reasonable, but not exact. Specifically, we will use the `survreg` function in the `survival` package in `R` to find initial values for $\widehat{\nu}$ and $\widehat{\boldsymbol{\beta}}$, and then use a grid search in order to find initial values for $\widehat{\tau}_j$ (Therneau, 2015).

We will first use the likelihood function corresponding to one change point, and then use a greedy algorithm that searches before and after that change point for a second change point, and so on. In optimization for the second change point, we fix the change point value of $\widehat{\tau}_1$ found in the first step, and then maximize the likelihood function corresponding to two change points. It is also possible to maximize by estimating both change points at the same time.

3.4 HYPOTHESIS TESTING

3.4.1 TESTING PROCEDURES

There are several well-known tests for the parameter vector $\boldsymbol{\theta}$ based on the likelihood. These include (but are not limited to) the Wald test and the likelihood ratio test. Each has proven asymptotic results - they are asymptotically chi-squared with test statistics of different forms. The Wald test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta_0}$ has the statistic

$$\chi^2_W = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\boldsymbol{I}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'$$

where $\boldsymbol{I}(\boldsymbol{\theta})$ is the Fisher information matrix. We use the observed Fisher information matrix $\boldsymbol{I}(\widehat{\boldsymbol{\theta}})$, which is simply the information matrix evaluated at the maximum likelihood estimator. From the optimization routine, the observed Fisher information matrix can be approximated by the inverse of the Hessian matrix. For large samples, $\widehat{\boldsymbol{\theta}}$ has a multivariate normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{I}^{-1}(\boldsymbol{\theta})$.

The likelihood ratio test for the same null hypothesis has a test statistic of the form

$$\chi^2_{LR} = -2\left[l(\boldsymbol{\theta}_0) - l(\widehat{\boldsymbol{\theta}})\right]$$

where $l(\boldsymbol{\theta})$ is the log-likelihood of the change point model as described in (3.3). Both of these tests can be used to test composite hypotheses. If the parameter vector $\boldsymbol{\theta}$ of length p, it is divided into two vectors $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ of lengths p_1 and p_2, respectively. If we would like to test the hypothesis $H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$, we treat $\boldsymbol{\phi}$ as a nuisance parameter. We let $\widehat{\boldsymbol{\phi}}(\boldsymbol{\psi}_0)$ be the maximum likelihood estimates of $\boldsymbol{\phi}$ obtained by maximizing the likelihood with respect to $\boldsymbol{\phi}$, with $\boldsymbol{\psi}$ fixed at $\boldsymbol{\psi}_0$. We also partition the information matrix such that

$$\boldsymbol{I} = \begin{pmatrix} \boldsymbol{I}_{\psi\psi} & \boldsymbol{I}_{\psi\phi} \\ \boldsymbol{I}_{\phi\psi} & \boldsymbol{I}_{\phi\phi} \end{pmatrix}$$

where $\boldsymbol{I}_{\psi\psi}$ has dimension $p_1 \times p_1$, $\boldsymbol{I}_{\phi\phi}$ has dimension $p_2 \times p_2$, $\boldsymbol{I}_{\psi\phi}$ has dimension $p_1 \times p_2$, and $\boldsymbol{I}'_{\phi\psi} = \boldsymbol{I}_{\psi\phi}$. Note that a partitioned information matrix has in inverse which is also a

partitioned matrix. The statistics for testing $H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$ are then given by

$$\chi^2_W = (\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) \left[\boldsymbol{I}^{\psi\psi}(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\phi}}) \right]^{-1} (\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)'$$

and

$$\chi^2_{LR} = -2 \left\{ l \left[\boldsymbol{\psi}_0, \widehat{\boldsymbol{\phi}}(\boldsymbol{\psi_0}) \right] - l(\widehat{\boldsymbol{\theta}}) \right\}.$$

As previously discussed, there are known issues with the distribution of the likelihood ratio test statistic when considering tests for change points. A naive approach to the distribution would be to assume a χ^2_2 distribution. Although Matthews and Farewell (1982) acknowledged the potential issues, the authors noted that the quantiles fit relatively well with a χ^2 distribution with two degrees of freedom. However, several authors (including Henderson (1990)) noted issues with the boundedness and distribution of the likelihood ratio test statistic. While the Wald test could circumvent these difficulties, there are times when finding the variance terms for the parameters is not trivial. Some details are given in Appendix B.3.

When comparing the proposed method (which can be interpreted as a Cox model, and can incorporate constant, increasing, or decreasing hazard rates), one area of interest might be whether the hazard rate is, indeed, constant. If the hazard rate is constant, the ν value could be set as 1 for optimization, resulting in a parameter reduction. This would be equivalent to reducing our method to that proposed by Goodman et al. (2011), with the exception that our method allows for covariate information to be included. Testing for a piecewise constant hazard is equivalent in our change point model for testing for $\nu = 1$. In that case, $\boldsymbol{\psi} = \nu$ and $\boldsymbol{\phi} = (\boldsymbol{\beta}, \lambda_1, \ldots, \lambda_{k+1}, \tau_1, \ldots, \tau_k)$. In simulation studies, the power and Type I error for testing the null hypothesis $H_0 : \nu = 1$ are reasonable for a Wald-type test statistic.

3.4.2 SEQUENTIAL TESTING APPROACH

Goodman et al. (2011) used results of independence and asymptotic distribution of a constrained maximum likelihood as proved by Yao (1986) in the case of a piecewise constant hazard. In this case, the value of τ_j are assumed to be independent of the λ_j, and a Wald-type statistic has the form

$$\chi^2_W = \frac{(\widehat{\lambda}_{k-1} - \widehat{\lambda}_k)^2}{\mathrm{Var}(\widehat{\lambda}_{k-1} - \widehat{\lambda}_k)}$$

to test the null hypothesis $H_0 : \lambda_{k-1} - \lambda_k = 0$ versus the alternative $H_1 : \lambda_{k-1} - \lambda_k \neq 0$. They approximate the Hessian matrix numerically and use only the section of the matrix needed

in order to estimate the variances in the denominator of the test statistic. The independence assumptions may not apply in our case (which has the additional shape parameter ν), and in simulation studies, we did not find that the power and Type I error of this Wald-type test were reasonable for our method.

Instead, we follow a similar method to that of He et al. (2013) and Lee et al. (2020) and propose using a likelihood ratio test statistic to develop a hypothesis test based on the number of change points. More specifically, in step m ($m = 0, 1, 2, \dots$) of the testing procedure, we consider the following hypotheses:

$$H_{0,m} : k = m \quad \text{versus} \quad H_{1,m} : k = m + 1.$$

Let $\boldsymbol{\theta}_{0,m}$ be the vector of unknown parameters in the null model in step m of sequential testing and $\boldsymbol{\theta}_{1,m}$ be the vector of unknown parameters in the alternative model. Then, the likelihood ratio test statistic at step m is of the form:

$$LR_m = -2\left[\sup l(\boldsymbol{\theta}_{0,m}) - \sup l(\boldsymbol{\theta}_{1,m})\right].$$

At step $m = 0$ of the procedure, we start by testing for the existence of change points in the model using the null hypothesis $H_{0,0} : k = 0$ compared with the alternative $H_{1,0} : k = 1$. For this step, $\boldsymbol{\theta}_{0,0} = (\boldsymbol{\beta}, \nu, \lambda)$ and $\boldsymbol{\theta}_{1,0} = (\boldsymbol{\beta}, \nu, \lambda_1, \lambda_2, \tau_1)$. We use the decreasing alpha spending function suggested by Goodman et al. (2011) so that each additional test has a more conservative confidence level. For the overall significance level α, we want $\alpha^*(m) = \alpha/2^m$, where $\alpha^*(m)$ is the significance level in step $m = 0, 1, 2, \dots$ of the hypothesis testing. Therefore, $\alpha^*(1) > \alpha^*(2) > \cdots > \alpha^*(K)$. In this manner, it is not necessary to specify K, the maximum number of change points, before model selection begins, which promotes flexibility during model selection. When using the decreasing alpha spending function, we make each sequential test more conservative. To find a parsimonious model, there should be stronger evidence for choosing a more complex model with more change points over a simpler model with fewer change points.

Because of the known complexities in the likelihood ratio statistics in the change point setting, we apply the following bootstrap scheme to approximate the distribution of the test statistic LR_m under the null hypothesis $H_{0,m}$:

1. For an observed dataset of size N, calculate the cumulative hazard estimate $\widehat{H}(t_i)$, as described in Section 3.3, using the maximum likelihood estimates $\widehat{\boldsymbol{\theta}}_{0,m}$ (found by maximizing the likelihood function under the null hypothesis) and the Kaplan–Meier

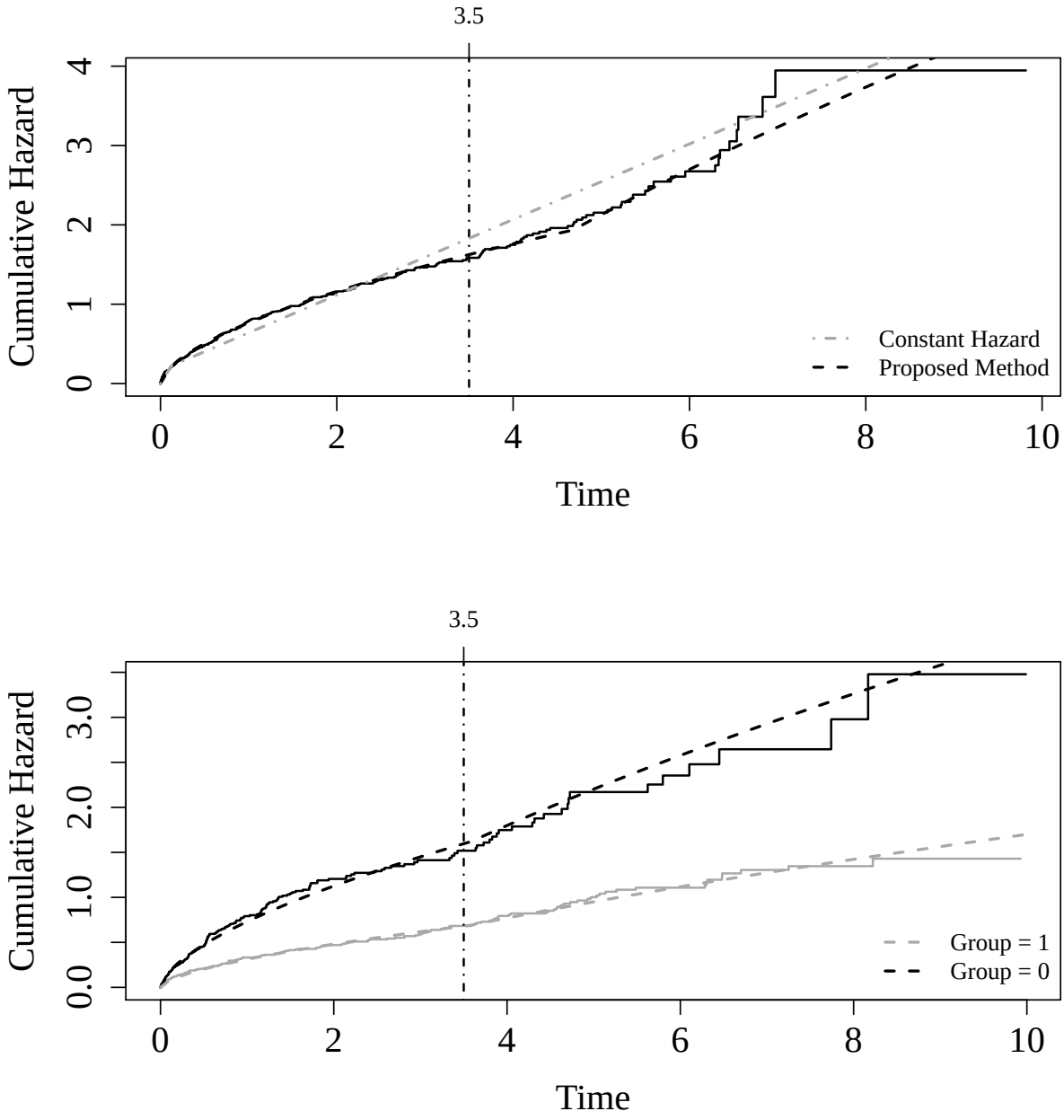

Figure 1: Estimation from one simulation with settings for 40% censoring and sample size 500 as in Table 2. The solid lines are the Nelson-Aalen estimate of cumulative hazard, and the vertical dash-dotted line shows the value of the change point.

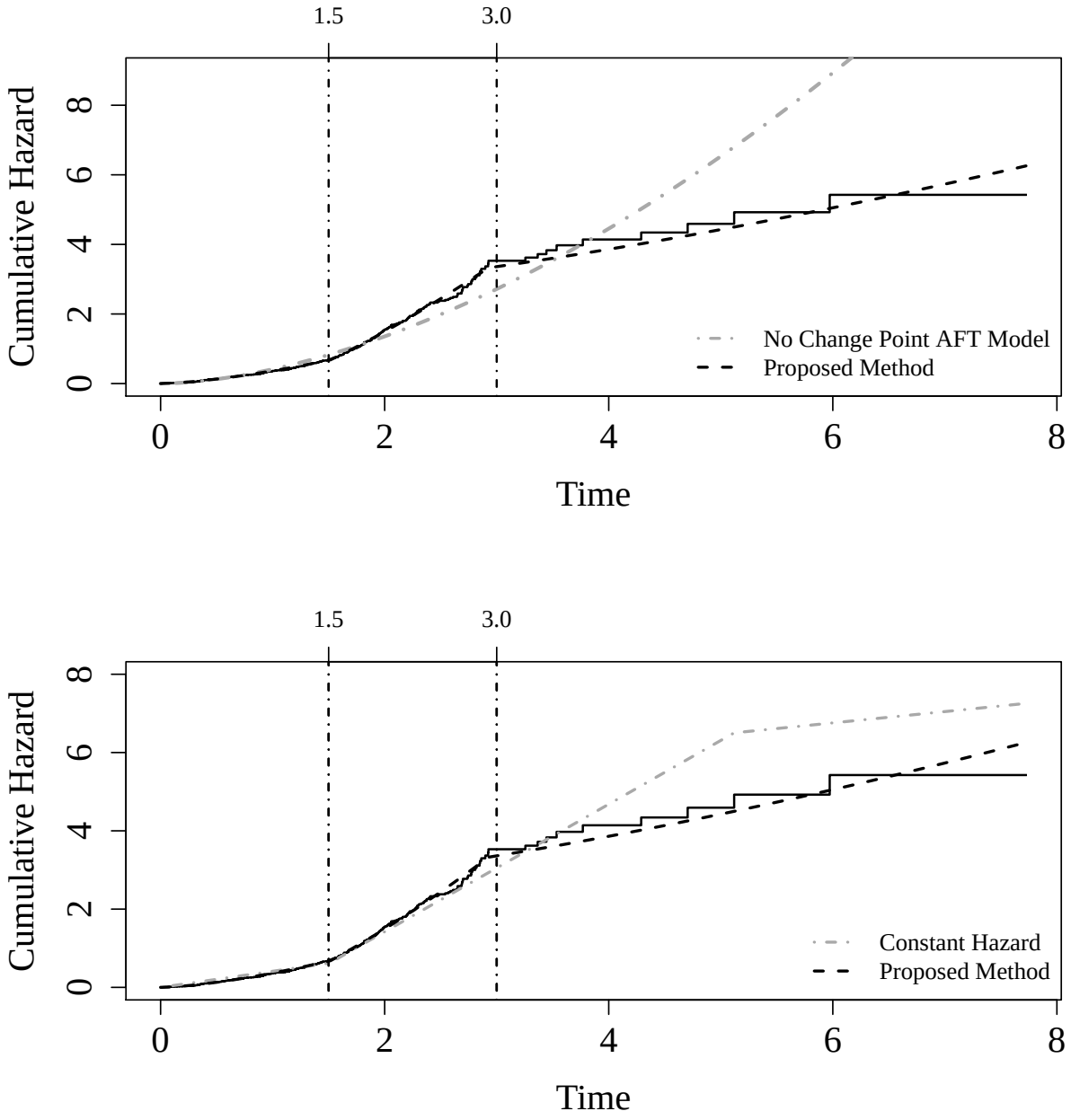

Figure 2: Estimation from one simulation with settings for 0% censoring and sample size 500 as in Table 3. The solid lines are the Nelson-Aalen estimate of cumulative hazard, and the vertical dash-dotted lines show the values of the change points.

estimate $\widehat{S}_c(t_i)$ for $S_c(t_i)$ (the survival function of the censoring variable C_i). The estimated survival function for the observed time is given by $\widehat{S}_{H_{0,m}}(t_i) = \exp\left\{-\widehat{H}(t_i)\right\}$. If covariates are included in the model, we hold them constant during the bootstrap procedure.

2. Generate B simulated datasets (e.g., $B = 1000$) based on $\widehat{S}_{H_{0,m}}$ corresponding to a true model of m change points and the censoring distribution $\widehat{S}_c(t_i)$. Calculate the likelihood ratio statistic $LR_m^b, b = 1, \cdots, B$ for each resampled data.

3. Reject the null hypothesis if LR_m, the likelihood ratio statistic calculated from the data, is larger than the $(1 - \alpha^*(m)) \times 100^{th}$ percentile of $\left\{LR_m^b, b = 1, \cdots, B\right\}$.

If we fail to reject the null hypothesis of no change points, then we conclude that the model has no change points. If the initial null hypothesis of no change points is rejected, we test the second null hypothesis that there is one change point in the model compared with the alternative hypothesis that there are two change points. This sequential testing procedure continues in this manner until the null hypothesis cannot be rejected.

3.5 SIMULATION STUDY

We examine the effectiveness of the method proposed in Section 3.2 using an extensive simulation study. First, we examine the results for models with a single change point, both with and without covariates and with varying values for the shape parameter, ν, including the cases of increasing ($\nu > 1$), decreasing ($\nu < 1$), and constant hazard rates ($\nu = 1$). Then, we present the results for the models with two change points, as this is the simplest multiple change point case, and finally provide results showing the power of our proposed test as well as the overall Type I error rate.

In the simulation, the survival times for the Weibull AFT model with change points are generated by inverting the cumulative distribution function and using the probability integral transformation (Austin, 2012). The censoring times are generated from a uniform distribution in the interval $(0, c)$, where c is adjusted to reach a pre-specified censoring rate. Three censoring rates are considered: no censoring, 20% censoring, and 40% censoring. We consider the two sample sizes of 500 and 1000. For comparison purposes, besides the proposed method, the piecewise constant hazard model of Goodman et al. (2011) is also considered here (see Section 3.1 for a description).

The initial values for the optimization were found as described in Section 3.3. For all the parameters except the change point, the initial values come from the `survreg` function in the

Table 1: Averaged point estimates (MEAN) and MSE values of the estimated parameters of the proposed model and piecewise constant hazard model based on 1000 replicated simulations for (3.4).

Sample size	Censoring rate	Parameters	Parameter values	Constant Hazard Model		Proposed Model	
				MEAN	MSE	MEAN	MSE
500	0%	ν	0.80	–	–	0.801	0.002
		λ_1	0.15	0.112	0.002	0.150	< 0.001
		λ_2	0.55	0.280	0.075	0.561	0.009
		τ	6.00	7.089	19.059	6.010	0.007
	20%	ν	1.60	–	–	1.615	0.005
		λ_1	0.75	0.552	0.078	0.752	0.002
		λ_2	0.35	1.355	1.440	0.331	0.013
		τ	2.00	1.313	2.318	2.020	0.036
	40%	ν	1.00	–	–	0.999	0.006
		λ_1	0.35	0.347	0.001	0.351	0.001
		λ_2	0.55	0.591	0.022	0.601	0.018
		τ	2.50	2.532	0.408	2.549	0.269
1000	0%	ν	0.80	–	–	0.799	0.001
		λ_1	0.15	0.112	0.002	0.150	< 0.001
		λ_2	0.55	0.276	0.077	0.559	0.005
		τ	6.00	7.089	23.696	6.004	0.001
	20%	ν	1.60	–	–	1.611	0.003
		λ_1	0.75	0.570	0.072	0.748	0.001
		λ_2	0.35	1.362	1.312	0.338	0.003
		τ	2.00	1.439	2.448	2.014	0.013
	40%	ν	1.00	–	–	0.997	0.002
		λ_1	0.35	0.349	< 0.001	0.348	< 0.001
		λ_2	0.55	0.565	0.004	0.576	0.005
		τ	2.50	2.510	0.130	2.550	0.081

`survival` package of `R` (R Core Team, 2019; Therneau, 2015). A grid search was performed to find a reasonable initial value for the change point estimate. Then, the likelihood function described in Section 3.3 was maximized using the `opm` function from the `optimx` package (Nash, 2014). Since Goodman et al. (2011) similarly note that reasonable starting values

are required to find accurate results when using the Nelder–Mead simplex algorithm, we also used a grid search for their method to find the initial values.

We begin with the simplest model, namely, that with one change point and without any covariates:

$$h(t_i; \boldsymbol{\zeta_i}) = \nu\lambda_1 t_i^{\nu-1} I(0 < t_i \leq \tau_1) + \nu\lambda_2 t_i^{\nu-1} I(t_i > \tau_1). \tag{3.4}$$

We evaluate the accuracy of the parameter estimates. Based on 1000 replicated simulations, Table 1 summarizes the averaged point estimates along with the mean square error (MSE) values of the estimated parameters. Because the piecewise constant hazard model of Goodman et al. (2011) does not have a shape parameter, it cannot be estimated. The table shows that the proposed method estimates the values of the change point, scale, and shape parameters well, even under increased censoring. In general, the proposed method provides more reliable results for the scale parameters than the piecewise constant hazard model, even when $\nu = 1$. Although the bias of the piecewise constant hazard estimates can be smaller in these cases, the MSE is usually larger. In simulation settings in which $\nu \neq 1$ (e.g., an increasing or decreasing hazard rate), the flexibility of the proposed method leads to an accurate estimation. While the piecewise constant hazard method sometimes provides reasonable estimates, it is clear that the bias and MSE are greater.

For a visual representation of the method, Figure 1 shows the cumulative hazard estimates for both the proposed method and the piecewise constant hazard method for a particular replication with a 20% censoring rate as in Table 2 with sample size 500. The solid lines represent the Nelson-Aalen estimate of cumulative hazard. The top figure shows the comparison of the constant hazard method with the proposed method when covariates are not considered. The bottom figure shows the estimation of the proposed method when the group covariate is included. We see that the fit is almost exact for the proposed method in both figures, and is closer to the Nelson-Aalen estimate in the top figure.

Figure 2 shows the fit for a particular replication with a 0% censoring rate as in Table 4, also with sample size 500. The top compares the null model with no change points to the model fit with two change points. We can see that the model with two change points is a better fit for the data than that with no change points. In the bottom, we compare the constant hazard method with our proposed method, each with two change points. We see that, especially in the second half of the graph, our proposed method has a better fit to the data.

Next, we consider the case in which the model includes a binary covariate:

$$h(t_i; \boldsymbol{\zeta_i}) = \nu\lambda_1 \exp(\beta\zeta_i) t_i^{\nu-1} I(0 < t_i \leq \tau_1) + \nu\lambda_2 \exp(\beta\zeta_i) t_i^{\nu-1} I(t_i > \tau_1), \tag{3.5}$$

Table 2: Averaged point estimates (MEAN) and MSE values of the estimated parameters of the proposed model and piecewise constant hazard model based on 1000 replicated simulations for (3.5).

Sample size	Censoring rate	Parameters	Parameter values	Constant Hazard Model		Proposed Model	
				MEAN	MSE	MEAN	MSE
500	0%	ν	1.00	–	–	1.006	0.003
		β	0.50	–	–	0.501	0.010
		λ_1	0.15	0.195	0.002	0.150	<0.001
		λ_2	0.55	0.647	0.014	0.556	0.009
		τ	5.00	5.036	0.298	5.015	0.004
	20%	ν	1.25	–	–	1.259	0.003
		β	1.20	–	–	1.192	0.011
		λ_1	0.75	1.245	0.249	0.762	0.004
		λ_2	0.25	0.417	0.046	0.238	0.006
		τ	2.00	1.941	0.069	1.975	0.041
	40%	ν	0.60	–	–	0.594	0.001
		β	-0.75	–	–	-0.765	0.014
		λ_1	0.70	0.819	0.094	0.699	0.003
		λ_2	1.20	0.238	0.926	1.327	0.090
		τ	3.50	0.673	8.806	3.346	0.430
1000	0%	ν	1.00	–	–	1.006	0.002
		β	0.50	–	–	0.504	0.005
		λ_1	0.15	0.194	0.002	0.149	<0.001
		λ_2	0.55	0.644	0.012	0.548	0.005
		τ	5.00	5.032	0.410	5.006	0.001
	20%	ν	1.25	–	–	1.256	0.001
		β	1.20	–	–	1.200	0.006
		λ_1	0.75	1.238	0.240	0.754	0.002
		λ_2	0.25	0.418	0.037	0.245	0.003
		τ	2.00	1.980	0.037	1.985	0.013
	40%	ν	0.60	–	–	0.598	0.001
		β	-0.75	–	–	-0.755	0.007
		λ_1	0.70	0.865	0.080	0.698	0.001
		λ_2	1.20	0.251	0.901	1.261	0.038
		τ	3.50	0.439	9.616	3.443	0.153

where ζ_i is generated from a Bernoulli distribution. In this way, the data are randomly assigned to a group covariate with an equal probability. The results in Table 2 show that the accuracy of the proposed method is not affected by the inclusion of a covariate. As expected, since the piecewise constant hazard method does not include information from the covariates, even when $\nu = 1$, the estimates have a larger bias and MSE than when the true dataset did not contain any covariates.

For two change-points, we first consider the simpler model that does not involve estimating any covariates:

$$\begin{aligned} h(t_i; \boldsymbol{\zeta_i}) =& \nu\lambda_1 t_i^{\nu-1} I(0 < t_i \leq \tau_1) + \nu\lambda_2 t_i^{\nu-1} I(\tau_1 < t_i \leq \tau_2) \\ &+ \nu\lambda_3 t_i^{\nu-1} I(t_i > \tau_2). \end{aligned} \tag{3.6}$$

Here, we find convergence issues for the method of Goodman et al. (2011). The † in Table 3 denotes such issues, and the simulated trials that had this were removed from calculations.

Table 3: Averaged point estimates (MEAN) and MSE values of the estimated parameters of the proposed model and piecewise constant hazard model based on 1000 replicated simulations for (3.6).

Sample size	Censoring rate	Parameters	Parameter values	Constant Hazard Model		Proposed Model	
				MEAN	MSE	MEAN	MSE
500	0%	ν	1.55	–	–	1.574	0.011
		λ_1	0.35	$0.302^\dagger$	$0.016^\dagger$	0.348	0.001
		λ_2	0.75	$1.134^\dagger$	$0.424^\dagger$	0.748	0.011
		λ_3	0.15	$0.963^\dagger$	$1.752^\dagger$	0.147	0.003
		τ_1	1.50	$1.021^\dagger$	$0.512^\dagger$	1.516	0.008
		τ_2	3.00	$3.13^\dagger$	$3.226^\dagger$	3.016	0.025
	20%	ν	1.00	–	–	1.013	0.004
		λ_1	0.75	0.763	0.003	0.758	0.003
		λ_2	0.55	0.542	0.021	0.553	0.060
		λ_3	0.15	0.153	0.006	0.140	0.004
		τ_1	2.00	1.868	0.488	1.899	0.341
		τ_2	4.00	3.895	0.243	3.892	0.156
	40%	ν	0.75	–	–	0.776	0.012
		λ_1	0.25	0.298	0.014	0.254	0.001
		λ_2	0.45	0.288	0.039	0.454	0.013

Continued from previous page.

Sample size	Censoring rate	Parameters	Parameter values	Constant Hazard Model		Proposed Model	
				MEAN	MSE	MEAN	MSE
500	40%	λ_3	0.65	0.398	1.084	0.708	0.105
		τ_1	1.00	1.745	2.066	1.376	0.832
		τ_2	4.00	4.474	1.664	4.174	0.929
1000	0%	ν	1.55	–	–	1.567	0.006
		λ_1	0.35	$0.315^\dagger$	$0.014^\dagger$	0.350	0.001
		λ_2	0.75	$1.174^\dagger$	$0.455^\dagger$	0.742	0.006
		λ_3	0.15	$1.076^\dagger$	$4.549^\dagger$	0.148	0.001
		τ_1	1.50	$1.066^\dagger$	$0.481^\dagger$	1.502	0.005
		τ_2	3.00	$3.493^\dagger$	$4.691^\dagger$	3.006	0.005
	20%	ν	1.00	–	–	1.004	0.001
		λ_1	0.75	0.756	0.001	0.751	0.001
		λ_2	0.55	0.524	0.010	0.520	0.014
		λ_3	0.15	0.152	0.003	0.147	0.003
		τ_1	2.00	2.005	0.308	2.084	0.247
		τ_2	4.00	3.972	0.182	3.995	0.092
	40%	ν	0.75	–	–	0.759	0.004
		λ_1	0.25	0.328	0.064	0.251	< 0.001
		λ_2	0.45	0.282	0.033	0.451	0.004
		λ_3	0.65	0.341	0.110	0.689	0.029
		τ_1	1.00	1.601	2.017	1.116	0.211
		τ_2	4.00	4.489	1.600	4.141	0.536

Instead of removing the simulated trials with the convergence issues for comparison of the mean and mean squared error for the piecewise constant hazard method, we can include the trials and instead compare the median and mean absolute deviance of the two methods.

From the results in Table 4, it is clear that the proposed method performs better even when considering the median and mean absolute deviance instead of the mean and mean squared error. This is the same parameterization as the no censoring cases for Table 3, excepting that in Table 3 for sample size 500 (1000), four (two) cases were removed that had the convergence issue. These values show an alternative way to compare the two methods when considering cases in which the piecewise constant hazard method has some

convergence issues. Overall, the bias of the median is lower for the proposed method than is it for the method of Goodman et al. (2011). Additionally, even when the bias is quite similar, the MAD indicates that the proposed method performs more accurately.

Table 4: Median of point estimates (MEDIAN) and mean absolute deviation (MAD) values of the estimated parameters of the proposed method and piecewise constant hazard method based on 1000 replicated simulations for (3.6).

Sample size	Parameters	Parameter values	Constant Hazard Model		Proposed Model	
			MEDIAN	MAD	MEDIAN	MAD
500	ν	1.55	–	–	1.557	0.097
	λ_1	0.35	0.368	0.076	0.348	0.026
	λ_2	0.75	1.384	0.470	0.753	0.100
	λ_3	0.15	0.641	0.523	0.144	0.046
	τ_1	1.50	1.380	0.199	1.502	0.019
	τ_2	3.00	2.954	2.157	2.989	0.048
1000	ν	1.55	–	–	1.549	0.067
	λ_1	0.35	0.385	0.046	0.349	0.017
	λ_2	0.75	1.453	0.335	0.754	0.071
	λ_3	0.15	0.686	0.491	0.148	0.031
	τ_1	1.50	1.460	0.066	1.501	0.010
	τ_2	3.00	2.990	2.216	2.995	0.025

We also consider the following more general model with two change points containing continuous covariates:

$$\begin{aligned} h(t_i;\boldsymbol{\zeta_i}) =& \nu\lambda_1 \exp(\boldsymbol{\beta'\zeta_i}) t_i^{\nu-1} I(0 < t_i \leq \tau_1) + \nu\lambda_2 \exp(\boldsymbol{\beta'\zeta_i}) t_i^{\nu-1} I(\tau_1 < t_i \leq \tau_2) \\ &+ \nu\lambda_3 \exp(\boldsymbol{\beta'\zeta_i}) t_i^{\nu-1} I(t_i > \tau_2), \end{aligned} \tag{3.7}$$

where $\boldsymbol{\zeta_i}$ is generated from a standard normal distribution. The estimate of the first change point, $\widehat{\tau}_1$, is found as described for the single change point simulation study. For the second change point, we hold $\widehat{\tau}_1$ constant and perform a grid search for the second change point before and after $\widehat{\tau}_1$. Then, we maximize the likelihood function to find the parameter estimates for the two-change point model. It is also possible to maximize τ_1 and τ_2 simultaneously. Again, similarly to the single change point results, we also perform grid searches for the piecewise constant hazard model.

Table 5 presents the results of the point estimates of the parameters based on 1000 replicated simulations. Similar conclusions to those in Table 2 can be drawn. The proposed AFT model performs well for the two change points. By comparison, the piecewise constant hazard method, which does not include information from the covariates, tends to largely overestimate the second change point. The † for the 0% censoring results indicates convergence issues with the piecewise constant hazard model. To enable comparability, the simulated trials with convergence issues were removed for that method. For a sample size of 500 (1000), three (two) cases were removed on this basis. During our simulations, we found that these issues occurred in settings where the hazard rate was not constant, covariates were not able to be included, or both. This illuminates the strict limitations of the method of Goodman et al. (2011). Altogether, the maximum likelihood for the proposed AFT change point model accurately estimates the shape, scale, and change point parameters in a wide range of simulation settings, showing the advantage of a more flexible model.

We performed a simulation study to examine how well the sequential hypothesis testing correctly identifies the true model. We generated 1000 trials with no change points, one change point, and two change points at two censoring rates (0% and 20%) and using two sample sizes (500 and 1000). For no censoring, the null model had $\nu = 1.55$ and $\lambda = 0.35$; the single change point model had $\nu = 1.55, \lambda_1 = 0.35, \lambda_2 = 0.75$, and $\tau = 1.5$; and the two-change point model had $\nu = 1.55, \lambda_1 = 0.35, \lambda_2 = 0.75, \lambda_3 = 0.15, \tau_1 = 1.5, \tau_2 = 3$. For 20% censoring, the null model had $\nu = 1$ and $\lambda = 0.15$; the single change point model had $\nu = 1, \lambda_1 = 0.15, \lambda_2 = 0.55$, and $\tau = 1$; and the two-change point model had $\nu = 1, \lambda_1 = 0.15, \lambda_2 = 0.55, \lambda_3 = 0.95, \tau_1 = 1, \tau_2 = 4$.

Table 5: Averaged point estimates (MEAN) and MSE values of the estimated parameters of the proposed model and piecewise constant hazard model based on 1000 replicated simulations for (3.7).

Sample size	Censoring rate	Parameters	Parameter values	Constant Hazard Model		Proposed Model	
				MEAN	MSE	MEAN	MSE
500	0%	ν	1.000	–	–	1.021	0.008
		β_1	-0.10	–	–	-0.100	0.002
		β_2	0.20	–	–	0.202	0.002
		λ_1	0.15	$0.152^{\dagger}$	$< 0.001^{\dagger}$	0.148	< 0.001
		λ_2	0.55	$0.477^{\dagger}$	$0.008^{\dagger}$	0.548	0.011
		λ_3	0.35	$0.350^{\dagger}$	$0.319^{\dagger}$	0.335	0.008

Continued from previous page.

Sample size	Censoring rate	Parameters	Parameter values	Constant Hazard Model		Proposed Model	
				MEAN	MSE	MEAN	MSE
500	0%	τ_1	2.00	$1.998^{\dagger}$	$0.001^{\dagger}$	2.006	0.002
		τ_2	4.00	$10.242^{\dagger}$	$62.442^{\dagger}$	4.072	0.370
	20%	ν	0.75	–	–	0.788	0.017
		β_1	0.50	–	–	0.509	0.004
		β_2	1.15	–	–	1.162	0.005
		λ_1	0.05	0.071	0.001	0.049	< 0.001
		λ_2	0.35	0.164	0.035	0.338	0.009
		λ_3	0.55	0.052	0.248	0.536	0.048
		τ_1	2.50	2.496	0.007	2.520	0.002
		τ_2	6.00	12.129	48.489	6.001	0.616
	40%	ν	1.40	–	–	1.429	0.016
		β_1	0.10	–	–	0.100	0.004
		β_2	0.10	–	–	0.102	0.004
		λ_1	0.75	0.276	0.249	0.781	0.014
		λ_2	0.35	0.539	0.043	0.340	0.003
		λ_3	0.15	0.329	0.640	0.132	0.004
		τ_1	0.50	0.255	0.213	0.538	0.039
		τ_2	2.00	2.200	0.263	2.069	0.068
1000	0%	ν	1.000	–	–	1.009	0.003
		β_1	-0.10	–	–	-0.100	0.001
		β_2	0.20	–	–	0.200	0.001
		λ_1	0.15	$0.152^{\dagger}$	$< 0.001^{\dagger}$	0.149	< 0.001
		λ_2	0.55	$0.478^{\dagger}$	$0.008^{\dagger}$	0.549	0.005
		λ_3	0.35	$0.349^{\dagger}$	$0.893^{\dagger}$	0.343	0.004
		τ_1	2.00	$1.997^{\dagger}$	$< 0.001^{\dagger}$	1.999	< 0.001
		τ_2	4.00	$10.836^{\dagger}$	$77.866^{\dagger}$	4.019	0.082
	20%	ν	0.75	–	–	0.771	0.008
		β_1	0.50	–	–	0.505	0.002
		β_2	1.15	–	–	1.156	0.003
		λ_1	0.05	0.071	< 0.001	0.049	< 0.001

Continued from previous page.

Sample size	Censoring rate	Parameters	Parameter values	Constant Hazard Model		Proposed Model	
				MEAN	MSE	MEAN	MSE
1000	20%	λ_2	0.35	0.163	0.035	0.341	0.004
		λ_3	0.55	0.055	0.245	0.535	0.021
		τ_1	2.50	2.504	0.037	2.509	0.001
		τ_2	6.00	11.831	40.469	6.144	0.380
	40%	ν	1.40	–	–	1.408	0.007
		β_1	0.10	–	–	0.096	0.002
		β_2	0.10	–	–	0.097	0.002
		λ_1	0.75	0.241	0.272	0.760	0.005
		λ_2	0.35	0.543	0.039	0.346	0.001
		λ_3	0.15	0.289	0.033	0.144	0.001
		τ_1	0.50	0.143	0.180	0.509	0.010
		τ_2	2.00	2.138	0.133	2.025	0.026

Table 6 displays the power and Type I error results for the sequential hypothesis testing. Power measures the accuracy of the proposed sequential testing approach for identifying the true model and the Type I error is the number of times our approach chooses the wrong model. In the case of the sequential testing, the Type I error is specifically the rejection of a null hypothesis that has more change points than the number of change points truly present. The Type I error rate of 0.05 was typically successfully controlled for our method. As the sample size increased, the power of the test increased and the Type I error decreased. However, while the Type I error was still controlled for increased censoring, the power of the test began to decrease. Goodman et al. (2011) find that for their Wald-type test, power was the most affected by the sample size, with a difference between λ_{k-1} and λ_k as well as between τ_{k-1} and τ_k. We tended to find the same to be true for our proposed method. Some additional simulation results for smaller sample size are given in Appendix B.4.

3.6 REAL DATA APPLICATIONS

We will apply the proposed method to two different data sets. The first is a data set on the recurrence of bladder tumours first collected by Byar (1980) in a study from the Veteran's Administration. This data set has been previously analyzed in a regression

Table 6: Power and Type I error results for the sequential hypothesis testing.

Sample size	Censoring rate	True Model	Detected Model			
			0 changes	1 change	2 changes	3 changes
500	0%	0 changes	0.945	0.045	0.010	0.000
		1 change	0.005	0.948	0.047	0.000
		2 changes	0.000	0.004	0.933	0.063
	20%	0 changes	0.920	0.080	0.000	0.000
		1 change	0.006	0.909	0.086	0.000
		2 changes	0.000	0.150	0.810	0.040
1000	0%	0 changes	0.964	0.034	0.002	0.000
		1 change	0.000	0.956	0.043	0.001
		2 changes	0.000	0.000	0.943	0.057
	20%	0 changes	0.934	0.066	0.000	0.000
		1 change	0.000	0.912	0.082	0.006
		2 changes	0.000	0.082	0.864	0.055

context by Wei et al. (1989), and was published in its entirety in Andrews and Herzberg (2012). The second data application is the Surveillance, Epidemiology, and End Results (SEER) database. Specifically, we use the data collected on prostate cancer cases diagnosed between 1973 - 2015. The SEER database is widely used to analyze cancer incidences and is implemented by the American Cancer Society to report results to the public.

3.6.1 BLADDER TUMOR RECURRENCE

The bladder tumor data first collected and reported by Byar (1980) includes 118 subjects, with a maximum number 9 tumor recurrences. We use the subset of this data of the 85 subjects who had nonzero follow-up times. The follow-up for the patients was five years, or sixty months. In this case, we consider the censoring indicator to be 1 for recurrence and 0 for everything else. Three-fourths of patients who are diagnosed with high-risk bladder cancer will recur, progress, or die within ten years of their diagnosis (Chamie et al., 2013). It is of interest to us to determine if a change point (or change points) exist in the hazard rate for tumor recurrence, and where these might occur.

Of the covariates available, we choose to include the treatment indicator based on model

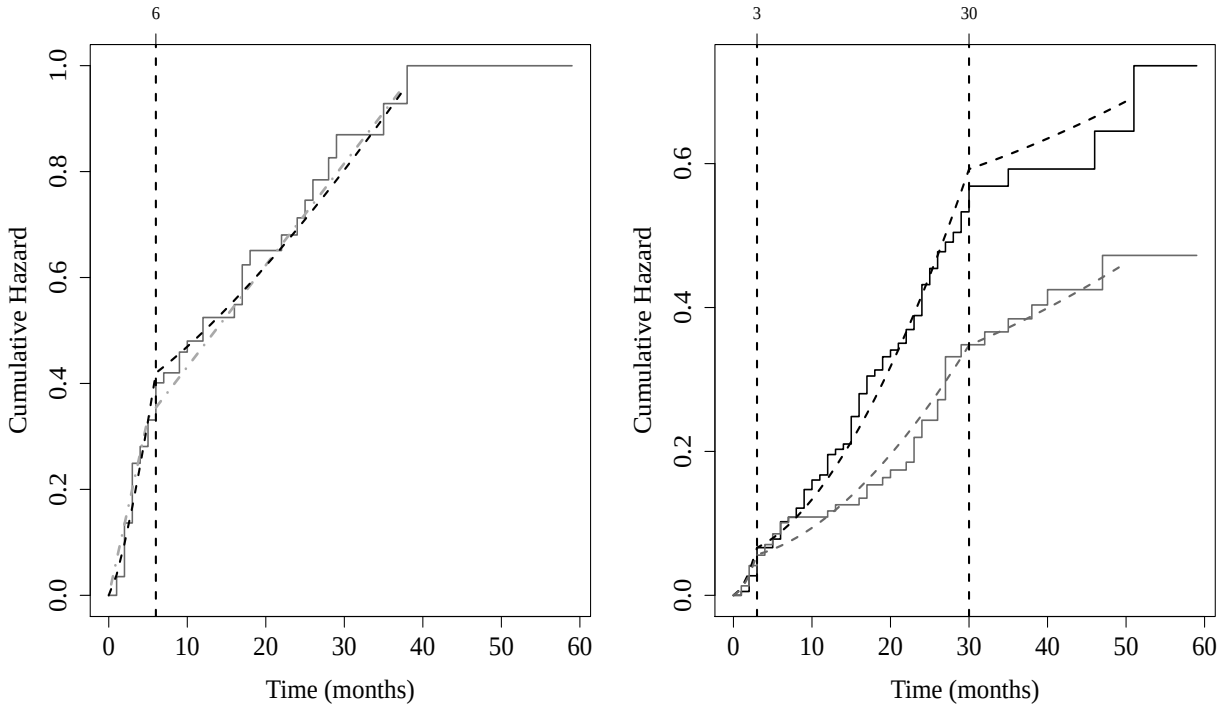

Figure 3: Vertical lines represent change point estimates from the proposed method. Left: Represents estimation without the consideration of any covariates. Right: Represents the estimation when considering the treatment covariate.

selection criteria as proposed by Akaike (1974). The placebo group contained 47 subjects while the treatment group was made of the remaining 38 subjects. There is approximately 67% censoring in this data set. The treatment group received thiotepa, which was historically used as an intravesical chemotherapy for several cancers, including Stage I bladder cancer (Fallah et al., 2012). The final model selected is

$$h(t_i; \boldsymbol{\zeta_i}) = \nu\lambda_1 \exp(\beta\zeta_i) t_i^{\nu-1} I(0 < t_i \leq \tau_1) + \nu\lambda_2 \exp(\beta\zeta_i) t_i^{\nu-1} I(t_i > \tau_1),$$

where ζ_i denotes the binary covariate indicating whether the patient received treatment or placebo. It should be noted that when no covariates are included in the analysis, only one significant change point with $LR_1 = 17.8246$ is found at $\widehat{\tau_1} = 6$ months. A second change point at $\widehat{\tau_2} = 29$ months is not significant with $LR_2 = 4.7942$. Without covariates, we find $\widehat{\nu} = 1.3306$, $\widehat{\lambda}_1 = 0.0387$, and $\widehat{\lambda}_2 = 0.0047$. The method of Goodman et al. (2011), which assumes $\nu = 1$, estimates $\widehat{\lambda}_1 = 0.0670$ and $\widehat{\lambda}_2 = 0.0192$ with a change point at $\widehat{\tau}_1 = 5$

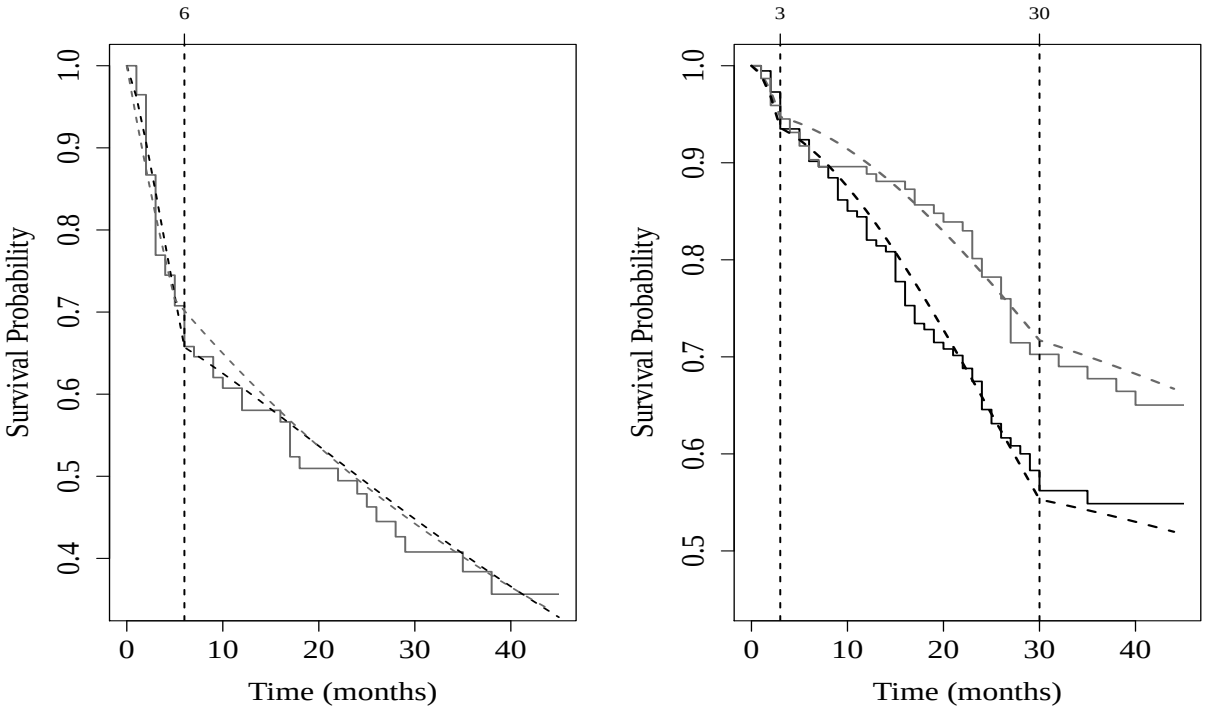

Figure 4: Vertical lines represent change point estimates from the proposed method. Left: Represents estimation without the consideration of any covariates. Right: Represents the estimation when considering the treatment covariate.

months. Both methods estimate a decrease in hazard for tumor recurrence after 5-6 months.

We see that (as in the simulation study), when ν is different from 1, the method of Goodman et al. (2011) gives a slightly different estimate for change point, and also gives a more pronounced difference in the estimates for the $\widehat{\lambda}_j$.

When the treatment covariate is included, we find two significant change points using the proposed method. We have $\widehat{\nu} = 1.7742$ with 95% confidence interval from regression (1.3577, 2.1907), indicating that the hazard rate is not constant. Also, the method estimates $\widehat{\lambda}_1 = 0.0105$, $\widehat{\lambda}_2 = 0.0012$, and $\widehat{\lambda}_3 = 0.0002$ each with standard errors from regression < 0.0001. The estimate for the covariate effect is $\widehat{\beta} = -0.4511$ with 95% confidence interval (-0.8383, -0.0638). The two change points are estimated at $\widehat{\tau}_1 = 3$ with confidence interval (2.9822, 3.0178) and finally $\widehat{\tau}_2 = 30$ with corresponding confidence interval (29.5699, 30.4301). The first change point is found significant with $LR_1 = 18.7696$ and the second change point is found significant with $LR_2 = 20.1944$. We see that there are two decreases in hazard

after 3 months and 30 months, respectively. The significance of a second change point when covariates are included shows the importance of our method's ability to include covariate information.

In Figure 3, the left graph represents the estimation without covariates with the solid line indicating the Nelson-Aalen estimate of cumulative hazard, the black dashed line the estimate from our method, and the grey dashed line the estimate from the piecewise constant hazard method. The graph on the right represents estimation including the treatment covariate. The solid black line is the Nelson-Aalen estimate for standard-of-care and the solid grey line indicates the Nelson-Aalen estimate for the treatment group. The dashed lines then represent the estimates of cumulative hazard using our proposed method. Figure 4 shows a different approach of visually examining the methods, with the solid lines indicating the Kaplan-Meier estimates of survival, and the dashed lines representing the estimates from the proposed method and the piecewise constant hazard method. We can see from the left plot in Figure 3 that the proposed method and the piecewise constant method have a similar fit to the Nelson-Aalen cumulative hazard estimate. The proposed method tends to have a better estimation until the end of the curve, at which point the two methods have a very similar estimation. When looking at the Nelson-Aalen cumulative hazard estimates in the right plot, we see that the proposed method has a good fit. The placebo group has a higher hazard than the treatment group, which is expected.

3.6.2 SEER PROSTATE CANCER

To examine prostate cancer mortality, we use the Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) database. The data includes cases diagnosed from 1973 - 2015 and follow up continues through to December 31, 2015. We are interested in finding in finding if change points exist, as well as their location. It is also of interest to include information from covariates, where appropriate. While some cancer follow-ups focus on a five year time period after diagnosis, there is interest in long-term survival and quality of life (Shrestha et al., 2019). It has been found that as diagnostic tests and treatments become more advanced, the prognosis for prostate cancer has improved (Brawley, 2012; Tewari et al., 2006).

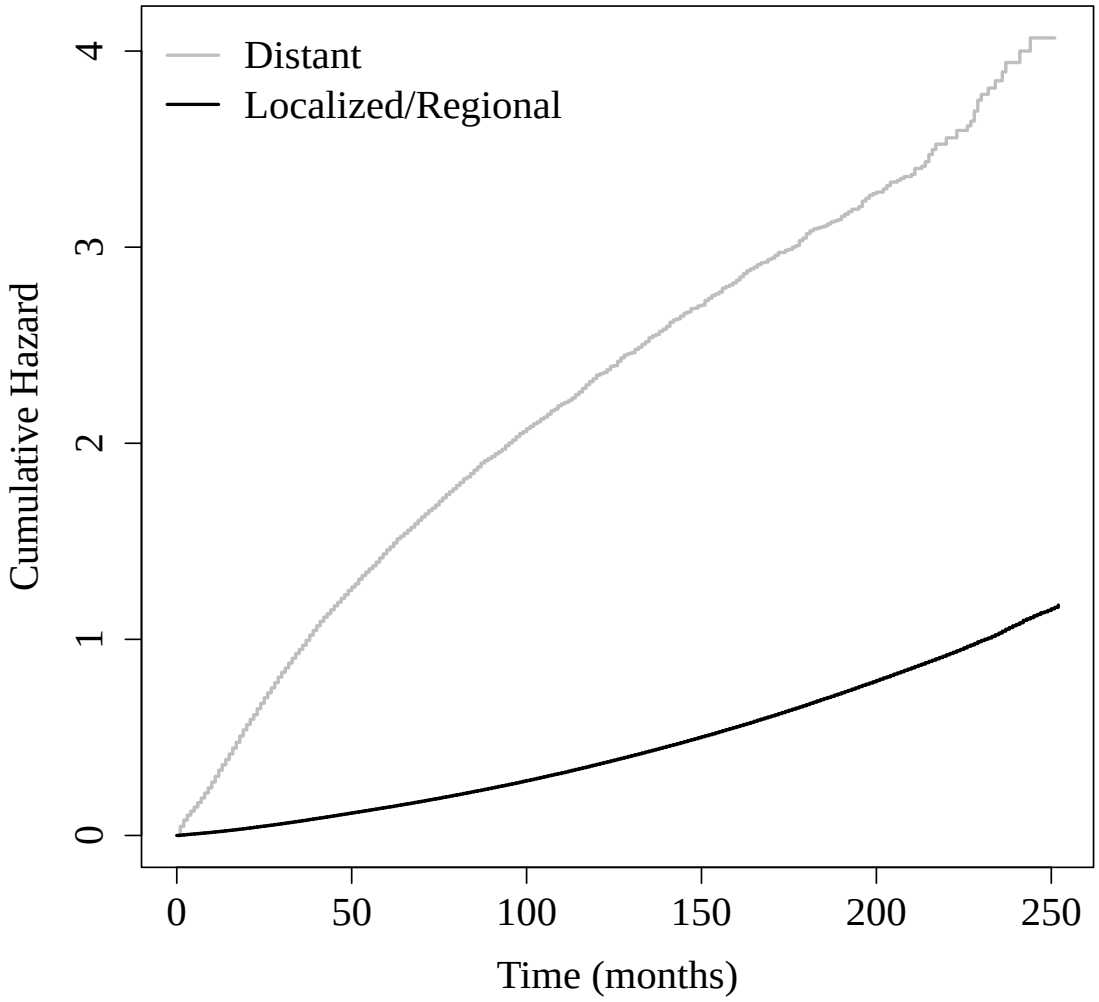

Figure 5: Nelson-Aalen cumulative hazard estimates of distant and localized/regional stages for SEER prostate cancer incidence data.

The covariates that we include are age, race, and stage of disease. All of these covariates have been of interest in past studies (Jemal et al., 2017). The SEER database codes the race covariate as "Black," "Other (American Indian/AK Native, Asian/Pacific Islander)," and "White." Stage of disease in prostate cancer has historically been described as localized, regional, or distant, where localized prostate cancer has not left the prostate gland, regional has spread outside the prostate gland to nearby structures or lymph nodes, and distant prostate cancer has spread to parts of the body farther away from the prostate. Localized and regional prostate cancers are generally grouped because they have such similar survival

expectancies (Jemal et al., 2017).

We restrict the data to cases diagnosed between 1995 - 2010 to allow for at least five years of follow-up. We begin with cases diagnosed in 1995 because this is the first year for which the covariate for stage is available in the selected database. The average age at diagnosis is 67.128 years. Out of 275,302 cases, 36,834 (13.379%) were "Black," 16,292 (5.918%) were "Other," and 222,176 (80.703%) were "White." The full data set has 58.7% censoring.

We can see in Figure 5 that the cases of prostate cancer that are localized have a much different hazard than cases that are distant. Due to this, we analyze the two stages separately. For the distant cases, men had an average age at diagnosis of 72.454 years. Out of the initial 275,302 cases, 13,091 were distant, and among these men 2,246 (17.157%) were "Black," 1,082 (8.265%) were "Other," and 9,763 (74.578%) were "White." The distant cases have 8.53% censoring. Using the model selection procedure discussed by Akaike (1974) the selected final model is

$$\begin{aligned} h(t_i;\boldsymbol{\zeta_i}) =& \nu\lambda_1 t_i^{\nu-1}\exp(\boldsymbol{\beta'\zeta_i})I(0<t_i\leq\tau_1)+\nu\lambda_2\exp(\boldsymbol{\beta'\zeta_i})t_i^{\nu-1}I(\tau_1<t_i\leq\tau_2) \\ &+\nu\lambda_3\exp(\boldsymbol{\beta'\zeta_i})t_i^{\nu-1}I(t_i>\tau_2), \end{aligned} \tag{3.8}$$

where $\boldsymbol{\zeta_i}$ includes age as a continuous covariate and race as a categorical covariate.

Table 7: Point estimates and 95% confidence intervals in parentheses for SEER prostate cancer data considering only distant cases by the proposed model and the constant hazard model by Goodman et al. (2011).

Parameters	Constant Hazard Model	Proposed Model
ν	–	1.0271 (1.0086, 1.0456)
Age	–	0.0295 (0.0276, 0.0313)
Race (Other)	–	-0.3522 (-0.4296, -0.2748)
Race (White)	–	-0.0945 (-0.1430, -0.0460)
λ_1	0.0268 (0.0264, 0.0272)	0.0032 (0.0030, 0.0034)
λ_2	0.0176 (0.0168, 0.0184)	0.0022 (0.0020, 0.0024)
λ_3	0.0128 (0.0120, 0.0136)	0.0016 (0.0014, 0.0018)
τ_1	45 (44.8312, 45.1688)	44 (43.8431, 44.1569)
τ_2	87 (86.4483, 87.5517)	87 (86.4051, 87.5950)

Table 7 shows that our method estimates that the hazard rate is very close to constant. As in the simulation studies, when the hazard rate is constant the change point estimates

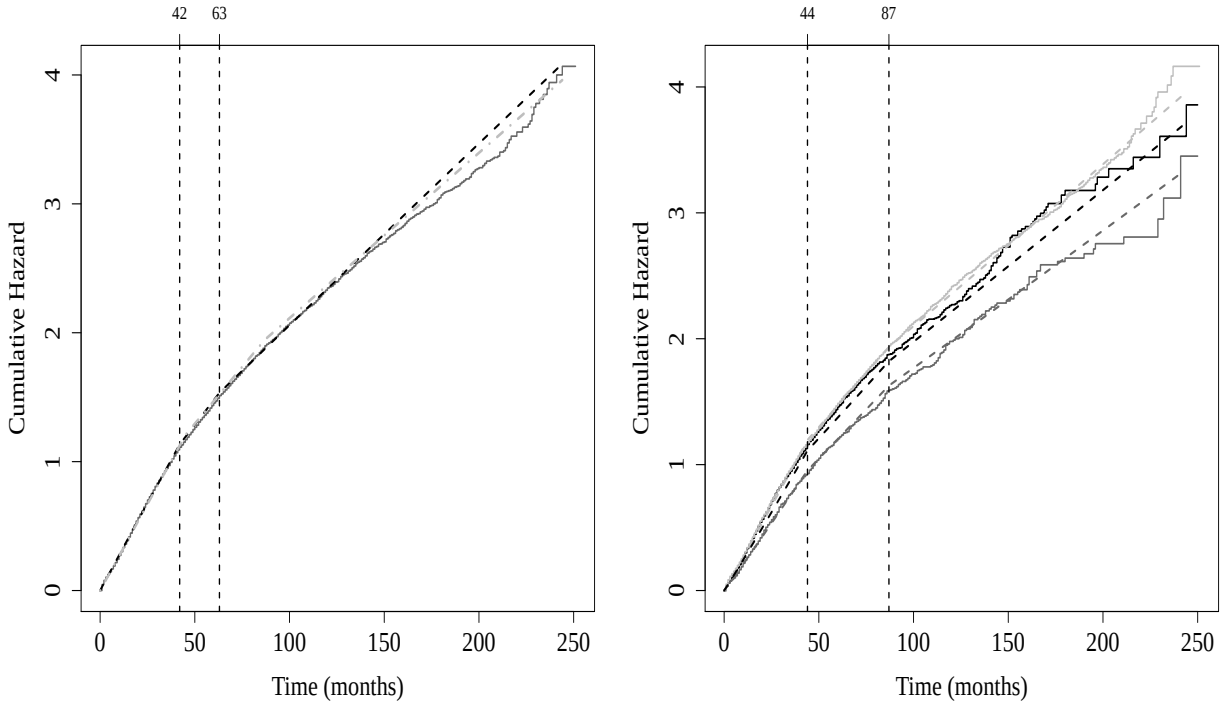

Figure 6: The vertical dotted lines represent the location of the change point estimates for the proposed method. Left: The figure represents estimation with no covariates included. Right: The figure represents estimation with the race covariate included.

can be very similar to each other. Because our proposed approach includes information from covariates and the piecewise constant method does not, the estimates for the scale parameters function are different from one another, which is also consistent with the simulation study. In both cases, the hazard decreases after about 3.7 years and again after 7.3 years. For the proposed method, we found that $LR_0 = 1463.986$. This was much larger than the 95^{th} percentile of the empirical distribution found using the bootstrap. We found $LR_1 = 54.059$, which was larger than the 97.5^{th} percentile, indicating that the two change points at 44 and 87 are significant. A third change point at 210 was found to be insignificant with $LR_2 = 6.073$.

If we do not consider the age and race covariates, our method gives the estimates $\widehat{\nu} = 0.990, \widehat{\lambda}_1 = 0.028, \widehat{\lambda}_2 = 0.020, \widehat{\lambda}_3 = 0.015, \widehat{\tau}_1 = 42,$ and $\widehat{\tau}_2 = 63$. Figure 6 shows that when covariates are not included, our method performs in a very similar manner to the piecewise constant method. This is to be expected, as our method would reduce to that of Goodman

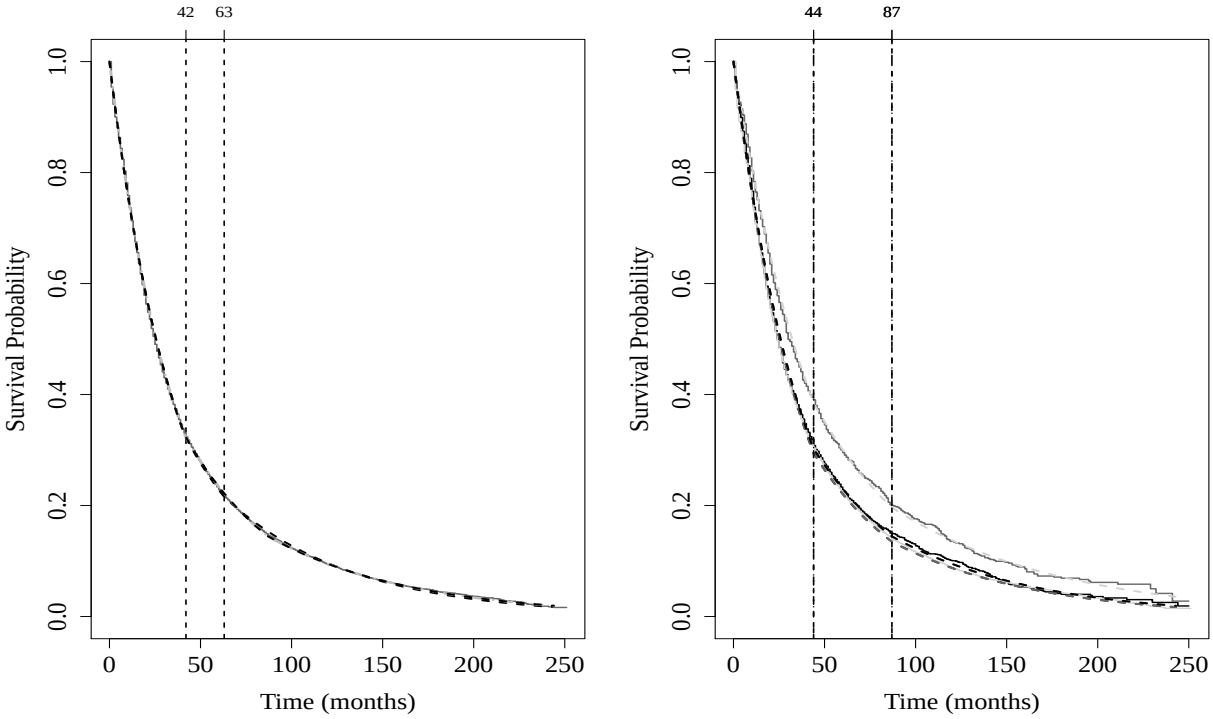

Figure 7: The vertical dotted lines represent the location of the change point estimates for the proposed method. Left: The figure represents estimation with no covariates included. Right: The figure represents estimation with the race covariate included.

et al. (2011) when the true value of ν is, indeed, 1. The simulation study shows a good level of accuracy in estimating the value of ν, so the closeness of estimation between the two methods in this case reasonable. In order to calculate the hazard estimates shown in the second plot of Figure 6, we take age to be the mean age at diagnosis and then calculate the cumulative hazard estimate for each race. We see that the "Black" hazard is highest, and "Other" is lowest. This is consistent with the findings analyzed by Brawley (2012). In the left-hand figure, the Nelson-Aalen estimate of the cumulative hazard of the data without covariates is represented by the solid black line. The dark grey dashed line represents the cumulative hazard estimated by our method and the light grey dotted line represents the estimate of the piecewise constant method. The Nelson-Aalen estimates by race are the solid lines, while the proposed method are the dashed lines. The black represents "White," dark grey represent "Other," and light grey represents "Black." Figure 7 shows a different visual representation of fit, but with the solid lines indicating the Kaplan-Meier estimates of

survival and the dashed lines representing the estimates of survival using the proposed and piecewise constant hazard methods. In the left graph, the dashed black line is the survival estimate using our method and the dashed grey line is the estimate using the piecewise constant hazard method. We can see that the survival estimate does not visually show as much difference between the methods in this case as the cumulative hazard. We expect this, since the hazard rate is extremely close to being constant. In the right-hand graph, we see results by race. The black dashed line shows survival estimate of our method for "White," the dark grey is "Other," and the light grey is "Black."

We give an analysis of the localized/regional cases in Appendix B.5.

3.7 CONCLUSION AND FURTHER RESEARCH

The inference of change-points in survival models has been discussed previously in literature (for example - Lee et al. (2020); Liu et al. (2008); Goodman et al. (2011); Yao (1986)). We use a likelihood ratio-type test statistic in order to sequentially test for change points using an alpha spending function and select a parsimonious model rather than an over-fitted one. Our methodology makes use of the Nelder-Mead simplex optimization algorithm, which requires the user to provide initial values for parameter estimates. As stated by Goodman et al. (2011), we found that these values should be reasonable, but need not be precise. There are other optimization methods readily available in software packages with differing advantages and disadvantages (Nash, 2014).

Our method uses the AFT model with a Weibull distribution, and is more flexible than other models. The method is parametric, and requires the assumption of distribution. It is reasonable to say that if the assumptions are not met, the results may be incorrect or unreliable. Non-parametric modifications to the AFT model have been suggested, but have not been widely used. If the shape parameter is set to 1, the method reduces to that of Goodman et al. (2011), and the Weibull distribution is also the only distribution that yields both an accelerated failure time model and a proportional hazards model. The simulation study shows that our approach is effective in choosing the correct number of change points and accurate in estimating the values of parameters.

While it is known that the likelihood ratio test statistic in the presence of a change point does not directly follow a Chi-squared distribution (see Henderson (1990); Matthews and Farewell (1982)), we avoid this issue by finding the empirical distribution of the test statistic through model-based bootstrap methods. We have found that, while the empirical distribution does not directly follow a Chi-squared distribution, the quantiles are relatively

similar.

We implemented the method on bladder tumor recurrence data and prostate cancer incidence data, but it could be of interest to other applications. The change-point hazard function allows for a more realistic and adaptable estimation of survival, which can be important in health care policy decisions. With a better understanding of mortality rates, we can seek to improve quality and longevity of life.

CHAPTER 4

A NON-PARAMETRIC APPROACH TO EVALUATION THE POINT OF TREATMENT TIME-LAG EFFECT

4.1 INTRODUCTION

In essence, the problem faced in this chapter is a change point analysis problem - we are trying to estimate the time of treatment time-lag effect, which in our case can be interpreted as the time point at which the survival curves of two groups change distribution. There has been a large variety of research into the change point analysis problem for different applications. In the scope of survival analysis, the problem can concern a change in distribution for a set of time-ordered observations. There are both parametric and non-parametric methods of analysis, with parametric analysis having the necessity to assume underlying distributions.

Survival analysis is a method for analyzing time-to-event data, where survival probabilities and times are sometimes presented in tables or graphs in a time-ordered fashion with indications as to which observations are censored or truncated. A major area of survival analysis includes testing for differences between groups by comparing hazard rate functions or, equivalently, survival probability functions (Klein and Moeschberger, 2003). Many procedures have been developed that non-parametrically test for differences between treatment groups. These procedures are applications of the Wilcoxon rank sum test, the most notable of which is the log-rank test first proposed by Mantel (1966). There have been adjustments proposed that are appropriate in a variety of situations, including different weighting functions that can give weight to earlier or later observations. These include, but are not limited to, Breslow (1970), Harrington and Fleming (1982), Gehan (1965), Peto and Peto (1972), and Tarone and Ware (1977). There have also been a number of methods proposed to deal with cases that have multiple or crossing hazard functions (for example, see Cheng et al. (2009); Chen et al. (2016, 2017); Liu et al. (2007); Qiu and Sheng (2008)). Additionally, there have been a few proposals to deal with cases that specifically have survival probabilities, that are initially quite similar between groups, and then differ (Dinse et al., 1993; Park and Qiu, 2018; Zucker and Lakatos, 1990). There have also been non-parametric and

semi-parametric approaches to estimating a change in the hazard rate when considering one group (Brazzale et al., 2019; Chen and Baron, 2014).

While there has been much research into testing whether survival is significantly different between groups for censored and truncated data, less work has been done to test for differences when the two groups are similar up to some point (say, $t = \tau$) and differ afterwards, specifically in terms of estimating the time point at which the two groups begin to differ. In literature, this time has been referred to as the treatment time-lag point (Dinse et al., 1993; Park and Qiu, 2018; Zucker and Lakatos, 1990). This type of analysis can be important in many areas - parts reliability, treatment effect, etc. - since it can be crucial to not only know if two groups differ but also at what point in time some treatment begins to take effect. In methods such as the non-parametric estimation proposed by Brazzale et al. (2019), and the semi-parametric approach suggested by Chen and Baron (2014) that find the change point in distribution of one survival curve, the lag effect could be found by finding the estimate for each group separately. However, it may be preferential in the case of a treatment time-lag effect when a treatment group is begin compared to a control group (for example) to use information from both groups to see when the treatment group changes in distribution from the control group.

Zucker and Lakatos (1990) proposed weighted log-rank statistics for comparing two survival curves when there is a time lag. Dinse et al. (1993) proposed an estimate of the time-lag point based on Kaplan-Meier estimates of survival. Park and Qiu (2018) suggested a semi-parametric model to estimate the time-lag point by using maximization of log partial likelihood. There is also research into this area using the terminology "change-point analysis." Chen and Baron (2014) reviewed some MLE methods as well as introducing some least-square estimation of the Cox proportional hazard model. More recently, Brazzale et al. (2019) proposed a non-parametric method to estimate the time point of change in a single survival curve, based on fitting a stump regression to p-values for testing hazards rates over small time intervals.

We will propose a way to estimate the time-lag point non-parametrically by adapting an approach for change point analysis initially implemented for multivariate time series data as initially proposed by Székely et al. (2005) and later expanded by Matteson and James (2014). We will focus not on multivariate data, but instead adapt the method for a univariate approach to compare the curves of two groups from survival data. Specifically, we want a point estimate $\widehat{\tau}$ for τ, where τ is the time point of change between two survival curves that have different distributions. In our case, we would like the survival curves to be

quite similar up until $t = \tau$, and differ after that time point. Since this is a non-parametric approach, we only say that the two survival curves have different distributions but can not assume specifically what those distributions are (i.e. $F_1 \neq F_2$ where F_1 and F_2 are some unknown distributions). The chapter will be structured as follows. In Section 4.2, we give an overview of the method and some theoretical properties. In Section 4.3, we describe the overall data simulation methods, and give simulation study results. In Section 4.4, we give a real data example to show an application of the method. Finally, Section 4.5 gives some concluding remarks and suggestions for further research. The proofs for some theoretical results are found in Appendix C.

4.2 PROPOSED EMPIRICAL DIVERGENCE METHOD

4.2.1 BACKGROUND INFORMATION AND NOTATION

For our method, we consider N independent subjects. If we let U_i be the event time (or, equivalently, survival time) for subject i and C_i be the censoring time, then we have $t_i = \min(U_i, C_i)$ as the observed event times. This means that if the data is subject to right-censoring, t_i is observed instead of U_i. Assume we have $t_1 < t_2 < \cdots < t_T$ as the distinct event times in the pooled sample of $k = 1, 2$ independent groups. At time t_i, we observe d_{ik} events in the k^{th} sample out of y_{ik} individuals at risk, $i = 1, \ldots, T$, where $T \leq N$ is the number of unique time point within the N subjects. We also have a censoring indicator δ_i such that

$$\delta_i = \begin{cases} 1 & \text{if } U_i \leq C_i, \\ 0 & \text{if } U_i > C_i. \end{cases}$$

One of the most commonly used estimates for the survival function was proposed by Kaplan and Meier (1958). They proposed a survival estimate such that

$$\widehat{S}_k(t) = \prod_{i:t_i<t} \left(1 - \frac{d_{ik}}{y_{ik}}\right)^{\delta_i}, \tag{4.1}$$

which is a step function with jumps at observations t_i for which $\delta_i = 1$. This estimate is non-parametric and can be applied in the presence of censoring. No assumptions are required for the probability distribution other than the independence between the survival and censoring variables.

Since we will be comparing our method to a previous non-parametric method for finding the point of time-lag effect that was proposed by Dinse et al. (1993), it is important to note

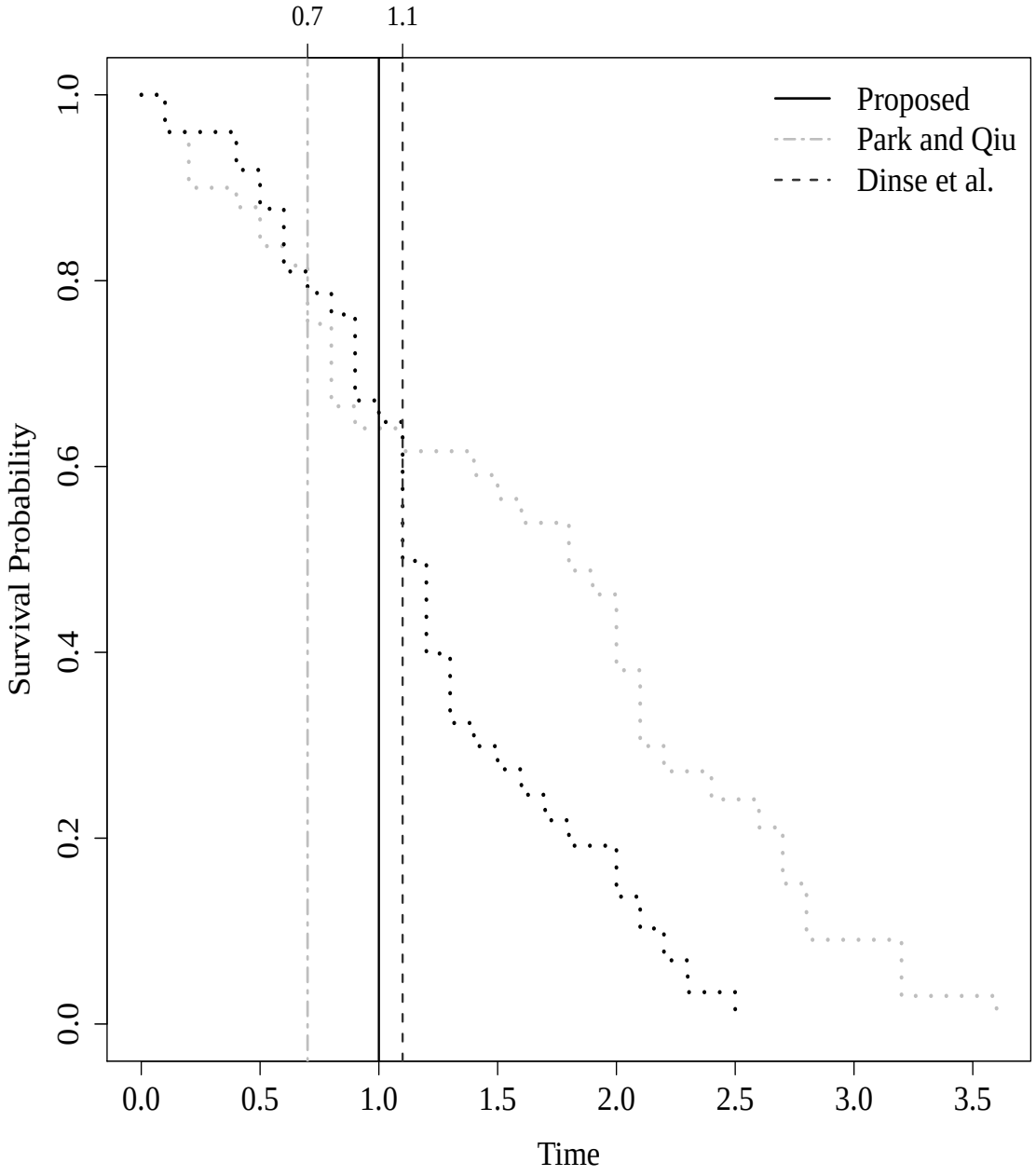

Figure 8: Kaplan-Meier curves with the black dotted line showing the control group and the grey dotted line showing the treatment group. Data was simulated from a Weibull distribution with $N = 100, \beta = 1, \lambda = 0.5, \nu = 1.5,$ and $\tau = 1$.

that the method uses the statistic

$$D(t) = \frac{\widehat{S}_1(t) - \widehat{S}_2(t)}{\sqrt{\widehat{V}_1(t) + \widehat{V}_2(t)}},$$

where, for $k = 1, 2$, $\widehat{V}_k(t) = \widehat{S}_k(t)^2 \sum_{t_i \leq t} \frac{d_{ik}}{y_{ik}(y_{ik} - d_{ik})}$ is the variance of $\widehat{S}_k(t)$ estimated by Greenwood's formula. Then, the proposed estimate of τ is $\sup\{t : t \leq t_T, D(t) \leq z_{1-\alpha}\}$ where $z_{1-\alpha}$ is the $(1-\alpha)^{th}$ quantile of the standard normal distribution. Moving forward, we will refer to this method as the Dinse et al. method. The method proposed by Park and Qiu (2018) uses a semi-parametric method to estimate τ which utilizes a model that is a generalization of the conventional Cox proportional hazards model and maximum partial likelihood estimation.

4.2.2 PROPOSED ESTIMATOR

In our method, we will be using a divergence measure based on the Euclidean distance between the two distributions to calculate the estimate $\widehat{\tau}$ (Székely et al., 2005; Matteson and James, 2014). In order to be most accurate, we should trim the data set so that we are only including time points until one of the groups reaches a survival probability estimate of 0 (i.e one of the groups has a number of subjects at risk of 0). If this does not occur before the final event time recorded, we will include all time points.

For random variables $X, Y \in \mathbb{R}$, let a primed variable X' be an independent copy of X and primed variable Y' be an independent copy of Y. Then, let ϕ_x and ϕ_y denote the characteristic functions of X and Y, respectively. Now, suppose $X, X' \overset{iid}{\sim} F_1$ and $Y, Y' \overset{iid}{\sim} F_2$ and that X, X', Y and Y' are mutually independent. If $E|X|^\alpha, E|Y|^\alpha < \infty$ for some fixed constant $\alpha \in (0, 2)$, a divergence measure between the distributions may defined by

$$\mathcal{D}(X, Y; \alpha) = \int_{\mathbb{R}} |\phi_x(t) - \phi_y(t)|^2 w(t) dt$$

where the weight function $w(t)$ is proposed by Székely et al. (2005):

$$w(t; \alpha) = \left(\frac{2\pi^{1/2}\Gamma(1 - \alpha/2)}{\alpha 2^\alpha \Gamma((1+a)/2)} |t|^{1+a} \right)^{-1}.$$

An alternate divergence measure based on Euclidean distances was then suggested by Székely et al. (2005) as

$$\mathcal{E}(X, Y; \alpha) = 2E|X - Y|^\alpha - E|X - X'|^\alpha - E|Y - Y'|^\alpha.$$

For our proposed method, let $Z_l = S_1(t_l) - S_2(t_l)$ and $\widehat{Z}_l = \widehat{S}_1(t_l) - \widehat{S}_2(t_l)$ for $l = 1, \ldots, T$ where $\widehat{S}_1(t_l)$ and $\widehat{S}_2(t_l)$ are the Kaplan-Meier estimators as in (4.1) of survival functions at time point t_l of groups 1 and 2, respectively. Note that the divergence measure is therefore based on the differences between the survival probabilities of two independent groups. Also, we define sets $\boldsymbol{X_r} = \left\{\widehat{Z}_i : i = 1, \ldots, r\right\}$ of size r and $\boldsymbol{Y_r} = \left\{\widehat{Z}_j : j = r+1, \ldots, T\right\}$ of size $T - r$ where r is defined so that $t_r \leq \tau < t_{r+1}$. Then, our empirical divergence measure based on U-statistics is obtained by:

$$\widehat{\mathcal{E}}(\boldsymbol{X_r}, \boldsymbol{Y_r}; \alpha) = \frac{2}{r(T-r)} \sum_{i=1}^{r} \sum_{j=r+1}^{T} |X_i - Y_j|^\alpha - \binom{r}{2}^{-1} \sum_{1 \leq i < k \leq r} |X_i - X_k|^\alpha - \binom{T-r}{2}^{-1} \sum_{(r+1) \leq j < k \leq T} |Y_j - Y_k|^\alpha,$$

and the scaled sample measure of divergence is as follows:

$$\widehat{\mathcal{Q}}(\boldsymbol{X_r}, \boldsymbol{Y_r}; \alpha) = \frac{r(T-r)}{T} \widehat{\mathcal{E}}(\boldsymbol{X_r}, \boldsymbol{Y_r}; \alpha).$$

Then, we can estimate the time point of treatment time-lag effect by

$$\widehat{\tau} = t_{\widehat{r}} \quad \text{where} \quad \widehat{r} = \underset{r}{argmax}\ \widehat{\mathcal{Q}}(\boldsymbol{X_r}, \boldsymbol{Y_r}; \alpha).$$

Note that by Lévy's Continuity Theorem and the Strong Law of Large Number for U-statistics, (Hoeffding, 1961) and properties of stochastic integrals for censored data, (Gill, 1980) $\widehat{\mathcal{E}}(\boldsymbol{X_r}, \boldsymbol{Y_r}; \alpha) \xrightarrow{a.s.} \mathcal{E}(X, Y; \alpha)$ as $T \to \infty$. Our statistic gives a consistent approach for estimating the treatment time-lag point by adapting clustering change point methods for multivariate time series. In simulations, the results are quite similar for values of $\alpha \in (0, 2)$, so for all calculations and simulations in this chapter, we will set $\alpha = 1$ for simplicity.

4.2.3 THEORETICAL PROPERTIES

We now give the assumptions and theorems showing the strong consistency of the estimator proposed in the previous section. For full proofs, please see the Appendix C.

Assumption 1. *Begin by assuming that we have a heterogeneous sequence of independent observations from two different distributions. Then, let $\eta \in (0, 1)$ signify the fraction of observations belonging to one of the distributions such that $Z_1, \ldots, Z_{\lfloor \eta T \rfloor} \sim F_1$ and $Z_{\lfloor \eta T \rfloor + 1}, \ldots, Z_T \sim F_2$ for every sample of size T. Let $r = \lfloor \eta T \rfloor$ and $s = T - r$. Let η be bounded away from 0 and 1 such that $r, s \to \infty$ as $T \to \infty$. Also, let $\mu_1^\alpha = E|X - X'|^\alpha$,*

$\mu_2^\alpha = E|Y - Y'|^\alpha$, *and* $\mu_{12}^\alpha = E|X - Y|^\alpha$. *Here,* $X, X' \overset{iid}{\sim} F_1$ *and* $Y, Y' \overset{iid}{\sim} F_2$ *and* X, X', Y, Y' *are all mutually independent. Further, suppose that* $E(|X|^\alpha + |Y|^\alpha) < \infty$ *for some* $\alpha \in (0, 2)$. *Therefore,* $\mu_1^\alpha, \mu_2^\alpha, \mu_{12}^\alpha, \mathcal{E}(X, Y; \alpha) < \infty$. *Finally, let* $\{\kappa_T\}$ *be a sequence of positive numbers such that* $\kappa_T \to 0$ *and* $T\kappa_T \to \infty$ *as* $T \to \infty$.

Lemma 1. *If Assumption 1 holds:*

$$\sup_{\eta \in [\kappa_T, 1-\kappa_T]} \left| \binom{T}{2}^{-1} \sum_{i<j} |Z_i - Z_j|^\alpha - [\eta^2 \mu_1^\alpha + (1-\eta)^2 \mu_2^\alpha + 2\eta(1-\eta)\mu_{12}^\alpha] \right| \overset{a.s.}{\to} 0,$$

as $T \to \infty$.

The proof follows from the Strong Law of Large Numbers for U-statistics as shown by Hoeffding (1961) and Matteson and James (2014) as well as the triangle inequality and properties of stochastic integrals for censored data proven by Gill (1980).

Theorem 1. *Suppose Assumption 1 holds. Let* $\widehat{\tau}$ *found at time point* $t_{\widehat{r}}$ *be the point estimate of treatment time-lag effect for a pooled sample with* T *distinct survival times, where* $\widehat{r} = \underset{r}{argmax}\ \widehat{\mathcal{Q}}(\boldsymbol{X}_{\boldsymbol{r}}, \boldsymbol{Y}_{\boldsymbol{r}}; \alpha)$. *Then, if* T *is large enough so that* $\eta \in [\kappa_T, 1 - \kappa_T]$, *we have:*

$$\widehat{r}/T \overset{a.s.}{\to} \eta \quad as \quad T \to \infty.$$

This implies that $\widehat{\tau} \to \tau$ as $T \to \infty$ and proves the almost sure convergence and strong consistency of the estimator. If we wish to consider specific rates of convergence, there must be additional information available about the distribution of the estimators. This, in turn, depends on the unknown distribution of the data (Matteson and James, 2014).

4.2.4 PERMUTATION TEST FOR SIGNIFICANCE

While it is of interest to estimate the location of the time point of treatment lag-effect, it can also be important to test for significance. In that case, the hypothesis we are interested in is $H_0 : F_1 = F_2$ verses $H_1 : F_1 \neq F_2$. In the simplest case of continuous univariate distributions with no censoring, the familiar non-parametric Kolmogorov-Smirnov test can be applied, as suggested initially by Kolmogorov (1933) and Smirnov (1948). When the data is right-censored, there are a variety of options that have been proposed as an alternative to the Kolmogorov-Smirnov test. The most widely-used of these may be the test of Fleming et al. (1980), known popularly as the Fleming-Harrington test.

We suggest a test statistic corresponding to the point estimate, which can be denoted as

$$\widehat{q} = \widehat{\mathcal{Q}}(\mathbf{X}_{\widehat{r}}, \mathbf{Y}_{\widehat{r}}; \alpha). \tag{4.2}$$

Large values of the test statistic indicate a significant change in distribution. However, finding a critical value with an acceptable level of precision would require knowledge of the underlying distributions, which are generally unknown. For this reason, we will use a permutation test. Under the null hypothesis of no change point, we will conduct a test with the following scheme. The observations before $\widehat{\tau}$ are permuted separately from those following $\widehat{\tau}$, creating a new data set of the same length. As in common in permutation tests for censored data, during permutation we keep the censoring and group indicators fixed (Sun and Sherman, 1996). We will apply the same estimation procedure as described in the previous section to each permuted data set, and after the permutation we record the value of the test statistic $\widehat{q}^{(b)}$, for $b = 1, \ldots, B$. There have been several forms of permutation tests that have been suggested for censored data (see, for example, Dallas and Rao (2000); Heller and Venkatraman (1996); Sun and Sherman (1996)).

Permutation tests will result in an exact p-value if all possible permutations are considered. In most cases this is computationally intractable, so we instead consider B random permutations, with a larger number of permutations reducing the error in the approximate p-value. We can estimate the p-value by $\#\left\{b : \widehat{q}^{(b)} \leq \widehat{q}\right\}/(B+1)$. In our simulations, we use $B = 1000$ permutations for all the testing. In practice, determining a suitably large number of permutations to obtain an acceptable approximation of the p-value depends primarily on sample size, as well as the number of $Z_l = S_1(t_l) - S_2(t_l)$ for $l = 1, \ldots, T$ (as described in the previous section) before and after τ.

In our extensive simulation study, we found that both point estimates (and corresponding confidence intervals) as well as the permutation tests are more accurate not only at a larger sample size but also when there are more time points both before and after the point of treatment time lag-effect, τ. Another factor that impacts the accuracy of the estimation method is the effect of treatment - with an increase in treatment effect, estimation is more exact. We should note that the proposed estimate is, by definition, limited to observed time points. As such, it will be more accurate to true values when intervals between time points are smaller. In the next section, we will present some results from our simulation study that indicate the efficacy of the proposed method.

4.3 SIMULATION STUDY

All computing for the methods presented was completed in `R` (R Core Team, 2019). We begin by simulating survival times from a Weibull distribution without time-invariant

Table 8: Mean point estimate (PE), variance, and mean square error (MSE) of the estimate for τ in model (4.3). In the table, CR denotes the censoring rate.

β	Sample size	Method	20% CR			40% CR		
			PE	Variance	MSE	PE	Variance	MSE
1.0	100	Proposed	1.109	0.044	0.055	1.124	0.046	0.062
		Park and Qiu (2018)	0.887	0.116	0.129	0.841	0.136	0.161
		Dinse et al.	1.668	0.428	0.874	1.672	0.424	0.875
	500	Proposed	1.077	0.006	0.012	1.111	0.007	0.019
		Park and Qiu (2018)	0.868	0.011	0.028	0.852	0.020	0.042
		Dinse et al.	1.134	0.257	0.275	1.102	0.192	0.202
1.5	100	Proposed	1.064	0.012	0.016	1.073	0.011	0.016
		Park and Qiu (2018)	0.822	0.041	0.073	0.792	0.050	0.093
		Dinse et al.	1.150	0.102	0.124	1.131	0.089	0.107
	500	Proposed	1.047	0.003	0.005	1.066	0.003	0.007
		Park and Qiu (2018)	0.886	0.002	0.015	0.873	0.005	0.021
		Dinse et al.	0.988	0.007	0.008	0.972	0.013	0.014

covariates using a model similar to the one used in Park and Qiu (2018):

$$h(t|g) = h_0(t)\exp\{\beta I(t > \tau)g\}, \tag{4.3}$$

where $h_0(t) = \lambda\nu t^{\nu-1}$ is the baseline hazard rate for the Weibull distribution with λ as the scale parameter and ν as the shape parameter, $I(\cdot)$ is an indicator function, g is the group indicator (1 if the i^{th} subject is in the treatment group, and 0 otherwise), τ is the time-lag point, and β is a regression coefficient vector. In (4.3), we assume $\lambda = 0.5$, $\nu = 1.5$, $\tau = 1$, $\beta = 1$ and $\beta = 1.5$, and also two sample sizes $N_1 = N_2 = 50$ and $N_1 = N_2 = 250$ (i.e the number of subjects in each group is equal). In this chapter, we simulate survival times by using results for simulating data with time-varying covariates from Austin (2012). We also discretize the survival times in order to make results more realistic in the sense that there will be more than one event per time point. Additionally, in practice most survival times are not measured on a continuous scale (Tutz and Schmid, 2016). In the following simulations, we have rounded to one decimal place. In order to simulate data that is subject to random right-censoring, we simulate survival times one at a time and also simulate a corresponding random censoring time from Uniform(a, b) where a is the minimum observed time and b is the maximum observed time. Once a desired number of censored observations has been

created, we simulate only uncensored survival times until the sample size has been reached. We consider two different censoring rates of 20% and 40% in the simulation.

The results for this simulation with 1000 replications are shown in Table 8 at different sample sizes for the non-parametric Dinse et al. method, the semi-parametric Park and Qiu (2018) method, and the proposed non-parametric method. The mean point estimate along with variance and mean square error values of the estimated time-lag point are summarized in Table 8. From the table, we can see that overall, the proposed method has a lower MSE than the other methods, even in cases where the bias is slightly larger. The proposed estimator also consistently performs better with larger samples sizes and with lower censoring rates, while this is not always true for the Dinse et al. estimator. In general, we find that the estimate for τ is over-estimated using the Dinse et al. method, and is slightly under-estimated using the Park and Qiu (2018) method. As expected, it is true that results are better for larger values of β (i.e when the two distributions are more different from each other after τ). It is also of note that sometimes the Dinse et al. method does not give an estimate, in which case there was no possible solution in that replication. In order to find an estimate for τ, we remove these values and calculate the estimate from the remaining values found.

In the case when the sample size of 500 and the censoring rate of 20% in model (4.3), the density curves of the estimated time-lag points by the proposed, Park and Qiu (2018), and Dinse et al. methods are shown in Figure 9. From the plots in the figure, it can be clearly seen that the results by Dinse et al. are generally more spread out with values that are further above the true value of τ, while the results from both proposed and Park and Qiu (2018) are typically much closer to each other and the true value of τ. The Dinse et al. results also show a cluster of time points that are quite over-estimated, which indicates bi-modality instead of the normal distribution originally assumed by the authors (Dinse et al., 1993).

In order to stay true to real data applications, we also simulate data including time-invariant covariates from the following more general model with an exponential distribution:

$$h(t|g) = h_0(t) \exp\{\beta I(t > \tau)g + \gamma_1 v_1 + \gamma_2 v_2\}. \tag{4.4}$$

For example, we simulate covariate v_1 (gender) from a Binomial($n = 1$, $p = 0.5$) distribution and v_2 (age) from a Uniform($a = 1$, $b = 25$) distribution rounded down to the year. We will show with these simulations that the addition of covariates does not largely change the results of the simulations. Further, we assume that we have $h_0(t) = 0.1$, $\beta = 0.75$ and 1.25,

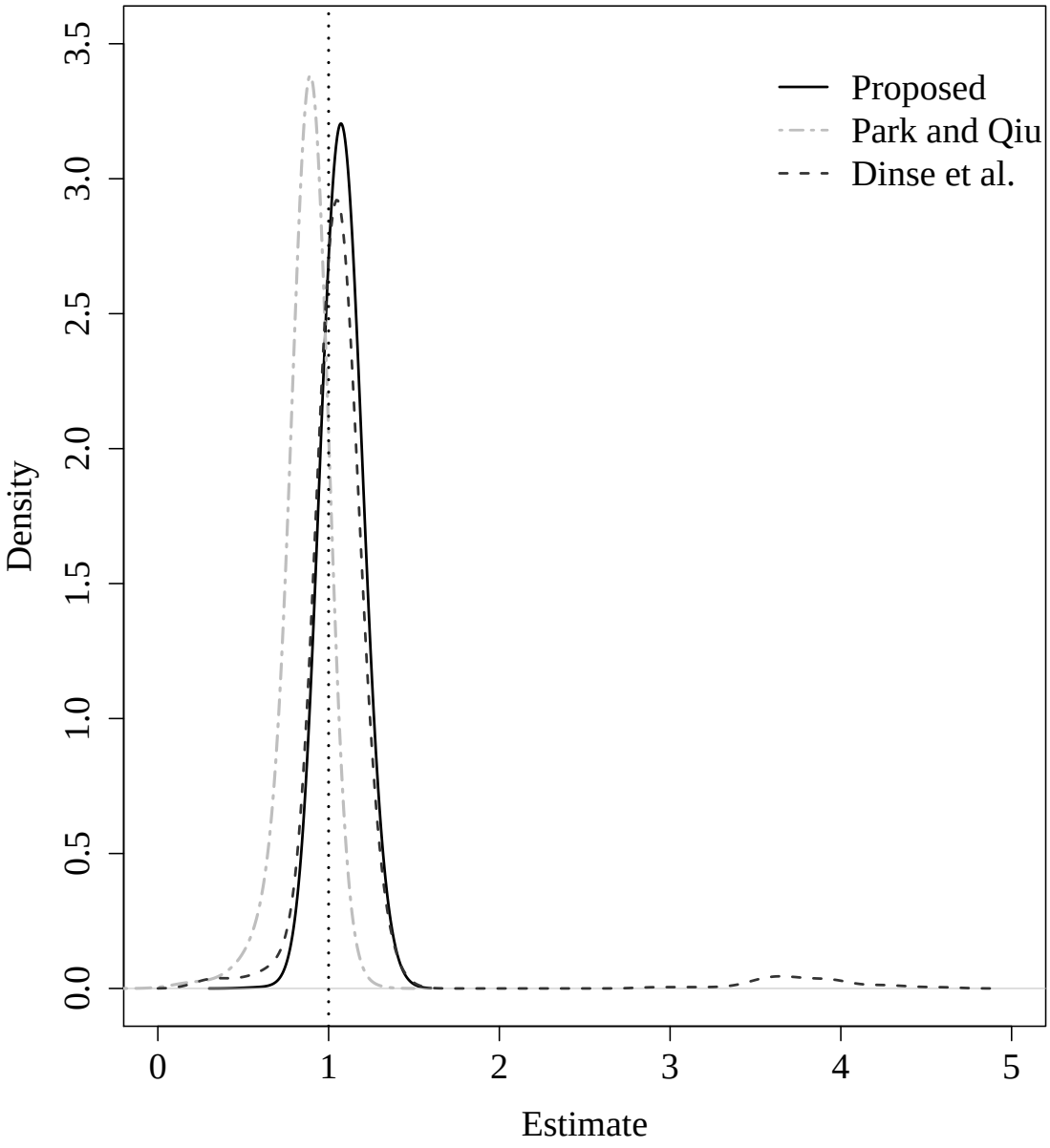

Figure 9: Density curves of the estimated time-lag points for 1000 replications, with the sample size of 500 and the censoring rate of 20% in model (4.3) with $\beta = 1$. The black dotted line represents the true value for $\tau = 1$.

Table 9: Mean point estimate (PE), variance, and mean square error (MSE) of the estimate for τ in model (4.4). In the table, CR denotes the censoring rate.

β	Sample size	Method	20% CR			40% CR		
			PE	Variance	MSE	PE	Variance	MSE
0.75	200	Proposed	3.303	1.642	1.732	3.384	1.610	1.755
		Park and Qiu (2018)	3.434	3.555	3.740	3.281	3.247	3.323
		Dinse et al.	8.176	1.433	28.226	8.009	1.748	26.838
	500	Proposed	3.590	0.948	1.295	3.639	0.958	1.366
		Park and Qiu (2018)	3.045	1.834	1.834	2.958	1.873	1.873
		Dinse et al.	8.894	1.600	36.340	8.589	2.505	33.741
1.25	200	Proposed	3.209	0.495	0.538	3.269	0.487	0.559
		Park and Qiu (2018)	2.777	0.693	0.742	2.655	0.685	0.803
		Dinse et al.	6.610	4.195	17.224	6.370	4.576	15.927
	500	Proposed	3.303	0.101	0.192	3.332	0.092	0.202
		Park and Qiu (2018)	2.793	0.102	0.145	2.758	0.186	0.245
		Dinse et al.	6.693	7.403	21.033	5.922	7.877	16.407

$\gamma_1 = 0.25$, $\gamma_2 = 0.1$, $\tau = 3$, and maximum allowed time of 10. The sample sizes are set to be $N_1 = N_2 = 100$ and $N_1 = N_2 = 250$ and the censoring rates are 20% or 40%. Table 9 presents the results about the mean point estimate, variance, and mean square error of the estimate for τ in model (4.4).

We can see in Table 9 that the results still seem reasonable, and overall are consistent with the previous simulation results with no time-invariant covariates. The proposed method is comparable to the Park and Qiu (2018) method. In the case where $\beta = 0.75$ (i.e. the two groups are more similar to each other) the proposed method performs best. The MSE and variance of the method consistently decrease with lower censoring and larger sample size. We have found through additional larger sample sizes (not shown) that the proposed estimator converges slower with covariates and with lower values for β. Although the method of Park and Qiu (2018) performs somewhat better at the larger value of β and larger sample sizes with the added information from covariates, this intuitively seems correct since the method is semi-parametric, and the proposed method does not use any of the information from the covariates in calculating the estimate. Also, the model used to simulate the data is the model used in the methods of Park and Qiu (2018), so these results are both reasonable and

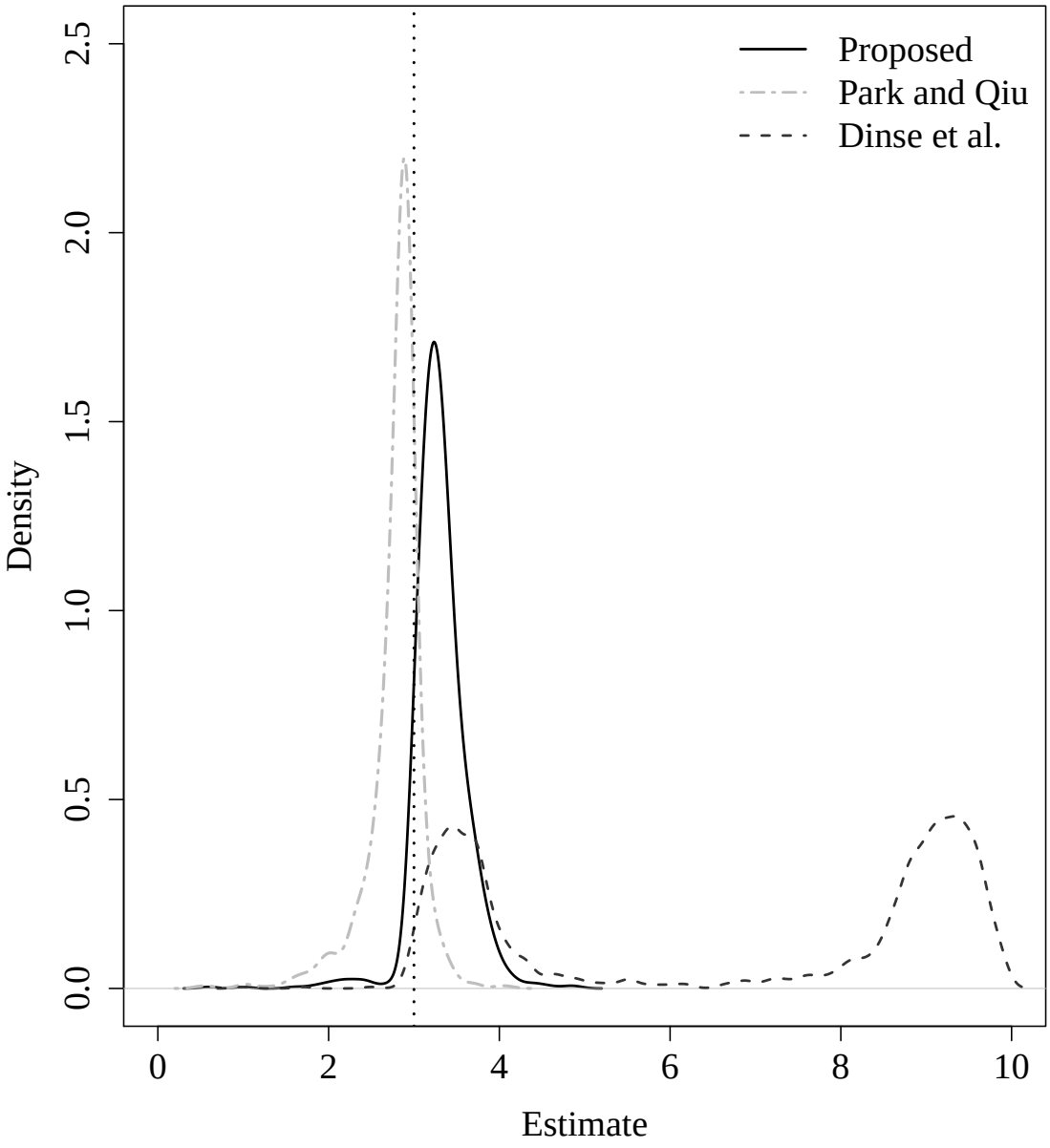

Figure 10: Density curves of the estimated time-lag points for 1000 replications, with the sample size of 500 and the censoring rate of 20% in model (4.4) with $\beta = 1.25$. The black dotted line represents the true value for $\tau = 3$.

expected. Despite this, there are still cases where the proposed method is more accurate. The Dinse et al. method still generally over-estimates, with the bias and MSE now being much larger than the previous simulation results. It should also be noted that the Dinse et al. method also sometimes has better results in these simulations with increased censoring, indicating that the method of Park and Qiu (2018) and the method proposed in this chapter are more reliable.

The density curves in Figure 10 show that the results for the data including covariates looks similar to the results that does not include covariates, but we can now much more obviously see the bi-modality of the Dinse et al. estimator. We see that this parameterization gives results that are approximately as accurate as previous simulations. This indicates that the addition of covariates does not have a very large effect on the results of the method in estimating the point of treatment time-lag effect with the proposed method. It is also important to note that the method of Park and Qiu (2018) assumes that the models in (4.3) and (4.4) are known, while this is not true of the current proposed method since we do not use any model information. In cases where the assumed model is incorrect, the results obtained using the method of Park and Qiu (2018) could be misleading but the method proposed here would still be valid.

Finally, we perform simulations in order to assess the performance of bootstrap confidence intervals. Because the method of data simulation discussed in this section can be computationally expensive, performing bootstrap confidence intervals would be intractable. Instead of using a uniform censoring distribution to create an exact number of censored observations, we simply use a maximum cut-off time. In order to increase (decrease) censoring, we decrease (increase) the maximum cut-off time. In these simulations, we found that the results for the point estimates have a similar accuracy to the results reported in Table 8 and Table 9. Under change point conditions, the classical bootstrap initially suggested by Efron and Tibshirani (1993) has been found to be inconsistent in the construction of bootstrap confidence intervals (Xu et al., 2014; Canty and Ripley, 2017). In our simulation studies, we found this to be true as well. Therefore, we will use the non-parametric bias-corrected and accelerated (BCa) bootstrap confidence intervals of Efron and Narasimhan (2020) using the `bcaboot` package (Efron and Narasimhan, 2018). The BCa confidence intervals are second-order accurate, and correct for both bias and skewness in the distribution of bootstrap replicates by incorporating a bias-correction factor and an acceleration factor.

As in the point estimate simulations, we give results for two different censoring rates (20% and 40%) and for several different samples sizes in order to investigate the effects of

Table 10: Percent coverage and average length of 95% BCa confidence intervals for Weibull distribution simulations following (4.3). CR denotes the censoring rate.

Parameter Values	Sample Size	CR	Empirical Coverage	Average Length
$\tau = 3$, $\nu = 1.25$, $\lambda = 0.1$, $\beta = 1.3$	200	20	0.943	1.701
		40	0.890	1.722
	500	20	0.945	0.986
		40	0.891	1.628
$\tau = 2$, $\nu = 1$, $\lambda = 0.25$, $\beta = 1$	100	20	0.923	2.216
		40	0.904	2.705
	500	20	0.947	0.967
		40	0.938	1.015

increased sample sizes and censoring on the interval length and coverage percentage. We use $B = 1000$ bootstrap replicates to calculate each interval, with the acceleration factor estimated using the jackknife method. This newer method of Efron and Narasimhan (2020) is advantageous to similar prior methodology due to the fact that the proposed method can be time-consuming with an increase in sample size - in the newer method the B bootstrap replications can be carried out separately on a distributed basis.

For 1000 replications, the results in Table 10 show the empirical coverage and average interval length for some different simulation settings for a Weibull distribution without covariates (as in (4.3)). It should be noted that the second setting has a Weibull distribution with the shape parameter $\nu = 1$, which simplifies it to an Exponential distribution. In bootstrapping with covariates, the covariates themselves are held constant and we found that the results for models including covariates are quite similar in accuracy to those shown in Table 10. We see that for a lower censoring rate of 20%, the empirical coverage of the BCa confidence intervals are quite close to the expected 95%, and the coverage increases with an increase in sample size. The average interval length also decreases with an increased sample size. For the increased censoring of 40%, the coverage is not as precise as for the lower censoring cases but still increases for a larger sample size. Also, the average interval lengths are somewhat longer. In general, we see that the results for the bootstrap confidence intervals are quite reasonable.

Finally, we present a selection of results for the permutation test described in Section 4.2.4. We can see that for the smaller sample sizes and higher censoring rates, the number

of rejected null hypotheses is lower than what would be expected. Also, we can see that with a decrease in β the results are not as reliable. In general, this should be expected since the distributions are more similar to one another with smaller values for β. However, with an increase in sample size the results become much more acceptable. In Table 11, on the left, data was simulated as in (4.3) and parameter values as in Table 8. On the right, data was simulated as in (4.4) and parameter values as in Table 9.

Most importantly, as was noted by Matteson and James (2014), the number of observation before and after the change point have more impact on accuracy of results than sample size alone. Although the estimation will still be quite reasonable for data where there are fewer observations than ideal either before (or after) the change point, the result from the permutation test may begin to become more unreliable.

4.4 REAL DATA APPLICATIONS

In order to show an application of the proposed method, we will apply the method to two data sets. First, we analyze lung cancer data from the Veteran's Administration Lung Cancer Trial. Second, we will use a German breast cancer data set first used by Schumacher et al. (1994) and then Sauerbrei and Royston (1999).

4.4.1 VA LUNG CANCER DATA

First, we will use a data set from the Veteran's Administration Lung Cancer Trial on patients with advanced, inoperable lung cancer who were treated with chemotherapy. The data is taken from Kalbfleisch and Prentice (2011). The variables available in the full data

Table 11: Number of significant test statistics out of 1000 simulations with $B = 1000$ permutations for each at $\alpha = 0.05$ level of significance. CR denotes censoring rate.

		95% Confidence	
β	Sample Size	20% CR	40% CR
1	100	0.784	0.786
	500	0.968	0.953
1.5	100	0.918	0.909
	500	0.978	0.974

		95% Confidence	
β	Sample Size	20% CR	40% CR
0.75	200	0.756	0.714
	500	0.877	0.806
1.25	200	0.957	0.920
	500	0.996	0.972

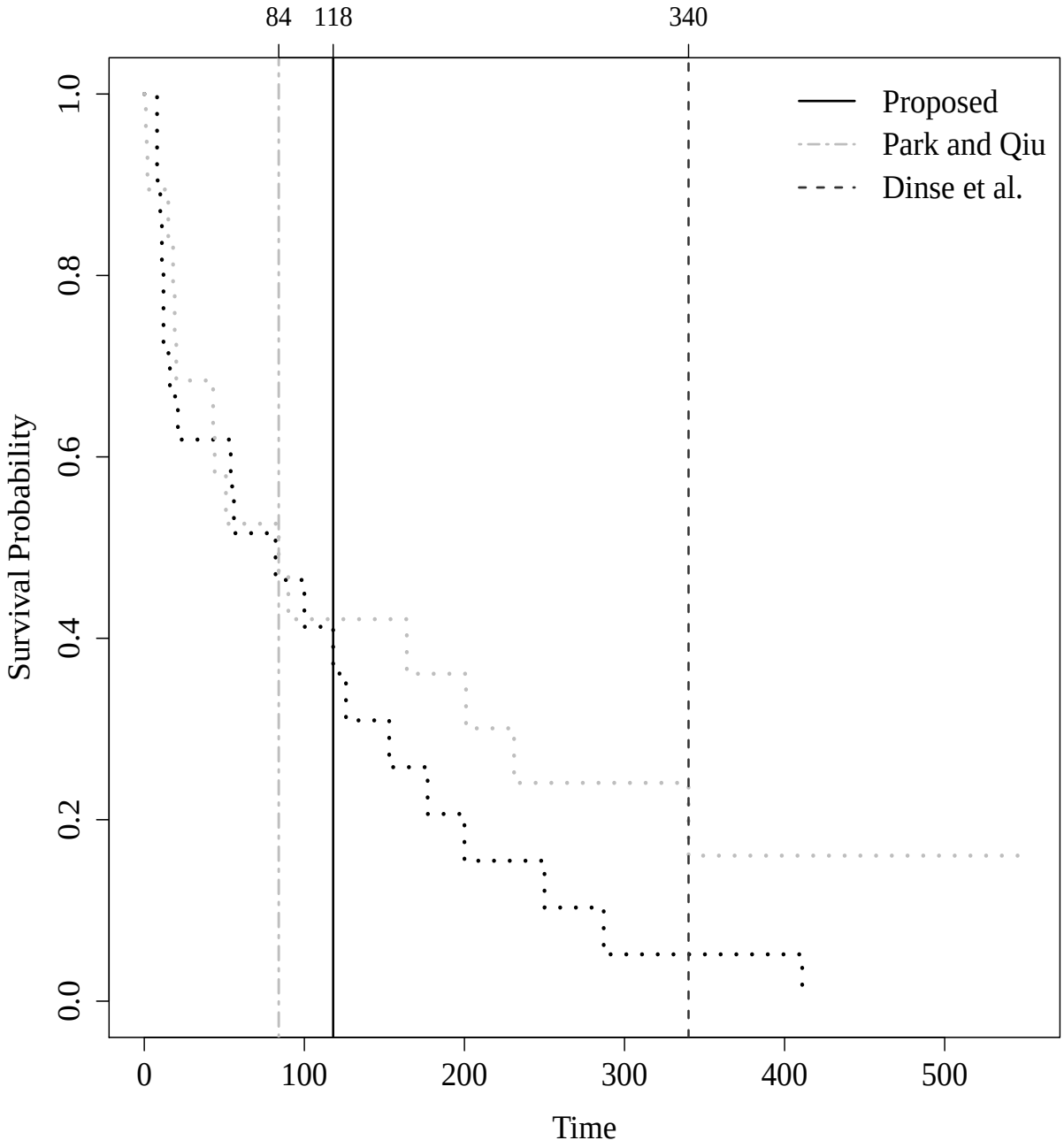

Figure 11: Kaplan-Meier curves for the distribution of event time in lung cancer according to two groups; the black dotted line denotes the standard treatment group and the grey dotted line denotes the test treatment group.

set are treatment (standard or test), cell type (squamous, small cell, adeno, large), survival (in days), status (dead or censored), a Karnofsky score as a measure of general performance, months from diagnosis, age in years, and a prior therapy indicator.

If we split the data set into two groups depending on whether the patient received standard or test treatment, we can implement the methods shown in the prior section to estimate the time point of change between the two groups. It may be of specific interest to see when patients who have had prior treatment begin to see a treatment lag-effect between the standard treatment and the test treatment. We can subset this data set and view only subjects in the study who have had previous treatment. In this case, we would have 40 patients in the study with an approximately 8% censoring rate. Then, the number of subjects in the standard treatment group is 21 and 19 in the test group.

Figure 11 shows the Kaplan-Meier curves for the distribution of time to event for lung cancer according to two groups; the black dotted line denotes the standard treatment group and the grey dotted line denotes the test treatment group. There seems to be treatment lag-effect between the two groups. It is clear that no distinction can be made between two groups until around time point 100, and they show different patterns right after the time point. Estimated time-lag points by our proposed method (solid line), the Park and Qiu (2018) method (grey two-dashed line), and the Dinse et al. method (dashed line) are also presented in the figure. From the proposed method, we find an estimate for τ of 118. From the Park and Qiu (2018) method, we find an estimate of 84, and finally from the Dinse et al. method, we find an estimate of 340. The results of these estimates initially seem consistent with the results from the simulation study. The Park and Qiu (2018) estimator is at a slightly earlier time point while the proposed estimator is at a slightly larger time point, and the Dinse et al. estimator is in this case quite far from the seeming time point of treatment lag-effect. This is consistent with the trends in over- and under-estimation found in the simulation study. The estimates of treatment lag-effect at 7 years for the method of Park and Qiu (2018) and approximately 9.8 years for the proposed method when considering the standard treatment versus test treatment are reasonably consistent when compared long-term survival rates (Boyer et al., 2017).

In order to find a confidence interval for the estimator, we use the non-parametric BCa bootstrap method as initially suggested by Efron and Narasimhan (2020). Because survival data is more complex in structure, and the size of the application data set is quite small this may be a case where the standard bootstrap fails and the percentile bootstrap confidence interval may not be appropriate to use (Efron and Tibshirani, 1993; Rizzo, 2008). We sample

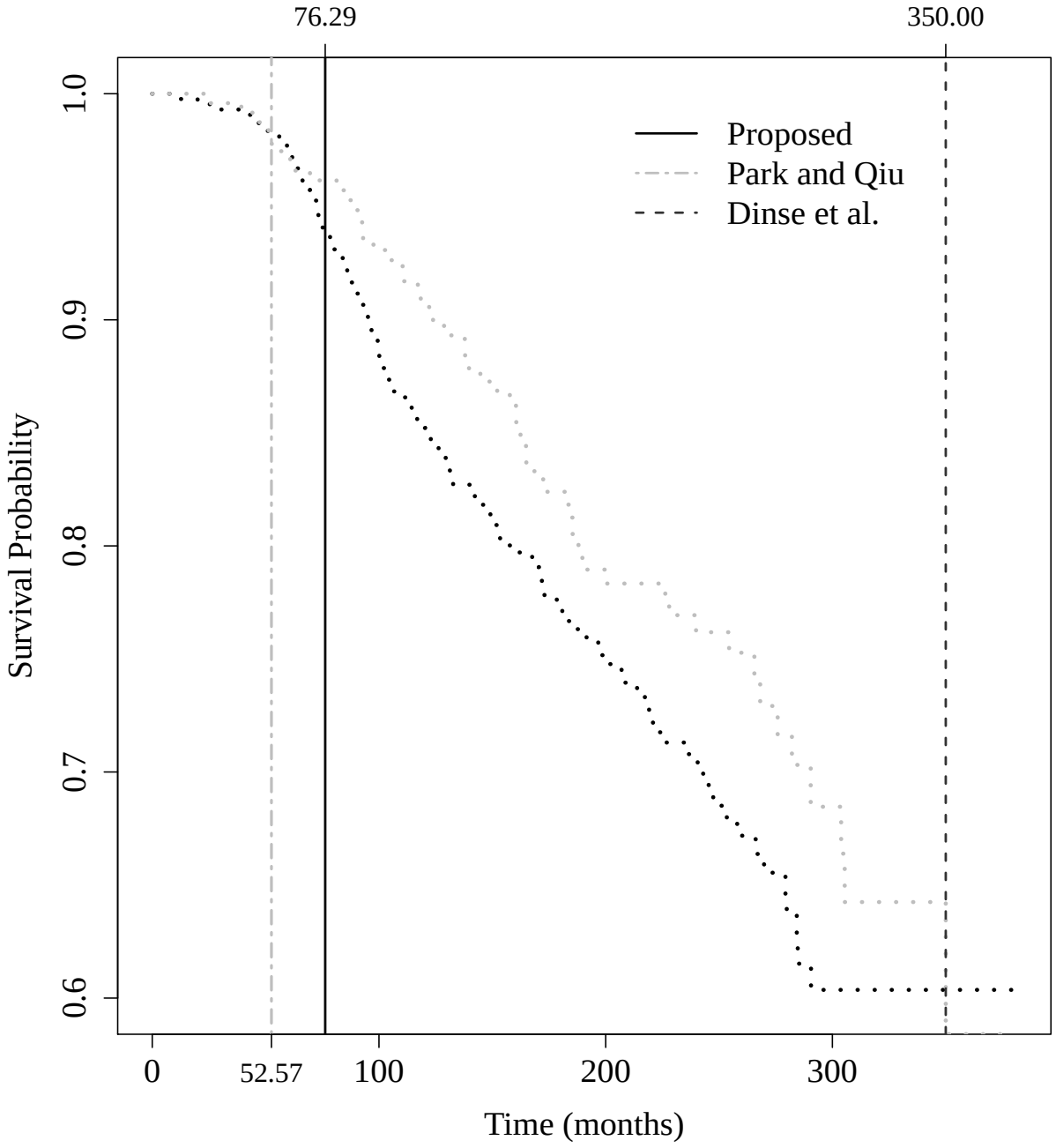

Figure 12: Kaplan-Meier curves for the distribution of time to death for breast cancer; the black dotted line denotes the standard treatment group without hormone replacement therapy and the grey dotted line denotes the hormone therapy group.

the observed data with replacement using the package `bcaboot` (Efron and Narasimhan, 2018).

From 1000 replications, the 95% percentile bootstrap confidence interval for the estimate for τ is $(51, 201)$. When looking at the plot of the survival curves, this makes sense since the sample size is very small, many of the events take place early on in the study and there are several early time points that early on that could be seen as the "true" value of τ.

4.4.2 GERMAN BREAST CANCER DATA

Next, we will apply our method to a German breast cancer data set collected and analyzed by Schumacher et al. (1994). The data was publicly published by Hosmer Jr et al. (2011). The data set has 686 observations and 16 variables considered, including age, a menopause indicator, size of tumor, grade of tumor, and number of progesterone and estrogen receptors that were positive. In our application, we consider the two groups as those patients who received hormone treatment versus those patients who did not. Overall, the data has a 75% censoring rate.

It is of interest here to see when the patients receiving hormone replacement therapy begin to see a treatment lag-effect compared to those who are not receiving the hormone replacement therapy. In the study, there are 440 patients who did not receive hormone therapy and 246 who did. We can see the Kaplan-Meier estimates of the survival curves along with the estimates of the time point of treatment lag-effect using the method proposed in this chapter, the method of Park and Qiu (2018), and the method of Dinse et al. in Figure 12. As in the simulation studies, the proposed method and the method of Park and Qiu (2018) are both reasonable estimates, following the trends of the Park and Qiu (2018) method being smaller and the proposed method being larger. The method of Dinse et al. clearly largely overestimates the value, with the estimate being the final uncensored time point. The method of Park and Qiu (2018) estimates the treatment lag-effect at 52.57 months, the proposed method at 76.29 months, and the Dinse et al. method at 350 months. We see that in the study, the method of Park and Qiu (2018) and the proposed method estimate the treatment lag-effect at approximately 4 years and 6 years, respectively. As noted by the American Cancer Society, hormone replacement therapy is typically administered for 5 to 10 years to help reduce the risk of breast cancer relapse. The treatment lag-effect estimators are in line with this observation.

We again find a confidence interval for the estimator using the BCa bootstrap method. From 1000 replications, the 95% percentile bootstrap confidence interval for the estimate

for τ is $(60.85, 85.86)$. We can see that this interval is narrower than the interval from the Veteran's Administration Lung Cancer Trial data, which is consistent with the simulation study. With an increase in the sample size, we expect that there is a decrease in the length of the confidence interval.

4.5 CONCLUSIONS AND FURTHER RESEARCH

While there has been much research into change point analysis, there has been little research on time-lag effects. Some authors have used change point methods to find the time point of change in one group (Brazzale et al., 2019; Chen and Baron, 2014; Park and Qiu, 2018), but there have been few non-parametric methods considering two groups. It should also be noted that non-parametric and semi-parametric methods that have already been proposed for survival analysis applications do not seem to be readily extendable to either identifying the time point of change between two groups, or to identify the time point of change non-parametrically. Throughout this chapter, we have presented a novel non-parametric method for estimating the time point of treatment lag-effect using change point methods adapted from previous work meant for implementation on multivariate time series proposed by Matteson and James (2014). Some theoretical properties of strong consistency of the proposed estimator are shown.

From the simulation study, we found that our method using the change point tends to give more accurate results than the previously proposed Dinse et al. estimator in several different cases and simulation settings, and gives results that are generally comparable to the method suggested by Park and Qiu (2018). We see that the distribution of the estimator seems to be empirically satisfactory, especially when compared to previously suggested non-parametric methods (primarily the Dinse et al. method). When compared to the methods of Park and Qiu (2018), we see that the proposed estimator performs better in terms of bias and MSE when there are no covariates included, and gives reasonable results in the case where the information from covariates is included. This makes sense since the previous method of Park and Qiu (2018) is semi-parametric and includes covariate information. Additionally, we see that the confidence intervals found using the non-parametric bias-corrected and accelerated method of Efron and Narasimhan (2020) provide a reasonably accurate empirical coverage and average interval length. Finally, we show some results for the suggested permutation test that show that with an increase in observations before and after the time point of treatment lag-effect and an increase in overall sample size, the permutation test gives accurate results.

Using real data sets for Veteran's Administration Lung Cancer Trial taken from Kalbfleisch

and Prentice (2011) and the German Breast Cancer data of Schumacher et al. (1994), we see that the estimator proposed here performs well when compared to the estimator proposed by Dinse et al., and gives results that are consistent with the simulation study in terms of the estimation bias. The 95% percentile confidence interval using BCa bootstrap methods seem to agree with the graphs visually. We see that (as in the simulation studies), the Park and Qiu (2018) method may slightly underestimate the treatment time-lag effect while the method of Dinse et al. clearly overestimates.

Overall, the method we present in this chapter gives results that are accurate both in a simulation setting and with real data application. The empirical results suggest that this non-parametric estimator found from an empirical divergence measure is applicable and appropriate for use to analyze survival data. For future research, it would be of interest to extend the method to multiple change points where the first point of change could be considered the treatment time-lag effect and the second could be the effect of long-term survivors (for example). It could also be possible to adapt the proposed method to be either semi-parametric or parametric by using applicable survival estimates.

CHAPTER 5

SUMMARY AND FUTURE RESEARCH

In this final chapter, we will briefly summarize the research presented in this dissertation and discuss some future research directions.

5.1 SUMMARY

Change point analysis tries to identify times when the probability distribution of a stochastic process or time series changes. In general the problem concerns both detecting whether or not a change has occurred, or whether several changes might have occurred, and identifying the times of any such changes. Data including information on time can be found across many areas of statistics. Censored or truncated data where it may be true that not all information is known for a data point is most commonly found in survival analysis. This type of analysis usually finds the most use in medical or clinical research studies. There has been a variety of research into change point analysis for time-to-event data.

In this dissertation, we have discussed two different methods for change point analysis in censored data. The first method we presented was an accelerated failure time (AFT) model with an underlying Weibull distribution incorporating a scale function that allows for change points in the hazard rate of the event of interest. We use a maximum likelihood estimation technique in order to find point estimates for the change points, the shape parameter of the Weibull, the scale parameters, and any potential covariate effects. We use a sequential testing procedure with a likelihood ratio test and an alpha spending function to test for each subsequent change point. Because of known issues with the distribution of the likelihood ratio test statistic under change point conditions, we use a model-based bootstrapping method in order to find the empirical distribution of the test statistic.

The second method we discuss is a non-parametric method to finding the time point of treatment lag effect. In literature, this has been referred to as the time that it takes any particular treatment to take effect. In this manner, a placebo (or standard of care) treatment group and an experimental treatment group might have the same distribution up until the change point, and then have a differing distribution afterwards. We use an empirical divergence measure based on U-statistics based on the difference in the Kaplan-Meier estimate of survival probability at each event time between the two groups in order

to calculate the empirical divergence. We discuss some asymptotic properties along with the construction of bootstrap confidence intervals and a permutation test.

5.2 FUTURE RESEARCH

There are many opportunities for ongoing research in this area, and there is a need for more flexible and accurate models in medical studies as treatments become more sophisticated.

5.2.1 EXTENSIONS FOR ACCELERATED FAILURE TIME MODEL

For the accelerated failure time (AFT) model, there are several areas in which more research could take place. Most simply, we could use a different underlying distribution (such as the log-logistic distribution, which also has desirable properties), or consider changes in a parameter other than the scale parameter. However, a more in-depth extension might be to use a least squares estimation instead of a maximum likelihood estimation. This is a semi-parametric regression approach for censored data that could be applied to this model for estimation in place of the fully parametric model which requires distributional assumptions. There are also non-parametric maximum likelihood estimation (NPMLE) methods and non-parametric AFT methods that have been proposed.

When considering the testing technique, as was discussed in Appendix B.3, there might be some research available in order to find the theoretical variances of all parameters in order to construct a Wald-type test statistic which has well-researched asymptotic properties and known distribution. Alternatively, there might be some more effective numerical methods to estimate the variance-covariance matrix and find the distribution empirically.

5.2.2 EXTENSIONS FOR NON-PARAMETRIC EMPIRICAL DIVERGENCE MEASURE

There are also many possible extensions to the research into the non-parametric method introduced in this dissertation. The first - and maybe most natural - would be to consider more than one change point, perhaps with a hierarchical testing scheme. A similar method was discussed for multivariate time series by Matteson and James (2014). There are data examples that have been suggested for a proportional hazards method using more than one change point in such a manner (such as in He et al. (2013)). In contrast to the fully parametric AFT model for which it is of interest to avoid assumptions, it might be of interest

to include a more comprehensive estimation to this treatment time-lag effect. Instead of using the non-parametric product limit estimator of survival probability suggested jointly by Kaplan and Meier (1958), we can instead use survival probabilities found from a semi-parametric proportional hazards model as suggested by Cox (1972) or even a fully parametric AFT model.

REFERENCES

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 701–726.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control 19*(6), 716–723.

Ali, Z., M. Hosseini, M. Mahmoodi, K. Mohammad, H. Zeraati, and K. H. Naieni (2015). A comparison between accelerated failure-time and cox proportional hazard models in analyzing the survival of gastric cancer patients. *Iranian journal of public health 44*(8), 1095.

Anderson, T. W. and D. A. Darling (1952). Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. *Annals of Mathematical Statistics 23*(2), 193–212.

Anderson, T. W. and D. A. Darling (1954, December). A Test of Goodness of Fit. *Journal of the American Statistical Association 49*(268), 765–769.

Andrews, D. F. and A. M. Herzberg (2012). *Data: a collection of problems from many fields for the student and research worker*. Springer Science & Business Media.

Austin, P. C. (2012). Generating survival times to simulate cox proportional hazards models with time-varying covariates. *Statistics in medicine 31*(29), 3946–3958.

Berk, R. H. and D. H. Jones (1979, January). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 47*(1), 47–59.

Berkelaar, M. and others (2015). *lpSolve: Interface to 'Lp_solve' v. 5.5 to Solve Linear/Integer Programs*. R package version 5.6.13.

Boyer, M. J., C. D. Williams, D. H. Harpole, M. W. Onaitis, M. J. Kelley, and J. K. Salama (2017). Improved survival of stage i non–small cell lung cancer: a va central cancer registry analysis. *Journal of Thoracic Oncology 12*(12), 1814–1823.

Brawley, O. W. (2012). Trends in prostate cancer in the united states. *Journal of the National Cancer Institute Monographs 2012*(45), 152–156.

Brazzale, A. R., H. Küchenhoff, S. Krügel, T. S. Schiergens, H. Trentzsch, and W. Hartl (2019). Nonparametric change point estimation for survival distributions with a partially constant hazard rate. *Lifetime data analysis 25*(2), 301–321.

Breslow, N. (1970). A generalized kruskal-wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika 57*(3), 579–594.

Buckley, J. and I. James (1979). Linear regression with censored data. *Biometrika 66*(3), 429–436.

Buja, A. and W. Rolke (2006). Calibration for Simultaneity: (Re)Sampling Methods for Simulataneous Inference with Applications to Function Estimation and Functional Data. Working paper.

Byar, D. (1980). The veterans administration study of chemoprophylaxis for recurrent stage i bladder tumours: comparisons of placebo, pyridoxine and topical thiotepa. In *Bladder tumors and other topics in urological oncology*, pp. 363–370. Springer.

Canty, A. and B. D. Ripley (2017). *boot: Bootstrap R (S-Plus) Functions*.

Chamie, K., M. S. Litwin, J. C. Bassett, T. J. Daskivich, J. Lai, J. M. Hanley, B. R. Konety, C. S. Saigal, and U. D. in America Project (2013). Recurrence of high-risk bladder cancer: a population-based analysis. *Cancer 119*(17), 3219–3227.

Chen, X. and M. Baron (2014, October). Change-Point Analysis of Survival Data with Application in Clinical Trials. *Open Journal of Statistics 4*, 663–677.

Chen, Z., H. Huang, and P. Qiu (2016). Comparison of multiple hazard rate functions. *Biometrics 72*(1), 39–45.

Chen, Z., H. Huang, and P. Qiu (2017). An improved two-stage procedure to compare hazard curves. *Journal of Statistical Computation and Simulation 87*(9), 1877–1886.

Cheng, M.-Y., P. Qiu, X. Tan, and D. Tu (2009). Confidence intervals for the first crossing point of two hazard functions. *Lifetime Data Analysis 15*(4), 441.

Cheng, R. C. and T. C. Iles (1983, February). Bands for Cumulative Distribution Functions of Continuous Random Variables. *Technometrics 25*(1), 77–86.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological) 34*(2), 187–202.

Csörgő, S. and J. T. Faraway (1996). The Exact and Asymptotic Distributions of Cramér-von Mises Statistics. *Journal of the Royal Statistical Society. Series B (Methodological) 58*(1), 221–234.

D'Agostino, R. B. and M. A. Stephens (Eds.) (1986). *Goodness-of-Fit Techniques*. Dekker.

Dallas, M. J. and P. Rao (2000). Testing equality of survival functions based on both paired and unpaired censored data. *Biometrics 56*(1), 154–159.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press.

Dehling, H. (2006). Limit theorems for dependent u-statistics. In *Dependence in probability and statistics*, pp. 65–86. Springer.

Denker, M. and G. Keller (1983). On u-statistics and v. mise'statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 64*(4), 505–522.

Dinse, G. E., W. W. Piegorsch, and D. D. Boos (1993). Confidence statements about the time range over which survival curves differ. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 42*(1), 21–30.

Dümbgen, L., P. Kolesnyk, and R. A. Wilke (2017, May). Bi-Log-Concave Distribution Functions. *Journal of Statistical Planning and Inference 184*, 1–17.

Efron, B. and B. Narasimhan (2018). *bcaboot: Bias Corrected Bootstrap Confidence Intervals*. R package version 0.2-1.

Efron, B. and B. Narasimhan (2020). The automatic construction of bootstrap confidence intervals. *Journal of Computational and Graphical Statistics*, 1–12.

Efron, B. and R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC.

Elandt-Johnson, R. C. and N. L. Johnson (1980). *Survival models and data analysis*, Volume 110. John Wiley & Sons.

Erdman, C. and J. W. Emerson (2008). A fast bayesian change point analysis for the segmentation of microarray data. *Bioinformatics 24*(19), 2143–2148.

Faddy, M. J., R. J. Wilson, and G. M. Winter (2003). The Determination of the Design Strength of Granite. In W. R. Blischke and D. N. Prabhakar Murthy (Eds.), *Case Studies in Reliability and Maintenance*, Chapter 5, pp. 111–134. Wiley.

Fallah, F., M. Fallah, and R. S. S. Nia (2012). Thiotepa versus bacille calmette-guérin in non-muscle invasive bladder cancer. *Current urology 6*(3), 160–164.

Fisher, R. (1929). Moments and product moments of sampling distribution. *Proceedings. London Mathematical Society*.

Fleming, T. R., J. R. O'Fallon, P. C. O'Brien, and D. P. Harrington (1980). Modified kolmogorov-smirnov test procedures with application to arbitrarily right-censored data. *Biometrics*, 607–625.

Frey, J. (2008). Optimal distribution-free confidence bands for a distribution function. *Journal of Statistical Planning and Inference 138*, 3086 – 3098.

Gehan, E. A. (1965). A generalized wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika 52*(1/2), 203–223.

Gill, J. and G. King (2003). *Numerical Issues Involved in Inverting Hessian Matrices*, Chapter 6, pp. 143–176. Hoboken, NJ: John Wiley and Sons, Inc.

Gill, R. (1980). *Censoring and stochastic integrals*. Mathematical Centre tracts: Mathematisch Centrum.

Goodman, M. S., Y. Li, and R. C. Tiwari (2011). Detecting multiple change points in piecewise constant hazard functions. *Journal of applied statistics 38*(11), 2523–2532.

Greenwood, M. (1926). The natural duration of cancer. *Reports on public health and medical subjects* (33).

Harrington, D. P. and T. R. Fleming (1982). A class of rank test procedures for censored survival data. *Biometrika 69*(3), 553–566.

He, P., L. Fang, and Z. Su (2013). A sequential testing approach to detecting multiple change points in the proportional hazards model. *Statistics in medicine 32*(7), 1239–1245.

Heller, G. and E. Venkatraman (1996). Resampling procedures to compare two survival distributions in the presence of right-censored data. *Biometrics*, 1204–1213.

Henderson, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika 77*(4), 835–843.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 293–325.

Hoeffding, W. (1961). The Strong Law of Large Numbers for U-statistics. Technical Report 302, North Carolina State University, Dept. of Statistics.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science 14*(4), 382–417.

Holleczek, B. and H. Brenner (2014). Provision of breast cancer care and survival in germany–results from a population-based high resolution study from saarland. *BMC cancer 14*(1), 757.

Hosmer Jr, D. W., S. Lemeshow, and S. May (2011). *Applied survival analysis: regression modeling of time-to-event data*, Volume 618. John Wiley & Sons.

Jager, L. and J. A. Wellner (2004). A new goodness of fit test: the reversed Berk-Jones statistic. Technical report, Dept. Statistics, Univ. Washington.

Jager, L. and J. A. Wellner (2007). Goodness-of-Fit Tests via Phi-divergences. *The Annals of Statistics 35*(5), 2018–2053.

Jemal, A., E. M. Ward, C. J. Johnson, K. A. Cronin, J. Ma, A. B. Ryerson, A. Mariotto, A. J. Lake, R. Wilson, R. L. Sherman, et al. (2017). Annual report to the nation on the status of cancer, 1975–2014, featuring survival. *JNCI: Journal of the National Cancer Institute 109*(9), djx030.

Kalbfleisch, J. D. and R. L. Prentice (2011). *The statistical analysis of failure time data*, Volume 360. John Wiley & Sons.

Kaplan, D. M. and M. Goldman ((Forthcoming)). Comparing distributions by multiple testing across quantiles or CDF values. *Journal of Econometrics*, XXX–XXX.

Kaplan, E. L. and P. Meier (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association 53*(282), 457–481.

Killick, R., I. Eckley, P. Jonathan, and K. Ewans (2010). "detection of changes in the characteristics of oceanographic time-series using statistical change point analysis. *Ocean Engineering 37*(13), 1120–1126.

Kim, A. Y., C. Marzban, D. B. Percival, and W. Stuetzle (2009, December). Using labeled data to evaluate change detectors in a multivariate streaming environment. *Signal Process. 89*(12), 2529–2536.

Klein, J. P. and M. L. Moeschberger (2003). *Survival analysis: techniques for censored and truncated data* (2nd ed.). New York: Springer.

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *G. Ist. Ital. Attuari 4*, 83–91.

Lee, C. Y., X. Chen, and K. F. Lam (2020). Testing for change-point in the covariate effects based on the cox regression model. *Statistics in Medicine 39*(10), 1473–1488.

Liu, K., P. Qiu, and J. Sheng (2007). Comparing two crossing hazard rates by cox proportional hazards modelling. *Statistics in medicine 26*(2), 375–391.

Liu, M., W. Lu, and Y. Shao (2008). A monte carlo approach for change-point detection in the cox proportional hazards model. *Statistics in Medicine 27*(19), 3894–3909.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports 50*(3), 163–170.

Marsaglia, G. and J. C. Marsaglia (2004, February). Evaluating the Anderson-Darling Distribution. *Journal of Statistical Software, Articles 9*(2), 1–5.

Mason, D. M. and J. H. Schuenemeyer (1983). A modified Kolmogorov-Smirnov test sensitive to tail alternatives. *The Annals of Statistics 11*(3), 933–946.

Matteson, D. S. and N. A. James (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association 109*(505), 334–345.

Matthews, D. E. and V. T. Farewell (1982). On testing for a constant hazard against a change-point alternative. *Biometrics 38*(2), 463 – 468.

Muggeo, V. M. R. and G. Adelfio (2011, 1). Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics 27*(2), 161–166.

Nam, C. F., J. A. Aston, and A. M. Johansen (2012). Quantifying the uncertainty in change points. *Journal of Time Series Analysis 33*(5), 807–823.

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software 60*(2), 1–14.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics 14*(4), 945–966.

Nguyen, H., G. Rogers, and E. Walker (1984). Estimation in change-point hazard rate models. *Biometrika 71*(2), 299–304.

Noé, M. (1972). The Calculation of Distributions of Two-sided Kolmogorov Smirnov Type Statistics. *Annals of Mathematical Statistics 43*, 58–64.

Owen, A. B. (1995, June). Nonparametric likelihood confidence bands for a distribution function. *Journal of American Statistical Association 90*(430), 516–521.

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

Park, K. and P. Qiu (2018). Evaluation of the treatment time-lag effect for survival data. *Lifetime data analysis 24*(2), 310–327.

Peto, R. and J. Peto (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General) 135*(2), 185–207.

Qiu, P. and J. Sheng (2008). A two-stage procedure for comparing hazard rate functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(1), 191–208.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of applied meteorology and climatology 46*(6), 900–915.

Rizzo, M. L. (2008). *Statistical Computing with R*. Chapman & Hall /CRC.

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics 1*(3), 239–250.

Rosenkrantz, W. A. (2000, August). Confidence Bands for Quantile Functions: A Parametric and Graphic Alternative for Testing Goodness of Fit. *The American Statistician 54*, 185–190.

Rosenkrantz, W. A. (2013, May). Confidence bands for distribution functions when parameters are estimated from the data: A non-monte-carlo approach. *Biometrical Journal 55*(3), 370–385.

Sauerbrei, W. and P. Royston (1999). Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 162*(1), 71–94.

Schumacher, M., G. Bastert, H. Bojar, K. Huebner, M. Olschewski, W. Sauerbrei, C. Schmoor, C. Beyerle, R. Neumann, and H. Rauschecker (1994). Randomized 2 x 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. german breast cancer study group. *Journal of Clinical Oncology 12*(10), 2086–2093.

Sego, L. H., M. R. Reynolds Jr, and W. H. Woodall (2009). Risk-adjusted monitoring of survival times. *Statistics in medicine 28*(9), 1386–1401.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. Macmillan And Co Ltd, London.

Shorack, G. R. and J. A. Wellner (1986). *Empirical Processes With Applications To Statistics*. Wiley.

Shrestha, A., C. Martin, M. Burton, S. Walters, K. Collins, and L. Wyld (2019). Quality of life versus length of life considerations in cancer patients: a systematic literature review. *Psycho-oncology 28*(7), 1367–1380.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics 19*(2), 279–281.

Stephens, M. A. (1974, September). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association 69*(347), 730 – 737.

Sun, Y. and M. Sherman (1996). Some permutation tests for survival data. *Biometrics*, 87–97.

Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov). SEER*Stat Database: Incidence - SEER 9 Regs Research Data, Nov 2017 Sub (1973-2015). <Katrina/Rita Population Adjustment >- Linked To County Attributes - Total U.S., 1969-2016 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2018, based on the November 2017 submission.

Surveillance Research Program. National Cancer Institute SEER*Stat software (www.seer.cancer.gov/seerstat) version 8.3.6.1.

Székely, G. J., M. L. Rizzo, et al. (2005). Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method. *Journal of classification 22*(2), 151–184.

Talih, M. and N. Hengartner (2005). Structural learning with time-varying components: tracking the cross-section of financial time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(3), 321–341.

Tarone, R. E. and J. Ware (1977). On distribution-free tests for equality of survival distributions. *Biometrika 64*(1), 156–160.

Tewari, A., J. D. Raman, P. Chang, S. Rao, G. Divine, and M. Menon (2006). Long-term survival probability in men with clinically localized prostate cancer treated either conservatively or with definitive treatment (radiotherapy or radical prostatectomy). *Urology 68*(6), 1268–1274.

Therneau, T. M. (2015). *A Package for Survival Analysis in S.* version 2.38.

Tukey, J. W. (1950). Some sampling simplified. *Journal of the American Statistical Association 45*(252), 501–519.

Tutz, G. and M. Schmid (2016). *Modeling Discrete Time-to-Event Data.* Springer International Publishing.

van Valkenhoef, G. and T. Tervonen (2018). *hitandrun: "Hit and Run" and "Shake and Bake" for Sampling Uniformly from Convex Shapes.* R package version 0.5-4.

Wang, J., F. Cheng, and L. Yang (2013). Smooth simultaneous confidence bands for cumulative distribution functions. *Journal of Nonparametric Statistics 25*(2), 395 – 407.

Wei, L.-J. (1992). The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in medicine 11*(14-15), 1871–1879.

Wei, L.-J., D. Y. Lin, and L. Weissfeld (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American statistical association 84*(408), 1065–1073.

Wellner, J. A. and V. Koltchinskii (2003). A note on the asymptotic distribution of Berk-Jones type statistics under the null hypothesis. *High Dimensional Probability III*, 321–332.

Wilks, S. S. (1962). *Mathematical Statistics. Wiley Series in Probability and Mathematical Statistics*. Wiley.

Xu, G., B. Sen, and Z. Ying (2014). Bootstrapping a change-point cox model for survival data. *Electronic journal of statistics 8*(1), 1345.

Xu, X., X. Ding, and S. Zhao (2009a, September). A New Confidence Band for Continuous Cumulative Distribution Functions. *Australian and New Zealand Journal of Statistics 51*(3), 305–318.

Xu, X., X. Ding, and S. Zhao (2009b). The reduction of the average width of confidence bands for an unknown continuous distribution function. *Journal of Statistical Computation and Simulation 79*(4), 335–347.

Yang, S. and R. Prentice (2010, March). Improved Logrank-Type Tests for Survival Data Using Adaptive Weights. *Biometrics 66*(1), 30–38.

Yao, Y.-C. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Communications in Statistics-Theory and Methods 15*(8), 2455–2466.

Zeileis, A., A. Shah, and I. Patnaik (2010). Testing, monitoring, and dating structural changes in exchange rate regimes. *Computational Statistics & Data Analysis 54*(6), 1696–1706.

Zhang, W. (2014). *Detection of Multiple Change-Points in Hazard Models*. Ph. D. thesis, Florida Atlantic University.

Zucker, D. M. and E. Lakatos (1990). Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika 77*(4), 853–864.

APPENDIX A

SELECTED R CODES

All computations in this dissertation are written in R (R Core Team, 2019). We include a selection of R codes used in order to ensure reproducibility. The following code is for the method described in Chapter 3.

```
library(survival)
library(optimx)

ind <- function(cond) {ifelse(cond, 1L, 0L)}

x.of <- function(t, time, delta){
  delta*ind(time <= t)
}

aft.profile.tau.alpha <- function(theta, time, delta, Z, p.tau = NULL){
  beta <- tail(theta, 2)
  temp.theta <- head(theta, -2)
  alpha <- temp.theta[1]
  tau.temp <- sort(c(temp.theta[2], p.tau))
  tau <- c(0, tau.temp, max(time)+1)
  llik <- c()
  for(i in 1:(length(tau)-1)){
    llik[i] <- sum((x.of(tau[i+1], time, delta) - x.of(tau[i], time, delta
       ))*(log(alpha*(time^(alpha-1))*exp(Z%*%beta)) +
               log(sum(x.of(tau[i+1], time, delta) - x.of(tau[i], time,
                  delta))/sum(exp(Z%*%beta)
              *(pmin(time, tau[i+1])^alpha - tau[i]^alpha)*ind(time > tau[
                 i])))))
  }
  -sum(llik)
}

lambda.est <- function(theta, time, delta, Z, beta){
  alpha <- theta[1]
  if(theta[2] == 0){
    tau <- c(0, max(time)+1)
```

```
29    } else {
30      tau <- c(0,theta[-1],max(time)+1)
31    }
32    lambda <- c()
33    for(i in 1:(length(tau)-1)){
34      lambda[i] <- sum(x.of(tau[i+1], time, delta) - x.of(tau[i], time,
          delta))/
35        sum(exp(Z%*%beta)*(pmin(time, tau[i+1])^alpha - tau[i]^alpha)*ind(
            time > tau[i]))
36    }
37    lambda}
38
39  twocp.simulations <- function(theta, n){
40
41    null.model <- survreg(Surv(time, delta)~z1 + z2, data = dat)
42    alpha.hat <- 1/null.model$scale
43    lambda.hat <- as.numeric(exp(-null.model$coef[1] / null.model$scale))
44    beta.hat <- -null.model$coef[-1]
45
46    profiletime <- unique(time)[order(unique(time))]
47    result <- c()
48    for(j in profiletime){
49          out <- aft.profile.tau.alpha(theta = c(alpha.hat,j, beta.hat),
              time = time, delta = delta, Z = Z, p.tau = NULL)
50          result <- rbind(result, c(j, out))
51        }
52    colnames(result) <- c("tau.hat", "loglik")
53    result <- as.data.frame(result)
54    tau.hat <- result[which.min(result$loglik),1]
55
56    meth0 <- c("Nelder-Mead")
57    aft.opt <- opm(par = c(alpha.hat, tau.hat, beta.hat), fn = aft.profile.
        tau.alpha,
58                   time = time, delta = delta, Z = Z, hessian = T, method =
                       meth0, control = list(dowarn = F))
59    aft.mle <- as.numeric(aft.opt[,1:4])
60
61    result2 <- c()
62    for(i in profiletime[profiletime < (aft.mle[2] - 1) | profiletime > (aft
        .mle[2] + 1)]){
```

```
    out2 <- aft.profile.tau.alpha(theta = c(aft.mle[1], i, aft.mle[3:4]),
       time = time, delta = delta, Z = Z, p.tau = aft.mle[2])
    result2 <- rbind(result2, c(i,out2))
    colnames(result2) <- c("tau2.hat", "loglik")
    result2 <- as.data.frame(result2)
  }

  aft.opt2 <- opm(par = c(aft.mle[1], tau2.hat, aft.mle[3:4]), fn = aft.
     profile.tau.alpha, time = time, delta = delta, hessian = T,
                  Z = Z, method = meth0, p.tau = aft.mle[2])

  aft.mle2 <- as.numeric(aft.opt2[,1:4])
  aft.mle2 <- c(as.numeric(aft.opt2[,1]), sort(c(as.numeric(aft.opt2[,2]),
      aft.mle[2])), as.numeric(aft.opt2[,3:4]))
  aft.lambda2 <- lambda.est(aft.mle2[-(4:5)], time, delta, Z, aft.mle2
     [4:5]) #calculate lambda values

  output <- c(perc.cens, aft.mle2[c(1,4,5)],aft.lambda2,aft.mle2[c(2:3)])
  output

}
```

Next, we include some code for the method described in Chapter 4.

```
library(survival)
library(tidyverse)

survdat <- survdat[order(survdat$time),]
survdat <- cbind(1:length(survdat$time), survdat)

df <- spread(survdat, group, status) %>% dplyr::rename(Group0 = "0",
   Group1 = "1")
kmtable <- df[,2:4] %>%  mutate(n.c0 = if_else(is.na(Group0) == TRUE |
   Group0 == 1, 0, 1),
                                n.c1 = if_else(is.na(Group1) == TRUE |
                                   Group1 == 1, 0, 1),
                                d0 = if_else(is.na(Group0) == TRUE |
                                   Group0 == 0, 0, 1),
                                d1 = if_else(is.na(Group1) == TRUE |
                                   Group1 == 0, 0, 1)) %>%
    dplyr::select(1, 4:7) %>%
    ddply("time", numcolwise(sum)) %>%
```

```
    mutate(d = d0 + d1,
           y0 = n - cumsum(lag(n.c0, 1, default = 0) + lag(d0,1,default = 0)),
           y1 = n - cumsum(lag(n.c1, 1, default = 0) + lag(d1,1,default = 0)),
           y = y0 + y1,
           surv0 = if_else(is.na(cumprod(1 - d0/y0)) == T, 0, cumprod(1 - d0/y0)),
           surv1 = ifelse(is.na(cumprod(1 - d1/y1)) == T, 0, cumprod(1 - d1/y1))) %>%
    filter(d >= 1) %>%
    dplyr::select(1, 4:11)

new <- subset(kmtable, kmtable$surv0 > 0 & kmtable$surv1 > 0)
Z <- new$surv0 - new$surv1
difference <- outer(Z, Z, function(x,y) abs(x - y))

parts <- list()
  for(i in 1:(ncol(difference)-1)){
    parts[[i]] <- difference[1:i, (i+1):ncol(difference)]
  }
first <- unlist(lapply(parts, sum))

parts2 <- list()
for(i in 1:(ncol(difference)-1)){
    parts2[[i]] <- difference[1:(i-1), 2:i]
    parts2[[i]][lower.tri(parts2[[i]])] <- 0
    parts2[[1]] <- 0
  }
second <- unlist(lapply(parts2, sum))

parts3 <- list()
for(i in 1:(ncol(difference)-2)){
    parts3[[i]] <- difference[(i+1):(ncol(difference)-1),(i+2):ncol(
       difference)]
    parts3[[i]][lower.tri(parts3[[i]])] <- 0
    parts3[[(ncol(difference)-1)]] <- 0
  }
third <- unlist(lapply(parts3, sum))

N <- 1:(length(Z) - 1)
```

```
50 M <- (length(Z) - 1):1
51
52 e.hat <- ((2/(N*M))*first) - (choose(N,2)^(-1) * second) - (choose(M,2)
       ^(-1) * third)
53 q.hat <- ((M*N) / (M+N))*e.hat
54 tau.hat <- kmtable$time[which.max(q.hat)]
```

APPENDIX B

ACCELERATED FAILURE TIME DERIVATIONS

B.1 AFT CUMULATIVE HAZARD

B.1.1 DERIVATION OF CUMULATIVE HAZARD

We can derive the cumulative hazard function of a Weibull AFT model with multiple change points in the scale parameter. For the change point model, we have the hazard function

$$h(t_i;\boldsymbol{\zeta_i}) = \begin{cases} \nu\lambda_1 t_i^{\nu-1}\exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) & \text{if } 0 < t_i \leq \tau_1 \\ \nu\lambda_2 t_i^{\nu-1}\exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) & \text{if } \tau_1 < t_i \leq \tau_2 \\ \quad\vdots & \\ \nu\lambda_{k+1} t_i^{\nu-1}\exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) & \text{if } t_i > \tau_k \end{cases}$$

This can be simplified using indicator functions, $I(\cdot)$, such that

$$h(t_i;\boldsymbol{\zeta_i}) = \nu\lambda_j t_i^{\nu-1}\exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})I(\tau_{j-1} < t_i \leq \tau_j) = \left[\nu\lambda_j t_i^{\nu-1}\exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\right]\left[I(t_i \leq \tau_j) - I(t_i \leq \tau_{j-1})\right].$$

To see why this is true, define the set $Y = \{t_i\}$. Then, define subset $A = \{t_i | t_i \leq \tau_j\}$ and subset $B = \{t_i | t_i > \tau_{j-1}\}$. Note that $I(\tau_{j-1} < t_i \leq \tau_j) = I(A \cap B)$. So, we can write the indicator function as

$$\begin{aligned} I(\tau_{j-1} < t_i \leq \tau_{j-1}) &= I(A \cap B) = \min\{I(A), I(B)\} = I(A)I(B) \\ &= I(A)\left[1 - I(B^C)\right] \\ &= I(A) - I(A)I(B^C) \\ &= I(A) - \min\left\{I(A), I(B^C)\right\} \\ &= I(A) - I(A \cap B^C) \\ &= I(t_i \leq \tau_j) - I(t_i \leq \tau_{j-1}). \end{aligned}$$

For $t_i \leq \tau_1$ the cumulative hazard rate can be found as follows:

$$\begin{aligned} H(t_i;\boldsymbol{\zeta_i}) &= \int_0^{t_i} h(u;\boldsymbol{\zeta_i})du \\ &= \lambda_1 \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \int_0^{t_i} \nu u^{\nu-1} du \\ &= \lambda_1 \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) t_i^{\nu} \end{aligned}$$

Then, for $\tau_1 < t_i \leq \tau_2$:

$$\begin{aligned} H(t_i;\boldsymbol{\zeta_i}) &= \int_0^{t_i} h(u;\boldsymbol{\zeta_i})du \\ &= \int_0^{\tau_1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\lambda_1 \nu u^{\nu-1} du + \int_{\tau_1}^{t_i} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\lambda_2 \nu u^{\nu-1} du \\ &= \lambda_1 \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\tau_1^{\nu} + \lambda_2 \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})(t_i^{\nu} - \tau_1^{\nu}) \end{aligned}$$

and so on until $t_i > \tau_k$:

$$\begin{aligned} H(t_i;\boldsymbol{\zeta_i}) &= \int_0^{t_i} h(u;\boldsymbol{\zeta_i})du \\ &= \int_0^{\tau_1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\lambda_1 \nu u^{\nu-1} du + \cdots + \int_{\tau_k}^{t_i} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\lambda_{k+1} \nu u^{\nu-1} \\ &= \lambda_1 \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\tau_1^{\nu} + \cdots + \lambda_{k+1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})(t_i^{\nu} - \tau_k^{\nu}). \end{aligned}$$

More generally, we can express the cumulative hazard function as

$$H(t_i;\boldsymbol{\zeta_i}) = \begin{cases} \lambda_1 \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) t_i^{\nu} & t_i \leq \tau_1 \\ \lambda_1 \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\tau_1^{\nu} + \lambda_2 \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})(t_i^{\nu} - \tau_1^{\nu}) & \tau_1 < t_i \leq \tau_2 \\ \quad\vdots & \\ \lambda_1 \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})\tau_1^{\nu} + \cdots + \lambda_{k+1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})(t_i^{\nu} - \tau_k^{\nu}) & t_i > \tau_k \end{cases}$$

which can be rewritten

$$H(t_i;\boldsymbol{\zeta_i}) = \begin{cases} \lambda_1 \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_1)^{\nu} - (t_i \wedge \tau_0)^{\nu}\right] & t_i \leq \tau_1 \\ \sum_{j=1}^{2} \lambda_j \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^{\nu} - (t_i \wedge \tau_{j-1})^{\nu}\right] & \tau_1 < t_i \leq \tau_2 \\ \quad\vdots & \\ \sum_{j=1}^{k+1} \lambda_j \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^{\nu} - (t_i \wedge \tau_{j-1})^{\nu}\right] & t_i > \tau_k \end{cases}$$

This can more simply be expressed as $H(t_i;\boldsymbol{\zeta_i}) = \sum_{j=1}^{k+1} \lambda_j \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^{\nu} - (t_i \wedge \tau_{j-1})^{\nu}\right]$ since for $j > 1$, if $t_i < \tau_{j-1}$ then $(t_i \wedge \tau_j)^{\nu} - (t_i \wedge \tau_{j-1})^{\nu} = 0$. Here, $(t_i \wedge \tau_j) = \min(t_i, \tau_j)$. From here, it is easy to find the survival function since $S(t_i, \boldsymbol{\zeta_i}) = \exp\left[-H(t_i;\boldsymbol{\zeta_i})\right]$.

B.1.2 INVERSE TRANSFORM FOR SIMULATION

We generate from a Uniform(0,1) distribution and then use the inverse cumulative hazard in order to simulate the time points using the inverse transform method. We will follow the general simulation method suggested by Austin (2012). To simplify, we can work out the inverse cumulative hazard rate for the case with one change point, but this can be generalized to k change points.

In the case of one change point, the cumulative hazard rate can be expressed as

$$H(t_i;\boldsymbol{\zeta_i}) = \begin{cases} \lambda_1 \exp(\boldsymbol{\beta'\zeta_i}) t_i^{\nu} & \text{if } t_i \leq \tau_1 \\ (\lambda_1 - \lambda_2)\exp(\boldsymbol{\beta'\zeta_i})\tau_1^{\nu} + \lambda_2 \exp(\boldsymbol{\beta'\zeta_i}) t_i^{\nu} & \text{if } t_i > \tau_1. \end{cases}$$

We can partition the domain (and corresponding range) into two intervals, and so find the inverse of the hazard within each of these. In the case of multiple change points, we would partition the domain (and range) into $k+1$ intervals and find the inverse cumulative hazard rate for each interval. The inverse of the cumulative hazard function when $H(t_i;\boldsymbol{\zeta_i}) \leq \lambda_1 \exp(\boldsymbol{\beta'\zeta_i})\tau_1^{\nu}$ is given by

$$H^{-1}(t_i;\boldsymbol{\zeta_i}) = \left(\frac{t_i}{\lambda_1 \exp(\boldsymbol{\beta'\zeta_i})}\right)^{1/\nu} \quad \text{if } t_i \leq \lambda_1 \exp(\boldsymbol{\beta'\zeta_i})\tau_1^{\nu}.$$

Similarly, we can find the inverse when $H(t_i;\boldsymbol{\zeta_i}) > \lambda_1 \exp(\boldsymbol{\beta'\zeta_i})\tau_1^{\nu}$:

$$H^{-1}(t_i;\boldsymbol{\zeta_i}) = \left(\frac{t_i - (\lambda_1 - \lambda_2)\exp(\boldsymbol{\beta'\zeta_i})\tau_1^{\nu}}{\lambda_2 \exp(\boldsymbol{\beta'\zeta_i})}\right)^{1/\nu} \quad \text{if } t_i > \lambda_1 \exp(\boldsymbol{\beta'\zeta_i})\tau_1^{\nu}.$$

Therefore, we can simulate survival time as

$$T_i = \begin{cases} \left(\frac{-\log(u_i)}{\lambda_1 \exp(\boldsymbol{\beta'\zeta_i})}\right)^{1/\nu} & -\log(u_i) \leq \lambda_1 \exp(\boldsymbol{\beta'\zeta_i})\tau_1^{\nu} \\ \left(\frac{-\log(u_i) - (\lambda_1-\lambda_2)\exp(\boldsymbol{\beta'\zeta_i})\tau_1^{\nu}}{\lambda_2 \exp(\boldsymbol{\beta'\zeta_i})}\right)^{1/\nu} & -\log(u_i) > \lambda_1 \exp(\boldsymbol{\beta'\zeta_i})\tau_1^{\nu} \end{cases}$$

where $u_i \sim \text{Uniform}(0,1)$ as described in the introduction of the method. The extension of this work to multiple change points follows easily.

B.2 MAXIMUM LIKELIHOOD ESTIMATION DETAILS

Because $h(t_i;\boldsymbol{\zeta_i}) = f(t_i;\boldsymbol{\zeta_i})/S(t_i;\boldsymbol{\zeta_i})$, we can find the likelihood function of the Weibull

AFT change point model with $\boldsymbol{\theta} = (\boldsymbol{\beta}, \nu, \lambda_1, \ldots, \lambda_{k+1}, \tau_1, \ldots, \tau_k)$:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{N} \left[f(t_i; \boldsymbol{\zeta_i})\right]^{\delta_i} \left[S(t_i; \boldsymbol{\zeta_i})\right]^{1-\delta_i} \\ &= \prod_{i=1}^{N} \left[h(t_i; \boldsymbol{\zeta_i}) S(t_i; \boldsymbol{\zeta_i})\right]^{\delta_i} \left[S(t_i; \boldsymbol{\zeta_i})\right]^{1-\delta_i} \\ &= \prod_{i=1}^{N} \left[h(t_i; \boldsymbol{\zeta_i})\right]^{\delta_i} S(t_i; \boldsymbol{\zeta_i}). \end{aligned}$$

The log-likelihood function $l(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta}))$ is

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{i=1}^{N} \delta_i \log\left[h(t_i; \boldsymbol{\zeta_i})\right] + \sum_{i=1}^{N} \log\left[S(t_i; \boldsymbol{\zeta_i})\right] \\ &= \sum_{i=1}^{N} \sum_{j=1}^{k+1} \Big\{ \left[X(\tau_j) - X(\tau_{j-1})\right] \log\left[\nu \lambda_j t_i^{\nu-1} \exp(\boldsymbol{\beta}' \boldsymbol{\zeta_i})\right] \\ &\qquad\qquad -\lambda_j \exp(\boldsymbol{\beta}' \boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^{\nu} - (t_i \wedge \tau_{j-1})^{\nu}\right] \Big\} \end{aligned}$$

where $X(t) = \delta_i I(t_i \leq t)$, and $\sum_{i=1}^{N} X(t)$ can be interpreted as the number of deaths up until time t. If we fix $\tau_j, j = 1, \ldots, k$, the estimate $\widehat{\lambda}_j$ for $\lambda_j, j = 1, \ldots, k+1$ maximizes $l(\boldsymbol{\theta}; \tau_j, \ldots, \tau_k)$:

$$\sum_{i=1}^{N} \left\{ \left[X(\tau_j) - X(\tau_{j-1})\right] \log\left[\nu \lambda_j t_i^{\nu-1} \exp(\boldsymbol{\beta}' \boldsymbol{\zeta_i})\right] - \lambda_j \exp(\boldsymbol{\beta}' \boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^{\nu} - \tau_{j-1}^{\nu}\right] I(t_i > \tau_{j-1}) \right\}.$$

The score functions for λ_j are found by

$$\frac{\partial l}{\partial \lambda_j} = \sum_{i=1}^{N} \left\{ \left[X(\tau_j) - X(\tau_{j-1})\right] \frac{1}{\lambda_j} - \exp(\boldsymbol{\beta}' \boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^{\nu} - \tau_{j-1}^{\nu}\right] I(t_i > \tau_{j-1}) \right\}.$$

Setting the score function equal to zero, we find $\widehat{\lambda}_j$:

$$\begin{aligned} & \sum_{i=1}^{N} \left\{ \left[X(\tau_j) - X(\tau_{j-1})\right] \frac{1}{\widehat{\lambda}_j} - \exp(\boldsymbol{\beta}' \boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^{\nu} - \tau_{j-1}^{\nu}\right] I(t_i > \tau_{j-1}) \right\} = 0 \\ \Longrightarrow & \sum_{i=1}^{N} \left[X(\tau_j) - X(\tau_{j-1})\right] \frac{1}{\widehat{\lambda}_j} = \sum_{i=1}^{N} \exp(\boldsymbol{\beta}' \boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^{\nu} - \tau_{j-1}^{\nu}\right] I(t_i > \tau_{j-1}) \\ \Longrightarrow & \frac{1}{\widehat{\lambda}_j} = \frac{\sum_{i=1}^{N} \exp(\boldsymbol{\beta}' \boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^{\nu} - \tau_{j-1}^{\nu}\right] I(t_i > \tau_{j-1})}{\sum_{i=1}^{N} \left[X(\tau_j) - X(\tau_{j-1})\right]} \\ \Longrightarrow & \widehat{\lambda}_j = \frac{\sum_{i=1}^{N} \left[X(\tau_j) - X(\tau_{j-1})\right]}{\sum_{i=1}^{N} \exp(\boldsymbol{\beta}' \boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^{\nu} - \tau_{j-1}^{\nu}\right] I(t_i > \tau_{j-1})}. \end{aligned}$$

If we substitute this in to $l(\boldsymbol{\theta})$ we can find the profile likelihood of ν and τ_j:

$$\begin{aligned} l(\boldsymbol{\theta}) = & \sum_i \sum_j \Big\{ [X(\tau_j) - X(\tau_{j-1})] \left[\log(\nu t_i^{\nu-1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})) + \log\left(\widehat{\lambda}_j\right) \right] \\ & - \widehat{\lambda}_j \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - (t_i \wedge \tau_{j-1})^\nu \right] \Big\} \\ = & \sum_i \sum_j [X(\tau_j) - X(\tau_{j-1})] \left[\log(\nu t_i^{\nu-1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})) + \log\left(\widehat{\lambda}_j\right) \right] \\ & - \sum_i \sum_j \widehat{\lambda}_j \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - (t_i \wedge \tau_{j-1})^\nu \right]. \end{aligned}$$

If we consider the second term, we see that

$$\begin{aligned} & \sum_i \sum_j \left(\frac{\sum_i [X(\tau_j) - X(\tau_{j-1})]}{\sum_i \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - \tau_{j-1}^\nu \right] I(t_i > \tau_{j-1})} \right) \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - (t_i \wedge \tau_{j-1})^\nu \right] \\ & \implies \sum_j \left(\frac{\sum_i \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - (t_i \wedge \tau_{j-1})^\nu \right]}{\sum_i \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - \tau_{j-1}^\nu \right] I(t_i > \tau_{j-1})} \right) \sum_i [X(\tau_j) - X(\tau_{j-1})] \\ & \implies \sum_j \left(\frac{\sum_i \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - \tau_{j-1}^\nu \right] I(t_i > \tau_{j-1})}{\sum_i \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - \tau_{j-1}^\nu \right] I(t_i > \tau_{j-1})} \right) \sum_i [X(\tau_j) - X(\tau_{j-1})] \\ & \implies \sum_j \sum_i [X(\tau_j) - X(\tau_{j-1})] = \sum_i [X(\tau_{k+1}) - X(\tau_0)] \because \text{ this is a telescoping sum.} \end{aligned}$$

It is clear that $\sum_i X(\tau_0) = 0$, and $\sum_i X(\tau_{k+1})$ is the total number of deaths. Then, finally

$$\begin{aligned} & \sum_i \sum_j \left(\frac{\sum_i [X(\tau_j) - X(\tau_{j-1})]}{\sum_i \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - \tau_{j-1}^\nu \right] I(t_i > \tau_{j-1})} \right) \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - (t_i \wedge \tau_{j-1})^\nu \right] \\ & = \sum_i X(\tau_{k+1}). \end{aligned}$$

Since we are trying to maximize and $\sum_i X(\tau_{k+1})$ is a constant, we do not need to include this term. The location of the maximums $\tau_j, j = 1, \ldots, k$ will be the same - the values of the likelihood function will simply be shifted by $\sum_i X(\tau_{k+1})$. The final profile likelihood for τ_j and ν is

$$\begin{aligned} l(\boldsymbol{\theta}) = \sum_i \sum_j [X(\tau_j) - X(\tau_{j-1})] \Bigg\{ & \log(\nu t_i^{\nu-1} \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i})) \\ & + \log \left[\frac{\sum_i [X(\tau_j) - X(\tau_{j-1})]}{\sum_i \exp(\boldsymbol{\beta}'\boldsymbol{\zeta_i}) \left[(t_i \wedge \tau_j)^\nu - \tau_{j-1}^\nu \right] I(t_i > \tau_{j-1})} \right] \Bigg\}. \end{aligned}$$

B.3 WALD TEST VARIANCE

We use the single change point model without covariates for simplicity. Then, the estimates for $\lambda_j, j = 1, 2$ are

$$\widehat{\lambda}_j = \frac{\sum_{i=1}^{N} [X(\widehat{\tau}_j) - X(\widehat{\tau}_{j-1})]}{\sum_{i=1}^{N} \left[(t_i \wedge \widehat{\tau}_j)^{\widehat{\nu}} - \widehat{\tau}_{j-1}^{\widehat{\nu}}\right] I(t_i > \widehat{\tau}_{j-1})}.$$

Let $\psi_1 = \nu$ and $\psi_2 = \tau$ be the parameters of interest, and $g(\cdot)$ be a continuous function. Then, let $\widehat{\psi}_1$ and $\widehat{\psi}_2$ be the maximum likelihood estimates. For large samples with some regularity conditions met (including some constraints leaving out the largest two observations as in Yao (1986)), then $(\widehat{\psi}_1, \widehat{\psi}_2)$ has a bivariate normal distribution with mean (ψ_1, ψ_2) and variance-covariance matrix estimated by the observed Fisher information. Let $\theta_1 = \lambda_1$ and $\theta_2 = \lambda_2$. We can see that $\theta_1 = g_1(\psi_1, \psi_2)$ and $\theta_2 = g_2(\psi_1, \psi_2)$. For $h, l = 1, 2$, let $g_h^l = \frac{\partial g_h(\widehat{\psi}_1, \widehat{\psi}_2)}{\partial \widehat{\psi}_l}$. Then, we have the partial derivatives

$$g_1^1 = \frac{\partial g_1(\widehat{\psi}_1, \widehat{\psi}_2)}{\partial \widehat{\psi}_1} \qquad g_2^1 = \frac{\partial g_2(\widehat{\psi}_1, \widehat{\psi}_2)}{\partial \widehat{\psi}_1}$$
$$g_1^2 = \frac{\partial g_1(\widehat{\psi}_1, \widehat{\psi}_2)}{\partial \widehat{\psi}_2} \qquad g_2^2 = \frac{\partial g_2(\widehat{\psi}_1, \widehat{\psi}_2)}{\partial \widehat{\psi}_2}$$

Let $\Delta(x - a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ip(x-a)} dp$ be the Dirac delta function and then

$$g_1^1 = \frac{-\left\{\sum_{i=1}^{N} X(\widehat{\tau})\right\} \left\{\sum_{i=1}^{N} (t_i \wedge \widehat{\tau})^{\widehat{\nu}} \log(t_i \wedge \widehat{\tau})\right\}}{\left\{\sum_{i=1}^{N} (t_i \wedge \widehat{\tau})^{\widehat{\nu}}\right\}^2}$$

$$g_1^2 = \frac{\left\{\sum_{i=1}^{N} \delta_i \Delta(\widehat{\tau} - t_i)\right\} \left\{\sum_{i=1}^{N} (t_i \wedge \widehat{\tau})^{\widehat{\nu}}\right\} - \left\{\sum_{i=1}^{N} X(\widehat{\tau})\right\} \left\{\sum_{i=1}^{N} I(t_i > \widehat{\tau})\right\}}{\left\{\sum_{i=1}^{N} (t_i \wedge \widehat{\tau})^{\widehat{\nu}}\right\}^2}$$

$$g_2^1 = \frac{-\left\{\sum_{i=1}^{N} [\delta_i - X(\widehat{\tau})]\right\} \left\{\sum_{i=1}^{N} \left[t_i^{\widehat{\nu}} \log(t_i) - \widehat{\tau}^{\widehat{\nu}} \log(\widehat{\tau})\right] I(t_i > \widehat{\tau})\right\}}{\left\{\sum_{i=1}^{N} \left[t_i^{\widehat{\nu}} - \widehat{\tau}^{\widehat{\nu}}\right] I(t_i > \widehat{\tau})\right\}^2}$$

$$g_2^2 = \frac{-\left\{\sum_{i=1}^{N} \delta_i \Delta(\widehat{\tau} - t_i)\right\} \left\{\sum_{i=1}^{N} \left[t_i^{\widehat{\nu}} - \widehat{\tau}^{\widehat{\nu}}\right] I(t_i > \widehat{\tau})\right\}}{\left\{\sum_{i=1}^{N} \left[t_i^{\widehat{\nu}} - \widehat{\tau}^{\widehat{\nu}}\right] I(t_i > \widehat{\tau})\right\}^2}$$
$$+ \frac{\left\{\sum_{i=1}^{N} [\delta_i - X(\widehat{\tau})]\right\} \left\{\sum_{i=1}^{N} \left[\delta_i \left(t_i^{\widehat{\nu}} - \widehat{\tau}^{\widehat{\nu}}\right) \Delta(\widehat{\tau} - t_i) + \widehat{\nu} \widehat{\tau}^{\widehat{\nu}-1} I(t_i > \widehat{\tau})\right]\right\}}{\left\{\sum_{i=1}^{N} \left[t_i^{\widehat{\nu}} - \widehat{\tau}^{\widehat{\nu}}\right] I(t_i > \widehat{\tau})\right\}^2}$$

For large samples, $Cov\left[g_l(\widehat{\psi_1},\widehat{\psi_2}), g_m(\widehat{\psi_1},\widehat{\psi_2})\right] = g_l^1 g_m^1 Var(\widehat{\psi_1}) + (g_l^1 g_m^2 + g_l^2 g_m^1) Cov(\widehat{\psi_1},\widehat{\psi_2}) + g_l^2 g_m^2 Var(\widehat{\psi_2}), l, m = 1, 2$. If $l = m = 1$, this is clearly the variance of $\widehat{\lambda}_1$ and similarly if $l = m = 2$ for $\widehat{\lambda}_2$. If $l \neq m$, it is the covariance term.

We see that even if all assumptions and regularity conditions are met, finding the distribution of the scale parameters based on the Delta method still has some difficulties. This is potential for ongoing future research.

B.4 ADDITIONAL SIMULATION RESULTS

In this section, we show some results from a simulation study of a smaller sample size with some settings that are more similar to the data from the bladder recurrence example. We simulate the data from a model with two change points and one binary covariate. The model is then:

$$\begin{aligned} h(t_i;\boldsymbol{\zeta_i}) = {} & \nu\lambda_1 \exp(\beta\zeta_i) t_i^{\nu-1} I(0 < t_i \leq \tau_1) + \nu\lambda_2 \exp(\beta\zeta_i) t_i^{\nu-1} I(\tau_1 < t_i \leq \tau_2) \\ & + \nu\lambda_3 \exp(\beta\zeta_i) t_i^{\nu-1} I(t > \tau_2). \end{aligned} \tag{B.1}$$

The results from Table 12 show point estimates from 1000 replicated samples, each of size 100. We see that compared to the results with a larger sample size shown in Section 3.5, the point estimate results are still quite accurate. As expected, the bias tends to be a bit increased than at the larger sample sizes but overall the estimates are still acceptable. Additionally, we see that the MSE is also increased compared to the larger sample sizes, which is also to as is expected. However, the increases in MSE are not so large that the results would be considered unreliable. There were convergence issues for the piecewise constant hazard method at all censoring rates reported, and the simulated trials that had this issue were removed from the calculations for comparison purposes.

Overall, throughout our extensive simulation study for this method we found that the smaller sample sizes tend to follow this trend and be less accurate with higher MSE. If the sample size becomes too small (for example a sample size of 50), the results could be quite unreliable and the proposed method would not be preferred for a data set of this size.

B.5 SEER LOCALIZED/REGIONAL PROSTATE CANCER CASE ANALYSIS

Here, we analyze the localized/regional cases. There were 262,211 men with 61.22% censoring. The average age at diagnosis was 66.862 years and there were 34,588 (13.191%) "Black," 212,413 (81.008%) "White," and 15,210 (5.801%) were "Other." Using the same

selection procedure of Akaike (1974), we find the model to be the same as shown in (3.8). We find two significant change points at 11 and 160 with $LR_0 = 70096.37$ and $LR_1 = 1190.307$.

Table 12: Averaged point estimates (MEAN) and MSE values of the estimated parameters of the proposed model and piecewise constant hazard model based on 1000 replicated simulations for B.1.

Censoring Rate	Parameter	Parameter Value	Constant Hazard Model		Proposed Model	
			MEAN	MSE	MEAN	MSE
0%	ν	1.7500	–	–	1.8584	0.0518
	β	-0.5000	–	–	-0.5061	0.0029
	λ_1	0.2500	0.1462	0.0150	0.2396	0.0088
	λ_2	0.1000	0.5121	0.1824	0.0983	0.0381
	λ_3	0.0500	0.6229	6.6981	0.0690	0.2112
	τ_1	3.0000	1.0242	4.2488	2.9631	1.5683
	τ_2	6.0000	5.8975	6.8057	5.5675	0.0450
20%	ν	1.7500	–	–	1.831	0.0518
	β	-0.5000	–	–	-0.5083	0.0522
	λ_1	0.1000	0.0878	0.0016	0.0952	0.0009
	λ_2	0.0500	0.3273	0.0840	0.0739	0.0159
	λ_3	0.0100	0.2072	0.1103	0.0400	0.0282
	τ_1	6.0000	1.7474	18.6374	5.7002	1.0190
	τ_2	10.0000	7.5617	15.3814	8.9669	3.4490
40%	ν	1.7500	–	–	1.9536	0.2916
	β	-0.4500	–	–	-0.4594	0.0001
	λ_1	0.0100	0.0234	0.0003	0.0098	0.0009
	λ_2	0.0050	0.0722	0.0059	0.0119	< 0.0001
	λ_3	0.0001	0.0030	< 0.0001	0.0006	4.3884
	τ_1	5.0000	10.6196	54.362	4.9948	1.5242
	τ_2	20.0000	19.7148	4.5108	19.3403	1.0840

The estimation results for this analysis including covariates can be found in Table 13. In this case, we see that $\widehat{\nu}$ indicates that the hazard is not constant, and so the difference in estimating the location of change points corresponds to the results of the simulation study. We see that the piecewise constant method estimates an increase in hazard after 2.5 years and another increase after 7.5 years. Our proposed method, on the other hand, estimates an increase in hazard after about a year, and a subsequent decrease after 13.3 years.

Table 13: Point estimates and 95% confidence intervals in parentheses for SEER prostate cancer data considering only localized/regional cases by the proposed model and the constant hazard model by Goodman et al. (2011).

Parameters	Constant Hazard Model	Proposed Model
ν	–	1.5758 (1.5707, 1.5809)
Age	–	0.1020 (0.1016, 0.1024)
Race (Other)	–	-0.4572 (-0.4727, -0.4417)
Race (White)	–	-0.3664 (-0.3756, -0.3572)
λ_1	0.0019 (0.0018, 0.0020)	0.0000041 (< 0.0001)
λ_2	0.0030 (0.0029, 0.0031)	0.0000219 (< 0.0001)
λ_3	0.0048 (0.0047, 0.0049)	0.0000032 (< 0.0001)
τ_1	30 (29.9807, 30.0193)	11 (10.9936, 11.0064)
τ_2	90 (89.9409, 90.0591)	160 (159.4212, 160.5788)

If we do not consider covariates, our method estimates $\widehat{\nu} = 1.2876$, $\widehat{\lambda}_1 = 0.0007$, $\widehat{\lambda}_2 = 0.0009$, $\widehat{\lambda}_3 = 0.0011$, $\widehat{\tau}_1 = 118$ and $\widehat{\tau}_2 = 162$. Figure 13 shows that, when the hazard is not piecewise constant, the flexibility of the proposed model allows for a more reasonable fit. We can also see that our estimation method again matches with previous analyses that show the highest hazard for "Black," and the lowest for "Other." We can also see that the estimation is quite reasonable, despite a higher level of censoring. Additional figures showing the Kaplan-Meier survival curves and the estimates of survival using the proposed method are not visually as explanatory as the cumulative hazard estimates for this data. For the localized and regional prostate cancer cases, while race is a significant explanatory covariate, the survival curves look visually to be quite close to one another.

In Figure 13, on the left the Nelson-Aalen estimate of the cumulative hazard of the data is represented by the solid grey line. The black dashed line represents the cumulative hazard estimated by our method and the light grey dashed line represent the estimate of the piecewise constant method. On the right, the Nelson-Aalen estimates by race is the solid lines, while the proposed method is the dashed lines. The black represents "White," dark grey represent "Other," and light grey represents "Black." In both plots, the vertical dotted lines represent the location of the change point estimates for the proposed method.

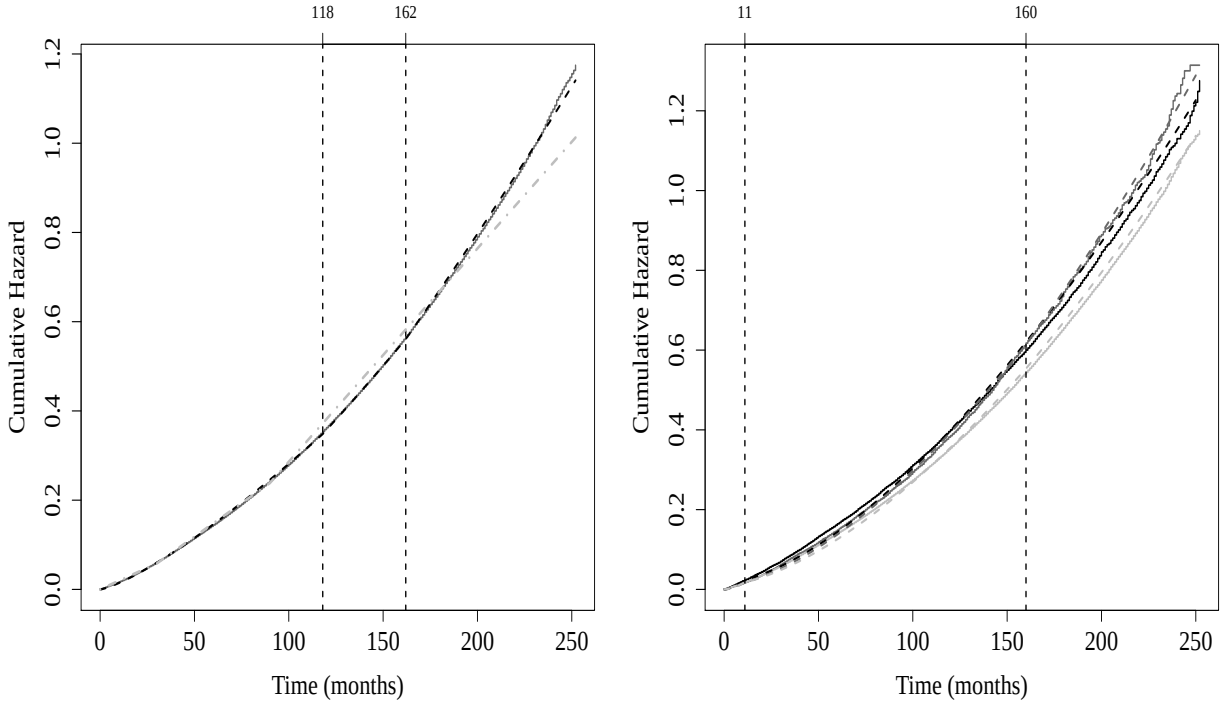

Figure 13: Left: Estimation without covariates. Right: This figure includes estimation with race and averaged age covariates.

APPENDIX C

DETAILED PROOFS FOR CONVERGENCE AND CONSISTENCY OF EMPIRICAL DIVERGENCE MEASURE

Lemma 1. *If Assumption 1 holds,*

$$\sup_{\eta\in[\kappa_T,1-\kappa_T]}\left|\binom{T}{2}^{-1}\sum_{i<j}|Z_i-Z_j|^\alpha-[\eta^2\mu_1^\alpha+(1-\eta)^2\mu_2^\alpha+2\eta(1-\eta)\mu_{12}^\alpha]\right|\overset{a.s.}{\to}0,\ \textit{as}\ T\to\infty.$$

Proof. Let $\epsilon^*>0$. Pick $\epsilon>0$ so that $\epsilon^3+\epsilon^2(2+3\mu_1^\alpha)+\epsilon<\epsilon^*$. Define sets $A_1=\{(i,j)|i<j;Z_i,Z_j\sim F_1\}$, $A_2=\{(i,j)|Z_i\sim F_1,Z_j\sim F_2\}$, and $A_3=\{(i,j)|i<j;Z_i,Z_j\sim F_2\}$. Then, let M_1,M_2, and M_3 be the number of elements in A_1,A_2, and A_3, respectively. Note that these sets are disjoint.

By the Strong Law of Large Numbers for U-Statistics proposed by Hoeffding, there exists an $N_1\in\mathbb{N}$ such that whenever $M_1>N_1$

$$\left|\binom{M_1}{2}^{-1}\sum_{A_1}|Z_i-Z_j|^\alpha-\mu_1^\alpha\right|<\epsilon.$$

We can similarly define $N_2,N_3\in\mathbb{N}$. Then, there also exists an $N_4\in\mathbb{N}$ such that, for $T>N_4$, $\frac{1}{T-1}<\epsilon/2$ (Hoeffding, 1961).

Let $N=\max\{N_1,N_2,N_3,N_4\}$. Then, $\forall\ T\kappa_T>N$ and $\forall\ \eta\in[\kappa_T,1-\kappa_T]$ it is true that $M_1=\lfloor\eta T\rfloor>N_1$, $M_2=\lfloor\eta T\rfloor(T-\lfloor\eta T\rfloor)>N_2, M_3=(T-\lfloor\eta T\rfloor)>N_3$, and also that each of $|\frac{r}{T}-\eta|,|\frac{r-1}{T-1}|,|\frac{s}{T}-(1-\eta)|$, and finally $|\frac{s-1}{T-1}-(1-\eta)|$ are less than ϵ.

It is then true that

$$\begin{aligned}\binom{T}{2}^{-1}\sum_{A_1}|Z_i-Z_j|^\alpha&=\frac{2}{T(T-1)}\sum_{A_1}|Z_i-Z_j|^\alpha\\&=\frac{2}{r(r-1)}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha\\&=\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha.\end{aligned}$$

Also,

$$\begin{aligned}
&\left|\frac{r}{T}-\eta\right|\left|\frac{r-1}{T-1}-\eta\right|<\epsilon^2 \\
\implies &\left|\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)-\eta\left(\frac{r}{T}-\frac{r-1}{T-1}\right)+\eta^2\right|<\epsilon^2 \\
\implies &\left|\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)+\eta^2\right|<\epsilon^2+\eta\left(\frac{r}{T}-\frac{r-1}{T-1}\right)<\epsilon^2+2\eta\epsilon \quad \because \epsilon>0 \text{ is arbitrary} \\
\implies &\left|\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)-\eta^2\right|<\left|\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)+\eta^2\right|<\epsilon^2+2\eta\epsilon \quad \because r,T\geq 1
\end{aligned}$$

Then, we can rearrange the inequalities so that

$$\begin{aligned}
&\left|\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)-\eta^2\right|\left|\binom{r}{2}^{-1}\sum_{A_1}|Z_i-Z_j|^\alpha-\mu_1^\alpha\right|<\epsilon^3+2\eta\epsilon^2 \\
\implies &\left|\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha-\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\mu_1^\alpha\right. \\
&\qquad\left.-\eta^2\binom{r}{2}^{-1}\sum_{A_1}|Z_i-Z_j|^\alpha+\eta^2\mu_1^\alpha\right|<\epsilon^3+2\eta\epsilon^2 \\
\implies &\left|\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha+\eta^2\mu_1^\alpha\right|<\epsilon^3+2\eta\epsilon^2+\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\mu_1^\alpha \\
&\qquad+\eta^2\binom{r}{2}^{-1}\sum_{A_1}|Z_i-Z_j|^\alpha \\
&\qquad<\epsilon^3+2\eta\epsilon^2+\epsilon^2\mu_1^\alpha+\eta^2\epsilon<\epsilon^3+2\eta\epsilon^2+\mu_1^\alpha\epsilon^2(1+2\eta)+\eta^2\epsilon \\
\implies &\left|\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha-\eta^2\mu_1^\alpha\right| \\
&\qquad<\left|\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha+\eta^2\mu_1^\alpha\right| \\
&\qquad<\epsilon^3+\epsilon^2(2\eta+(1+2\eta)\mu_1^\alpha)+\eta^2\epsilon \\
&\qquad<\epsilon^3+\epsilon^2(2+3\mu_1^\alpha)+\epsilon.
\end{aligned}$$

So, $T>N$ implies that

$$P\left(\left|\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha-\eta^2\mu_1^\alpha\right|<\epsilon^*\right)=1 \quad \text{for } \epsilon^*>0.$$

We can apply similar reasoning to the sets defined by A_2 and A_3. Then, using the triangle inequality, we complete the proof since $\epsilon^*>0$ is arbitrary, and we have uniform convergence.

□

Theorem 1. *Suppose Assumption 1 holds. Let $\widehat{\tau}$ found at time point $t_{\widehat{r}}$ be the point estimate of treatment time-lag effect for a pooled sample with T distinct survival times, where $\widehat{r} = \underset{r}{argmax}\ \widehat{\mathcal{Q}}(\boldsymbol{X}_{\boldsymbol{r}}, \boldsymbol{Y}_{\boldsymbol{r}}; \alpha)$. Then, if T is large enough so that $\eta \in [\kappa_T, 1-\kappa_T]$, we have for all $\epsilon > 0$*

$$P\left(\lim_{T\to\infty}\left|\eta - \frac{\widehat{r}}{T}\right| < \epsilon\right) = 1.$$

Proof. Let T be such that $\eta \in [\kappa_T, 1-\kappa_T]$. Then, for any $\tilde{\eta} \in [\kappa_T, 1-\kappa_T]$, let $\tilde{r} = \lfloor \tilde{\eta} T \rfloor$ and $\tilde{s} = T - \tilde{r}$. Then, $\boldsymbol{X}_{\tilde{\boldsymbol{r}}} = \{Z_1, \ldots, Z_{\tilde{r}}\}$ and $\boldsymbol{Y}_{\tilde{\boldsymbol{r}}} = \{Z_{\tilde{r}+1}, \ldots, Z_T\}$ for all T. Then

$$\begin{aligned}\widehat{\mathcal{E}}(\boldsymbol{X}_{\tilde{\boldsymbol{r}}}, \boldsymbol{Y}_{\tilde{\boldsymbol{r}}}; \alpha) &\xrightarrow{a.s.} \left(\frac{\eta}{\tilde{\eta}}\mathbb{I}(\tilde{\eta} \geq \eta) + \frac{1-\eta}{1-\tilde{\eta}}\mathbb{I}(\tilde{\eta} < \eta)\right)^2 \mathcal{E}(X, Y; \alpha) \\ &= h(\tilde{\eta}; \eta)\mathcal{E}(X, Y; \alpha),\end{aligned}$$

as $T \to \infty$, uniformly in $\tilde{\eta}$. The maximum of $h(\tilde{\eta}; \eta)$ is attained when $\tilde{\eta} = \eta$. We also see that

$$\frac{1}{T}\widehat{\mathcal{Q}}(\boldsymbol{X}_{\tilde{\boldsymbol{r}}}, \boldsymbol{Y}_{\tilde{\boldsymbol{r}}}; \alpha) \xrightarrow{a.s.} \tilde{\eta}(1-\tilde{\eta})h(\tilde{\eta}; \eta)\mathcal{E}(X, Y; \alpha),$$

as $T \to \infty$, uniformly in $\tilde{\eta}$. Additionally, the maximum of $\tilde{\eta}(1-\tilde{\eta})h(\tilde{\eta}; \eta)$ is also attained when $\tilde{\eta} = \eta$. Now, define

$$\widehat{r} = \underset{r \in \{\lceil T\kappa_T \rceil, \ldots, \lfloor T(1-\kappa_T) \rfloor\}}{argmax} \widehat{\mathcal{Q}}(\boldsymbol{X}_{\boldsymbol{r}}, \boldsymbol{Y}_{\boldsymbol{r}}; \alpha),$$

and the interval

$$\widehat{\Gamma} = \underset{\tilde{\eta} \in [\kappa_T, 1-\kappa_T]}{argmax}\ \widehat{\mathcal{Q}}(\boldsymbol{X}_{\tilde{\boldsymbol{r}}}, \boldsymbol{Y}_{\tilde{\boldsymbol{r}}}; \alpha).$$

Then, we can see that $\frac{\widehat{r}}{T} \in \widehat{\Gamma}$. Since

$$\frac{1}{T}\widehat{\mathcal{Q}}\left(\boldsymbol{X}_{\widehat{\boldsymbol{r}}/\boldsymbol{T}}, \boldsymbol{Y}_{\widehat{\boldsymbol{r}}/\boldsymbol{T}}; \alpha\right) > \frac{1}{T}\widehat{\mathcal{Q}}\left(\boldsymbol{X}_{\boldsymbol{\eta}}, \boldsymbol{Y}_{\boldsymbol{\eta}}; \alpha\right) - o(1),$$

we have that

$$\frac{1}{T}\widehat{\mathcal{Q}}\left(\boldsymbol{X}_{\widehat{\boldsymbol{r}}/\boldsymbol{T}}, \boldsymbol{Y}_{\widehat{\boldsymbol{r}}/\boldsymbol{T}}; \alpha\right) \geq \eta(1-\eta)h(\eta; \eta)\mathcal{E}(X, Y; \alpha) - o(1),$$

by the almost sure uniform convergence shown previously. Letting $\widehat{\eta} = \frac{\widehat{r}}{T}$, it follows that

$$\begin{aligned}0 &\leq \eta(1-\eta)h(\eta; \eta)\mathcal{E}(X, Y; \alpha) - \widehat{\eta}(1-\widehat{\eta})h(\widehat{\eta}; \eta)\mathcal{E}(X, Y; \alpha) \\ &\leq \frac{1}{T}\widehat{\mathcal{Q}}(\boldsymbol{X}_{\widehat{\boldsymbol{\eta}}}, \boldsymbol{Y}_{\widehat{\boldsymbol{\eta}}}; \alpha) - \widehat{\eta}(1-\hat{\eta})h(\widehat{\eta}; \eta)\mathcal{E}(X, Y; \alpha) + o(1).\end{aligned}$$

This tends to 0 as $T \to \infty$. For every $\epsilon > 0$, there exists an ϵ^* such that

$$\tilde{\eta}(1-\tilde{\eta})h(\tilde{\eta}; \eta)\mathcal{E}(X, Y; \alpha) < \eta(1-\eta)h(\eta; \eta)\mathcal{E}(X, Y; \alpha) - \epsilon^*$$

for all $\tilde{\eta}$ with $|\tilde{\eta} - \eta| \geq \epsilon$. Therefore,

$$P\left(\lim_{T\to\infty} |\widehat{\eta} - \eta| \geq \epsilon\right) \leq P\left(\lim_{T\to\infty} \widehat{\eta}(1-\widehat{\eta})h(\widehat{\eta};\eta)\mathcal{E}(X,Y;\alpha) < \eta(1-\eta)h(\eta;\eta)\mathcal{E}(X,Y;\alpha) - \epsilon^*\right)$$
$$= 0.$$

This proves the claim of uniform convergence and strong consistency of the estimator. To specifically consider the rates of convergence, we need additional information about the distribution of estimators which, in turn, depends on the distribution of the data (which is considered to be unknown or arbitrary). □

VITA

Kristine Gierz
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529

Education

Ph.D	Old Dominion University, Norfolk, VA, USA	December 2020
	Major: Computational and Applied Mathematics (Biostatistics)	
M.S.	Old Dominion University, Norfolk, VA, USA	December 2016
	Major: Computational and Applied Mathematics (Biostatistics)	
B.S.	University of California, Los Angeles, California, USA	June 2013
	Major: Biology	

Experience

Graduate Teaching Assistant, Old Dominion University — August 2015 – August 2019
Department of Mathematics and Statistics, Norfolk, VA, USA

Graduate Research Fellow, NIST — May – July 2018
Statistical Engineering Division, Gaithersburg, MD, USA

Awards

Department of Defense SMART Scholarship — August 2019 – December 2020
Sponsoring Facility: Pentagon, Air Force A9 HQ

NSF Mathematical Sciences Graduate Internship — May – July 2018
NIST, Gaithersburg, MD

Typeset using LaTeX.