

Durham E-Theses

Incorporating Imprecision into the Proportional Hazards Model

AHMAD OMAR S ALBAITY

How to cite:

ALBAITY, AHMAD OMAR S (2024) Incorporating Imprecision into the Proportional Hazards Model. Doctoral thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a <https://etheses.durham.ac.uk/id/eprint/15675/> is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

Incorporating Imprecision into the Proportional Hazards Model

Ahmad Omar S Albaity

A Thesis presented for the degree of
Doctor of Philosophy



Department of Mathematical Sciences
University of Durham
England

August 2024

Dedicated to

My parents

for their boundless support and prayers that have been my guiding light

My wife, Slma, and sons, Omar and Hattan

for filling my days with joy and purpose

My siblings

for their steadfast belief and good wishes that have bolstered my resolve

My friends

for their encouragement and unwavering faith in my journey

Incorporating Imprecision into the Proportional Hazards Model

Ahmad Omar Saeed Albaity

Submitted for the degree of Doctor of Philosophy
August 2024

Abstract

This thesis explores the potential and implications of incorporating imprecision into the Cox proportional hazards (PH) model, a widely endorsed method for examining the effects of covariates in survival analysis. Despite the fact that the PH model does not impose any parametric assumptions regarding the distribution of the baseline hazard function, it relies on the assumption of proportional hazards over time, which is often not sustainable in real-world scenarios. The research highlights the inherent limitations of the PH model, notably its vulnerability to deviations arising from factors like time-dependent covariates, which can compromise the integrity of statistical analyses in vital areas such as clinical research and public health. By introducing imprecision into the PH model, this thesis establishes advanced methodologies that effectively balance the trade-off between imprecision and validity when the conventional assumptions of the model are compromised. In response to these limitations, the thesis introduces novel methodologies to address the non-proportionality of hazards, proposing two innovative imprecise proportional hazards models: the individual-based model and the group-based model. These models offer a robust alternative to the conventional PH model by accommodating variability within the hazard functions and enabling the estimation of more reliable survival functions.

The thesis introduces another robust PH model designed for survival data with continuous covariates. Diverging from traditional measurement error approaches, the robust PH model integrates errors directly into covariate values as a strategy

to mitigate the proportional hazards assumption. This shifts the focus from merely diminishing estimation bias to enhancing model adaptability. The proposed model incorporates additive errors into continuous covariates which are distribution-free, but fluctuate strictly within a predefined small interval. Consequently, imprecise estimates can be derived for individuals' survival functions which enhances the flexibility and reliability of the robust PH model, particularly when the validity of the proportional hazards assumption is questioned.

This thesis concludes by introducing a novel imprecise estimation technique referred to as the Most Likely Data method (MLD) as an alternative to the well-known maximum likelihood estimation (MLE). Unlike the MLE, which offers precise point estimates through optimizing the likelihood function, the MLD focuses on interval estimates derived from the most likely observed data configurations. In this method, the parameter space is partitioned into intervals based on data that is most likely to be observed compared to others, resulting in a distinct interval for each possible observations. For discrete distributions, the MLD method can be applied seamlessly to both binomial and Poisson distributions, allowing partitioning the parameter space for different observations and providing a close-form technique for identifying imprecise estimates. In the context of the PH model, the MLD methodology revealed promising results as a means of relaxing the PH assumption. The objective of introducing the MLD method in this thesis is to pave the way for further investigation and development in the field of statistical inference.

Besides challenging the conventional application of the PH model in specific research contexts, the findings of this thesis offer significant methodological advancements that can enhance the robustness of conclusions drawn from survival data, thus influencing future research and practices in similar fields.

Declaration

The work in this thesis is based on research carried out in the Department of Mathematical Sciences at Durham University. No part of this thesis has been submitted elsewhere for any degree or qualification, and it is all my own work unless referenced to the contrary in the text.

Copyright © 2024 by Ahmad Omar Saeed Albaity.

“The copyright of this thesis rests with the author. No quotations from it should be published without the author’s prior written consent and information derived from it should be acknowledged”.

Acknowledgements

The culmination of this academic journey fills me with a deep sense of gratitude. This thesis represents a milestone not only for me, but also for all those who have accompanied me along this journey.

Allah, my God, has been a constant source of strength and wisdom for me throughout my journey, and I am humbled and grateful to Him for His blessings and guidance.

A heartfelt thank you to my academic supervisors, Frank Coolen and Peter Craig, for their invaluable guidance, patience, and expertise. Your mentorship has not only shaped this thesis, but has also immensely contributed to my personal and professional growth.

I cherish the memory of my father, who has been a constant source of inspiration and strength throughout my life. Though he is no longer with us, I continue to be guided by his teachings and values in every aspect of my life.

As a result of my mother's unwavering belief in my abilities, I am able to overcome every challenge. Your sacrifices, love, and unending support have been the cornerstone of my resolve. I am also deeply grateful to my dear aunts, Fottom, who is no longer with us but whose guidance continues to inspire me, and Khadijah, whose nurturing presence has been just like that of a second mother. Your love and encouragement have been instrumental in my journey.

My siblings, who have lightened my load with their continuous encouragement and moments of joy and struggle, have been my pillars of strength.

My deepest appreciation is extended to my wife, Slma Bahuwayrith, thank you for standing by me with limitless patience, love, and understanding. To my sons, Omar and Hattan, you are my biggest cheerleaders. Your smiles and unconditional

love are my daily inspiration and happiness.

My friends and colleagues, your encouragement during challenging times and celebrations in moments of triumph have made this journey all the more memorable.

In conclusion, I would like to offer my sincere thanks to my country, the Kingdom of Saudi Arabia, whose generous support has played an instrumental role in bringing this research to fruition through Jazan University. Your belief in my work has been both an honor and a responsibility that I have worked to achieve.

Contents

Abstract	iii
Declaration	v
Acknowledgements	vii
1 Introduction	1
1.1 Overview	1
1.2 Imprecise probability	2
1.3 Motivation	3
1.4 Contents and outline of the thesis	5
2 Background	7
2.1 Introduction	7
2.1.1 Empirical Likelihood	10
2.2 Non-parametric survival methods	12
2.2.1 Kaplan-Meier estimation	12
2.2.2 Nelson-Aalen estimation	13
2.3 Empirical full likelihood for survival data	14
2.3.1 Construction of the full likelihood for survival data	14
2.3.2 Poisson and Binomial empirical likelihoods	16
2.4 Proportional hazards model	19
2.4.1 Introduction	19
2.4.2 Estimating the PH Model	21
2.4.3 Tied Survival Times	23

2.4.4	Estimating the hazard and survival functions	24
2.5	Empirical full likelihood for the PH model	26
2.5.1	Poisson empirical likelihood for the PH model	27
2.5.2	Empirical full likelihood for the PH model F_0	29
2.6	Simulating survival data for the PH model	31
2.6.1	Parametric PH models	31
2.6.2	Survival times	32
2.6.3	Censoring times	34
2.7	Bootstrap methods for the PH model	35
2.7.1	Zelterman <i>et al.</i> 's bootstrap method for the PH model	36
2.7.2	Hjort's bootstrap method for the PH model	38
3	Imprecise PH model	40
3.1	Introduction	40
3.2	Individual-based Imprecise PH model (IPH)	41
3.2.1	Profiling out the baseline hazard function	43
3.2.2	Maximizing the likelihood function of the IPH model	44
3.2.3	Hazard and Survival functions for the IPH model	46
3.3	Group-based Imprecise PH model (GPH)	60
3.3.1	Maximizing the likelihood function of the GPH model	61
3.3.2	Hazard and survival functions for the GPH model	63
3.4	Assessing the IPH and GPH models	74
3.4.1	Bootstrap investigation	78
3.5	Concluding remarks	93
4	Robust PH model	96
4.1	Introduction	96
4.2	Robust PH model	98
4.3	Robust-Poisson PH likelihood (RP)	99
4.3.1	Imprecise hazard and survival functions	111
4.4	Robust-Empirical PH likelihood (RE)	128
4.5	Simulations	145

4.5.1	Inconsistency of the naive imprecise survival functions	145
4.5.2	Robust PH model with covariate measurement error	153
4.5.3	Bootstrap investigation	158
4.6	Concluding remarks	160
5	Most Likely Data	164
5.1	Introduction	164
5.2	MLD Method	165
5.2.1	MLD method for Binomial distribution	166
5.2.2	MLD method for Poisson distribution	168
5.3	MLD method for the the PH model	170
5.3.1	MLD for the PH model using marginal probability	171
5.3.2	MLD for the PH model using adjusted probability	173
5.3.3	Time-based MLD method for the PH model	175
5.4	Flexible MLD method	179
5.4.1	Flexible MLD for method Binomial distribution	180
5.4.2	Relaxed MLD for Poisson distribution	181
5.4.3	Flexible MLD method for the PH model	183
5.4.4	Flexible time-based MLD method for the PH model	185
5.4.5	Simulation study	186
5.5	Concluding remarks	191
6	Conclusions	193
	Appendix	197
A	Additional materials	197
A.1	η for PH and NPH data	197
A.2	Proof of Theorem 2.5.2	200
A.3	Robust PH model (Poisson)	205
A.4	Bootstrap investigation (Chapter 4)	208

B R codes	210
B.1 Simulating PH and NPH survival data	210
B.2 IPH model	212
B.3 GPH model	213
B.4 Zelterman <i>et al.</i> [91]’s Bootstrap	215
B.5 Robust PH model	221
B.6 MLD Method	225
 Bibliography	 229

Chapter 1

Introduction

1.1 Overview

Research in many fields involves analyzing the occurrence of time-related events as well as investigating the relationship between a variety of covariates, independent variables, and event times of a particular subject. In medical research, survival analysis is the umbrella of all statistical methods that are used to determine and analyze death times, the recurrence of a disease, or to evaluate the effects of different treatments on the survival experience of individuals. Survival analysis is primarily used in medical research, but these statistical techniques are employed across a wide range of disciplines under other terminology [21]. In engineering, for example, the statistical methods associated with analysing the life history of machines or manufactured items are referred to as reliability theory. Similarly, these methods can be applied in many different fields such as sociology, biology, economics, demography, criminology, and epidemiology [75].

Survival data are often affected by the presence of censorship, which prevents other statistical methods, such as regression, from being implemented. Censoring occurs when the time of the event of interest for a particular subject is unobserved [46]. Censoring can be classified into different schemes, including right, left and interval [55]. In right censoring, the survival time is greater than the observed time, which is the most common type of censoring. Left censoring refers to the case where the event of interest had already occurred prior to the beginning of a study. Interval

censoring occurs when a subject or individual experiences an event within an interval of time, so the exact time may not be known. The censoring mechanism is assumed to be independent of survival time, and to be non-informative [21, 49].

1.2 Imprecise probability

The term imprecise probability, according to Augustin [4], refers to all approaches that replace precise, traditional probabilities with non-empty sets of precise probabilities as their primary modeling entity, including all approaches that can be translated into an equivalent set of precise probabilities. These approaches include the robust Bayes analysis [79], linear partial information [51], Levi's approach to epistemology [56], as well as approaches based on non-linear functionals and non-additive set-functions, covering lower and upper previsions in tradition of Walley's book [87], Weichselberger's type of interval probabilities [89], or those which are based on capacities including Shafer theory [28], random sets [12] and p-boxes[35]. Imprecise probabilities are represented by using lower and upper bounds for probabilities rather than the standard theory of precise or single-valued probability [24]. Researchers in a variety of areas of statistics have been motivated by a growing desire to model complex uncertainties more comprehensively. There has been evidence that imprecise probabilities are capable of providing excellent solutions to some of the most difficult foundational problems in probability and statistics. Even though imprecise probability theories are rarely explored in statistics since many are still in their infancy, a wide range of methodological and philosophical foundations have been developed, along with their specific applications, in somewhat divergent literature [4]. For those interested in learning more about imprecise probabilities, we recommend Walley's book [87], Augustin and his colleagues' book [7] entitled "Introduction to Imprecise Probabilities" as well as the website of the Society for Imprecise Probability, *sipta.org*. The book includes both the core theory of imprecise probabilities as well as recent developments that can be applied to a variety of application areas. The Nonparametric Predictive Inference (NPI), for instance, provides a notable example of such a statistical methodology in which Augustin and

Coolen combine Weichselberger's interval probability and Hill's $A_{(n)}$ assumption to construct an imprecise probability based on frequentist principles [6, 22]. Recently, Coolen and his collaborators have applied the NPI lower and upper probabilities to analyzing various types of data in the context of statistical hypothesis testing and prediction. For right-censored data, Coolen and Yan [23] presented the NPI framework to construct lower and upper survival functions as an alternative to the well-known Kaplan-Meier (KM) estimator. This approach inspired us to investigate the possibility of incorporating imprecision into the PH model. In spite of the fact that the NPI method is primarily used for prediction, the purpose of this thesis is to investigate the effect of incorporating imprecision into the PH model in order to find out how it affects the estimated regression parameter and survival functions of populations.

1.3 Motivation

The Cox proportional hazards model (PH) is the most common method to study the effects of covariates in survival analysis [48]. Despite the fact that the PH model does not impose any parametric assumptions regarding the distribution of the baseline hazard function, it relies on the proportionality of hazards assumption [25]. As a result, the proportional hazards assumption implies that the hazard ratios between any different individuals or objects remain constant over time. While validating the PH assumption is crucial, the assumption is often unrealistic in real-life situations [21, 84]. The PH assumption can be violated for a variety of reasons such as time-dependent covariates, missed covariates, and natural changes in the hazard function [50]. The violation of the PH assumption results in misleading hazard ratios that provide an oversimplified understanding of complex survival data structures [90].

This compromises the integrity of statistical conclusions and subsequent decisions in clinical research, public health policy, and other sectors [40]. Although the PH assumption can be assessed by a number of methods [37, 39], including the standard K-M plot, log-minus-log plot, Schoenfeld's residuals, and scaled Schoenfeld's resid-

uals test, etc., it has been shown that this assumption has not been tested or is not reported in most studies [53]. Lack of consideration and discussion of the underlying causes of apparent non-proportionality could result in biased hazard ratio estimates and undermine the overall research effort as these causes could have substantial implications for future studies. Based on a study conducted by Rahman [74] on the deviations from proportional hazards in cancer clinical trials published between 2014 and 2016, almost a quarter of these clinical trials had evidence of deviations from proportional hazards, yet only a few publications explicitly test for the PH assumption in detail. Kuitunen [53] performed similar study to investigate the validity of the PH assumption in Total Joint Arthroplasty studies retrieved from the PubMed database, demonstrating that 45% of these studies had KM survival curves crossed, while only 20% reported and addressed violations of the PH assumption. Further, over seventy percent of phase III trials involving immune checkpoint inhibitors have been analyzed using the proportional hazards assumption, even though these trials exhibit crossing hazards [20].

Researchers are strongly encouraged to consider alternative models that can account for such complexities that are not adequately represented by the Cox PH model if the PH assumption is violated. A number of alternative models are available, including extended Cox models which incorporate time-dependent covariates or stratification to account for non-proportional hazards, Frailty models which consider heterogeneity within survival data [41], Accelerated Failure Time Models [88], Additive Hazard Models [60], parametric survival models such as the Weibull and Gompertz [21], flexible parametric models such as Royston-Parmar model [78], machine learning approaches [9], and others. Due to the fact that the available approaches are based on specific NPH scenarios, there is no agreement on the optimal best practices for dealing with violations of the PH assumption [9]. It is therefore imperative to explore alternative models to enhance the robustness of conclusions drawn from survival data. This thesis presents and investigates different approaches to incorporating imprecision into the PH model, so that the proportional hazards assumption can be relaxed.

1.4 Contents and outline of the thesis

This thesis explores the implications of incorporating imprecision into the Proportional Hazards model using various approaches. The thesis is structured as follows:

A comprehensive review of key concepts extracted from existing literature that are pertinent to this study is presented in Chapter 2, which serves as the foundation of the thesis. In the introduction, an overview of the PH model is provided, as well as a discussion of parameters estimation and baseline survival and hazard functions. This chapter introduces empirical likelihood methods, with a special focus on their application in the presence of right-censored observations. A concise review of methods for generating PH survival data and the application of bootstrap methods to PH data is provided in this chapter, which is designed to frame these concepts as essential tools for analyzing the data in the following chapters.

Chapter 3 presents two novel variations of imprecise PH models based on Poisson empirical likelihoods: the Individual-based Imprecise PH (IPH) model and the Group-based Imprecise PH (GPH) model. In the IPH model, each subject or individual in the data set has its own imprecision factor. The GPH model, on the other hand, expands this concept to groups allowing for shared imprecision factors within defined clusters. The IPH model can be viewed as a special case of the GPH model, in which each group has a unique member. The exploration of these models includes a detailed discussion of their properties, particularly the imprecise estimation of survival functions. The chapter further enriches the research by conducting a bootstrap study, which evaluates the advantages of utilizing these models and helps in determining the optimal level of imprecision to be used.

In Chapter 4, robust PH models are presented for continuous covariates. The model considers a surrogate version of the observed covariate values that raises doubts regarding the proportional hazards assumption. This model introduces a novel concept where observed covariate values are permitted to vary within small intervals. By including these errors, we relax the PH assumption, as opposed to measurement error modeling that increases the model's accuracy. Despite these errors remaining distribution-free, they are bounded by predefined intervals. Additionally,

a variety of likelihoods including partial likelihood and empirical likelihood were considered for this model. Furthermore, we examine the effects of including these errors when estimating regression parameter, likelihood value gain, and imprecise survival functions for specific individuals. We conclude this chapter by bootstrap investigations aim to identify the optimal level of imprecision.

Chapter 5 explores the Most Likely Data (MLD) method, an innovative early-stage statistical inference technique that can be used for estimation as an alternative to the maximum likelihood approach, and consider its application to relax the PH assumption. The MLD method involves considering all potential outcomes to identify the most probable parameter values yielding desired outcomes. The chapter begins withThe chapter begins by introducing the novel MLD method, followed by an exploration of its application across various contexts, including discrete distributions such as Binomial and Poisson, as well as within the PH model. As challenges emerged in applying the MLD method to the PH model, alternative formulas for marginal probabilities were considered, including a time-based probability approach. A flexible version of the MLD method was then proposed, providing an adaptive approach to address imprecise estimation, which is particularly advantageous in complex models like the PH model. The chapter concludes with a discussion of the challenges encountered and directions for future research.

The final chapter, Chapter 6, provides a comprehensive summary of the findings, underscoring the significance of the novel approaches developed in the thesis. It reflects on the challenges encountered throughout the research and suggests possible extensions and future research directions. In the appendix, we provide supplementary material and detailed R scripts designed specifically for the proposed methods. These scripts are compatible with R version 3.6.1 [72].

Chapter 2

Background

2.1 Introduction

The purpose of this chapter is to provide the foundational materials required to follow the thesis seamlessly. In the realm of survival analysis, the survival time can be modeled as a continuous or discrete positive random variable denoted by T , where t refers to the actual survival time of an individual. $F(t)$ is the distribution function representing the cumulative probability of an event occurs at or prior to time t . Three functions play a major role in survival analysis: the survival function, $S(t)$, the hazard function, $h(t)$, and the cumulative hazard function, $H(t)$. The survival function indicates the probability that the event occurs after a given time t . Since the survival function is the complement of distribution function, then

$$S(t) = P(T > t) = 1 - F(t) \tag{2.1}$$

The hazard function measures the instantaneous risk or hazard associated with an event, such as mortality, and represents the conditional probability of the event occurring at a specific time t for an individual conditional on having been survived to that time. Further, the cumulative hazard function quantifies the cumulative risk of an event occurring by time t .

Continuous distributions

Suppose that the random variable associated with the survival time is continuous. Now, let $f(t)$ be the probability density function of the random variable T . The cumulative distribution function of $f(t)$ is given by

$$F(t) = P(T \leq t) = \int_0^t f(u)du \quad (2.2)$$

Based on the probability that the random variable T for an individual survival time falls between $[t, t + \Delta t]$ conditional on T greater or equal to t , the hazard function can be defined as follows

$$h(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \quad (2.3)$$

By utilizing the properties of the conditional probability and the definition of derivative it's easy to see that

$$h(t) = \frac{f(t)}{S(t)} \quad (2.4)$$

Noteworthy is the fact that $S(t) = P(T > t) = P(T \geq t)$ since T belongs to a continuous distribution. The cumulative hazard function of the event occurring by time t becomes

$$H(t) = \int_0^t h(u)du \quad (2.5)$$

These functions can be determined from the others. For instance, the density function can be obtained by the survival function as follows

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}(1 - S(t)) = 0 - \frac{d}{dt}S(t) = -\frac{d}{dt}S(t) \quad (2.6)$$

By substituting 2.6 in 2.4, the hazard function can be derived by the survival function as follows

$$h(t) = \frac{-\frac{d}{dt}S(t)}{S(t)} = -\frac{d}{dt}(\ln S(t)) \quad (2.7)$$

Similarly, by taking the integral of both sides in Equation (2.7), it is obvious since $S(0) = 1$ and so $\ln S(0) = 0$, the cumulative hazard function can be determined by

the survival function by

$$H(t) = -\ln S(t) \quad (2.8)$$

Consequently, survival function can be written in terms of cumulative hazard function as follows

$$S(t) = e^{-H(t)} \quad (2.9)$$

Discrete distributions

In contrast to the continuous time, the hazard function related to the discrete time distribution is given as follows

$$h(t) = \frac{f(t)}{S(t^-)} \quad (2.10)$$

In this equation, $f(t)$ represents the probability mass function and $S(t^-)$ denotes the survival function at the time just prior to time t . To ensure clarity, the notation $f(\cdot)$ will be used interchangeably in this thesis based on the specific context being discussed to refer to either the probability density function or the probability mass function. The survival function for discrete time is given by

$$S(t) = \prod_{t_l \leq t} [1 - h(t_l)] \quad (2.11)$$

where t_l denote the times at which there is positive probability mass such that $t_l \leq t$. The cumulative hazard function is now defined as

$$H(t) = \sum_{t_l \leq t} h(t_l) \quad (2.12)$$

Notice that Equation (2.9) relating survival and cumulative hazard functions does not hold for a discrete distribution. Additionally, probability mass function at time t can be expressed, using Equation (2.10) and Equation (2.11) as

$$\begin{aligned} f(t) &= h(t)S(t^-) \\ &= h(t) \left(\prod_{t_l < t} [1 - h(t_l)] \right) \end{aligned} \quad (2.13)$$

2.1.1 Empirical Likelihood

A robust nonparametric approach known as empirical likelihood is widely studied in the literature as a method of statistical inference Owen [68]. The method extends the concept of maximum likelihood estimation to situations where the underlying distribution is unspecified [67]. Due to its nonparametric nature, it facilitates the construction of confidence intervals and hypothesis testing without relying on parametric assumptions [68]. Shortly after the introduction of the empirical likelihood method, Hall and Scala [38] provided an overview of its critical properties. A comprehensive book was published following this by Owen [68] provided a comprehensive review of significant developments in the empirical method and pointed out various variants and applications. This method is particularly notable for its ability to incorporate side information through moment conditions, which increases its applicability in various fields including econometrics, biostatistics, and environmental studies [71]. Furthermore, empirical likelihood has been successfully integrated into regression analysis, enhancing its utility in examining relationships between variables under very general conditions [86]. This section explains how the empirical cumulative distribution function (ECDF) can be viewed as a nonparametric estimate of the CDF based on the empirical likelihood.

Definition 2.1.1 (Owen [68, p. 7]) Assume X_1, \dots, X_n are independent and identically distributed random variables. Then, the empirical likelihood function for the cumulative distribution function (CDF) is defined as follows

$$\begin{aligned}
 L(F) &= \prod_{i=1}^n F(x_i) - F(x_i^-) \\
 &= \prod_{i=1}^n P(X \leq x_i) - P(X < x_i) \\
 &= \prod_{i=1}^n P(X = x_i)
 \end{aligned} \tag{2.14}$$

Recall the definition of the ECDF, for a sample consisting of the following i.i.d. random variables X_1, \dots, X_n . The ECDF for the observed sample values x_1, \dots, x_n

is defined as

$$F_n(x) = P(X \leq x) = \frac{\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}}{n} \quad (2.15)$$

where $\mathbb{1}_{\{X_i \leq x\}}$ is an indicator function that equals 1 if $X \leq x$ and 0 otherwise.

Theorem 2.1.2 (Owen [68, p. 8]) Let X_1, \dots, X_n be i.i.d. random variables, then the ECDF in Equation (2.15) is the non-parametric MLE of the CDF F based on empirical likelihood in Equation (2.14).

Proof.

Consider a scenario where X_1, X_2, \dots, X_n are i.i.d. from a discrete distribution, then empirical likelihood can be applied seamlessly. However, this likelihood is invalid for i.i.d. random variables that are continuously distributed and always equal zero due to the fact that $P(X = x_i) = 0$. A solution is to extend the parameter space to include discrete distributions [62]. Hence, the distribution function, F , should be carefully defined by such a distribution that places positive probability, jump, on every observed data and no probability elsewhere since assigning probability mass outside the observed data will lead to smaller likelihood. It follows that shifting the probability from an unobserved value to any observed data point will increase the overall likelihood because it increases the probability assigned to the observed data point, denoted by p_i , while leaving the others unchanged.

For the case where the i.i.d. random variables from a discrete distribution with the following k unique values x_1, x_2, \dots, x_k in the sample such that $k \leq n$. Let n_i denotes the number of times the value x_i appears in the sample. Then, the empirical log likelihood function can written as follows

$$\ell(F) = \ln \left(\prod_{i=1}^k p_i^{n_i} \right) = \sum_{i=1}^k n_i \ln(p_i) \quad (2.16)$$

where p_i is the probability assigned to each unique value x_i and $\sum_{i=1}^k p_i = 1$. Lagrange multipliers can be utilized to maximize $\ell(F)$ subject to the constraint that summation of these probabilities, p_i , is equal to 1 or $p_1 + p_2 + \dots + p_k - 1 = 0$. Consider the function $G(p_1, p_2, \dots, p_k, \lambda)$ such that $G(p_1, p_2, \dots, p_k, \lambda) = \sum_{i=1}^k n_i \ln(p_i) -$

$\lambda(p_1 + p_2 + \dots + p_k - 1)$, then

$$\begin{aligned}
\frac{\partial G}{\partial p_1} &= \frac{n_1}{p_1} - \lambda \implies p_1 = \frac{n_1}{\lambda} \\
\frac{\partial G}{\partial p_2} &= \frac{n_2}{p_2} - \lambda \implies p_2 = \frac{n_2}{\lambda} \\
&\vdots \\
\frac{\partial G}{\partial p_k} &= \frac{n_k}{p_k} - \lambda \implies p_k = \frac{n_k}{\lambda} \\
\frac{\partial G}{\partial \lambda} &= -(p_1 + p_2 + \dots + p_k - 1) \implies \sum_{i=1}^k p_i = 1
\end{aligned} \tag{2.17}$$

By substituting $p_i = \frac{n_i}{\lambda}$ in $\sum_{i=1}^k p_i = 1$ we obtain $\lambda = \sum_{i=1}^k n_i = n$. Hence, the estimates of these probabilities becomes

$$\hat{p}_i = \frac{n_i}{n}$$

where $i = 1, 2, \dots, k$. For continuous distributions, there are n unique observation in the sample which leads to same conclusion, but with $\hat{p}_i = \frac{1}{n}$. Thus, based on the empirical likelihood, the ECDF is the non-parametric MLE for the CDF. \square

2.2 Non-parametric survival methods

2.2.1 Kaplan-Meier estimation

Kaplan and Meier [33] suggested a standard estimator to the survival function using the method of maximum likelihood. The Kaplan-Meier estimator is considered to be a great alternative to the life-table estimator since the life-table is very sensitive to the choice of the constructed interval and it is inadequate when the actual event time is known as it might lose some information [21]. The Kaplan-Meier estimate is also called the product-limit estimate of the survival function that takes the censoring data into account. Suppose the event time, say death, is divided into intervals $t_1 < t_2 < \dots < t_k$, such that each interval begins with death time and contains at least one death, so k distinct intervals would be constructed. The estimated probability of surviving during the interval $[t_j, t_{j+1} - \epsilon)$ is $(n_j - d_j)/n_j$ where n_j is

the number of individuals at risk in this interval time including those who died or were censored at this period. On the other hand, d_j represents the individuals who died in this interval. Then, Kaplan-Meier estimate of the survivor function at time t , such that $t_j \leq t < t_{j+1}$ can be defined by

$$\hat{S}_{KM}(t) = \prod_{j:t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) \quad (2.18)$$

2.2.2 Nelson-Aalen estimation

Nelson [65] and Aalen [1] proposed another estimate to the survival function based on an empirical estimation of the cumulative hazard function, $\hat{H}_{NA}(t) = \sum_{j:t_j \leq t} \frac{d_j}{n_j}$, where k represents the number of unique event times. The Nelson-Aalen estimator of the survival function is given by

$$\hat{S}_{NA}(t) = \prod_{j:t_j \leq t} \exp\left(\frac{-d_j}{n_j}\right) \quad (2.19)$$

In the absence of covariates, these two estimates, particularly the Kaplan-Meier estimator, are commonly employed for nonparametric analysis of survival distributions. The main difference between the Kaplan-Meier estimator and the Nelson-Aalen estimator is that, in essence, the Kaplan-Meier is the generalization of the empirical survival function, which is given by $S_{KM}(t) = \prod_{j:t_j \leq t} (1 - h_j)$, with $h_j = d_j/n_j$. On the other hand, the Nelson-Aalen estimator utilizes the relation between the survival function and the cumulative hazard function in Equation (2.9). These two estimators are quite similar and both are considered as nonparametric maximum likelihood estimators [49], see also Section 2.3.1. Collett [21] indicated that since $\exp(-x) = 1 - x + \frac{x^2}{2} - \frac{x^3}{6} + \dots \approx 1 - x$ for small x , then the Kaplan-Meier survival function can be considered as a generalization of the Nelson-Aalen survival function when d_j is small relative to n_j as $\exp(-d_j/n_j) \approx 1 - (d_j/n_j)$. Andersen *et al.* [3] and Fleming and Harrington [36] elaborated the construction of these estimators based on the theory of counting processes.

2.3 Empirical full likelihood for survival data

2.3.1 Construction of the full likelihood for survival data

The general full likelihood function for right-censored survival data is discussed in this section. Suppose that a combination of event and right-censored observed times, for n individuals were collected. Denote the random variables corresponding to the event and censoring times of individuals by \mathbf{V} and \mathbf{C} , respectively. Assume that censoring times are non-informative, the actual survival time of an individual does not depend on its censoring time. Let \mathbf{X} be a set of k covariates that do not change with time and relate to the true event time, \mathbf{V} , recorded for each individual [69]. Suppose that the actual observed right-censored survival data are

$$(t_1, \delta_1, \mathbf{x}_1), (t_2, \delta_2, \mathbf{x}_2), (t_3, \delta_3, \mathbf{x}_3), \dots, (t_n, \delta_n, \mathbf{x}_n) \quad (2.20)$$

where $t_i = \min\{v_i, c_i\}$ represents the non-negative underlying event or right censoring time. In addition, the non-censoring indicator variable is denoted by $\delta_i = I\{v_i \leq c_i\}$ and $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{ik}\}$.

For clarity and to avoid potential confusion, this thesis assumes the absence of ties unless stated otherwise. Additionally, survival data as outlined in Equation (2.20) are considered to be time-ordered observations, such that $t_0 = 0 < t_1 < \dots < t_n < t_{n+1} = \infty$. In this context, the pair (x_i, δ_i) are associated with time point t_i , where t_i refers to survival times encompassing both event times and right-censored times. On some occasions, this thesis employs $t_{(i)}$, with $i = 1, 2, \dots, k \leq n$, to denote ordered event times, explicitly excluding survival times related to right-censored observations, even though these may contribute to the risk set at $t_{(i)}$. For simplicity, infinitesimal values will be introduced randomly to resolve any ties in the survival times, while maintaining the assumption that event times occur before right-censored times.

We are mainly interested in making inferences on $f(t, \delta|\mathbf{x})$ which represents the conditional probability density function of the survival data (2.20). Let $f_v(t|\mathbf{x})$, $S_v(t|\mathbf{x})$, $f_c(t)$, and $S_c(t)$ represent the density and survival functions for event and censoring times, respectively. Building on the methodology outlined by Collett [21]

for deriving the full likelihood in the absence of covariates, the likelihood function for survival data presented in Equation (2.20) can be established by determining the probability distributions for both survival and censoring times in the following phases:

I. Event time the conditional joint distribution of t_i and $\delta_i = 1$ given \mathbf{x}_i obtained by

$$\begin{aligned} P(t_i, \delta_i = 1 | \mathbf{x}_i) &= P(v_i, c_i > v_i | \mathbf{x}_i); \text{ since } \delta_i = 1 \Rightarrow t_i = v_i \text{ and } c_i > v_i \\ &= P(v_i | \mathbf{x}_i) P(c_i > v_i); V \text{ and } C \text{ are mutually independent} \\ &= f_v(v_i | \mathbf{x}_i) S_c(v_i); \text{ which can be written as} \\ &= f_v(t_i | \mathbf{x}_i) S_c(t_i) \end{aligned} \tag{2.21}$$

II. Censoring time the conditional probability related to censoring times is given by

$$\begin{aligned} P(t_i, \delta_i = 0 | \mathbf{x}_i) &= P(c_i, v_i > c_i | \mathbf{x}_i); \text{ since } \delta_i = 0 \Rightarrow t_i = c_i \text{ and } v_i > c_i \\ &= P(c_i) P(v_i > c_i | \mathbf{x}_i) \\ &= P(c_i) P(v_i > c_i | \mathbf{x}_i) \\ &= f_c(t_i) S_v(t_i | \mathbf{x}_i) \end{aligned} \tag{2.22}$$

Therefore, the likelihood function of (t_i, δ_i) given \mathbf{x}_i for $i = 1, 2, \dots, n$ can be represented as the product of Equations (2.21) and (2.22) for the n observations as follows

$$\begin{aligned} \prod_{i=1}^n f(t_i, \delta_i | \mathbf{x}_i) &= \prod_{i=1}^n [f_v(t_i | \mathbf{x}_i) S_c(t_i)]^{\delta_i} [f_c(t_i) S_v(t_i | \mathbf{x}_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n [f_v(t_i | \mathbf{x}_i)^{\delta_i} S_v(t_i | \mathbf{x}_i)^{1-\delta_i}] [f_c(t_i)^{1-\delta_i} S_c(t_i)^{\delta_i}] \end{aligned} \tag{2.23}$$

Careful readers of Equation (2.22) should note the existence of different versions of $S_v(t_i | \mathbf{x}_i)$ according to the nature of the distribution function, i.e. discrete or continuous. In the case of continuous distributions, we have

$$S_v(t_i | \mathbf{x}_i) = P(V_i > t_i | \mathbf{x}_i) = P(V_i \geq t_i | \mathbf{x}_i)$$

which indicate that there is no impact of using either $>$ or \geq . For discrete distribution, however, we must emphasize that the survival function is given by

$$S_v(t_i | \mathbf{x}_i) = P(V_i > t_i | \mathbf{x}_i)$$

Under the independent censoring assumption, the censoring functions $f_c(t_i)$ and $S_c(t_i)$ are uninformative [21, 46, 77], that is the censoring mechanism does not contain any information about the parameter relevant to the survival time distribution, V [77]. Hence, these censoring functions can be considered as a constant and the left side of Equation (2.23) is proportional to the event time functions as follows (see Kalbfleisch and Prentice [46, sec. 6.2])

$$\prod_{i=1}^n f(t_i, \delta_i | \mathbf{x}_i) \propto \prod_{i=1}^n f_v(t_i | \mathbf{x}_i)^{\delta_i} S_v(t_i | \mathbf{x}_i)^{1-\delta_i} \quad (2.24)$$

For simplicity, the subscript v will be removed from f_v and S_v in Equation 2.24. The next section describe two different variations of empirical likelihood for survival data.

2.3.2 Poisson and Binomial empirical likelihoods

Two different variations of the empirical likelihood for survival data are described in this section, namely Binomial and Poisson empirical likelihoods. In addition, this section shows that the Kaplan-Meier estimator of the survival function and the Nelson-Aalen estimator of the hazard function are the NPMLE maximizes the full empirical likelihood for survival data. Assume the absence of ties among survival times and consider the one-to-one relationship between the cumulative distribution function and the cumulative hazard function. Then, the empirical likelihood can be expressed in terms of the hazard function. In spite of this, there are various ways in which this likelihood can be expressed in terms of the hazards due to the fact that different formulas exist for discrete and continuous cumulative hazard functions [3, 62, 63, 92]. It is not our intent to delve into much detail about these types of empirical likelihood functions, but rather to provide a brief overview of the options available. The book by Andersen *et al.* [3, p. 180-229], which uses martingale theory and counting process to illustrate these representations in extensive detail, should be highly recommended for interested readers.

The binomial empirical likelihood, denoted by the subscript EB, is defined based on the purely discrete survival function $S(t) = \prod_{t_l \leq t} [1 - h(t_l)]$ with $f(t) = h(t)S(t^-)$ as indicated in Equations (2.11) and (2.10). To simplify the process, we replace $h(t_i)$

by p_i and substitute both $f(t)$ and $S(t)$ into the empirical likelihood as follows

$$\begin{aligned}
L_{EB}(p_i) &= \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \\
&= \prod_{i=1}^n \left[\left(p_i \prod_{l \leq i-1} [1 - p_l] \right)^{\delta_i} \prod_{l \leq i} [1 - p_l]^{1-\delta_i} \right] \\
&= \prod_{i=1}^n \left[p_i^{\delta_i} [1 - p_i]^{1-\delta_i} \prod_{l \leq i-1} [1 - p_l]^{\delta_i} \prod_{l \leq i-1} [1 - p_l]^{1-\delta_i} \right] \\
&= \prod_{i=1}^n \left[\left(\frac{p_i}{1 - p_i} \right)^{\delta_i} \prod_{l \leq i} [1 - p_l] \right]
\end{aligned} \tag{2.25}$$

Since we assume discrete cumulative hazards then, we expect $0 < p_i \leq 1$ for each event time and $p_n = 1$ for an event time t_n [92]. Thus, this likelihood can be rewritten as

$$L_{EB}(p_i) = \prod_{i=1}^n \left[p_i^{\delta_i} \prod_{l < i} [1 - p_l] \right] \tag{2.26}$$

To obtain the NPMLE for the Binomial likelihood, we differentiate the logarithmic function of Equation (2.26) w.r.t. a fixed, but arbitrary, p_i and setting the derivative equal to zero as follows

$$\begin{aligned}
\frac{\partial \ell_{EB}}{\partial p_i} &= \frac{\partial}{\partial p_i} \left(\sum_{j=1}^n \delta_j \ln p_j + \sum_{j=1}^n \sum_{l \leq j} \ln [1 - p_l] \right) \\
&= \frac{\partial}{\partial p_i} \left(\sum_{j=1}^n \delta_j \ln p_j + \sum_{j=1}^n \sum_{l \leq j} \ln [1 - p_l] \right) \\
&= \frac{\delta_i}{p_i} - \frac{\sum_{l=1}^n \mathbb{1}_{[l > i]}}{1 - p_i}
\end{aligned} \tag{2.27}$$

It can be seen that the log-likelihood function reaches its maximum by the Nelson-Aalen estimation of the hazard function, $\hat{p}_i = \frac{\delta_i}{n_i}$, where $n_i = \sum_{l=1}^n \mathbb{1}_{[l \geq i]}$ represents the number of individuals in the risk set prior to time t_i . When ties occur, Li [57] and Thomas and Grunkemeier [85] suggested, although in different notations, that the binomial empirical likelihood in Equation (2.26) should be expressed as follows

$$L_{EB}(p_i) = \prod_{i=1}^k [p_i^{d_i} (1 - p_i)^{n_i - d_i}] \tag{2.28}$$

where d_i reflects the number of events at the ordered time t_i , and k is the number of distinct event times. The resulted estimators when ties occur are $\hat{h}(t_{(i)}) = \frac{d_i}{n_i}$ and $\hat{S}(t) = \prod_{t_{(l)} \leq t} [1 - \hat{h}(t_{(l)})] = \prod_{l: t_{(l)} \leq t} [1 - \frac{d_l}{n_l}]$. This proves that the Kaplan-Meier estimator of the survival function and the Nelson-Aalen estimator of the hazard function are the NPMLE maximizes the full empirical likelihood for survival data

On the other hand, the Poisson empirical likelihood is determined based on the continuous relation between the survival and the cumulative hazard functions as in Equation (2.9), $S(t) = \exp(-H(t))$ with $f(t) = h(t) \exp(-H(t))$. As indicated by Kiefer and Wolfowitz [47] and Johansen [44], the assumption of a purely continuous cumulative hazard function does not lead to an MLE for this empirical likelihood [62]. Consequently, expanding the space of the cumulative hazard function by including discrete cumulative hazard functions is crucial to obtain an MLE [3]. The cumulative hazard function is determined by $H(t) = \sum_{t_i \leq t} h(t_i) = \sum_{t_i \leq t} p_i$ in this case. This may result in an inconsistency issue which will be addressed later. The asymptotic, Poisson, empirical likelihood for survival data, L_{EP} , is given by

$$\begin{aligned} L_{EP}(p_i) &= \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \\ &= \prod_{i=1}^n [p_i^{\delta_i} \exp(-H(t_i))] \\ &= \prod_{i=1}^n \left[p_i^{\delta_i} \exp \left(- \sum_{l \leq i} p_l \right) \right] \end{aligned} \quad (2.29)$$

One may verify that the NPMLE for the Poisson empirical likelihood is $\hat{p}_i = \frac{\delta_i}{n_i}$ by taking the derivative of the logarithmic function of Equation (2.29) with respect to a fixed p_i and setting it to zero as follows

$$\begin{aligned} \frac{\partial \ell_{EP}}{\partial p_i} &= \frac{\partial}{\partial p_i} \left(\sum_{j=1}^n \delta_j \ln p_j - \sum_{j=1}^n \sum_{l \leq j} p_l \right) \\ &= \frac{\delta_i}{p_i} - \sum_{l=1}^n \mathbb{1}_{[l \geq i]} \end{aligned} \quad (2.30)$$

A noteworthy finding is that the Nelson-Aalen estimation of the hazard function maximizes both, the Binomial and the Poisson, empirical likelihoods. As it has been indicated, changing from continuous to discrete cumulative hazard may cause inconsistent issue in the Poisson likelihood due to the fact that $\prod_{t_i \leq t} [1 - \frac{\delta_i}{n_i}] \neq \exp\left(-\sum_{t_i \leq t} \frac{\delta_i}{n_i}\right)$. With a continuous CDF, the issue will disappear, since these jumps will converge uniformly to zero as the number of observations increases; however, the issue will persist when the distribution function is discrete regardless of the number of observations [92].

In general, hazard function of a discrete distribution imposes the constraint that jumps must be strictly less than one, except for the last event time which is equal one. Therefore, it is recommended to consider the binomial version if the CDF is discrete since the estimates obtained using the binomial are equal to the estimates obtained from the empirical likelihood based on the CDF [62]. As opposed to the Binomial version, the Poisson empirical likelihood does not impose any restrictions on the height of jumps in the hazard function, which makes it an attractive option when analyzing complex models [92].

2.4 Proportional hazards model

2.4.1 Introduction

A semiparametric proportional hazards model was presented by Cox [25], which had a profound impact on the development of survival analysis, specifically in medical research. Unlike the previous methods, the proportional hazards model provides a methodology that allows researchers to investigate the relationship between survival experience and other covariates such as age, gender, heart rate, etc. In order to implement this model, an essential assumption must be met, namely that proportional hazards exist [25]. A comparison of the survival curves obtained by Kaplan-Meier with the survival curves generated by the PH model for a data set from the literature [21] is illustrated in Figure 2.1. This figure illustrates how the PH model uses the entire event times associated with both groups in order to estimate the survival function for each group. On the other hand, the Kaplan-Meier estima-

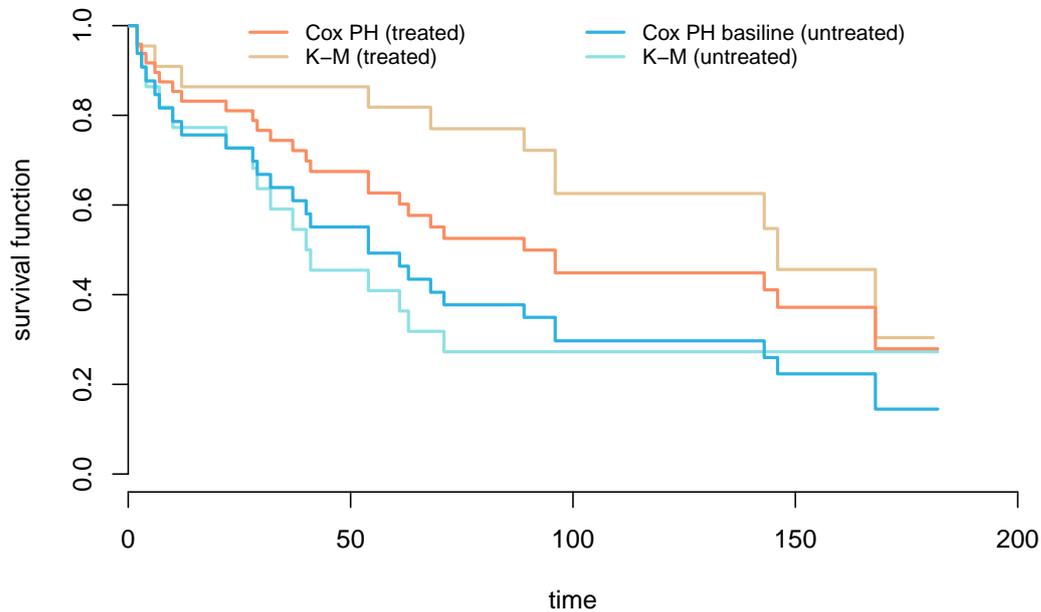


Figure 2.1: Kaplan-Meier versus Cox estimations of the survival function

tor uses event times that correspond to each group separately in order to estimate survival functions.

Let us consider a comparison between two groups in which $h_a(t)$ represents the hazard function of patients under treatment, while $h_b(t)$ represents a placebo group. It follows that the proportional hazards model assumes that $h_a(t) = \omega h_b(t)$, where ω does not depend on time. Consequently, the quantity ω is referred to as the hazard ratio, which indicates the relative risk of a patient receiving treatment as compared to a placebo patient. Suppose that X_1, X_2, \dots, X_p are covariates recorded for each individual such that $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})$ corresponds to the values of these covariates for i th individual. Then, the proportional hazards model for an individual i is defined by

$$h_i(t | \mathbf{x}_i) = h_0(t)\phi(\mathbf{x}_i; \boldsymbol{\beta}) \text{ or } h_i(t) = h_0(t)\phi_i \quad (2.31)$$

where $\phi_i = e^{\boldsymbol{\beta}^T \mathbf{x}_i} = e^{\sum_{j=1}^p \beta_j x_{ji}}$ which clearly does not depend on the time, $\boldsymbol{\beta}$ is a $p \times 1$ vector of Cox regression coefficients, and the term $h_0(t)$ is unknown function called

the baseline hazard function which can be considered as the hazard for individuals with $\boldsymbol{x} = \mathbf{0}$ [21]. Since the baseline hazard function, h_0 , is completely unknown, this model is referred to as semiparametric [25, 27].

2.4.2 Estimating the PH Model

Cox [26] illustrated the estimation the β -coefficients obtained by a new approach called Partial Likelihood which is based on the information that does not engage either the nuisance function $h_0(t)$ or the actual data times (censored and failure) directly, but only failure rank [21, 26, 61]. In the construction of the partial likelihood estimation, it is assumed that the model in Equation (2.31) is based on event times "failure times" and individuals who failed at these times. In other words, the likelihood function is uninformative in the interval between any successive failure times. Suppose a data set consisting of n individuals and consider $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ distinct ordered failure times where $k \leq n$. In addition, let R_j represent the set of individuals at risk at time $t_{(j)}$, i.e., individuals who are surviving at $t_{(j)} - \varepsilon$. Then, the probability that the j^{th} individual with \boldsymbol{x}_j fails at time t_j given one individual from R_j failed at time t_j is given by [21, 27]

$$\left(\frac{h_j(t)}{\sum_{l \in R_j} h_l(t)} \right)^{\delta_j} \quad (2.32)$$

By utilizing Equation (2.31) and canceling out the nuisance function, $h_0(t)$ from the numerator and the denominator

$$\left(\frac{\phi_j}{\sum_{l \in R_j} \phi_l} \right)^{\delta_j} \quad (2.33)$$

Thus, Cox [25] proposed the following likelihood function for the model introduced in Equation (2.31)

$$PL(\beta) = \prod_{j=1}^k \left(\frac{\phi_j}{\sum_{l \in R_j} \phi_l} \right)^{\delta_j} \quad (2.34)$$

In case there is censoring and death occur at the same, then it is assumed that all censoring occur after the death [21]. Without loss of generality, consider the case with only one covariate. Then, the parameter β can be estimated by taking the log of $PL(\beta)$ as follows

$$\begin{aligned}
 l(\beta) &= \sum_{j=1}^k \delta_j \ln \left[\frac{\phi_j}{\sum_{l \in R_j} \phi_l} \right] \\
 &= \sum_{j=1}^k \left[\delta_j \ln \phi_j - \delta_j \ln \sum_{l \in R_j} \phi_l \right] \\
 &= \beta \sum_{j=1}^k \delta_j x_j - \sum_{j=1}^k \delta_j \ln \left[\sum_{l \in R_j} e^{\beta x_l} \right]
 \end{aligned} \tag{2.35}$$

The score function of the log partial likelihood is

$$U(\beta) = \frac{\partial}{\partial \beta} l(\beta) = \sum_{j=1}^k \delta_j x_j - \sum_{j=1}^k \delta_j \left[\frac{\sum_{l \in R_j} x_l e^{\beta x_l}}{\sum_{l \in R_j} e^{\beta x_l}} \right] \tag{2.36}$$

where the negative second derivative, information matrix, of the log partial likelihood is

$$\begin{aligned}
 I(\beta) &= -\frac{\partial^2}{\partial \beta^2} l(\beta) \\
 &= \sum_{j=1}^k \delta_j \left[\frac{\left(\sum_{l \in R_j} x_l^2 e^{\beta x_l} \right) \left(\sum_{l \in R_j} e^{\beta x_l} \right) - \left(\sum_{l \in R_j} x_l e^{\beta x_l} \right)^2}{\left(\sum_{l \in R_j} e^{\beta x_l} \right)^2} \right]
 \end{aligned} \tag{2.37}$$

Now, one can utilize any optimization standard derivative-based method, such as Newton-Raphson, to obtain the maximum likelihood estimate of β . There are alternative methods of estimating the PH model's parameters which will be discussed later.

2.4.3 Tied Survival Times

Partial likelihood assumes the absence of ties in survival times. However, this assumption is frequently violated in real-life data due to data collection methods or purely discrete survival times [58]. Observations censored at the time of a failure have lived at least beyond that time, so it is reasonable to assume that they occurred after the failure [27]. The issue arises when tied observations are related to event times. Several approaches have been proposed to deal with tied observations across event times in the PH model. This section presents three of the most common approaches, namely the exact method [46], Breslow approximation [13], and Efron's approximation [29]. While the exact method is the most accurate approach due to the inclusion of all possible permutations of those tied times, it is not practical given its computational cost [21, 81]. Breslow suggested a crude approximation to resolve the tied survival times issue. Consider k distinct event times, and let D_j denote the set of all individuals who had an event at time $t_{(j)}$, d_j be the number of observations in D_j , and suppose that $\varphi_j = \prod_{i \in D_j} \phi_i$. Then, Breslow's approximation is given by

$$PL_b(\beta) = \prod_{j=1}^k \frac{\varphi_j}{\left[\sum_{l \in R_j} \phi_l \right]^{d_j}} \quad (2.38)$$

While Efron's numerator shares the same numerator as Breslow's, Efron's denominator is differs by using the summation over a simplified version of the permutations as follows

$$PL_e(\beta) = \prod_{j=1}^k \frac{\varphi_j}{\prod_{s=1}^{d_j} \left[\sum_{l \in R_j} \phi_l - \left(\frac{s-1}{d_j} \right) \varrho_j \right]} \quad (2.39)$$

where $\varrho_j = \sum_{i \in D_j} \phi_i$. Both approximations result in computationally efficient solutions. However, Efron's approach performs better, in terms of the accuracy, when the number of tied event times is large with respect to the size of the data set [2]. Consequently, Efron's method is the default method for fitting the proportional hazards model when ties occur in the R and the S software [81, 83]. Where tied observations appear in this thesis, ties will be broken by adding very small fractions to tied observations.

2.4.4 Estimating the hazard and survival functions

The emphasis in the previous subsection has been on the regression coefficients, β , of the proportional hazards model. These estimates of β 's are informative to study the relation between the covariate and the risk as well as to compare the risk of different individuals or groups utilizing the hazard ratio. However, information about the nuisance parameter, baseline hazard, is necessary to draw inferences about the survival distribution of an individual with specific covariates values. Suppose that a Cox PH model has been fitted to a given data with p covariates $\mathbf{X}_1, \dots, \mathbf{X}_p$ and the estimated coefficients of the fitted Cox PH model were $\hat{\beta}_1, \dots, \hat{\beta}_p$, so the hazard function of the i th individual as mentioned in Equation (2.31) is given by

$$h_i(t) = h_0(t)\phi_i \quad (2.40)$$

Assume that we are interested to estimate the survival function of individual with a specific combination of covariates vector, \mathbf{x}_i . By integrating both sides of Equation (2.40) we obtain

$$\begin{aligned} \int_0^t h_i(u)du &= \int_0^t h_0(u)\phi_i du \\ &= \phi_i \int_0^t h_0(u)du \end{aligned} \quad (2.41)$$

From the definition of cumulative hazard function

$$H_i(t) = H_0(t)\phi_i \quad (2.42)$$

Consequently, the survival function of the i th individual is given by

$$\begin{aligned} S_i(t) &= e^{-H_i(t)} \\ &= e^{-H_0(t)\phi_i} \\ &= S_0(t)^{\phi_i} \end{aligned} \quad (2.43)$$

which forms a Lehmann family of distributions that has the form of $S_0(t)^m$; where $0 < m < \infty$ [46]. There are several methods for estimating the baseline survival function in the proportional hazards model. Cox [25] proposed an estimation of the survival function using an iterative method utilizing the parameters that obtained by

the maximum partial likelihood; however, it is a complicated estimate [13, 25]. Alternatively, Breslow [13] and Kalbfleisch and Prentice [45] estimators are considered as the most common nonparametric approaches to estimate the baseline survival function in the Cox PH model. The Breslow estimator, b , is constructed based on estimating the cumulative baseline hazard function, $H_0^b(t)$, which is analogous to the Nelson-Aalen estimator in the absence of a covariate, see Section 2.2.2. Breslow's cumulative baseline hazard estimator is determined by utilizing the estimated parameters obtained by the partial likelihood as follows

$$\hat{H}_0^b(t) = \sum_{j:t_j \leq t} \frac{\delta_i}{\sum_{l \in R_j} \hat{\phi}_l} \quad (2.44)$$

where R_j is the set of all individuals under risk at time t_j , and $\hat{\phi}_l = e^{\hat{\beta}x_l}$. Hence, the baseline survival function can be obtained by substituting $\hat{H}_0^b(t)$ in Equation (2.9) as follows

$$\hat{S}_0^b(t) = \exp[-\hat{H}_0^b(t)] = \exp \left[- \sum_{j:t_j \leq t} \frac{\delta_i}{\sum_{l \in R_j} \hat{\phi}_l} \right] \quad (2.45)$$

In the case of ties, the summation will be active over only events times, $t_{(j)}$, and δ_i will be replaced by where d_j is the number of failures at time $t_{(j)}$.

In contrast, Kalbfleisch and Prentice estimator, kp , extends the concept of the Kaplan-Meier estimator, see Section 2.2.1, by the discretization of failure times to approach a continuous function as follows [21, 45]

$$\hat{S}_0^{kp}(t) = \prod_{j:t_j \leq t} \hat{\zeta}_j \quad (2.46)$$

where $\hat{\zeta}_j = 1 - \hat{h}_0(t_j)$ is the conditional probability of surviving at time t_j for the baseline individual which is zero at non event times. In the absence of ties, we have

$$\begin{aligned} \frac{\hat{\phi}_j}{1 - \hat{\zeta}_j^{\hat{\phi}_j}} = \sum_{l \in R_j} \hat{\phi}_l &\Leftrightarrow 1 - \hat{\zeta}_j^{\hat{\phi}_j} = \frac{\hat{\phi}_j}{\sum_{l \in R_j} \hat{\phi}_l} \\ &\Leftrightarrow \hat{\zeta}_j^{\hat{\phi}_j} = 1 - \frac{\hat{\phi}_j}{\sum_{l \in R_j} \hat{\phi}_l} \\ &\Leftrightarrow \hat{\zeta}_j = \left[1 - \frac{\hat{\phi}_j}{\sum_{l \in R_j} \hat{\phi}_l} \right]^{\hat{\phi}_j^{-1}} \end{aligned}$$

Hence, the solution of Equation (2.48) is given by

$$\zeta_j = \left[1 - \frac{\hat{\phi}_j}{\sum_{l \in R_j} \hat{\phi}_l} \right]^{\hat{\phi}_j^{-1}} ; \quad \hat{\phi}_j^{-1} = e^{-\beta x_j} \quad (2.47)$$

When ties occur, the conditional probability ζ_j has no closed form and can be estimated as a solution for the following equation

$$\sum_{k \in D_j} \frac{\hat{\phi}_k}{1 - \zeta_j^{\hat{\phi}_k}} = \sum_{l \in R_j} \hat{\phi}_l \quad (2.48)$$

where D_i is the set of individuals who failed at $t_{(i)}$.

2.5 Empirical full likelihood for the PH model

A brief description of the empirical full likelihood of the PH model is provided in this section. Pan [69] suggested an empirical likelihood approach for Cox PH model to estimate the regression coefficient, β , and $H_0(t)$, the baseline cumulative hazard function, by profiling out $H_0(t)$. As a result, the estimated baseline cumulative hazard is Breslow's estimator with a well known drawback when paired with $S(t) = \prod_{t_i \leq t} [1 - h(t_i)]$ that its baseline hazard can exceed one which leads to negative values of the estimated baseline survival function, $S_0(t)$, in case of discrete distributions [27, 46, 77]. Furthermore, the regression parameter estimate obtained from the Poisson empirical likelihood under the PH model is identical to the estimate derived from Cox's partial likelihood, differing only in their log-likelihood values [92]. In contrast, Efron [29] proposed the full empirical likelihood for the PH model to show that Cox's partial likelihood function contains nearly all of the information related to the regression parameters. Ren and Zhou [77] proposed the same empirical likelihood for the PH model corresponding to β and $F_0(t)$ where the baseline distribution function is profiled out to attain the full profile likelihood function for β . For small or moderate sample sizes, this method performs better than Cox's partial likelihood estimator, in terms of simulation standard error, simulation relative bias, and simulation relative MSE [77]. The following subsections describe the construction of the PH model using both the Poisson empirical likelihood and the empirical likelihood based on F_0 .

2.5.1 Poisson empirical likelihood for the PH model

This section illustrates the Poisson full empirical likelihood for the PH model, as described by Pan [69] and Zhou [92]. The Poisson likelihood for the PH model is constructed using the same methodology as presented in Section 2.3.2 for survival data, but with the presence of a covariate. Based on the PH model, the hazard function and cumulative hazard function of the i th individual are

$$h_i(t) = h_0(t)\phi_i, \quad H_i(t) = H_0(t)\phi_i \quad (2.49)$$

with $\phi_i = e^{\beta x_i}$ as used in Equation (2.31). According to Equation (2.29), the full likelihood associated with the proportional hazards model is given by

$$\begin{aligned} L(\beta, h_0(t_1), h_0(t_2), \dots, h_0(t_n)) &= \prod_{i=1}^n [h_i(t_i)^{\delta_i} \exp(-H_i(t_i))] \\ &= \prod_{i=1}^n [(h_0(t_i)\phi_i)^{\delta_i} \exp(-H_0(t_i)\phi_i)] \end{aligned} \quad (2.50)$$

While using the analogy described in Section 2.3.2, the baseline cumulative hazard should be considered to be the sum of the masses assigned to observed times such that $h_0(t_i) = \Delta H_0(t_i) = p_i$ and $H_0(t_i) = \sum_{l \leq i} \Delta H_0(t_l) = \sum_{l \leq i} p_l$ [92]. Consequently, the Poisson empirical likelihood for the PH model can be expressed as

$$L(\beta, p_1, \dots, p_n) = \prod_{i=1}^n \left[(p_i \phi_i)^{\delta_i} \exp \left(-\phi_i \sum_{l \leq i} p_l \right) \right] \quad (2.51)$$

Rewriting the exponential terms of the likelihood function with reference to p_i will be useful, so

$$\begin{aligned} \prod_{i=1}^n \exp \left(-\phi_i \sum_{l \leq i} p_l \right) &= \exp(-\phi_1 p_1) \times \dots \times \exp(-\phi_n (p_1 + \dots + p_n)) \\ &= \prod_{i=1}^n \exp \left(-p_i \sum_{l \geq i} \phi_l \right) \\ &= \prod_{i=1}^n \exp(-p_i r_i) \end{aligned} \quad (2.52)$$

where $r_i = \sum_{l \geq i} \phi_l$, representing the summation of the covariate effects for individuals in the risk set at time $t_i - \varepsilon$. Consequently, the Poisson empirical likelihood for

the PH model is given by

$$L(\beta, p_1, \dots, p_n) = \prod_{i=1}^n \left[(p_i \phi_i)^{\delta_i} \exp(-p_i r_i) \right] \quad (2.53)$$

In order to maximize the likelihood function, let β be arbitrary, but fixed and consider the term involving p_i in Equation (2.53). If $\delta_i = 0$, the term $\exp(-p_i r_i)$ reaches its maximum value at $p_i = 0$, since r_i represents a summation of positive quantities. When $\delta_i = 1$, then the derivative of the log-likelihood function with respect to a particular p_i is given by

$$\begin{aligned} \frac{\partial \ell}{\partial p_i} &= \frac{\partial}{\partial p_i} \sum_{j=1}^n \ln [(p_j \phi_j) \exp(-p_j r_j)] \\ &= \frac{1}{p_i} - r_i \end{aligned} \quad (2.54)$$

Now, by setting the derivative of the log-likelihood to zero we obtain the estimation $\hat{p}_i = \frac{1}{r_i}$. Since the second derivative of log-likelihood function with respect to p_i is negative, the likelihood function in Equation (2.53) attains its maximum for

$$\hat{p}_i = \frac{\delta_i}{r_i} \quad (2.55)$$

The profile Poisson likelihood function of the PH model can be obtained by substituting \hat{p}_i into Equation (2.53) as follows

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n [\hat{p}_i \phi_i]^{\delta_i} \exp(-\hat{p}_i r_i) \\ &= \prod_{i=1}^n \left[\frac{\phi_i \exp(-1)}{r_i} \right]^{\delta_i} \end{aligned} \quad (2.56)$$

In the absence of a covariate, this likelihood is reduced to the likelihood function in Equation (2.29). In the following theorem, it is shown how this likelihood relates to the Cox's partial likelihood.

Theorem 2.5.1 (Zhou [92, p. 95])

Consider the case where there is no tie in the observed times. The estimate obtained from maximizing the profiled log-likelihood in Equation (2.56) is equivalent to the estimate obtained from the Cox's partial log-likelihood in Equation (2.34).

Additionally, the difference between the log-likelihood values is equal to the number of times the event has occurred in the sample. Thus, the likelihood ratio will be the same regardless of which likelihood is used.

Proof.

This can be verified by observing that the term $\exp(-1)$ in Equation (2.56) adds no information about the regression parameter, so it can be excluded. It follows that the Poisson profiled likelihood is proportional to the partial likelihood, leading to a difference of $-\sum_{i=1}^n \mathbb{1}_{[\delta_i=1]}$, the number of events occurred in the sample, between the partial log-likelihood value and the Poisson empirical log-likelihood value for the PH model. \square

As a consequence of using the Poisson empirical likelihood for the PH model, it should be noted that the resulting cumulative baseline hazard and baseline survival functions will be Breslow's estimates as in Equations (2.44) and (2.45).

2.5.2 Empirical full likelihood for the PH model F_0

Our objective in this section is to explain the basic idea behind the construction of the empirical likelihood for PH model based on the baseline CDF, F_0 , which was demonstrated in Ren and Zhou [77]. Recall Equation (2.43), which is an essential tool to acquire the exact full likelihood function for the PH model;

$$S(t | x_i) = S_0(t)^{\phi_i}$$

By differentiating both sides of Equation (2.43) w.r.t. t and using $S(t) = 1 - F(t)$ we have

$$\begin{aligned} \frac{d}{dt} S(t | x_i) &= \frac{d}{dt} S_0(t)^{\phi_i} \Rightarrow -f(t | x_i) = \phi_i [-f_0(t)] [S_0(t)]^{\phi_i - 1} \\ &\Rightarrow f(t | x_i) = \phi_i f_0(t) S_0(t)^{\phi_i - 1} \end{aligned} \quad (2.57)$$

Then, the likelihood function of $(t_i, \delta_i | x_i)$ under the proportional hazards model can be obtained by substituting Equations (2.43) and (2.57) into Equation (2.24),

as follows

$$\begin{aligned}
\prod_{i=1}^n f(t_i, \delta_i | x_i) &\propto \prod_{i=1}^n f(t_i | x_i)^{\delta_i} S(t_i | x_i)^{1-\delta_i} \\
&\propto \prod_{i=1}^n [\phi_i f_0(t_i) S_0(t_i)^{\phi_i-1}]^{\delta_i} [S_0(t_i)^{\phi_i}]^{1-\delta_i} \\
&\propto \prod_{i=1}^n [\phi_i f_0(t_i)]^{\delta_i} [S_0(t_i)^{\phi_i-1}]^{\delta_i} [S_0(t_i)^{\phi_i}]^{1-\delta_i} \\
&\propto \prod_{i=1}^n [\phi_i f_0(t_i)]^{\delta_i} [S_0(t_i)]^{\phi_i \delta_i - \delta_i + \phi_i - \phi_i \delta_i} \\
&\propto \prod_{i=1}^n [\phi_i f_0(t_i)]^{\delta_i} [S_0(t_i)]^{\phi_i - \delta_i}
\end{aligned}$$

Accordingly, the full likelihood function for (β, F_0) under the proportional hazards model with the survival data (2.20) is given by

$$L(\beta, F_0) = \prod_{i=1}^n [\phi_i f_0(t_i)]^{\delta_i} [S_0(t_i)]^{\phi_i - \delta_i} \quad (2.58)$$

It is noteworthy that in the absence of covariates, i.e., $\phi_i = 1$ for every individual, the empirical likelihood in Equation (2.58) reduces to the binomial likelihood discussed in Section 2.3.2, see Zhou [92].

Theorem 2.5.2 (Ren and Zhou [77])

The baseline survival function in Equation (2.58) can be maximized for any fixed value of β by

$$\hat{S}_n(t) = 1 - \hat{F}_n(t) = \prod_{i:t_i \leq t} \frac{r_i - \delta_i}{r_i} \quad (2.59)$$

where $r_i = \sum_{j=i}^n \phi_j$. Consequently, the profile full likelihood function for the PH model is given by

$$L_p(\beta) = \prod_{i=1}^n \left[\frac{\phi_i}{r_i} \right]^{\delta_i} \left[\frac{r_i - \delta_i}{r_i} \right]^{r_i - \delta_i} \quad (2.60)$$

The proof of Theorem 2.5.2 can be found in Ren and Zhou [77] and in Appendix A.2. It should be recognized that the profile likelihood function in Equation (2.60) remains valid only when $r_i \geq 1$ for event times. Consequently, optimization of the profile likelihood function is conducted under the constraint that $\phi_n \geq 1$ to ensure the validity of $r_i \geq 1$ for event times. This constraint can be satisfied by

adjusting the covariate x for each individual such that $\tilde{x}_i = x_i - x_n$ before optimizing the profile likelihood. Therefore, the subscript n in Equation (2.59) indicates that the resulting survival function, obtained by substituting the MLE of Equation (2.60) into Equation (2.59), represents the survival function for individuals with $x = x_n$. As a result, the baseline survival function can be estimated by $S_0(t) = \hat{S}_n^{\exp[-\hat{\beta}x_n]}$.

2.6 Simulating survival data for the PH model

2.6.1 Parametric PH models

This section presents some of the well known parametric PH models and their related functions in order to facilitate later simulation studies. When the assumed theoretical distribution for survival times is plausible, parametric models provide additional insight into the nature survival data, particularly the hazard rate [59]. The exponential distribution, for instance, is appealing to model the survival time for a population with constant hazard. On the other hand, populations with monotonously increasing or decreasing hazards can be modeled using the Weibull model. However, if the distribution assumption is incorrect, the results will be misleading. Therefore, when the assumed distribution is plausible, parametric tests are more effective than nonparametric tests in this situation [21]. These models are particularly suited for analyzing the effects of certain covariates on survival experience due to their inclusion of covariates. In survival analysis, a variety of parametric models are available, such as exponential, Weibull, gamma, lognormal, and Gompertz. Once a particular model has been selected, the distribution function can be used to determine other functions such as the hazard and the survival functions, see Section 2.1. In this thesis, three of the most common parametric distributions that exhibit the proportional hazards aspect are used. Table 2.1 provides a summary of various features of these three distributions.

Feature	Distribution			
	Exponential	Weibull	Gompertz	
	$T \sim \text{Exp}(\lambda)$	$T \sim \text{Weibull}(\lambda, \rho)$	$T \sim \text{Gompertz}(\lambda, \rho)$	
Parameter	$\lambda > 0$	$\lambda > 0$ and $\rho > 0$	$\lambda > 0$ and $\rho \in (-\infty, \infty)$	
Hazard	$h(t)$	λ	$\lambda \rho t^{\rho-1}$	$\lambda \exp(\rho t)$
Cumulative hazard	$H(t)$	λt	λt^ρ	$\frac{\lambda}{\rho} (\exp(\rho t) - 1)$
Density	$f(t)$	$\lambda \exp(-\lambda t)$	$\lambda \rho t^{\rho-1} \exp(-\lambda t^\rho)$	$\lambda \exp(\rho t) \exp\left(\frac{\lambda}{\rho} (1 - \exp(\rho t))\right)$
Survival	$S(t)$	$\exp(-\lambda t)$	$\exp(-\lambda t^\rho)$	$\exp\left(\frac{\lambda}{\rho} (1 - \exp(\rho t))\right)$

Table 2.1: Summary of the exponential, the Weibull and the Gompertz distributions

2.6.2 Survival times

This section illustrates how to generate survival data for Cox PH model as described by Bender, Augustin and Blettner [11]. The distribution function of the survival times in the PH model is given by

$$\begin{aligned}
 F(v|x_i) &= 1 - S(v|x_i) \\
 &= 1 - S_0(v)^{\phi_i}; \quad \text{by Equation (2.43)} \\
 &= 1 - e^{-H_0(v)\phi_i}
 \end{aligned} \tag{2.61}$$

For any continuous random variable Y with distribution function F , $W = F(Y)$ and $U = S(Y) = 1 - W$ are uniformly distributed from 0 to 1. Therefore, the survival time V of the PH model in Equation (2.31) can be generated by the following given that the cumulative baseline hazard function is invertible (i.e., positive baseline hazard for all time $v \geq 0$)

$$\begin{aligned}
 U &= e^{-H_0(V)\phi_i} \\
 \ln(U) &= -H_0(V)\phi_i \\
 H_0(V) &= \frac{-\ln(U)}{\phi_i} \\
 V &= H_0^{-1}\left(\frac{-\ln(U)}{\phi_i}\right)
 \end{aligned} \tag{2.62}$$

where $U \sim \text{Unif}(0, 1)$ can be easily generated in any statistical software such as R using the built-in `runif` function. There is no further assumption for β and X , but it is essential to specify β , X , and appropriate form of H_0^{-1} before generating the

survival times for the PH model using Equation (2.62). The hazard function can be modeled parametrically in a PH model using Exponential, Weibull and Gompertz distributions since the parameter λ affects the hazard multiplicatively in each family [54]. However, a more flexible parametric model might be considered too, for more details see Bender *et al.* [11, Section 2.5]).

Example 2.6.1 Based on the PH model, we may consider a Weibull distribution for baseline hazards, where the hazards may be fixed or monotonically increasing or decreasing. The Weibull distribution is constructed based on two positive parameters: the scale parameter, λ , and the shape parameter, ρ . As a special case of the Weibull distribution, survival times are exponentially distributed with a constant baseline hazard when $\rho = 1$. On the other hand, $\rho > 1$ or $0 < \rho < 1$ describe Weibull distribution survival times with monotonically increasing or decreasing baseline hazards, respectively. The cumulative hazard function for the Weibull model can be determined as follows

$$H_0(v) = \lambda v^\rho \quad (2.63)$$

Consequently, the inverse of the cumulative hazard function is given by

$$H_0^{-1}(z) = \left(\frac{z}{\lambda}\right)^{1/\rho} \quad (2.64)$$

The survival times of the PH model can now be calculated by substituting H_0^{-1} in Equation (2.62) as follows

$$V = \left(\frac{-\ln(U)}{\lambda\phi_i}\right)^{1/\rho} \quad (2.65)$$

Example 2.6.2 Similar to those used in Weibull distribution, we have two parameters in Gompertz distribution: a positive scale parameter denoted by λ and a shape parameter denoted by $\rho \in (-\infty, \infty)$. According to the Gompertz distribution, the cumulative hazard function is given by

$$H_0(v) = \frac{\lambda(e^{\rho v} - 1)}{\rho} \quad (2.66)$$

The inverse of the cumulative hazard function of the Gompertz distribution is therefore

$$H_0^{-1}(z) = \frac{1}{\rho} \ln \left(\frac{\rho z}{\lambda} + 1 \right) \quad (2.67)$$

To obtain the survival times of the PH model, substitute H_0^{-1} into Equation (2.62) as follows

$$V = \frac{1}{\rho} \ln \left(1 - \frac{\rho \ln(U)}{\lambda \phi_i} \right) \quad (2.68)$$

2.6.3 Censoring times

Censoring times for the PH model can be generated such that the proportion of censored observations is unaffected by either the choice of distribution of X or the conditional distribution of $V|X$. We first assume any distribution for X , regressing parameter β , and an appropriate form of the baseline hazards $h_0(t)$. The joint resulting distribution for (X, V) determines marginal distribution for V . Suppose that F_v represents the cumulative distribution function of the marginal distribution for V . We can now define the the random censoring time C which is independent of V and X as follows

$$C = F_v^{-1}(\tilde{C}) \quad (2.69)$$

where \tilde{C} is a random variable taking values in $[0, 1]$ and independent of X and V (e.g. one may consider $\tilde{C} \sim \text{Beta}(a, b)$ for some choice of a and b). In practice, one could replace the marginal distribution for V by the empirical distribution of V , \hat{F}_v , based on a large random sample from the marginal distribution of V . Then, C can be generated by Equation (2.69), using values of a and b that lead to the desired proportion of censored observations noting that $\tilde{C} \sim \text{Beta}(a, b)$ leads to approximately the fraction $b/(a + b)$ of right censored observations. For instance, using $\tilde{C} \sim \text{Beta}(4, 1)$ leads to approximately 20% of right censoring proportion.

2.7 Bootstrap methods for the PH model

Bootstrap is an excellent tool for generating pseudo-samples with specified properties, particularly when dealing with complex data structures, for a variety of purposes, such as estimating standard errors and confidence intervals. This section discusses bootstrap methods for the PH model, particularly, Zelterman's method which we have utilized in this thesis to assess the imprecise proportional hazards model. Efron [30] developed a bootstrapping technique for survival data that incorporates censoring observations into estimation of certain properties of the Kaplan-Meier survival function in the absence of covariates. For the purpose of accommodating covariates, Reid [76] proposed a conditional bootstrap method for generating survival times based on each covariate's value using the Kaplan-Meier function. A straightforward bootstrap methods for sampling survival data with covariates was described by Efron and Gong [31] as well as Efron and Tibshirani [32]. These two methods can be applied directly to Cox's model by resampling the triples (T_i, δ_i, X_i) non-parametrically from the original data either within each group or assuming random covariate. There is a possibility that bootstrap samples do not follow the proportional hazards assumption, which is considered a disadvantage of this method. Hjort [66] describes a bootstrap method for evaluating the sampling accuracy of the estimated coefficient associated with the proportional hazards model. The bootstrap scheme of Hjort assumes that the censoring variables are known precisely, due to a fixed endpoint [16]. It has been pointed out by Burr that the sensitivity of influential points might be a drawback of Hjort's method. That is, individuals with higher values of a covariate when $\hat{\beta}$ is positive tend to be censored, leading to a new bootstrap method by Burr. In fact, both Burr's and Hjort's methods generate survival and censored times separately with survival times conditioned on the exact values of covariates in the original data. Essentially, using the survival function $\hat{S}(t|x_i) = \hat{S}_0(t)^{\hat{\phi}_i}$ based on fitting the PH model to the original data one can generate event times, v^* , for bootstrap samples in these methods. In the case of Hjort's bootstrap, the resampled censoring times c^* can be determined for each individual by the random common distribution \hat{G} based on the Kaplan-Meier estimates for the survival function correspond to the censored observations. The only difference between

Hjort and Burr's methods is that $c^* = t$ for observed right censored observations, in Burr's. As opposed to these methods, the variation in survival and censoring times of the original data is better reflected in the bootstrap samples obtained by Zelterman *et al.*'s method [91]. In other words, the proportion of censored observations in all bootstrap samples is identical to that in the original data, since any right-censored observation in the original data will remain censored in all bootstrap samples as well. Bootstrap observations in this approach are generated by choosing a pair of time and censoring indicator (t^*, δ^*) from the observed data with replacements, then determine the covariate x^* based on a conditional distribution given the selected pair. The next two sections illustrate the two bootstrap approaches as described in Zelterman *et al.* [91] and N.L. Hjort [66], respectively.

2.7.1 Zelterman *et al.*'s bootstrap method for the PH model

There are two types of bootstrap approaches proposed by Zelterman *et al.* [91], a restricted one in which all the observed values of x must be contained in the bootstrap pseudo-samples, and an unrestricted one. In light of the fact that the restricted approach can be extremely time-consuming and computer-intensive, we have opted to use an unrestricted bootstrap method instead. In order to ensure clarity, we will limit our attention to observations related to event times. The bootstrap observations are denoted by an asterisk. Then, the bootstrap technique begins by considering τ_i as the probability of selecting the distinct pair $(t^* = t_i, \delta_i^* = 1)$ such that $\tau_i = P[t^* = t_i | \delta_i^* = 1] = n_i / \tilde{n}$ where n_i and \tilde{n} represent the number of events at time t_i and the total number of events in the original data, respectively. Additionally, assume the probability of selecting $x^* = x_j$ denoted by $\omega_j = P[x^* = x_j] = \sum_{l=1}^n 1_{[x_l=x_j]} / n$. Regarding the conditional probability of choosing $x^* = x_j$ given the selected pair $(t^* = t_i, \delta^* = 1)$, Zelterman *et al.* [91] utilized the approximation of the hazard function that $h(t|x) = \frac{P[(t^*, \delta^*)=(t_i, 1)|x]}{P[t^* \geq t_i, \delta^* = 1|x]}$ and applied Bayes' theorem to show that

$$\pi_{ij} = P[x^* = x_j | (t^*, \delta^*) = (t_i, 1)] = \frac{\phi_j (\tilde{n}\omega_j - \sum_{l < i} n_l \pi_{lj})}{\sum_{k=1}^n [\phi_k (\tilde{n}\omega_k - \sum_{l < i} n_l \pi_{lk})]} \quad (2.70)$$

Note that the interval sums of Equation (2.70) are zero for $l = 1$ and only run

over event times. The exterior summation, on the other hand, run over all possible values of x_j . In a similar manner, the conditional distributions of $x^* = x_j$ given the chosen pair $(t^* = t_i, \delta^* = 0)$ has been shown to be

$$\pi_{ij}^c = P[x^* = x_j | (t^*, \delta^*) = (t_i, 0)] = \frac{\sum_{l=u}^n \pi_{lj} \tau_l}{\sum_{l=u}^n \tau_l} \quad (2.71)$$

where $u = \min\{t : t \geq t_i\}$. As stated by Zelterman *et al.*, these conditional probabilities have some issues regarding the last observed time, t_n . First, π_{nj}^c is undefined when t_n corresponds to censored observation, so π_{nj}^c will be arbitrarily defined as equal to π_{ij} associated with the longest-lived non-censored observations. In addition, the authors suggested that this remedy could also be applied if $\pi_{nj} \notin [0, 1]$, for non-censored observation. Hence, this unrestricted bootstrap can be explained as follows [91]:

1. Fit the PH model to the original data and obtain an estimate for the regression parameter using either Cox's partial likelihood or any other estimator.
2. Assign τ_i as the probability of selecting the distinct pair $(t^* = t_i, \delta_i = 1)$ such that $\tau_i = n_i / \tilde{n}$ where n_i and \tilde{n} are the number of observed events at time $t_{(i)}$ and the total number of events in the dataset for the case of ties observations. The thesis assumes the absence of ties, so $\tau_i = 1 / \tilde{n}$ and $\tau_i^c = 1 / (n - \tilde{n})$ with n is the total number of observations.
3. Assign $\omega_j = \sum_{l=1}^n 1_{[x_l = x_j]} / n$, note x_j can be either a single value or a vector.
4. Use results from Steps 1-3 in Equation (2.70) to compute $\pi_{11}, \pi_{12}, \dots$, for all π_{ij} related to event times.
5. Use π_{ij} obtained from Step 4 in Equation (2.71) to determine π_{ij}^c .
6. Resample pairs of observations $(t_i^*, 1)$, $i = 1, \dots, \tilde{n}$ according to τ_i 's and use τ_i^c 's to resample pairs of observations $(t_i^*, 0)$, $i = \tilde{n} + 1, \dots, n$, both with replacement.
7. According to π_{ij} 's and π_{ij}^c 's from Steps 4 and 5 assign x^* for each selected pairs in Step 6.
8. Perform Steps 6-7 for B times to obtain B resamples.

2.7.2 Hjort's bootstrap method for the PH model

This section describes the bootstrap approach developed by Hjort. This bootstrap approach is relatively straightforward to implement than Zelterman *et al.*'s method. Although Hjort's bootstrap is acknowledged for its sensitivity to influential factors, its simplicity, particularly with continuous covariates, is one of its most compelling features. By contrast, Zelterman *et al.*'s method may yield negative probabilities when the covariate is continuous due to Equation (2.70). The resampled covariate values in Hjort's method are identical to those observed in the original data. Event times for individuals with covariate value x_i are resampled according to the following distribution function

$$\hat{F}(t_j|x_i) = 1 - \prod_{l \leq j} (1 - \Delta \hat{H}_0(t_l))^{\hat{\phi}_i} \quad (2.72)$$

where the estimates of the baseline hazard are given by

$$\Delta \hat{H}_0(t_j) = \min \left\{ \frac{d_j}{\sum_{l \geq j} \hat{\phi}_l}, 1 \right\} \quad (2.73)$$

Note that these baseline hazard estimates are a refined version of Breslow's estimates, and are intended to eliminate the possibility of negative probabilities resulting from $\Delta \hat{H}_0(t_j) > 1$. These distribution function estimates determine the resampled event times v_i^* related to individuals with $x = x_i$. Hjort proposed various schemes for resampling the right-censored times, c_i^* , for the i th individual. Among these schemes, assuming a fixed endpoint for the censoring time. An alternative approach involves assuming random censorship that is independent of event times [66]. In this approach, the right-censored times for all individuals are resampled from a common distribution function G , which is equivalent to Kaplan-Meier estimates based on the right-censored observations, with $1 - \delta_i$ replacing δ_i . The latter approach is employed in this thesis. Given $t_i^* = \min\{v_i^*, c_i^*\}$ and $\delta_i^* = I\{v_i^* \leq c_i^*\}$, the resampled data for the i th individual comprise (t_i^*, δ_i^*, x_i) . The procedure for conducting Hjort bootstrap is outlined as follows:

1. Fit the proportional hazards model to the original data and derive an estimate for the regression parameter, $\hat{\beta}$.

2. Substitute $\hat{\beta}$ into Equation (2.73) to estimate the baseline hazard function $\Delta\hat{H}_0(t_j)$.
3. Use $\hat{\beta}$ and $\Delta\hat{H}_0(t_j)$ to derive estimates of the CDF functions, \hat{F}_i , for each individual as described in Equation (2.72).
4. Use the pairs $(t_1, 1 - \delta_1), (t_2, 1 - \delta_2), \dots, (t_n, 1 - \delta_n)$ to estimate the distribution function for the resampled right-censoring times \hat{G} .
5. Resample pairs of observations (v_i^*, c_i^*) according to \hat{F}_i and \hat{G} , $i = 1, \dots, \tilde{n}$.
6. For each pair in Step 5, employ $t_i^* = \min v_i^*, c_i^*$ and $\delta_i^* = I\{v_i^* \leq c_i^*\}$ to construct the triplets (t_i^*, δ_i^*, x_i) , $i = 1, 2, \dots, n$.
7. Repeat Steps 5 – 6 for B times to obtain B resamples.

Chapter 3

Imprecise PH model

3.1 Introduction

The main intention of this chapter is to investigate opportunities for developing an imprecise proportional hazards model which can be employed when the PH assumption is questionable. The Poisson full empirical likelihood function for the PH model will be used to construct an imprecise proportional hazards model. To accomplish this, the hazard function for each individual will be augmented by including imprecision as an additional factor. Those additional factors can be incorporated either based on individual imprecision or based on group imprecision. In the individual-based imprecise model, IPH, the imprecision terms are incorporated into the hazard function of the i th individual at time t and permitted to vary throughout the lifespan of each individual separately. The group-based imprecise model, GPH, enables individuals within each group to share the same imprecision through dividing the data into groups according to similar characteristics, e.g., gender. These imprecision terms associated with the IPH or the GPH models are constrained to lie in a predetermined small interval.

This chapter is organized as follows. Section 3.2 discusses the construction of the Poisson full empirical likelihood function for individual-based imprecise PH model. Section 3.3 derives the Poisson full empirical likelihood related to the group-based imprecise PH model. In Section 3.4, the asymptotic behaviour as imprecision increases will be illustrated and the benefits of using the GPH model over the PH

model when the PH assumption is violated will be assessed. Section 3.5 includes comments on our findings as well as suggestions for possible future work.

3.2 Individual-based Imprecise PH model (IPH)

In the spirit of Section 2.5.1, we are constructing an individual-based imprecise PH, hereafter IPH model, model assuming that imprecision applies to each individual separately. The IPH model includes an extra positive multiplicative factor in the hazard function of the i th individual at time t such that the resulting model can be expressed as follows

$$h_i(t) = h_0(t) \exp(\epsilon_i(t)) \phi_i \quad (3.1)$$

The imprecision factors $\exp(\epsilon_i(t))$ in Equation 3.1 can vary throughout the lifespan of each individual, but are constrained by a prespecified value ϵ^* , such that

$$|\epsilon_i(t)| \leq \epsilon^* \quad (3.2)$$

Accordingly, the survival function, the cumulative hazard function, and the baseline hazard function at time t using the IPH model are given by

$$S_i(t) = \exp[-H_i(t)] \quad (3.3)$$

$$H_i(t) = \sum_{l:t_l \leq t} h_i(t_l) = \sum_{l:t_l \leq t} h_0(t_l) \exp(\epsilon_i(t_l)) \phi_i \quad (3.4)$$

$$h_0(t) = \Delta H_0(t) = H_0(t) - H_0(t^-) \quad (3.5)$$

Note that this model related to continuous time, so the continuous relation between the survival and the cumulative hazard functions was employed as in Equation (3.3). Due to the fact that the empirical likelihood will be written in term of hazard function and that the empirical likelihood can not optimized under the continuous distributions, the cumulative hazard function for particular individual at time t can be determined as summation for the hazard estimates related to that individual from time $t_0 = 0$ up to t . Consequently, the Poisson empirical likelihood function

for IPH model is given by

$$\begin{aligned} L(\boldsymbol{\psi}) &= \prod_{i=1}^n [h_i(t_i)^{\delta_i} \exp(-H_i(t_i))] \\ &= \prod_{i=1}^n \left[[h_0(t_i) \exp(\epsilon_i(t_i)) \phi_i]^{\delta_i} \exp \left(- \sum_{l \leq i} h_0(t_l) \exp(\epsilon_i(t_l)) \phi_i \right) \right] \end{aligned} \quad (3.6)$$

where $\boldsymbol{\psi} = \{\beta, h_0(t_1), \dots, h_0(t_n), \epsilon_1(t_1), \dots, \epsilon_1(t_n), \dots, \epsilon_n(t_1), \dots, \epsilon_n(t_n)\}$. The contribution of the i th individual to the empirical likelihood function under the IPH model is given by

$$\begin{aligned} L_i(\psi_i) &= h_i(t_i)^{\delta_i} \exp[-H_i(t_i)] \\ &= [h_0(t_i) \exp(\epsilon_i(t_i)) \phi_i]^{\delta_i} \exp \left(- \sum_{l \leq i} h_0(t_l) \exp(\epsilon_i(t_l)) \phi_i \right) \end{aligned} \quad (3.7)$$

where $\psi_i = \{\beta, h_0(t_1), \dots, h_0(t_i), \epsilon_i(t_1), \dots, \epsilon_i(t_i)\}$. Please note that throughout this thesis, we consider the convention that any expression of the form y^{δ_i} evaluates to zero when $y = 0$, even when $\delta_i = 0$. Therefore, when $\delta_i = 0$ the notation should be interpreted to mean $L_i = \exp(-\sum_{l \leq i} h_0(t_l) \exp(\epsilon_i(t_l)) \phi_i)$. For ease of notation, $h_0(t_i)$ has been replaced by p_i in what follows. In accordance with Equation (3.7), the likelihood function is

$$\begin{aligned} L(\boldsymbol{\psi}) &= [p_1 \exp(\epsilon_1(t_1)) \phi_1]^{\delta_1} \exp[-p_1 \exp(\epsilon_1(t_1)) \phi_1] \\ &\quad \times [p_2 \exp(\epsilon_2(t_2)) \phi_2]^{\delta_2} \exp[-p_1 \exp(\epsilon_2(t_1)) \phi_2 - p_2 \exp(\epsilon_2(t_2)) \phi_2] \\ &\quad \times [p_3 \exp(\epsilon_3(t_3)) \phi_3]^{\delta_3} \exp[-p_1 \exp(\epsilon_3(t_1)) \phi_3 - p_2 \exp(\epsilon_3(t_2)) \phi_3 - p_3 \exp(\epsilon_3(t_3)) \phi_3] \\ &\quad \times \dots \\ &\quad \times [p_n \exp(\epsilon_n(t_n)) \phi_n]^{\delta_n} \exp[-p_1 \exp(\epsilon_n(t_1)) \phi_n - \dots - p_n \exp(\epsilon_n(t_n)) \phi_n] \end{aligned}$$

By multiplying the contribution of all individuals and rearranging the likelihood function with reference to p_i , for $i = 1, 2, \dots, n$, one can write the full likelihood of the IPH model as

$$L(\boldsymbol{\psi}) = \prod_{i=1}^n [p_i \exp(\epsilon_i(t_i)) \phi_i]^{\delta_i} \exp(-p_i r_i^*) \quad (3.8)$$

where $r_i^* = \sum_{l \geq i} \exp(\epsilon_l(t_i)) \phi_l$ represents the sum of the imprecision factors times the covariate effect for all individuals at the risk at time t_i .

3.2.1 Profiling out the baseline hazard function

The next step is to profile out the terms p_i from the likelihood function. Consider arbitrary, but, fixed β and $\epsilon_i(t)$, and let $\delta_i = 0$, so the likelihood function of Equation (3.8) only involves the terms $\exp(-p_i r_i^*)$. Hence, the log-likelihood function can be maximized by taking $\hat{p}_i = 0$ when $\delta_i = 0$ since r_i^* is a sum of positive quantities. The derivative of the log-likelihood function with respect to a particular p_i when $\delta_i = 1$ is given by

$$\begin{aligned} \frac{\partial \ell}{\partial p_i} &= \frac{\partial}{\partial p_i} \left[\sum_{j=1}^n \delta_j \ln p_j + \delta_j \epsilon_j(t_j) + \delta_j \beta x_j - p_j r_j^* \right] \\ &= \frac{1}{p_i} - r_i^* \end{aligned} \quad (3.9)$$

Now, by setting the derivative of the log-likelihood to zero and solving for p_i we obtain $\hat{p}_i = \frac{1}{r_i^*}$. This maximizes the log-likelihood function when $\delta_i = 1$, because the second derivative of Equation (3.9) with respect to p_i is negative. The empirical likelihood function of the IPH model reaches its maximum at

$$\hat{p}_i = \frac{\delta_i}{r_i^*} \quad (3.10)$$

for $i = 1, 2, \dots, n$. Hence, the full profile likelihood function of the IPH model can be obtained by substituting \hat{p}_i from Equation (3.10) into Equation (3.8),

$$\begin{aligned} L^p &= \prod_{i=1}^n [\hat{p}_i \exp(\epsilon_i(t_i)) \phi_i]^{\delta_i} \exp(-\hat{p}_i r_i^*) \\ &= \prod_{i=1}^n \left[\frac{\delta_i \exp(\epsilon_i(t_i)) \phi_i}{r_i^*} \right]^{\delta_i} \exp(-\delta_i) \\ &= \prod_{i=1}^n \left[\frac{\delta_i \exp(\epsilon_i(t_i)) \phi_i \exp(-1)}{r_i^*} \right]^{\delta_i} \end{aligned} \quad (3.11)$$

Notice that the product in this likelihood is taken effectively only over event times and that the constant $\exp(-1)$ can be excluded from the likelihood function. Additionally, the full profile likelihood function in Equation (3.11) is well defined under the convention $0^0 = 1$, so the likelihood function is proportional to

$$L^p \propto \prod_{i=1}^n \left[\frac{\phi_i}{\exp(-\epsilon_i(t_i))r_i^*} \right]^{\delta_i} \quad (3.12)$$

3.2.2 Maximizing the likelihood function of the IPH model

Given a fixed value for β , the contribution of the likelihood function at time t_i can be expressed as follows

$$\begin{aligned} L_i^p &= \left[\frac{\phi_i}{\exp(-\epsilon_i(t_i))r_i^*} \right]^{\delta_i} \\ &= \left[\frac{\phi_i}{\exp(-\epsilon_i(t_i)) [\exp(\epsilon_i(t_i))\phi_i + \exp(\epsilon_{i+1}(t_i))\phi_{i+1} + \cdots + \exp(\epsilon_n(t_i))\phi_n]} \right]^{\delta_i} \\ &= \left[\frac{\phi_i}{\exp(\epsilon_i(t_i) - \epsilon_i(t_i))\phi_i + \exp(\epsilon_{i+1}(t_i) - \epsilon_i(t_i))\phi_{i+1} + \cdots + \exp(\epsilon_n(t_i) - \epsilon_i(t_i))\phi_n} \right]^{\delta_i} \\ &= \left[\frac{\phi_i}{\phi_i + \exp(\epsilon_{i+1}(t_i) - \epsilon_i(t_i))\phi_{i+1} + \cdots + \exp(\epsilon_n(t_i) - \epsilon_i(t_i))\phi_n} \right]^{\delta_i} \end{aligned} \quad (3.13)$$

The contribution L_i^p corresponding to event time t_i achieves its maximum by making the denominator as small as possible. That is, we consider the upper bound of imprecision for the i th individual where $i \neq n$ at time t_i , $\epsilon_i(t_i) = \epsilon^*$, and the lower bound of imprecision for all other individual in the risk set, $\epsilon_l(t_i) = -\epsilon^*$ for all $l > i$. Due to the fact that the profile likelihood is only be calculated at event times, imprecision terms associated with censoring times cannot be determined. Furthermore, the imprecision term corresponding to the last observation at t_n makes no contribution, since for $i = n$ we have

$$L_n^p = \left[\frac{\phi_n}{\exp(-\epsilon_n(t_n))r_n^*} \right]^{\delta_n} = \left[\frac{\phi_n}{\exp(-\epsilon_n(t_n)) \exp(\epsilon_n(t_n))\phi_n} \right]^{\delta_n} = 1$$

The likelihood function in Equation (3.12) attaining its maximum by setting

$$\epsilon_l(t_i) = \begin{cases} \epsilon^* & ; \text{ for } l = i < n \\ -\epsilon^* & ; \text{ for } l > i \\ \text{Not determined} & ; \text{ for censored times and } t_n \end{cases}$$

Hence, the maximized contribution of L_i^p at time t_i is

$$\begin{aligned} \max(L_i^p) &= \max \left[\frac{\phi_i}{\phi_i + \exp(\epsilon_{i+1}(t_i) - \epsilon_i(t_i))\phi_{i+1} + \cdots + \exp(\epsilon_n(t_i) - \epsilon_i(t_i))\phi_n} \right]^{\delta_i} \\ &= \left[\frac{\phi_i}{\phi_i + \exp(-2\epsilon^*)\phi_{i+1} + \cdots + \exp(-2\epsilon^*)\phi_n} \right]^{\delta_i} \\ &= \left[\frac{\phi_i}{\phi_i + \exp(-2\epsilon^*) \sum_{l>i} \phi_l} \right]^{\delta_i} \end{aligned} \quad (3.14)$$

Consequently, the full likelihood function of Equation (3.12) can be profiled as

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n \left[\frac{\phi_i}{\phi_i + \exp(-2\epsilon^*)r_{i+1}} \right]^{\delta_i} \\ &= \prod_{i=1}^n \left[\frac{\exp(\beta x_i)}{\exp(\beta x_i) + \exp(-2\epsilon^*)r_{i+1}} \right]^{\delta_i} \end{aligned} \quad (3.15)$$

where $r_{i+1} = \sum_{l>i} \phi_l = \sum_{l>i} \exp(\beta x_l)$ with $r_{n+1} = 0$. This profile likelihood has a single parameter to be estimated, β , and the likelihood value converges to 1 as $\epsilon^* \rightarrow \infty$. It should be noted from the first line of Equation (3.14) that this model considers the length of the imprecision interval rather than the exact value of the bounds. For instance, considering either $|\epsilon_i(t)| \leq 0.2$ or $\epsilon_i(t) \in [0, 0.4]$ for the imprecision terms will yield the same estimates because both intervals have equal length. Additionally, the estimated regression parameter using this likelihood and the partial likelihood of the PH model are equal when $\epsilon^* = 0$; however, the denominators of these likelihoods differ when $\epsilon^* > 0$. The denominator of the partial likelihood is $\sum_{l>i} \phi_l = \phi_i + \sum_{l>i} \phi_l$ while the likelihood of the IPH model has denominator $\phi_i + \exp(-2\epsilon^*) \sum_{l>i} \phi_l$, which leads to higher value of the likelihood function according to the chosen level of imprecision, ϵ^* . The estimated parameter obtained from maximizing with respect to β in Equation (3.15) will be denoted by $\hat{\beta}(\epsilon^*)$, which represents the estimated regression parameter $\hat{\beta}$ for a given level of imprecision ϵ^* .

Example 3.2.1 The purpose of this example is to demonstrate how each observed event time contributes to the imprecision of the IPH model via $\epsilon_i(t)$. In order

t_i	δ_i	x
10	1	1
12	0	0
14	1	0
16	1	1
18	0	1
20	1	0

Table 3.1: Survival data (Example 3.2.1)

to serve our purpose, an artificial survival data consisting of six observations has been created, as shown in Table 3.1. Consider the contribution of L_1 according to Equation (3.13). Only the following imprecision terms are needed in the evaluation of L_1 ,

$$\begin{aligned} \epsilon_{x_1}(t=10) &= \epsilon^* & \epsilon_{x_2}(t=10) &= -\epsilon^* & \epsilon_{x_3}(t=10) &= -\epsilon^* \\ \epsilon_{x_4}(t=10) &= -\epsilon^* & \epsilon_{x_5}(t=10) &= -\epsilon^* & \epsilon_{x_6}(t=10) &= -\epsilon^* \end{aligned}$$

It is obvious that $\delta_2 = 0$ when $i = 2$ results in $L_2 = 1$, thereby eliminating the need for imprecision terms at time $t = 12$. The same procedure may be followed to obtain the remaining imprecision terms, these are shown in Table 3.2. The number of imprecision factors that are required for evaluating the likelihood function is $\sum_{i=1}^{n-1} \delta_i(n-i+1)$.

◇

3.2.3 Hazard and Survival functions for the IPH model

In this section we illustrate the estimation of the hazard and survival functions for a particular individual using the IPH model. Equation (3.10) can be employed in conjunction with $\hat{\beta}(\epsilon^*), \hat{\epsilon}_1(t_1), \dots, \hat{\epsilon}_1(t_n), \dots, \hat{\epsilon}_n(t_1), \dots, \hat{\epsilon}_n(t_n)$ obtained through the optimization process to determine the baseline hazard function $h_0(t)$. Hence, the baseline hazard function for the IPH model can be derived as follows

x_i	1	0	0	1	1	0
t_i						
10	ϵ^*	$-\epsilon^*$	$-\epsilon^*$	$-\epsilon^*$	$-\epsilon^*$	$-\epsilon^*$
12 ⁺						
14			ϵ^*	$-\epsilon^*$	$-\epsilon^*$	$-\epsilon^*$
16				ϵ^*	$-\epsilon^*$	$-\epsilon^*$
18 ⁺						
20						

Table 3.2: Imprecision terms related to the likelihood function (Example 3.2.1) with the superscript, +, refers to right-censored observations

$$\begin{aligned}
h_0(t_j) &= \frac{\delta_j}{r_j^*} \\
&= \frac{\delta_j}{\sum_{l \geq j} \exp(\epsilon_l(t_j)) \phi_l} \\
&= \frac{\delta_j}{\exp(\epsilon_j(t_j)) \phi_j + \sum_{l > j} \exp(\epsilon_l(t_j)) \phi_l} \\
&= \frac{\delta_j}{\exp(\epsilon^*) \phi_j + \exp(-\epsilon^*) \sum_{l > j} \phi_l}
\end{aligned} \tag{3.16}$$

At right-censored and non-observed times, the baseline hazard function in the IPH model equals zero as a result of the profiled maximization of the empirical likelihood, see Equation (3.10). Due to the adherence to a Poisson empirical likelihood, the survival functions can be estimated by the continuous time relationship between cumulative hazard and survival functions as in Equation (2.9). Different variants of the hazard and survival functions can be constructed depending upon how imprecision values are assigned to the individual being studied. These variants can be classified into two classes: restricted and unrestricted functions.

I. Restricted hazard functions for the IPH model

Recall the hazard function of the IPH model for an individual with $x = x_i$ at time t

$$h_i(t) = h_0(t) \exp(\epsilon_i(t)) \phi_i$$

As a result of the optimization process, it is known that $\exp(\epsilon_i(t)) = \epsilon^*$ if that individual is observed to have had the event at time t and $\exp(\epsilon_i(t)) = -\epsilon^*$ otherwise. Consequently, the hazard function for the individual with $x = x_i$ is $h_0(t) \exp(-\epsilon^*) \phi_i$ in all survival times except for the time when the individual had the event, in which case the hazard function will be $h_0(t) \exp(\epsilon^*) \phi_i$. Suppose that there are several individuals with the same covariate value, x_i , then each will have its own hazard function. The hazard function for individuals with $x = x_i$ who did not experience the event will be unique and equal to $h_0(t) \exp(-\epsilon^*) \phi_i$ in all observed times. Meanwhile, the hazard functions for individuals who have experienced the event will be distinct. These individuals will have hazard functions equivalent to those right-censored individuals in all observed times, with the exception of a spike equal to $h_0(t) \exp(\epsilon^*) \phi_i$ reflecting the occurrence time of the event. To better understand this aspect, recall the artificial survival data in Table 3.1 and consider estimating the hazard function for individuals with $x = 1$ given $\epsilon^* = 0.5$. The estimated hazard functions related to individuals with $x = 1$, namely the first, fourth, and fifth individuals, are presented in Figure 3.1, along with Breslow's estimate of the hazard function using the PH model. The estimated hazard functions derived from the IPH model increase at times $t = 10$ and $t = 16$ for the first individual and the fourth individual, respectively. Additionally, the hazard functions of these individuals decrease if no events occur, and equal to the hazard function of the fifth individual, right-censored. Thus, the imprecise restricted hazard functions for the IPH model corresponding to individuals with $x = x_i$ can therefore be defined as follows

$$\bar{h}_i(t) = \max\{\mathfrak{H}(t|x = x_i)\} \quad (3.17)$$

$$\underline{h}_i(t) = \min\{\mathfrak{H}(t|x = x_i)\} \quad (3.18)$$

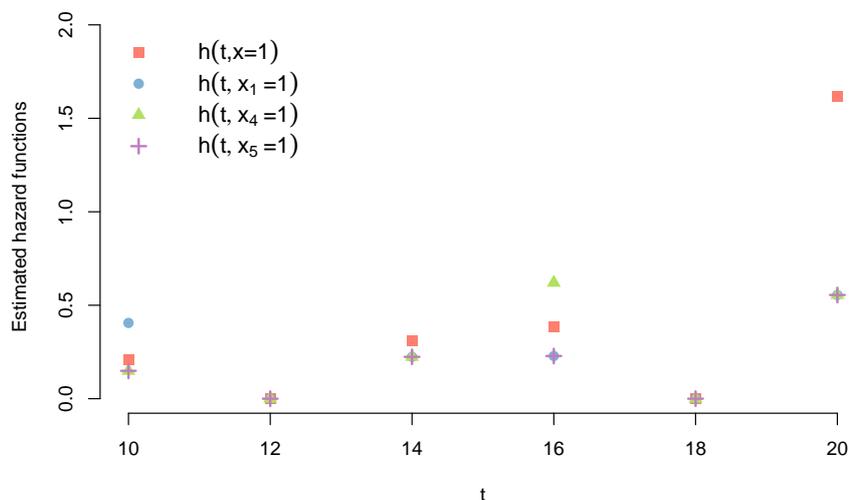


Figure 3.1: The estimated IPH restricted hazard functions for $x_1 = 1$, blue circles, $x_4 = 1$, green triangles, and $x_5 = 1$, indigo crosses using $\epsilon^* = 0.5$ along with Breslow estimates of the hazard function for individual with $x = 1$, red squares.

Where $\mathfrak{H}(t|x = x_i)$ represents the set of all hazard functions at time t for all individuals with $x = x_i$. As the estimated hazard function for the right-censored individual is always the lowest among others, it will serve as the lower hazard function for all individuals with $x = x_i$. In contrast, the upper hazard function is equal to the lower hazard function, but increases only if an individual with $x = x_i$ has the event, therefore the term restricted is used. Accordingly, the lower hazard function can be interpreted as the lower bounds for all of these hazard functions related to individuals with $x = x_i$ conditioned on the assumed level of imprecision, while the upper hazard function will be regarded as the upper bounds of these functions as illustrated in Figure 3.2. Hence, the PH assumption is relaxed by considering these functions as imprecise estimates of the hazard functions which include all possible hazard functions for the population of individuals with $x = x_i$.

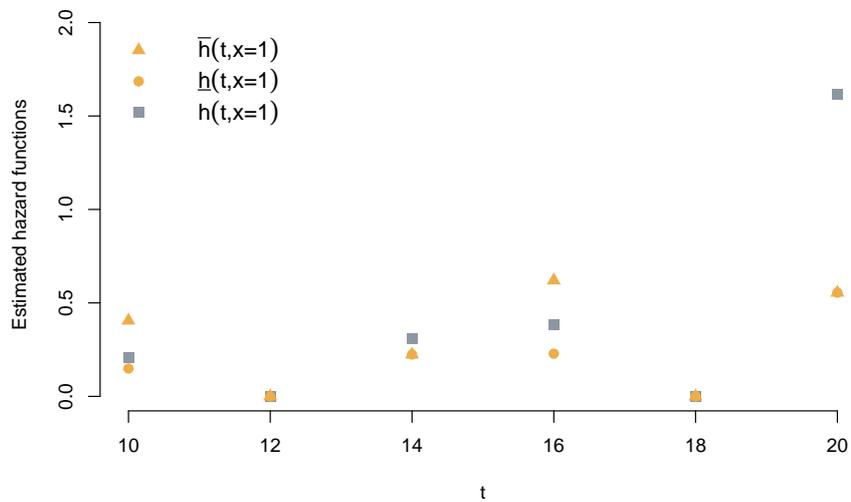


Figure 3.2: The imprecise restricted hazard estimates related to the IPH model for individual $x = 1$ using $\epsilon^* = 0.5$ along with Breslow estimates of the hazard function for the same individual.

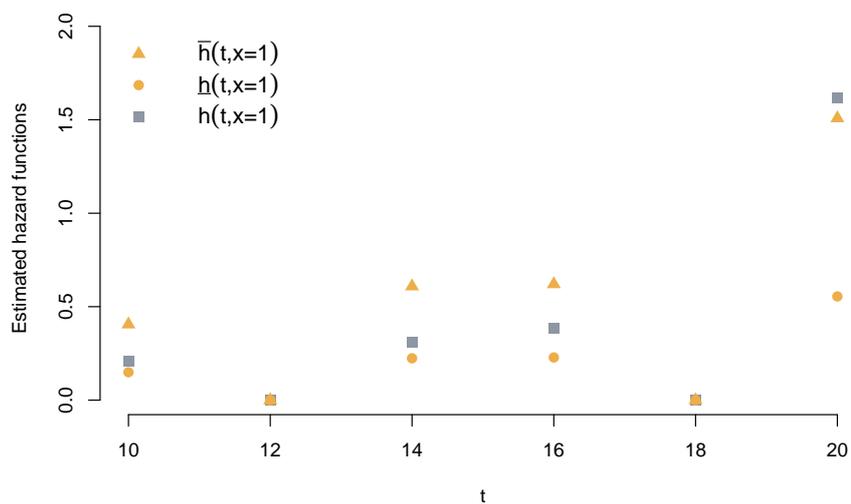


Figure 3.3: The imprecise unrestricted hazard functions related to the IPH model for individual $x = 1$ using $\epsilon^* = 0.5$ along with Breslow estimates of the hazard function for the same individual.

II. Unrestricted hazard functions for the IPH model

Based on the definition of the restricted hazard, in the case of continuous covariates, those individuals who have not experienced the event during the study period or whose covariate values have not been observed in the data will only have a lower hazard function. In other words, the estimated upper hazard functions for these individuals are equal to their estimated lower hazard functions. Therefore, another simple version of hazard functions, namely unrestricted, were introduced in which the assumed limits of imprecision are assigned to the hazard function directly, regardless of whether or not an individual with the same covariate value has experienced the event. Therefore, the upper hazard function will increase at all event times. Nevertheless, the estimated baseline hazard function will still be constrained to the values of imprecision factors that maximize the likelihood function. Hence, the estimated unrestricted imprecise hazard functions are given by

$$\bar{h}_i(t) = h_0(t) \exp(\epsilon^*)\phi_i \quad (3.19)$$

$$\underline{h}_i(t) = h_0(t) \exp(-\epsilon^*)\phi_i \quad (3.20)$$

The lower hazard functions resulting from either type are identical, whether restricted or unrestricted. This is because according to the definition of the restricted lower hazard function, the minimum restricted lower hazard function is always related to individuals who are right-censored or unobserved where the imprecision factor is equal to the lower imprecision bound. Based on the imprecision constraint, the estimated imprecise hazard functions may be interpreted as upper and lower bounds for all possible hazard functions for the entire population with the same covariate value, irrespective of whether individuals experienced the event or not.

III. Restricted and unrestricted survival functions for the IPH model

The imprecise restricted and unrestricted survival functions for the IPH model are determined by taking into account the imprecise upper and lower hazard functions using Equations (3.17) and (3.18) for the restricted type or Equations (3.19)

and (3.20) for the unrestricted as follows

$$\bar{S}(t; x_i) = \exp[-\underline{H}_i(t)] = \exp\left[-\sum_{j:t_j \leq t} \underline{h}_i(t_j)\right] \quad (3.21)$$

$$\underline{S}(t; x_i) = \exp[-\bar{H}_i(t)] = \exp\left[-\sum_{j:t_j \leq t} \bar{h}_i(t_j)\right] \quad (3.22)$$

A key feature of the IPH survival functions of either types is that the lower survival function tends to be highly effected by increasing the level of imprecision, ϵ^* , in comparison to the upper survival function. In the case of the restricted hazard functions, this feature can be justified by recalling the lower and upper hazard functions for individuals with $x = x_i$ at event time t_j and substituting the baseline hazard from Equation (3.16) as follows

$$\begin{aligned} \underline{h}_i(t_j) &= h_0(t_j) \exp(-\epsilon^*) \phi_i \\ &= \frac{\delta_j}{\exp(\epsilon^*) \phi_j + \exp(-\epsilon^*) \sum_{l>j} \phi_l} \exp(-\epsilon^*) \phi_i \\ &= \frac{\delta_j \phi_i}{\exp(2\epsilon^*) \phi_j + \sum_{l>j} \phi_l} \end{aligned} \quad (3.23)$$

$$\bar{h}_i(t_j) = \frac{\delta_j \phi_i}{\phi_j + \exp(-2\epsilon^*) \sum_{l>j} \phi_l} \quad (3.24)$$

The imprecision factor $\exp(2\epsilon^*)$ in Equations (3.23) appears to affect only ϕ_j related to the individual who experienced the event at time t_k . This results in a slight decrease in the lower restricted hazard function which leads to a slight increase in the upper survival function. Unlike the upper restricted hazard function in Equations (3.24), where the imprecision effect $\exp(-2\epsilon^*)$ impacts the summation of ϕ_l related to all other individuals in the risk set at time t_j . This explains why increasing the imprecision has a relatively smaller impact on the upper survival function than on the lower survival function. The same reasoning can be used to explain this feature for the unrestricted functions. As a conclusion to this section, we will examine how fitting the IPH model and increasing imprecision levels affect the regression parameter and survival functions. This analysis will be conducted in Example 3.2.2 and 3.2.3 using simulated data with a binary covariate and the Stanford Heart Transplant data with a continuous covariate, respectively.

	$n = 30$		$n = 60$		$n = 200$	
ϵ^*	$\hat{\beta}$	ℓ	$\hat{\beta}$	ℓ	$\hat{\beta}$	ℓ
0	-0.4399	-87.470	-1.0339	-199.45	-0.5208	-858.06
0.2	-0.4316	-79.115	-1.0227	-181.99	-0.5198	-796.71
0.4	-0.4198	-71.095	-1.0072	-165.01	-0.5176	-735.98
1	-0.3634	-50.093	-0.9258	-118.63	-0.4982	-560.33
8	-5×10^{-6}	-24.000	-4×10^{-5}	-47.000	-0.0003	-157.00

Table 3.3: Estimated β using the IPH model given $\epsilon^* = 0, 0.2, 0.4, 1,$ and 8

Example 3.2.2 The goal of this example is to investigate the effect of the IPH model on the estimated regression parameter and on the estimated survival functions. We generate survival data that follow the PH model, as discussed in Section 2.6. The Weibull distribution is assumed for the survival times with shape parameter $\rho = 3$ and scale parameter $\lambda = 2$, based on the corresponding functions in Table 2.1. These survival times are dependent on the pre-specified distribution of the covariates and the value of the regression coefficient. To keep the example simple, we assume X to be a single time-independent covariate following the Bernoulli distribution with probability equal to 0.5 and a PH regression coefficient $\beta = -0.5$. Within each group, we assume 20% of observations are right censored as illustrated in Section 2.6.3. The IPH model was fitted to three simulated data sets with sample sizes of 30, 60, and 200.

Table 3.3 presents the estimates of the regression coefficients and the log-likelihood values for five different levels of imprecision, $\epsilon^* = 0, 0.2, 0.4, 1,$ and 8 , for each simulated sample. According to Table 3.3, as the imprecision level is increased, the estimated parameter converges to zero and the log-likelihood value increases. By decreasing the imprecision level, the estimate approaches the partial likelihood estimate, so when $\epsilon^* = 0$, we obtain the MLE of the PH model. The IPH restricted and unrestricted survival functions were estimated for both groups based on the simulated sample of size $n = 60$ and level of imprecision $\epsilon^* = 0.2$. These survival functions, along with Breslow's estimates are shown in Figure 3.4 for the restricted survival functions and Figure 3.5 for the unrestricted survival functions. Super-

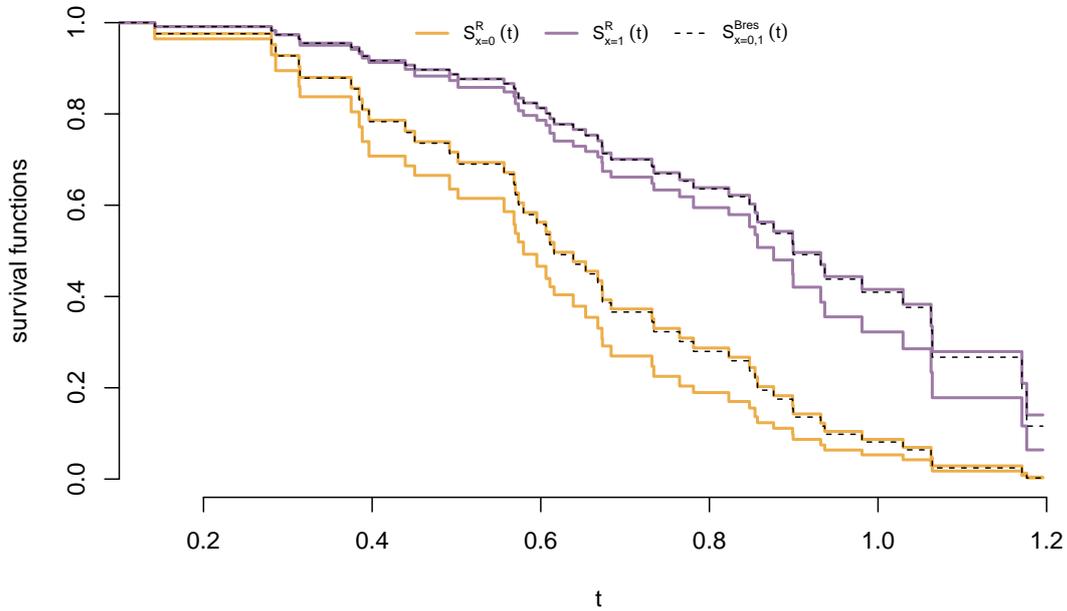


Figure 3.4: The IPH restricted survival functions for $x = 0$, amber, and $x = 1$, indigo, using $\epsilon^* = 0.2$ along with Breslow's estimates, dashed lines.

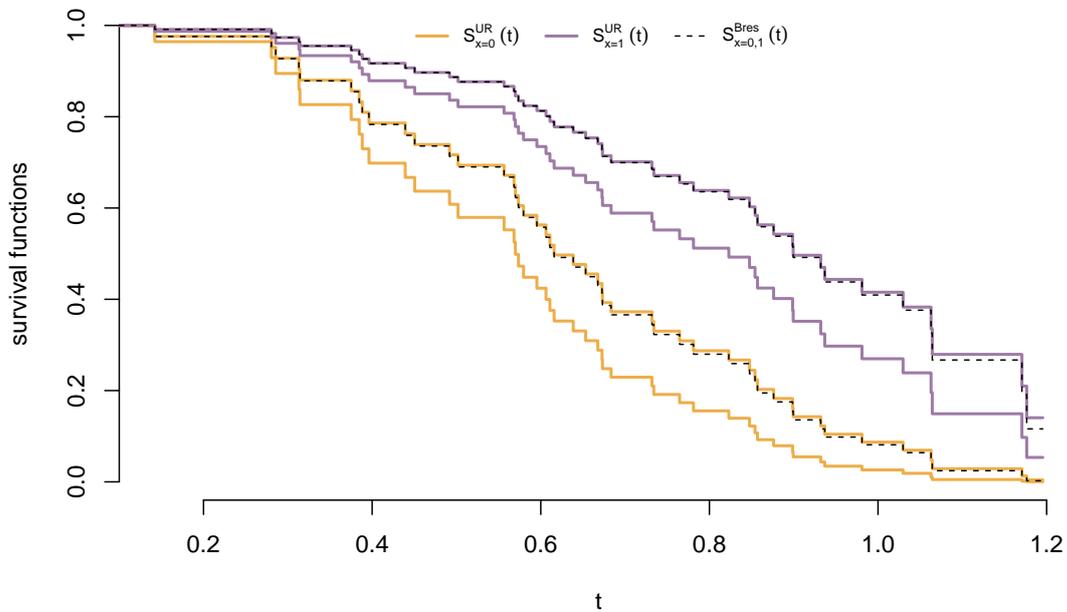


Figure 3.5: The IPH unrestricted survival functions for $x = 0$, amber, and $x = 1$, indigo, using $\epsilon^* = 0.2$ along with Breslow's estimates, dashed lines.

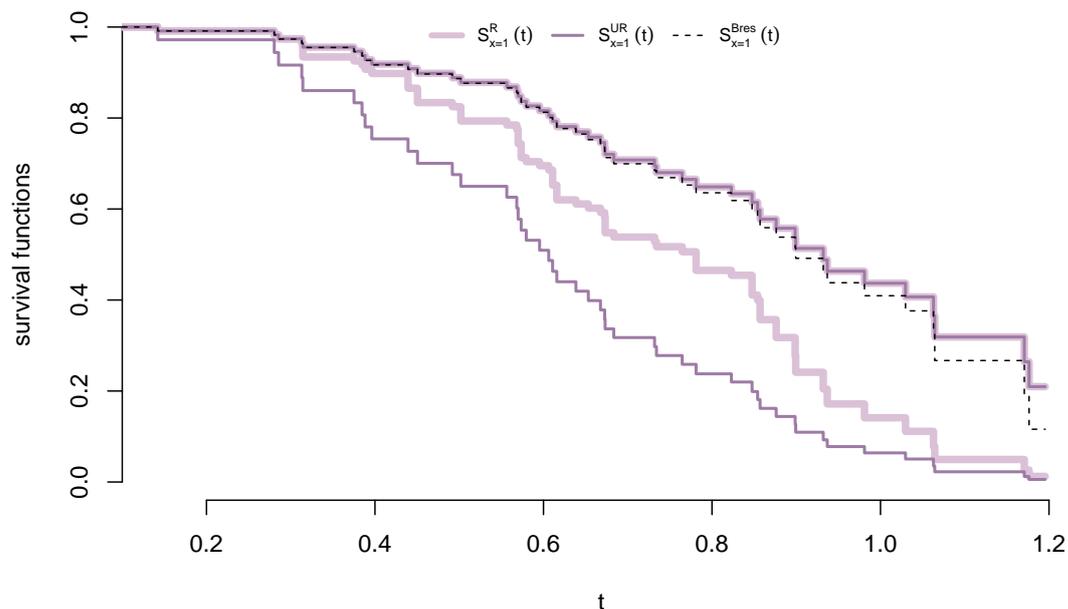


Figure 3.6: The estimated IPH restricted, light-indigo , and unrestricted, indigo , survival functions for individuals with $x = 1$ using $\epsilon^* = 0.6$ along with Breslow estimate, dashed line.

scripts have been used in the legend to distinguish between these functions, with R, UR, and Bres representing restricted, unrestricted, and Breslow's estimates of the PH survival functions, respectively. To clarify, Breslow's estimates refer to the PH survival function for an individual in which the corresponding baseline survival function is derived from the exponential of the negative Breslow's estimator of the baseline cumulative hazard function, though it may also be referred to interchangeably as PH survival function, unless otherwise specified.

Figure 3.4 illustrates that the survival functions derived from the PH model for both groups are very close to the restricted upper survival functions obtained from the IPH model. Even though it may not be obvious from the figure, it should be noted that the PH survival functions may not fall within the restricted IPH upper and lower survival functions. The unrestricted functions behave similarly, as shown in Figure 3.5, with the exception that the PH survival functions will lie within the corresponding unrestricted upper and lower survival functions. The reason for this is that the estimates of the PH hazard function is always located within the

unrestricted upper and lower hazard functions for binary covariate, except for the last survival time.

Both figures indicate that the restricted and unrestricted upper survival functions are less affected by imprecision changes than their corresponding lower survival functions, as we demonstrated earlier in Section 3.2.3. Figure 3.6 illustrates this characteristic when the imprecision level is increased to 0.6 and the survival functions of both types are estimated for individuals with the covariate value $x = 1$. Figure 3.6 also indicates that the unrestricted survival functions exhibit larger differences between the upper and lower survival functions compared to the restricted survival functions. It is due to the fact that the unrestricted upper hazard function increases at all event times, even if the event occurs for an individual from the other group, unlike the restricted upper hazard function which increases only when an individual from the group with $x = 1$ has the event. Furthermore, the upper survival functions of both types are identical, and either slightly greater than or equal to the survival function estimate obtained from the PH model.

◇

Example 3.2.3 Our objective in this example is to examine the effect of fitting the IPH model to survival data when a continuous covariate is present using the Stanford Heart Transplant data set [34], “stanford2” which is provided in the R *survival* package. The data record the survival experience of 184 transplant patients, including 71 right-censored observations. To simplify the analysis, we add infinitesimal values to break any ties in the survival times. To determine the effect of age on the survival of individuals with different levels of imprecision, we fitted the IPH model to the entire data set. Although the original data contains other covariates recorded for patients, they have not been considered in this example. Table 3.4 presents the estimates of the regression parameter and the associated logarithmic likelihood values for $\epsilon^* = 0, 0.2, \text{ and } 0.4$. As expected, the table shows that as the imprecision level increases, the estimated β decreases towards 0 and the log-likelihood value increases. Theorem 3.4.2 on page 77 illustrates the asymptotic behaviour of increasing the imprecision level.

ϵ^*	$\hat{\beta}$	ℓ
0	0.0291	-621.99
0.2	0.0288	-577.73
0.4	0.0285	-533.88

Table 3.4: Estimates of β obtained from fitting the IPH model to the Stanford Heart Transplant data using $\epsilon^* = 0, 0.2,$ and 0.4

time	274	54	66	263	265	279	538	547	729	834	1866	1996	2878	65	136
status	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1
x	31	49	49	49	49	49	49	49	49	49	49	49	49	55	55

Table 3.5: Observed survival times related to individuals with $x = \{31, 49, 55\}$ in the Stanford Heart Transplant data

The IPH survival functions were estimated for intentionally selected individuals at age 31, 37, 49, and 55, in order to highlight key aspects of the IPH model in relation to survival data with a continuous covariate. A summary of the records for these individuals is provided in Table 3.5, noting the absence of individuals at the age of 37 which represent individuals who have not been observed. The table reveals that the individual at age 31 was observed only once, corresponding to an event time, while the individuals at age 49 exhibited twelve observations, ten of which were event times and two being right-censored. Furthermore, at age 55, the data comprises two individuals, both of whom experienced the event. In light of the large disparity between the upper and lower survival functions of the unrestricted survival functions, the imprecision levels were randomly selected for the restricted survival functions using $\epsilon^* = 0.4$, whereas $\epsilon^* = 0.04$ for the unrestricted survival functions.

The restricted and unrestricted IPH survival functions along with the corresponding PH survival functions are shown in Figure 3.7 and Figure 3.8, respectively. Figure 3.7 illustrates that the upper restricted survival functions obtained from the

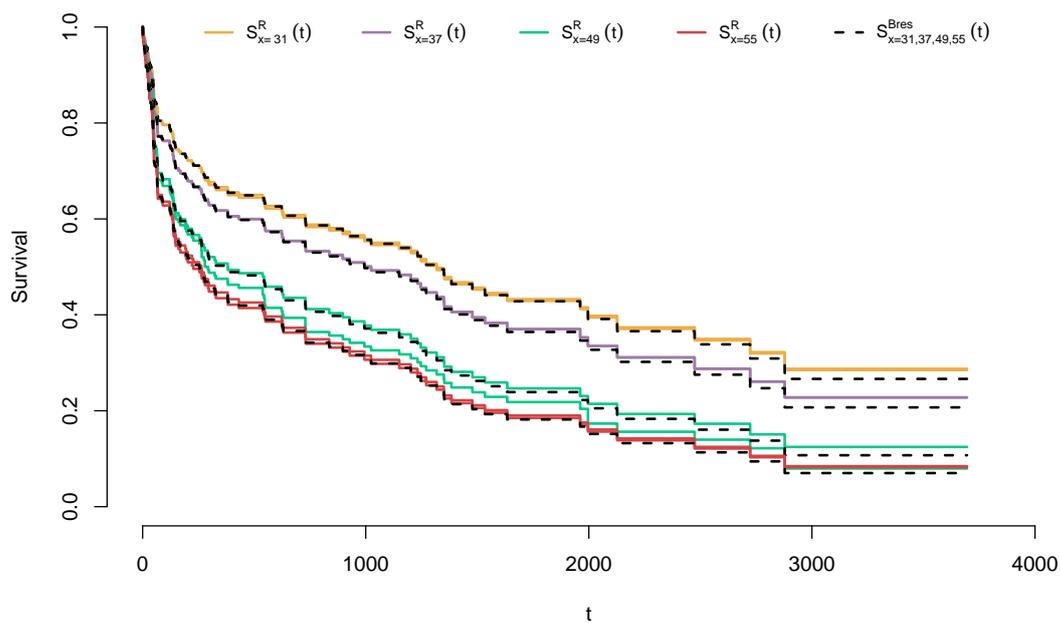


Figure 3.7: The restricted IPH survival functions for individuals with $x = 31, 37, 49, 55$ using $\epsilon^* = 0.4$ along with their corresponding PH survival functions based on Breslow's estimates.

IPH model are either closely aligned or slightly higher than the PH survival functions. The figure also reveals a correlation between the number of events that occur for individuals and the discrepancy between their upper and lower survival functions. For instance, at age 37, where there are no individuals in the data set, both the upper and lower survival functions are identical. Due to the occurrence number of events for individuals at age 49, ten events, the figure displays a larger disparity between the restricted upper and lower survival functions. Similar patterns were observed for individuals at ages 31 and 55. Notably, the PH survival function does not necessarily fall between the restricted upper and lower survival functions for a continuous covariate. This phenomenon can be attributed to the restricted upper hazard function which increases exclusively at event times corresponding to individuals sharing the same covariate value. Thus, as the number of events occurring for these individuals increases, the more likely the PH survival function will lie within the imprecise IPH survival functions, as shown in Figure 3.7 for individuals aged 49.

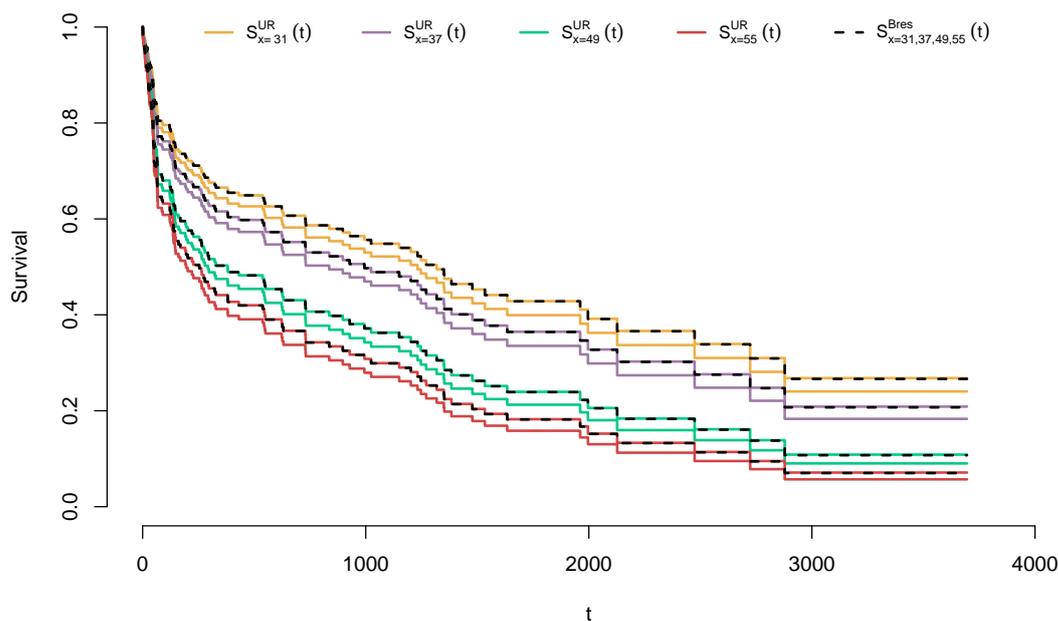


Figure 3.8: The unrestricted IPH survival functions for individuals with $x = 31, 37, 49, 55$ using $\epsilon^* = 0.04$ along with their corresponding PH survival functions based on Breslow's estimates.

The unrestricted upper and lower hazard functions exhibit an interesting characteristic of being constantly increasing at all event times. As a result, this characteristic ensures the existence of both upper and lower survival functions for all individuals, including those who may be absent from the available data set. Consequently, the difference between the unrestricted imprecise survival functions derived from the IPH model is not influenced by the frequency of events reported for individuals, as demonstrated in Figure 3.8. It is worth noting that despite not being explicitly visible in Figure 3.8, the estimates of the PH survival functions may not fall within their corresponding unrestricted upper and lower survival functions.

◇

The IPH model formalization incorporates individual-specific imprecision factors. However, the estimation process of the hazard or survival functions involves considering the highest and lowest hazard functions associated with all individuals sharing the same covariate value, as demonstrated in Equations (3.17) and (3.18).

Building upon the analogy of the IPH model and the insights gained from Examples 3.2.3 and 3.2.2, we propose a novel generalized group-based imprecise model, which imposes the natural sharing of the imprecision factors among sets of individuals. A more detailed analysis of this generalized model is presented in the subsequent section, Section 3.3.

3.3 Group-based Imprecise PH model (GPH)

The GPH model is primarily proposed to avoid any unnecessary imprecision that may arise as a result of individual imprecision in the IPH model, especially in the case of binary covariates. Through dividing the data according to similar characteristics, the GPH model imposes individuals within each group to share the same level of imprecision during the study. The resulting hazard function at time t for the i th individual under the GPH model is given by

$$h_i(t) = h_0(t) \exp(\epsilon_{[i]}(t))\phi_i \quad (3.25)$$

where $[i]$ indicates the group that the i th candidate belongs to, with $|\epsilon_{[i]}(t)| \leq \epsilon^*$. Consider the case of survival data with a binary covariate representing gender of patients, with a value of zero for females and one for males. By fitting the GPH model with groups corresponding to the binary covariate, $\epsilon_{[i]}(t) = \epsilon_0(t)$ if the i th individual is female and $\epsilon_{[i]}(t) = \epsilon_1(t)$ if the i th individual belongs to the male group. Hence, all females will exhibit the same imprecision effect at all times, $\epsilon_0(t)$, and similarly for males, $\epsilon_1(t)$.

According to this definition, the IPH model can be viewed as a special case of the GPH model when there is only one candidate in each group. Based on similar steps to those described in Section 3.2, it can be seen that the contribution of the i th individual to the likelihood function is given by

$$L_i = [p_i \exp(\epsilon_{[i]}(t_i))\phi_i]^{\delta_i} \exp\left(-\sum_{t_l \leq t_i} p_l \exp(\epsilon_{[i]}(t_l))\phi_i\right) \quad (3.26)$$

where $p_i = h_0(t_i)$. The likelihood function corresponding to the GPH model can be

determined by multiplying the contributions L_1, \dots, L_n as follows

$$\begin{aligned} L(\boldsymbol{\psi}) &= [p_1 \exp(\epsilon_{[1]}(t_1))\phi_1]^{\delta_1} \exp[-p_1 \exp(\epsilon_{[1]}(t_1))\phi_1] \\ &\times [p_2 \exp(\epsilon_{[2]}(t_2))\phi_2]^{\delta_2} \exp[-p_1 \exp(\epsilon_{[2]}(t_1))\phi_2 - p_2 \exp(\epsilon_{[2]}(t_2))\phi_2] \\ &\times [p_3 \exp(\epsilon_{[3]}(t_3))\phi_3]^{\delta_3} \exp[-p_1 \exp(\epsilon_{[3]}(t_1))\phi_3 - p_2 \exp(\epsilon_{[3]}(t_2))\phi_3 - p_3 \exp(\epsilon_{[3]}(t_3))\phi_3] \\ &\times \dots \\ &\times [p_n \exp(\epsilon_{[n]}(t_n))\phi_n]^{\delta_n} \exp[-p_1 \exp(\epsilon_{[n]}(t_1))\phi_n - \dots - p_n \exp(\epsilon_{[n]}(t_n))\phi_n] \end{aligned}$$

where $\boldsymbol{\psi} = \{\beta, p_1, \dots, p_n, \dots, \epsilon_{[1]}(t_1), \dots, \epsilon_{[n]}(t_1), \dots, \epsilon_{[n]}(t_n)\}$. Following this multiplication across all individuals and rearranging the likelihood function with reference to p_i , the Poisson empirical likelihood of the GPH model can be written as

$$L = \prod_{i=1}^n [p_i \exp(\epsilon_{[i]}(t_i))\phi_i]^{\delta_i} \exp(-p_i r_{[i]}^*) \quad (3.27)$$

where $r_{[i]}^* = \sum_{l \geq i} \exp(\epsilon_{[l]}(t_i))\phi_l$ represents the sum of the shared imprecision factors times the covariate effects for all individuals at risk at time t_i . As with the IPH model discussed in Section 3.2, the profile likelihood function for the GPH model can be directly determined from Equation (3.12) by assuming fixed β and $\epsilon_{[l]}(t_i)$. Thus, the profile likelihood function for the GPH model is proportional to

$$L^p \propto \prod_{i=1}^n \left[\frac{\phi_i}{\exp(-\epsilon_{[i]}(t_i))r_{[i]}^*} \right]^{\delta_i} \quad (3.28)$$

with $\hat{p}_i = \frac{\delta_i}{r_{[i]}^*}$ and $r_{[i]}^* = \sum_{l \geq i} \exp(\epsilon_{[l]}(t_i))\phi_l$.

3.3.1 Maximizing the likelihood function of the GPH model

To simplify the analyses, let us consider the case where the survival data is limited to two groups, so we have two shared imprecision factors with $\epsilon_0(t_i)$ for the group with $x = 0$ and $\epsilon_1(t_i)$ for $x = 1$. Thus, $r_{[i]}^*$ can be represented by

$$r_{[i]}^* = \exp(\epsilon_0(t_i)) \sum_{l \geq i} \mathbb{1}_{[x_l=0]} \phi_l + \exp(\epsilon_1(t_i)) \sum_{l \geq i} \mathbb{1}_{[x_l=1]} \phi_l$$

Generalizing this representation to incorporate m distinct groups is straightforward. Consequently, the contribution L_i^p of the event time t_i to the above likelihood

function can be written as follows

$$\begin{aligned}
 L_i^p &= \left[\frac{\phi_i}{\exp(-\epsilon_{[i]}(t_i)) r_{[i]}^*} \right]^{\delta_i} \\
 &= \left[\frac{\phi_i}{\exp(-\epsilon_{[i]}(t_i)) (\exp(\epsilon_0(t_i)) \sum_{l \geq i} \mathbb{1}_{[x_l=0]} \phi_l + \exp(\epsilon_1(t_i)) \sum_{l \geq i} \mathbb{1}_{[x_l=1]} \phi_l)} \right]^{\delta_j} \\
 &= \left[\frac{\phi_i}{(\exp(\epsilon_0(t_i) - \epsilon_{[i]}(t_i)) \sum_{l \geq i} \mathbb{1}_{[x_l=0]} \phi_l + \exp(\epsilon_1(t_i) - \epsilon_{[i]}(t_i)) \sum_{l \geq i} \mathbb{1}_{[x_l=1]} \phi_l)} \right]^{\delta_j}
 \end{aligned} \tag{3.29}$$

If $x_i = 0$, then $\epsilon_{[i]}(t_i) = \epsilon_0(t_i) = \epsilon^*$ and $\epsilon_1(t_i) = -\epsilon^*$ maximize the contribution of L_i , and vice versa when $x_i = 1$. Hence, L_i corresponding to a non-censored observation achieves its maximum at $\epsilon_{[i]}(t_i) = \epsilon_j(t_i) = \epsilon^*$ with $j \in \{0, 1\}$ for the group that experience the event at time t_i , and the lower bound of imprecision for other groups in the risk set, $-\epsilon^*$. Similar to the IPH model in Section 3.2, the imprecision terms corresponding to the censoring times and also to t_n have no impact on the likelihood, because the product in the likelihood is taken only over failure times, and $L_n = 1$. Consequently, the full likelihood function of Equation (3.28) can be profiled as

$$L(\beta) = \prod_{i=1}^n \left[\frac{\phi_i}{\sum_{l \geq i} \phi_l [\mathbb{1}_{[x_l=x_i]} + \exp(-2\epsilon^*) \mathbb{1}_{[x_l \neq x_i]}]} \right]^{\delta_i} \tag{3.30}$$

zzz In order to demonstrate how the imprecision factors in Equation (3.29) are determined, we will provide a brief example.

Example 3.3.1 This example illustrates how each observed individual contributes to the imprecision of the GPH model via $\epsilon_{[i]}(t)$. In order to facilitate our purpose, recall the artificial survival data in Table 3.1. Consider the contribution of L_1 as in Equation (3.29), it can be seen that only the following imprecision terms are required when evaluating L_1

$$\epsilon_{[x=1]}(t = 10) = \epsilon^* \quad \epsilon_{[x=0]}(t = 10) = -\epsilon^*$$

The imprecision terms are not required at $t = 12$ since it is a right-censored observation, which implies that $L_2 = 1$. By following the same procedure, the remaining imprecision terms can be obtained, see Table 3.6.

x	1	0	0	1	1	0
t_i						
10	ϵ^*	$-\epsilon^*$	$-\epsilon^*$	ϵ^*	ϵ^*	$-\epsilon^*$
12 ⁺						
14			ϵ^*	$-\epsilon^*$	$-\epsilon^*$	ϵ^*
16				ϵ^*	ϵ^*	$-\epsilon^*$
18 ⁺						
20						

Table 3.6: GPH's imprecision terms for the Survival data in Table 3.1

When assessing the imprecision terms that maximize the likelihood functions, Examples 3.2.1 and 3.3.1 demonstrate the differences between the GPH and IPH models. In particular, in the IPH model, only the imprecision factor associated with the individual who experienced the event equals the upper bound of the imprecision level, whereas the imprecision factor associated with other individuals in the risk set equals the lower bound of the imprecision level. This contrasts with the GPH model, which assigns an upper bound of imprecision to all individuals in a risk set who belong to the same group as the individual who experienced the event, and a lower bound to those from other groups.

◇

3.3.2 Hazard and survival functions for the GPH model

This section illustrates the estimation of the hazard and survival functions for individuals with particular covariate values using the GPH model. The results of the profile likelihood function in Equation (3.28) can be used as a basis for determining the baseline hazard function h_0 , along with $\hat{\beta}(\epsilon^*)$, $\hat{\epsilon}_{[1]}(t_1), \dots, \hat{\epsilon}_{[n]}(t_n)$ obtained through maximizing the likelihood function. In view of this, the baseline hazard function of the GPH model can be derived as follows

$$\begin{aligned}
h_0(t_j) &= \frac{\delta_j}{r_{[j]}^*} \\
&= \frac{\delta_j}{\exp(\epsilon^*) \sum_{l \geq j} \mathbb{1}_{[x_l = x_j]} \phi_l + \exp(-\epsilon^*) \sum_{l \geq j} \mathbb{1}_{[x_l \neq x_j]} \phi_l}
\end{aligned} \tag{3.31}$$

Note that the baseline hazard function is zero at right-censoring or non-observed times due to the profiled maximization of the empirical likelihood. Similar to the IPH model, the imprecise hazard and survival functions associated with the GPH model can be introduced in different forms, restricted and unrestricted.

As a result of the optimization process, it is known that $\exp(\epsilon_{[i]}(t)) = \epsilon^*$ whenever an member of the group with covariate value $x = x_i$ is observed to have had the event at time t , and $\exp(\epsilon_{[i]}(t)) = -\epsilon^*$ otherwise. The restricted hazard function relies on this result, so the upper hazard function for the group with $x = x_i$ is $h_0(t) \exp(-\epsilon^*) \phi_i$ in all survival times except for the time when a member of that group had the event, in which case the hazard function will increase to $h_0(t) \exp(\epsilon^*) \phi_i$. The lower hazard function will reflect the scenario that none of these members has experienced the event. Consequently, the restricted hazard functions for the GPH model are

$$\bar{h}_i(t) = \begin{cases} h_0(t) \exp(\epsilon^*) \phi_i & ; \text{if an individual with } x = x_i \text{ had the event at } t \\ h_0(t) \exp(-\epsilon^*) \phi_i & ; \text{otherwise} \end{cases} \tag{3.32}$$

$$\underline{h}_i(t) = h_0(t) \exp(-\epsilon^*) \phi_i \tag{3.33}$$

The unrestricted hazard functions assign the assumed imprecision limits directly to the hazard function. This is regardless of whether or not members who share the same covariate value have experienced an event. Consequently, the upper hazard function will increase at all event times. Therefore, the imprecise unrestricted hazard functions are

$$\bar{h}_i(t) = h_0(t) \exp(\epsilon^*) \phi_i \tag{3.34}$$

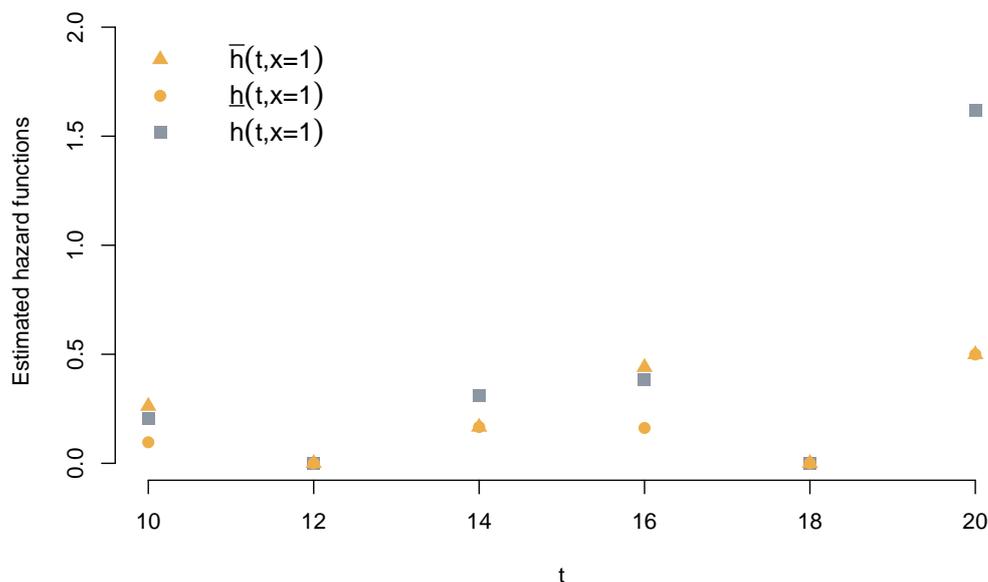


Figure 3.9: The estimated imprecise GPH restricted hazard functions for the group with $x = 1$ using $\epsilon^* = 0.5$ along with Breslow estimates of the hazard function.

$$\underline{h}_i(t) = h_0(t) \exp(-\epsilon^*) \phi_i \quad (3.35)$$

According to the definition of the restricted and unrestricted lower hazard functions in the GPH model, these functions are the same. In light of the imprecision constraint, the estimated imprecise hazard functions can be interpreted as upper and lower bounds for the set of all possible hazard functions for the entire population with the same covariate value, regardless of whether members of that population experienced the event. Despite the fact that the lower hazard functions of the IPH and GPH models are defined in the same way, their estimations will differ as a result of the different likelihood functions.

Based on fitting the GPH model to the artificial survival data in Table 3.1, Figures 3.9 and 3.10 display the imprecise restricted and unrestricted hazard functions, respectively, for individuals with $x = 1$. According to Figure 3.9, the restricted upper hazard function for individuals with $x = 1$ increases only when the event occurs at $t = 10$ and $t = 16$ for individuals with $x = 1$. As opposed to the upper unrestricted hazard function in Figure 3.10, which increases at all event times regardless

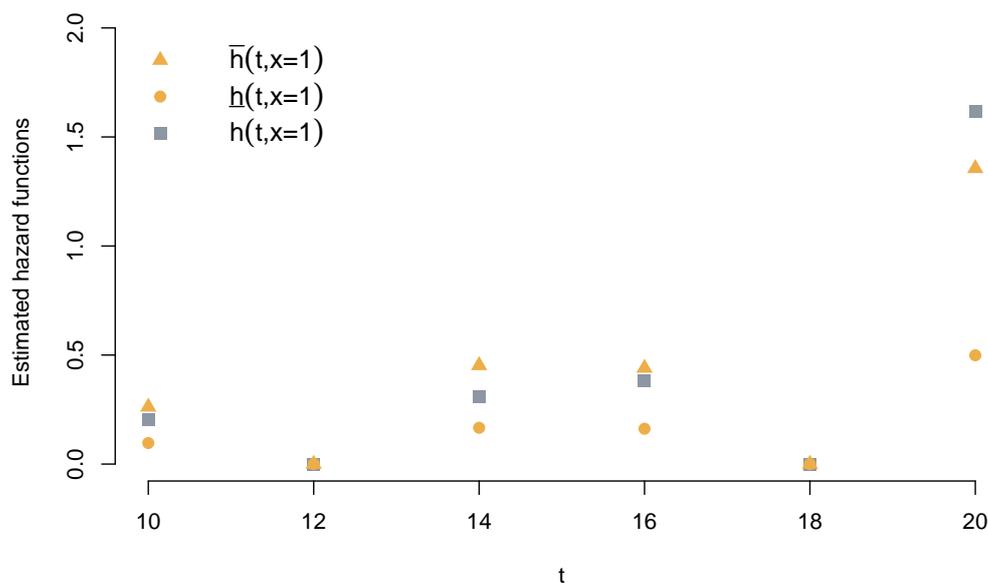


Figure 3.10: The estimated imprecise GPH unrestricted hazard functions for the group with $x = 1$ using $\epsilon^* = 0.5$ along with Breslow estimates of the hazard function.

of which individual has experienced the event. The lower hazard functions in both types decrease in all event times.

To determine the imprecise survival function for the GPH model, we substitute the imprecise upper and lower hazard functions using Equations (3.34) and (3.35) for the restricted type or Equations (3.32) and (3.33) for the unrestricted type as follow

$$\bar{S}(t; x_i) = \exp[-\underline{H}_i(t)] = \exp\left[-\sum_{j:t_j \leq t} \underline{h}_i(t_j)\right] \quad (3.36)$$

$$\underline{S}(t; x_i) = \exp[-\bar{H}_i(t)] = \exp\left[-\sum_{j:t_j \leq t} \bar{h}_i(t_j)\right] \quad (3.37)$$

As opposed to the IPH model, increasing the imprecision level has relatively similar effects on GPH upper and lower survival functions. Essentially, this feature can be justified by considering the baseline hazard function related to the GPH model in Equation (3.31), where individuals in each group share the same imprecision effect. As a result of the fact that the unrestricted upper hazard function increases in all event times regardless of whether an individual from the same group experiences the

ϵ^*	$n = 30$		$n = 60$		$n = 200$	
	β	ℓ	β	ℓ	β	ℓ
0	-0.4399	-87.470	-1.0339	-199.45	-0.5208	-858.06
0.2	-0.4246	-83.551	-0.9781	-193.21	-0.5164	-831.80
0.4	-0.4142	-80.468	-0.9351	-188.30	-0.5154	-811.07
1	-0.3998	-75.153	-0.8646	-179.77	-0.5140	-774.95
8	-1×10^{-6}	-72.372	-3×10^{-6}	-175.25	-9×10^{-6}	-755.68

Table 3.7: Estimated β using the GPH model given $\epsilon^* = 0, 0.2, 0.4, 1, \text{ and } 8$

event or not, the unrestricted lower survival function appears to be affected more than the restricted lower survival function.

Based on the data sets described in Example 3.2.2 and 3.2.3, the remainder of this section will evaluate the impact of increasing imprecision levels on the GPH regression parameter and survival functions.

Example 3.3.2 In this example we investigate the impact of the GPH model on the estimated regression parameter and on the estimated survival functions. Consider the simulated data sets that were presented in Example 3.2.2 in which the IPH model was used. For this example, however, the simulated data sets were fitted with the GPH model. The regression coefficients along with the likelihood values are shown in Table 3.7 for each simulated sample based on the following levels of imprecision: 0, 0.2, 0.4, 1, 8. Similar to the behavior observed in Example 3.2.2 for the IPH model, the GPH model demonstrates a consistent impact on both the log-likelihood value and the parameter estimate when increasing the imprecision level. Specifically, an increase in the imprecision level corresponds to an increase in the log-likelihood, while the parameter estimate converges towards zero, as seen in Table 3.7. Additionally, the GPH regression coefficient estimate converges towards the estimate derived from the PH model as ϵ^* approaches zero.

The GPH restricted and unrestricted survival functions were estimated for both groups based on the simulated sample of size $n = 60$ using the level of imprecision $\epsilon^* = 0.2$. These survival functions, along with the PH survival function based on

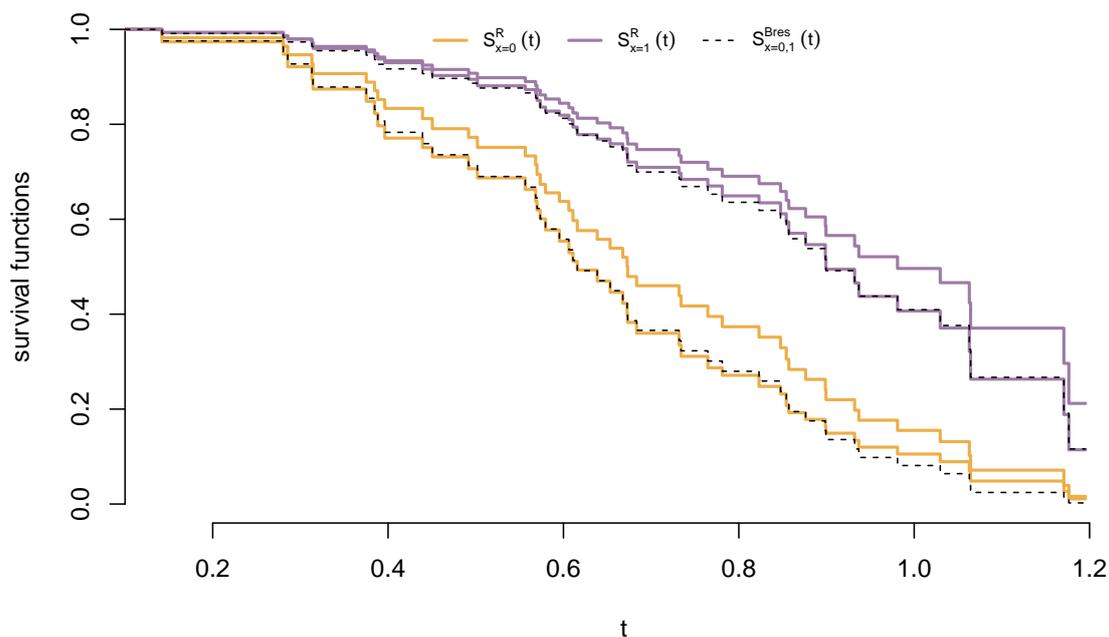


Figure 3.11: The GPH restricted survival functions for $x = 0$, amber, and $x = 1$, indigo, using $\epsilon^* = 0.2$ along with Breslow estimates, dashed lines, based on the simulated data with size $n = 60$.

Breslow's estimates are shown in Figure 3.11 for the restricted survival functions and Figure 3.12 for the unrestricted survival functions. Figure 3.11 illustrates that the survival estimates derived from the PH model for both groups are relatively similar to the restricted lower survival functions obtained from the GPH model. However, the PH survival estimates may not be within the range of GPH restricted survival estimates due to the fact that the upper hazard estimates increase only at event times related to an individual from the same group. Moreover, the figure reveals that restricted upper survival functions are more influenced by imprecision changes than lower survival functions since the lower hazard function is always decreases at all event times. In terms of the unrestricted survival functions, Figure 3.5 demonstrates that the PH survival functions consistently fall within the GPH model's unrestricted survival functions. Furthermore, the figure shows that imprecision levels have a similar influence on the unrestricted survival functions. Figure 3.13 highlights this aspect by estimating the GPH restricted and unrestricted survival functions using an imprecision level $\epsilon^* = 0.6$ for individuals with the covariate value $x = 1$. The

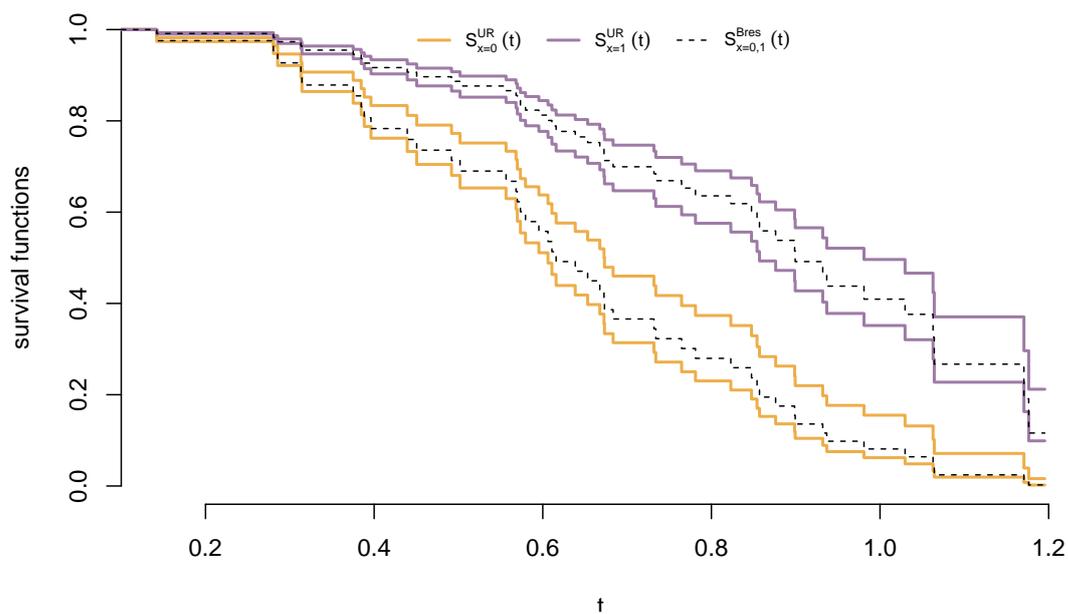


Figure 3.12: The GPH unrestricted survival functions for $x = 0$, amber, and $x = 1$, indigo, using $\epsilon^* = 0.2$ along with Breslow estimates, dashed lines, based on the simulated data with size $n = 60$.

figure also indicates that the upper survival functions of both types are identical, while the unrestricted survival functions exhibit larger differences between the upper and lower survival functions than the restricted survival functions because the restricted upper hazard function increases only when an individual from the group with $x = 1$ has the event.

In Figure 3.14, the GPH and IPH models are compared using $\epsilon^* = 0.2$ for the restricted survival estimates (top) and the unrestricted survival estimates (bottom) along with the corresponding PH survival estimates. The comparison of the restricted type survival estimates shown in Figure 3.14 (top) indicates a tendency for the GPH survival estimates to expand upward as the imprecision level increases, implying higher survival probability, as opposed to the IPH estimates. Possibly, this is due to differences in how imprecision is incorporated into these models, with the GPH model including a collective imprecision effect that affects all individuals within each group, while the IPH model imposes a distinct imprecision effect on each individual. With regard to the unrestricted survival estimates, Figure 3.14 (bottom)

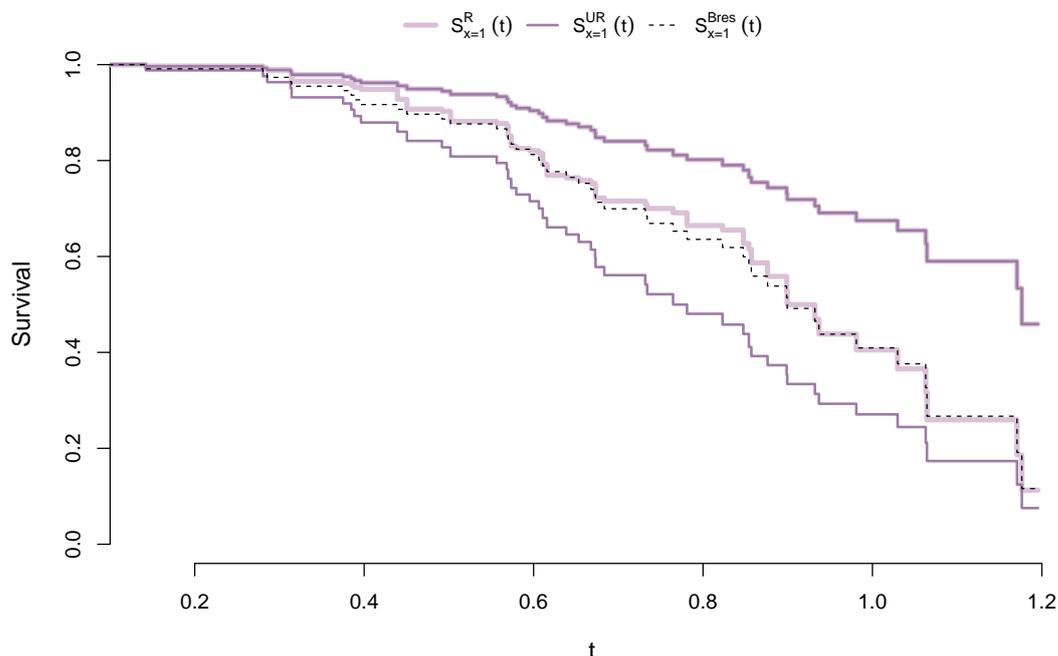


Figure 3.13: The GPH restricted, light-indigo , and unrestricted, indigo , survival estimates for individuals with $x = 1$ using $\epsilon^* = 0.6$, based on the simulated data with size $n = 60$.

illustrates that the unrestricted GPH survival estimates expand in both directions as a result of the relatively similar impact of the imprecision level on the upper and lower survival functions. In contrast, the unrestricted IPH survival estimates exhibit similar attributes to restricted IPH survival estimates. These attributes have been found to hold true when analysing survival data with a binary covariate.

◇

Example 3.3.3 In this example we examine the effect of fitting the GPH model to survival data for continuous covariate based on the Stanford Heart Transplant data set [34], also used in Example 3.2.3. Notice that the similarity between the GPH and the IPH models is directly influenced by the diversity of covariate values in the survival data since the IPH model is a special case of the GPH model. In other words, the GPH and IPH models tend to exhibit greater similarities when there is limited repetition of covariate values among individuals. This suggests that

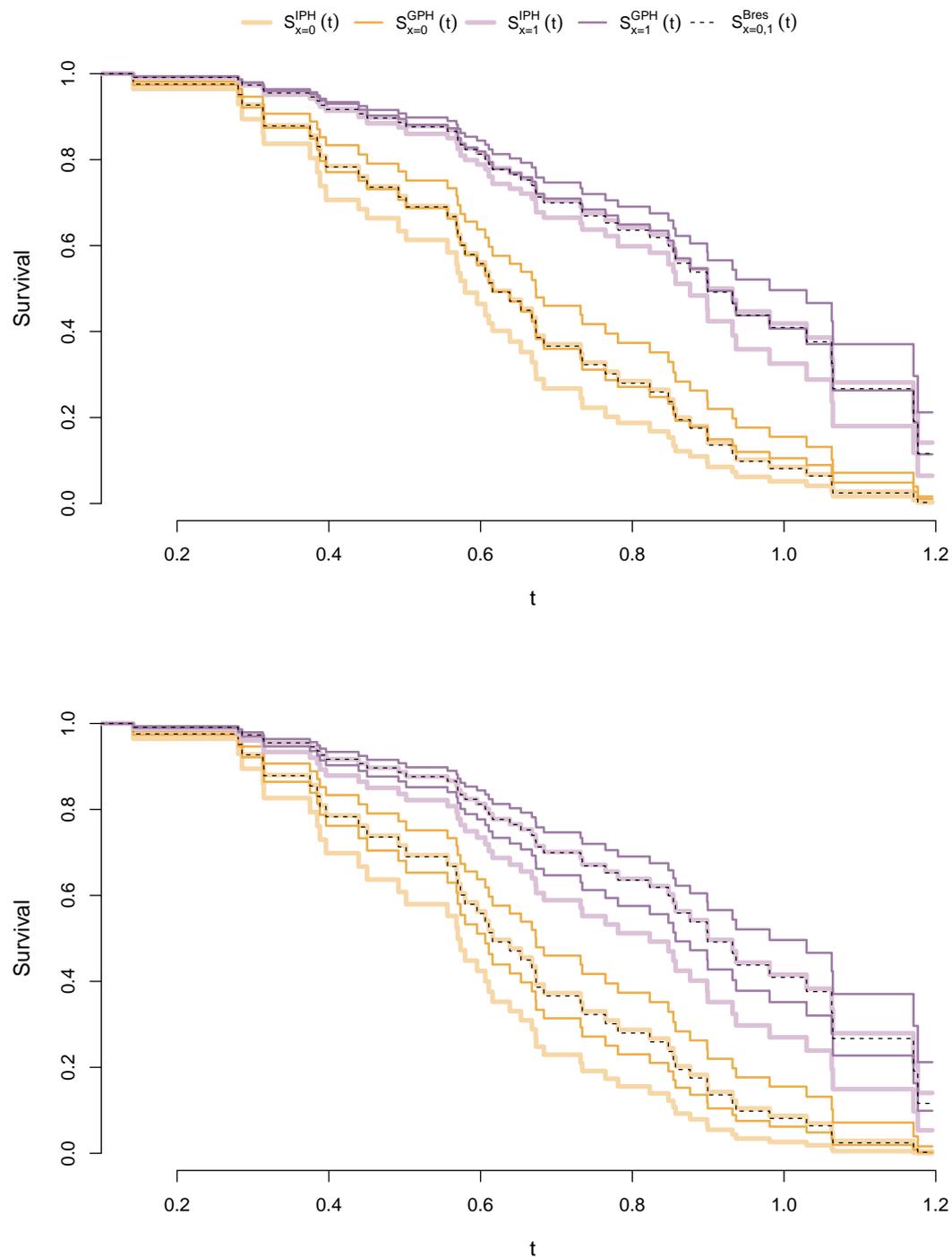


Figure 3.14: IPH and GPH survival estimates for individuals with $x = 0, 1$, using $\epsilon^* = 0.2$ for both the restricted (top) and unrestricted (bottom) functions, based on the simulated data with size $n = 60$.

ϵ^*	$\hat{\beta}$	ℓ
0	0.0291	-508.99
0.2	0.0287	-464.45
0.4	0.0284	-420.19

Table 3.8: Estimates of β obtained from fitting the GPH model to the Stanford Heart Transplant data using $\epsilon^* = 0, 0.2,$ and 0.4

the results of using both models should be fairly comparable when the covariate is continuous. Table 3.8 presents the estimates of the regression parameter and the associated log-likelihood values obtained from fitting the GPH model to the Stanford Heart Transplant data, grouped by the same age, using $\epsilon^* = 0, 0.2,$ and 0.4 . As expected, the table shows that as the imprecision level increases, the estimated regression parameter converges to zero and the log-likelihood value increases. Additionally, Table 3.8 indicates that the estimates of the regression parameter are relatively similar to those obtained from using the IPH model, as seen in Table 3.4. Theorem 3.4.1 on page 74 illustrates the asymptotic behaviour of increasing the imprecision level.

Figure 3.15 presents the survival estimates derived from the GPH and IPH models for individuals aged 37 and 49, using $\epsilon^* = 0.8$ for the restricted survival functions (top), and $\epsilon^* = 0.5$ for the unrestricted survival functions (bottom) along with the corresponding PH survival functions. Similar to what have been observed for survival data with a binary covariate in Example 3.3.2, survival estimates obtained from the GPH model are marginally higher than those obtained from the IPH model. Perhaps this is due to the differences in regression coefficient estimates and how each model incorporates the imprecision effect. Figure 3.15 (top) highlights that the higher number of observed events for a particular covariate value is associated with a wider imprecision in the restricted survival estimates obtained from both models. Consequently, the absence of individuals aged 37 in the data will result in unique survival estimates since the lower and upper survival estimates will be identical in each model. On the basis of the unrestricted survival estimates,

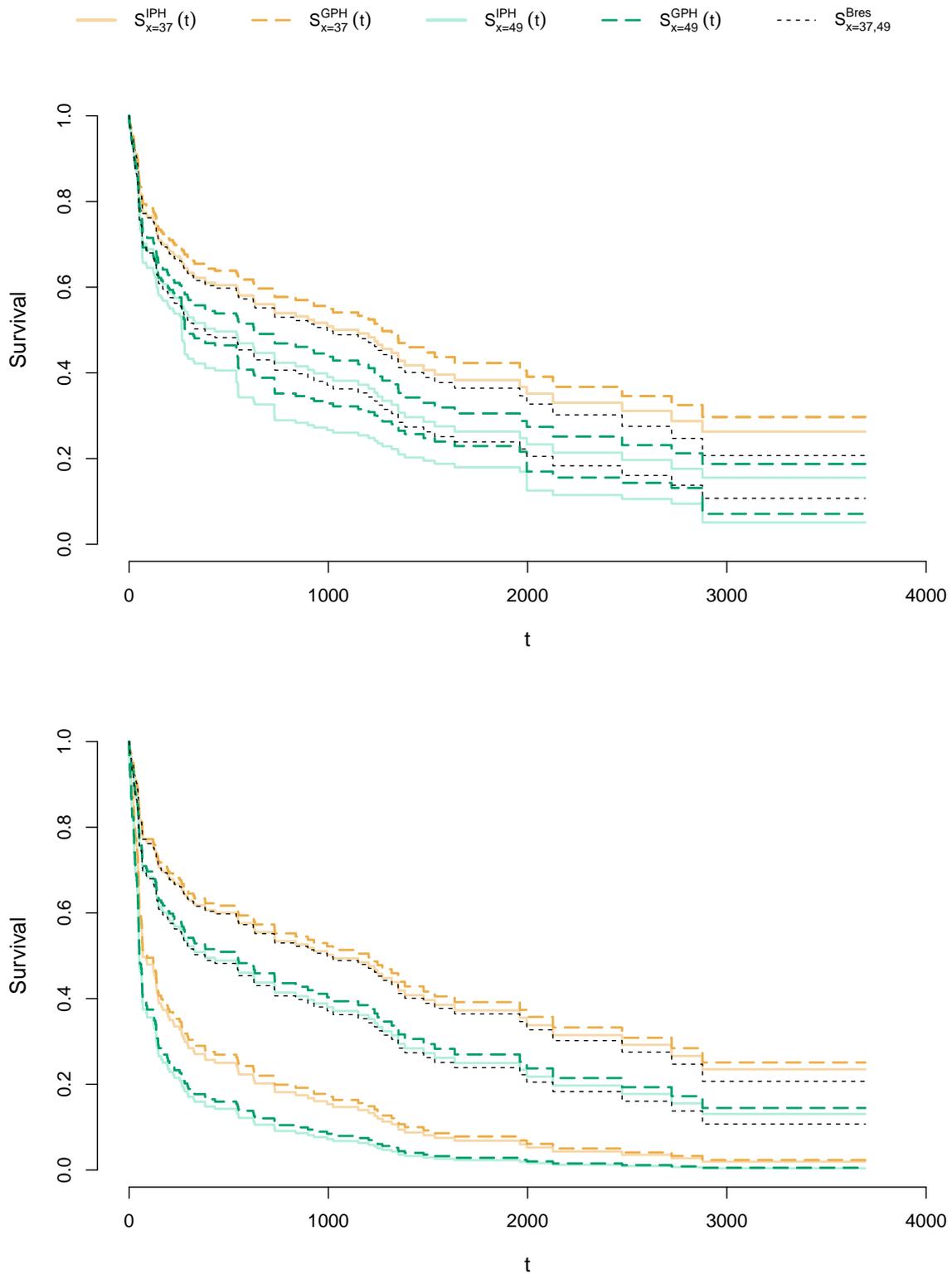


Figure 3.15: IPH and GPH survival estimates for individuals with $x = 37, 49$, using $\epsilon^* = 0.8$ for the restricted functions (top), and $\epsilon^* = 0.5$ for the unrestricted functions (bottom).

Figure 3.15 (bottom) demonstrates that the lower survival estimates derived from the IPH and GPH models are almost indistinguishable. However, the unrestricted upper estimates obtained from the IPH model progressively diverge upwards from their GPH counterparts as the level of imprecision increases.

◇

3.4 Assessing the IPH and GPH models

In this section, a variety of approaches are presented for exploring and gaining insight into the effects of fitting the IPH and the GPH models. It has been observed in Table 3.3 that, as the imprecision level, ϵ^* , increases the parameter estimate converges to zero and the log-likelihood value increases for both the IPH and the GPH models. The following theorems illustrate the effect of increasing ϵ^* to infinity on the likelihood contribution of t_i and the imprecise hazard functions of an individual based on both the GPH and the IPH models.

Theorem 3.4.1 Let $\tilde{n}_{ij} = \tilde{n}_j(t_i)$ denote the number individuals in the risk set R_i with $x = x_j$. Additionally, let $\tilde{n}_{i[i]}$ denote the number of observations in the risk set R_i with the same covariate value $x_{[i]}$ as the individual who experiences an event at time t_i . By fitting the GPH model to survival data and assuming $\epsilon^* \rightarrow \infty$, we get

- (i) The restricted upper hazard function of an individual with $x = x_j$ reduces to the Nelson-Aalen estimator, as given in Section 2.2.2, while the lower hazard function will be zero, such that

$$\bar{h}(t_i|x_j) = \begin{cases} \frac{1}{\tilde{n}_{ij}} & ; \text{if an individual with } x = x_j \text{ has the event at } t_i \\ 0 & ; \text{censoring times or event times for an individual with } x \neq x_j \end{cases}$$

$$\underline{h}(t_i|x_j) = 0$$

(ii) The unrestricted hazard functions of an individual with $x = x_j$ will be

$$\bar{h}(t_i|x_j) = \begin{cases} \frac{1}{\tilde{n}_{i[i]}} & ; \text{ for event time } t_i \\ 0 & ; \text{ censoring times} \end{cases}$$

$$\underline{h}(t_i|x_j) = 0$$

(iii) The contribution of an event time t_i to the likelihood function becomes

$$L_i = \frac{1}{\tilde{n}_{i[i]}}$$

Proof.

(i) The restricted upper hazard function for an individual with $x = x_j$ who has the event at t_i , using the GPH model as given in Equation (3.32), is

$$\begin{aligned} \bar{h}(t_i|x_j) &= h_0(t_i) \exp(\epsilon^*) \phi_j \\ &= \left(\frac{1}{\exp(\epsilon^*) \sum_{l \geq i} \mathbb{1}_{[x_l=x_j]} \phi_l + \exp(-\epsilon^*) \sum_{l \geq j} \mathbb{1}_{[x_l \neq x_j]} \phi_l} \right) \exp(\epsilon^*) \phi_j \\ &= \frac{\phi_j}{\tilde{n}_{ij} \phi_j + \exp(-2\epsilon^*) \sum_{l \geq i} \mathbb{1}_{[x_l \neq x_j]} \phi_l} \end{aligned} \tag{3.38}$$

As $\epsilon^* \rightarrow \infty$ the restricted upper hazard function for an individual with $x = x_j$ who has the event at t_i converges to $\frac{1}{\tilde{n}_{ij}}$.

If an individual with $x = x_k \neq x_j$ experienced the event at t_i , then from the optimization $\exp(\epsilon_{[i]}(t)) = \exp(\epsilon_k(t)) = \epsilon^*$ and $\exp(\epsilon_j(t)) = -\epsilon^*$. Thus, the restricted upper hazard function for an individual with $x = x_j$ at t_i is

$$\begin{aligned} \bar{h}(t_i|x_j) &= h_0(t_i) \exp(-\epsilon^*) \phi_j \\ &= \left(\frac{1}{\exp(\epsilon^*) \sum_{l \geq i} \mathbb{1}_{[x_l=x_k]} \phi_l + \exp(-\epsilon^*) \sum_{l \geq i} \mathbb{1}_{[x_l \neq x_k]} \phi_l} \right) \exp(-\epsilon^*) \phi_j \\ &= \frac{\phi_j}{\exp(2\epsilon^*) \tilde{n}_{ik} \phi_k + \sum_{l \geq i} \mathbb{1}_{[x_l \neq x_k]} \phi_l} \end{aligned} \tag{3.39}$$

The restricted upper hazard function for an individual with $x = x_j$ given that an individual from other groups had the event at time t_i converges to zero as

$\epsilon^* \rightarrow \infty$. For right-censored survival times the upper hazard is equal to zero. The restricted lower hazard function for an individual with $x = x_j$ at an event or right-censored time t_i , as given in Equation (3.33), is

$$\begin{aligned} \underline{h}(t_i|x_j) &= h_0(t_i) \exp(-\epsilon^*) \phi_j \\ &= \left(\frac{\delta_i}{\exp(\epsilon^*) \sum_{l \geq i} \mathbb{1}_{[x_l=x_i]} \phi_l + \exp(-\epsilon^*) \sum_{l \geq i} \mathbb{1}_{[x_l \neq x_i]} \phi_l} \right) \exp(-\epsilon^*) \phi_j \\ &= \frac{\delta_i \phi_j}{\exp(2\epsilon^*) \sum_{l \geq i} \mathbb{1}_{[x_l=x_i]} \phi_l + \sum_{l \geq i} \mathbb{1}_{[x_l \neq x_i]} \phi_l} \end{aligned} \quad (3.40)$$

Thus, the restricted lower hazard function for an individual with $x = x_j$ converges to zero as ϵ^* approaches infinity. \square

- (ii) The unrestricted upper hazard function for an individual with $x = x_j$ at the event time t_i , as given in Equation (3.34), is

$$\begin{aligned} \bar{h}(t_i|x_j) &= h_0(t_i) \exp(\epsilon^*) \phi_j \\ &= \frac{\phi_j}{\tilde{n}_{i[i]} \phi_i + \exp(-2\epsilon^*) \sum_{l \geq i} \mathbb{1}_{[x_l \neq x_i]} \phi_l} \end{aligned} \quad (3.41)$$

In conjunction with the result that $\hat{\beta}$ converges to zero as ϵ^* increases, the restricted upper hazard function for an individual with $x = x_j$ at the event time t_i converges to $\frac{1}{\tilde{n}_{i[j]}}$ as ϵ^* approaches infinity.

The upper hazard for right-censored survival times is zero. Consequently, the unrestricted upper hazard function obtained from the GPH model converges to the baseline PH hazard function with $\beta = 0$ as ϵ^* approaches infinity. Due to the fact that the restricted and unrestricted lower hazard functions are equal, the unrestricted lower hazard function for an individual with $x = x_i$ converges to zero as ϵ^* approaches infinity. \square

- (iii) According to Equation (3.28) the contribution of an event time t_i to the like-

likelihood function is given by

$$\begin{aligned}
 L_i &= \frac{\phi_i}{\exp(-\epsilon_{[i]}(t_i)r_{[i]}^*)} \\
 &= \frac{\phi_i}{\exp(-\epsilon^*)r_{[i]}^*} \\
 &= \frac{\phi_i}{\tilde{n}_{i[i]}\phi_i + \exp(-2\epsilon^*) \sum_{l \geq i} \mathbb{1}_{[x_l \neq x_i]}\phi_l}
 \end{aligned} \tag{3.42}$$

Therefore, the contribution of event time t_i to the likelihood function converges to $\frac{1}{\tilde{n}_{i[i]}}$ if $\epsilon^* \rightarrow \infty$. \square

Theorem 3.4.2 Fitting the IPH model to survival data and assume $\epsilon^* \rightarrow \infty$, then

(i) The restricted hazard functions of an individual with $x = x_j$ will be

$$\begin{aligned}
 \bar{h}(t_i|x_j) &= \begin{cases} 1 & \text{; if an individual with } x = x_j \text{ has the event at } t_i \\ 0 & \text{; censoring times or event times for an individual with } x \neq x_j \end{cases} \\
 \underline{h}(t_i|x_j) &= 0
 \end{aligned}$$

(ii) The unrestricted hazard functions of an individual with $x = x_j$ will be

$$\begin{aligned}
 \bar{h}(t_i|x_j) &= \begin{cases} 1 & \text{; event times} \\ 0 & \text{; censoring times} \end{cases} \\
 \underline{h}(t_i|x_j) &= 0
 \end{aligned}$$

(iii) The contribution of an event time t_i to the likelihood function becomes

$$L_i = 1$$

Proof.

In the same manner as Theorem 3.4.1, the proof for this theorem is straightforward by using the definition of the IPH model and its corresponding likelihood and hazard functions, as in Equations (3.1), (3.16), (3.18), (3.17), (3.20), and (3.19). \square

3.4.1 Bootstrap investigation

The purpose of the section is to investigate whether including imprecision to the PH model by fitting either the GPH or the IPH model is beneficial using Zelterman bootstrap method, as illustrated in Section 2.7. The bootstrap investigation is to give us insights into when these models could be applied, and what value should be used for the imprecision level, ϵ^* .

Three types of survival data sets were considered: proportional hazards data, non-proportional data with crossing hazard functions, and non-proportional data with diverging hazard functions, which are represented by PH, CNPH, and DNPH, respectively. The PH data were generated by assuming a single covariate X that follows a Binomial distribution with $p = 0.5$, the regression parameter $\beta = -0.5$, and the baseline hazard follows the Weibull distribution with scale parameter $\lambda = 2$ and shape parameter $\rho = 3$. Based on two non-proportional hazards distributions, the DNPH data were generated such that the set with $x = 0$ drawn from $\text{Exp}(\lambda = 1)$ while the other set is drawn from $\text{Gompertz}(\lambda = 2, \rho = 0.8)$. Similarly, the CNPH data assumed $\text{Weibull}(\lambda = 0.6, \rho = 0.8)$ for observations related to $x = 0$ and $\text{Weibull}(\lambda = 0.5, \rho = 1.5)$ for observations corresponding to $x = 1$. Figure 3.16 represents the true survival (left) and hazard (right) functions associated with these three types of survival data.

Various factors are taken into account during the bootstrap analysis, including sample size, N , percentage of right-censored observations, and level of imprecision, ϵ^* . For each type of survival data, two sample sizes were considered to ensure computational efficiency. These included generating original samples of moderate size with $N = 60$, and relatively large samples with $N = 1000$ to capture a more comprehensive representation. In addition, two types of censoring are employed, namely complete (uncensored) data and 20% of right-censored data. When fitting either the GPH or IPH model, we consider the following levels of imprecision: $\epsilon^* = 0.0001, 0.1, 0.5, 1, 2$, and 4.

The objectives and procedures of the study will be briefly described before we move onto the details. The objective of this study is to evaluate the feasibility of using either the GPH or the IPH model in place of the PH model when the PH

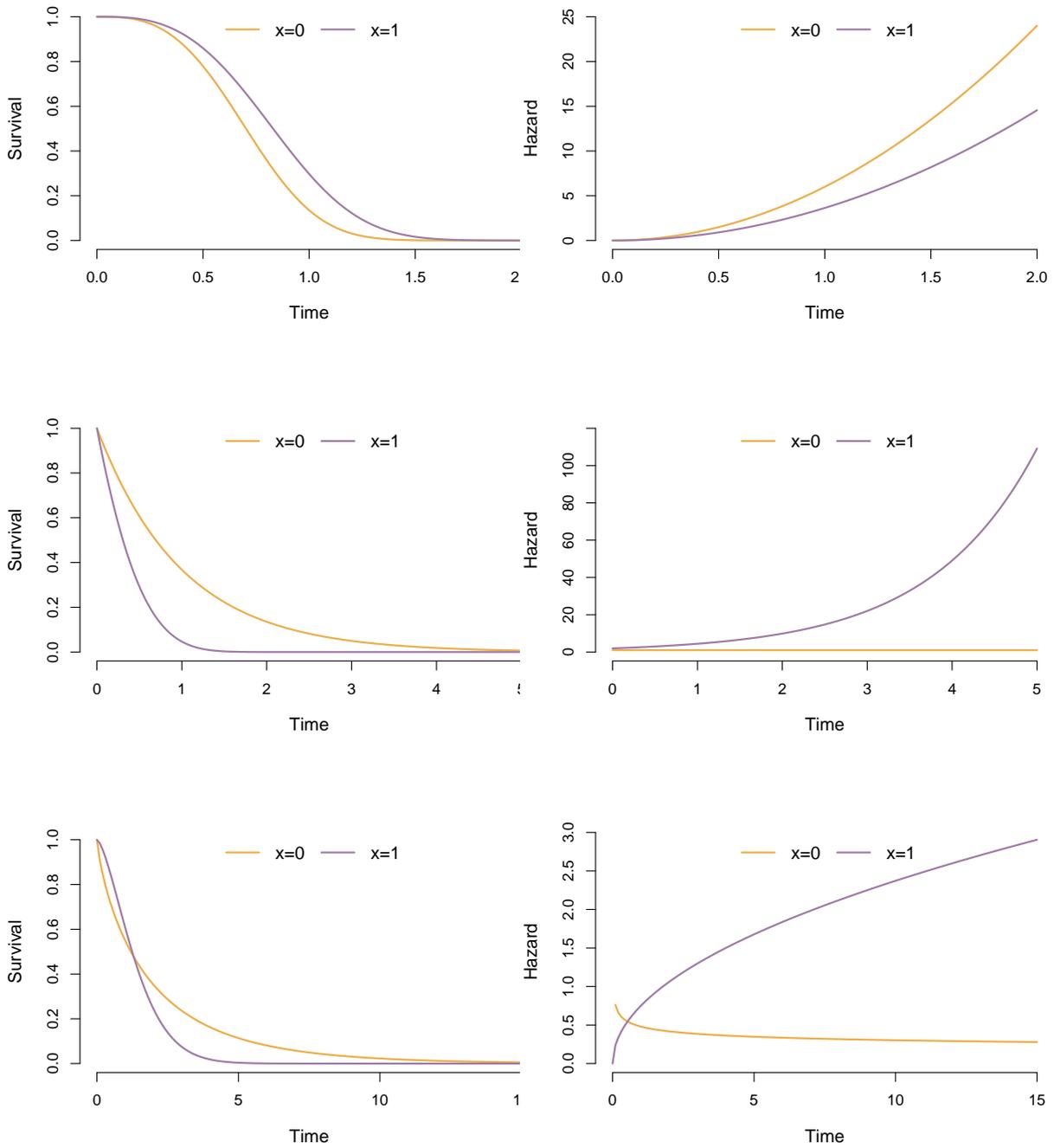


Figure 3.16: Survival distributions (left) and hazard functions (right) for $x = 0$ and $x = 1$ related to the PH, DNPH and CNPH data, respectively

assumption is not valid for the observed data. The evaluation involves bootstrap hypothesis testing based on the null hypothesis that the PH assumption is valid, regardless of evidence to the contrary, as opposed to the alternative hypothesis that the PH assumption is invalid. As part of the bootstrap study, we attempt to determine an appropriate test statistic, denoted by γ . The bootstrap distribution of the test statistic is obtained by resampling under the null hypothesis using the Zelterman method, and computing the corresponding test statistics. The selection of the appropriate test statistic will be based on the relative position of the observed test statistic, γ , in its bootstrap distribution. In particular, the observed test statistic γ should be located in the upper tail if the data set does not follow the PH assumption. This indicates substantial evidence against the null hypothesis. Consequently, the level of imprecision will be gradually increased until the null hypothesis can no longer be rejected. We will investigate the suitability of two test statistics, including ℓ_{ϵ^*} and $\ell_{\epsilon^*} - \ell_0$, where ℓ_{ϵ^*} and ℓ_0 denote the maximum log-likelihood values resulting from our models with imprecision and without imprecision, respectively.

The following procedure will be followed for each scenario of the generated data outlined in this section:

1. Generate $M = 100$ original data sets, D_1, \dots, D_M , according to the assumed scenario, such that the number of observations is N for each data set.
2. For each of the M original data sets fit the target model, either the IPH or GPH, and evaluate the test statistic, leading to $\gamma_1^*, \dots, \gamma_M^*$. Note that γ depends on ϵ^* and according to the choice of γ we will fit the imprecise PH model several times according to the selected levels of ϵ^* . If we consider, for instance, assessing the GPH model using $\gamma = \ell_{\epsilon^*} - \ell_0$ given $\epsilon^* = 0.1$, then we fit the GPH model twice for each of the M data sets once without imprecision assuming $\epsilon^* = 0$ to obtain ℓ_0 and once with imprecision using $\epsilon^* = 0.1$ to obtain $\ell_{0.1}$.
3. Resample $R = 100$ bootstrap samples under the null hypothesis that the original data sets follow the PH assumption using the method proposed by

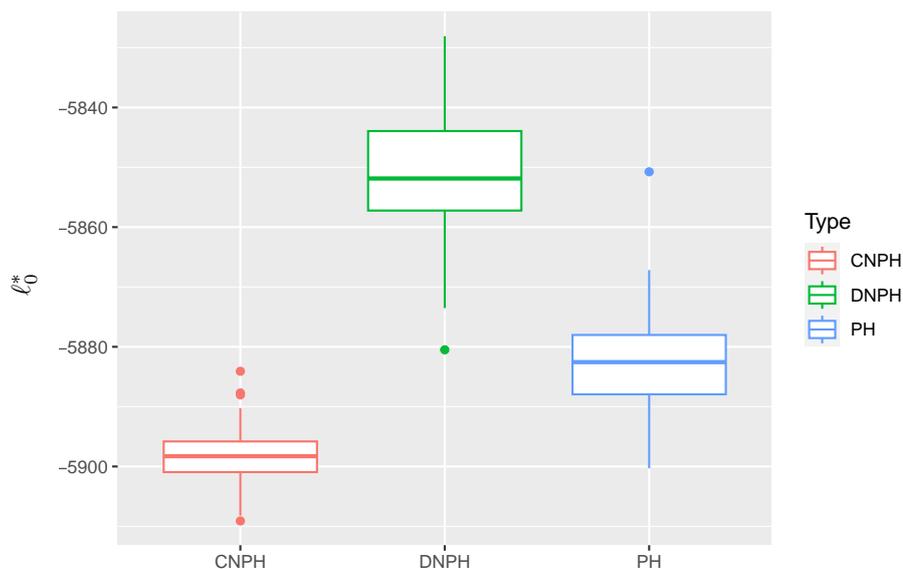


Figure 3.17: The distributions of ℓ_0^* resulted from fitting the GPH model with $\epsilon^* = 0$ to the original (non-censored) survival data with $N = 1000$.

Zelterman *et al.* [91], as illustrated in Section 2.7, along with the estimated parameters obtained from fitting the models without imprecision. In other words, for each original data set, a traditional PH model will be fitted, and the resulting parameter estimate is used to resample R bootstrap samples according to the Zelterman *et al.* [91] method. This procedure will be applied even if the original data do not follow the PH assumption, due to the null hypothesis that the original data sets follow the PH assumption. For the i th original sample we obtain bootstrap samples $B_i^1, B_i^2, \dots, B_i^R$, where $i = 1, 2, \dots, M$. Note that Zelterman bootstrapping may result in ties among the observations of the bootstrap samples; therefore, very small fractions are added to break any ties.

4. Calculate the test statistics $\gamma_i^1, \dots, \gamma_i^R$ for the bootstrap samples $B_i^1, B_i^2, \dots, B_i^R$, where $i = 1, 2, \dots, M$. Upon completion of this step, the result will be a matrix of size $M \times R$, in which the i th row represents the bootstrap distribution of γ_i^* for each of the M original data sets.
5. Record the quantiles where $\gamma_1^*, \dots, \gamma_M^*$ lie on their bootstrap distributions.

Consider investigating the impact of fitting the GPH model using $\gamma = \ell_{\epsilon^*}$, as

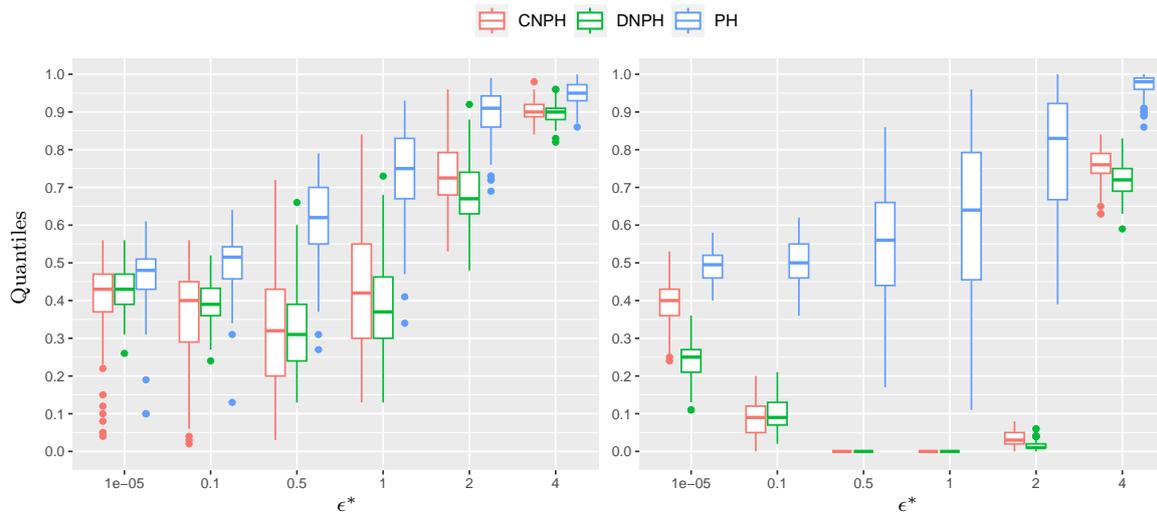


Figure 3.18: Quantiles of $\hat{\ell}_\epsilon^*$ for non-censored survival data using the GPH model for $N = 60$ (left) and $N = 1000$ (right).

a test statistic. We know that increasing the imprecision level will increase the contribution of each event time to the likelihood function which leads to a higher likelihood value for both PH and non-PH data. Theorem 3.4.1 and Example 3.3.2 demonstrate that fitting the GPH model and increasing the imprecision level yield to a higher likelihood value compared to fitting the PH model. When $\epsilon^* = 0$, the majority of likelihood values of ℓ_0^* obtained from fitting the GPH model to the DNPH data are higher than those obtained from fitting the GPH model to either PH or CNPH data, see Figure 3.17. The boxplots in Figure 3.18 illustrate the quantiles of $\ell_{\epsilon^*}^*$ in their corresponding bootstrap distributions for $N = 60$ (left) and $N = 1000$ (right). For small ϵ^* in the PH data, the quantiles of $\ell_{\epsilon^*}^*$ are around the median, and they gradually increase with increasing ϵ^* . Non-PH data, on the other hand, showed initial quantiles of $\ell_{\epsilon^*}^*$ relatively less than the median, which then slightly decreased until $\epsilon^* = 2$, at which point the quantiles of $\ell_{\epsilon^*}^*$ began to increase. Similar patterns can be seen for $N = 1000$; however, the quantiles decrease sharply to zero and remained until $\epsilon^* = 2$. Consequently, $\gamma = \ell_{\epsilon^*}^*$ may not be an adequate statistic to assess the benefits of fitting the GPH model to non-proportional hazards data, in comparison to proportional hazards data.

Based on standard mathematical statistics, especially the Neyman-Pearson lemma

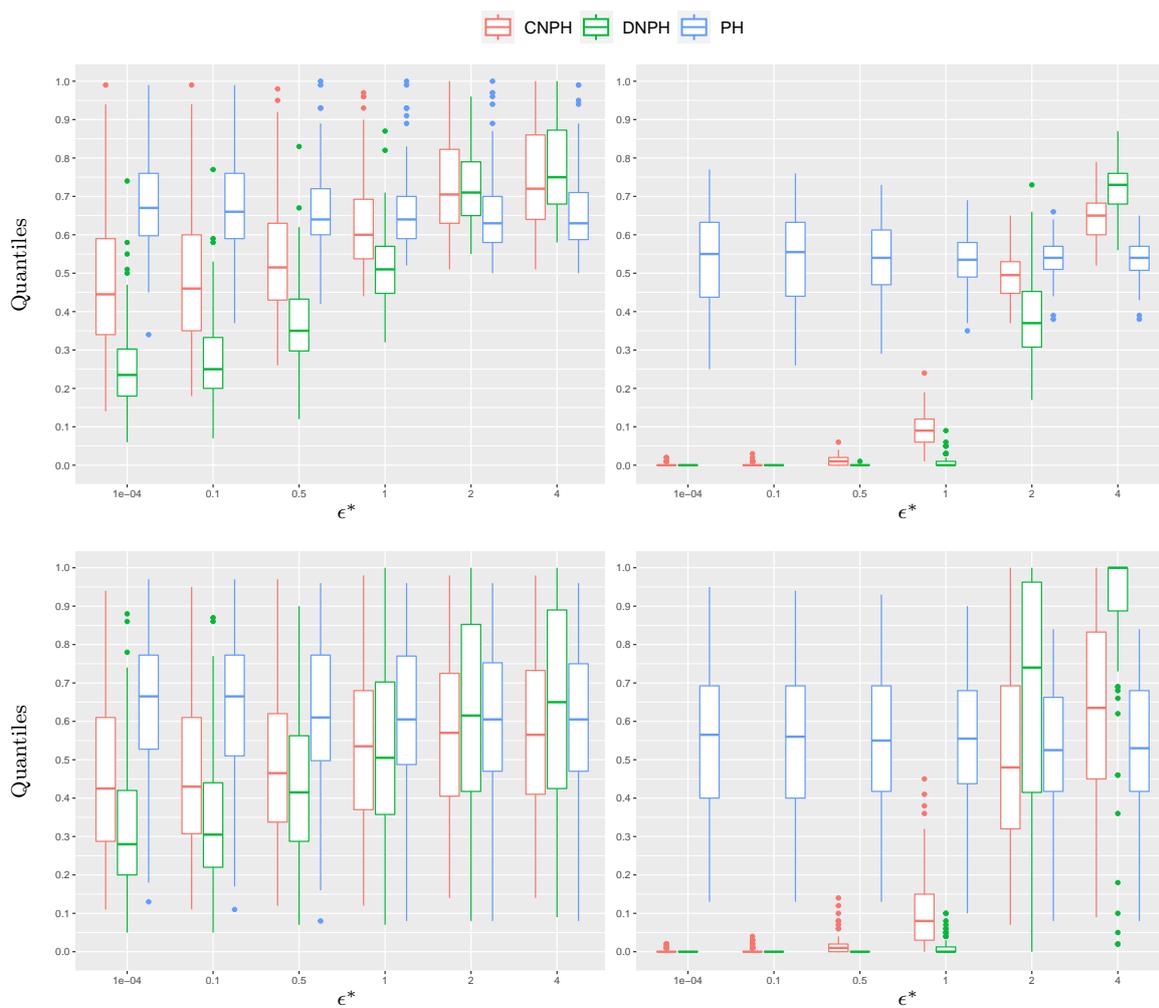


Figure 3.19: Quantiles of $(\hat{\ell}_\epsilon - \hat{\ell}_0)^*$ for non-censored survival data (top) and for survival data with 20% of right-censored observations (bottom) using the GPH model for $N = 60$ (left) and $N = 1000$ (right).

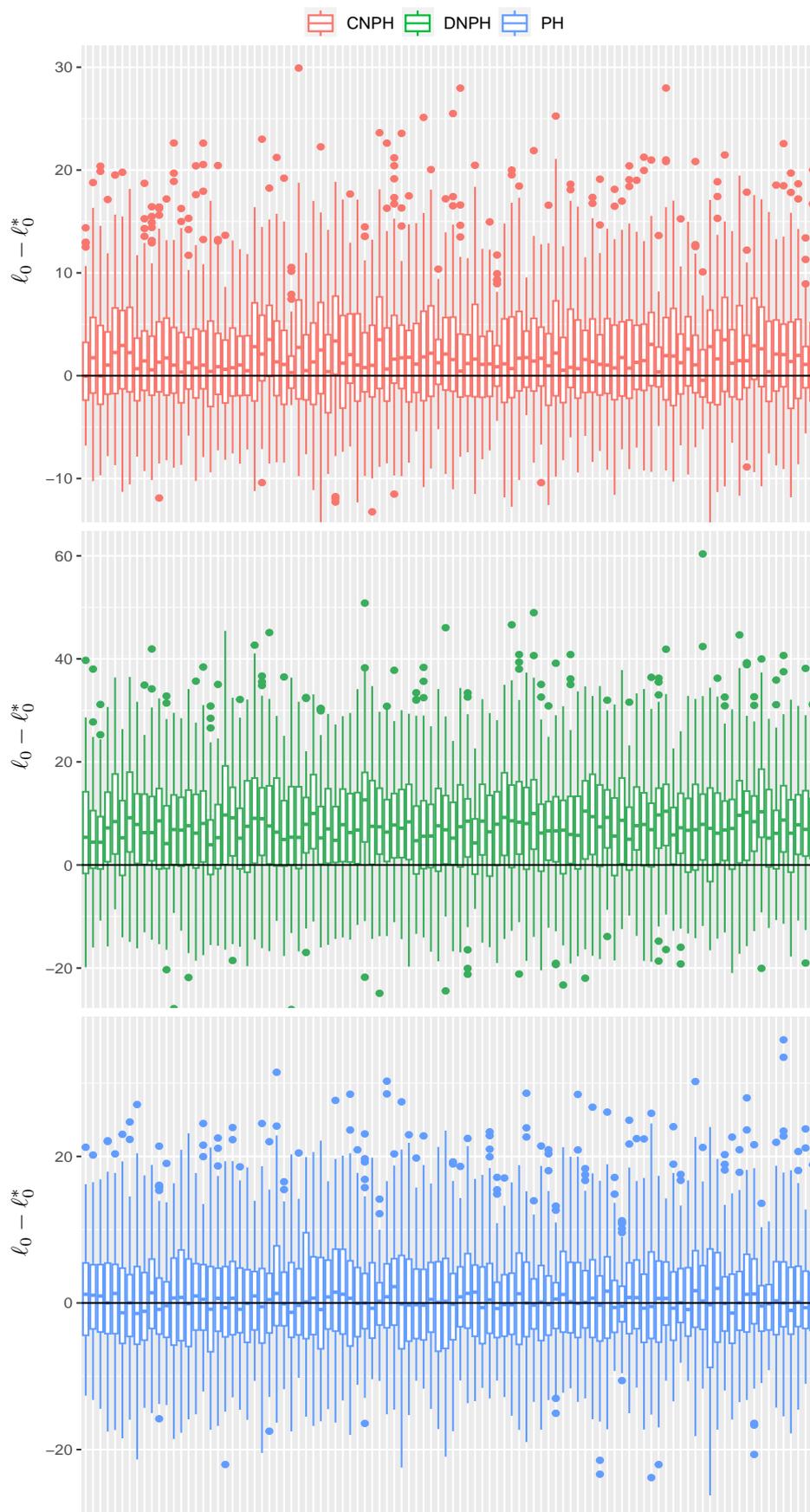


Figure 3.20: The differences between the log-likelihood values obtained by fitting the GPH model to the bootstrap samples and their corresponding 100 original data sets with $N = 1000$

and the generalized likelihood ratio test, we consider the log-likelihood increment that results from imprecision instead of the log-likelihood value. Hence, the differences in log-likelihood values between GPH models with and without imprecision, $\ell_{\epsilon^*} - \ell_0$, are used as test statistic. Let $(\ell_{\epsilon^*} - \ell_0)^*$ denotes the value of $\ell_{\epsilon^*} - \ell_0$ related to the original M data sets. Interestingly, as illustrated in Figure 3.19 (right), the quantiles of $(\ell_{\epsilon^*} - \ell_0)^*$ are in the opposite direction from what we anticipated for both complete survival data (top) and survival data with 20% of right-censored observations (bottom). According to Figure 3.19 (right), the quantiles of the log-likelihood increments for the DNPH and CNPH data are significantly lower than their bootstrap counterparts where the PH assumption is held. For the PH data, the quantiles of $(\ell_{\epsilon^*} - \ell_0)^*$ are between 0.25 – 0.75 in their bootstrap distributions for small imprecision level, and variability decreases as the imprecision level increases. Similar results were demonstrated for survival data with right-censored observations, but with higher variability. This result may not be visible for small or moderate-sized data, as shown in Figure 3.19 (left). The remainder of this section investigates why the bootstrap quantiles of $(\ell_{\epsilon^*} - \ell_0)^*$ for non-proportional hazards data are lower than those for proportional hazards data, particularly at small levels of imprecision. Further, we will briefly comment on determining the optimal value of ϵ^* .

In order to validate the bootstrap results associated with $\ell_0 = \{\ell_0^1, \dots, \ell_0^R\}$, we first verify whether Zelterman bootstrapping is functioning correctly. This can be done by inspecting the distribution of the difference between $\ell_0 = \{\ell_0^1, \dots, \ell_0^R\}$, resulting from the bootstrap samples, and their corresponding ℓ_0^* for the original 100 data sets with $N = 1000$. The boxplots in Figures 3.20 illustrate that for the PH data sets, the medians of the differences are generally close to zero, with some variations on either side. Furthermore, both the CNPH and DNPH data sets have positive medians, perhaps a gain of two to three in the log-likelihood values for the former and slightly more for the latter, although the original log-likelihood values for fitting the GPH model without imprecision were lower for the CNPH compared to the PH, as shown in Figure 3.17. This is relevant because it is consistent with the concept that Zelterman bootstrapping produces PH-like samples, thereby increasing the log-likelihood value. Having confirmed that Zelterman bootstrapping

is functioning as intended, we will now investigate the impact of including small ϵ^* to the log-likelihood value. The following theorem approximates the value of $\ell_{\epsilon^*} - \ell_0$ for small ϵ^* .

Theorem 3.4.3 For small ϵ^* , the difference between the log-likelihood values resulted from fitting the GPH model with imprecision level ϵ^* and without imprecision can be approximated by

$$\hat{\ell}_{\epsilon^*} - \hat{\ell}_0 \approx 2\epsilon^*\eta \quad (3.43)$$

where $\eta = \sum_{i=1}^n \frac{\delta_i r'_i}{r_i}$, r_i is as used previously, and $r'_i = \sum_{l \geq i} \mathbb{1}_{[x_l \neq x_i]} \phi_l$.

Proof.

Recall the GPH profile likelihood function in Equation (3.30), then the profile log-likelihood function is

$$\ell_{\epsilon^*}(\beta) = \sum_{i=1}^n \delta_i \ln \phi_i - \sum_{i=1}^n \delta_i \ln \left(\sum_{l \geq i} \phi_l [\mathbb{1}_{[x_l = x_i]} + \exp(-2\epsilon^*) \mathbb{1}_{[x_l \neq x_i]}] \right)$$

The derivative of this function with respect to ϵ^* is

$$\begin{aligned} \frac{\partial \ell_{\epsilon^*}(\beta)}{\partial \epsilon^*} &= \frac{\partial}{\partial \epsilon^*} \sum_{i=1}^n \delta_i \ln \phi_i - \frac{\partial}{\partial \epsilon^*} \sum_{i=1}^n \delta_i \ln \left(\sum_{l \geq i} \phi_l [\mathbb{1}_{[x_l = x_i]} + \exp(-2\epsilon^*) \mathbb{1}_{[x_l \neq x_i]}] \right) \\ &= - \sum_{i=1}^n \delta_i \frac{\partial}{\partial \epsilon^*} \ln \left(\sum_{l \geq i} \phi_l [\mathbb{1}_{[x_l = x_i]} + \exp(-2\epsilon^*) \mathbb{1}_{[x_l \neq x_i]}] \right) \\ &= - \sum_{i=1}^n \delta_i \frac{\frac{\partial}{\partial \epsilon^*} \sum_{l \geq i} \phi_l \mathbb{1}_{[x_l = x_i]} + \frac{\partial}{\partial \epsilon^*} \sum_{l \geq i} \phi_l \exp(-2\epsilon^*) \mathbb{1}_{[x_l \neq x_i]}}{\sum_{l \geq i} \phi_l [\mathbb{1}_{[x_l = x_i]} + \exp(-2\epsilon^*) \mathbb{1}_{[x_l \neq x_i]}}} \\ &= 2 \sum_{i=1}^n \frac{\delta_i \sum_{l \geq i} \phi_l \exp(-2\epsilon^*) \mathbb{1}_{[x_l \neq x_i]}}{\sum_{l \geq i} \phi_l [\mathbb{1}_{[x_l = x_i]} + \exp(-2\epsilon^*) \mathbb{1}_{[x_l \neq x_i]}}} \end{aligned}$$

Hence, the derivative of the GPH profile log-likelihood function evaluated at $\epsilon^* = 0$ is

$$\left. \frac{\partial \ell_{\epsilon^*}(\beta)}{\partial \epsilon^*} \right|_{\epsilon^*=0} = 2 \sum_{i=1}^n \frac{\delta_i r'_i}{r_i}$$

Where $r_i = \sum_{l \geq i} \phi_l$, as used previously, and $r'_i = \sum_{l \geq i} \phi_l \mathbb{1}_{[x_l \neq x_i]}$. It is obvious that changing ϵ^* leads to change on $\hat{\beta}$, so we can consider $\hat{\beta}$ as a function of ϵ^* . Let $\hat{\ell}_{\epsilon^*}$ denote the log-likelihood function at its maximum such that $\hat{\ell}_{\epsilon^*} = \ell_{\epsilon^*}(\hat{\beta}(\epsilon^*))$. Now,

we can use first-order Taylor series to approximate the value of $\hat{\ell}_{\epsilon^*}$ for small ϵ^* as follows

$$\begin{aligned}\hat{\ell}_{\epsilon^*} &\approx \hat{\ell}_0 + \frac{\partial}{\partial \epsilon^*} \ell_{\epsilon^*} \left(\hat{\beta}(\epsilon^*) \right) \cdot (\epsilon^* - 0) \\ &\approx \hat{\ell}_0 + 2\epsilon^* \sum_{i=1}^n \frac{\delta_i r'_i}{r_i}\end{aligned}$$

Consequently, the difference between the log-likelihood values resulting from fitting the GPH model with imprecision level ϵ^* and without imprecision can be approximated for small $\epsilon^* > 0$ by

$$\hat{\ell}_{\epsilon^*} - \hat{\ell}_0 \approx 2\epsilon^* \eta$$

where η denote the summation in the difference between the log-likelihood values, $\sum_{i=1}^n \frac{\delta_i r'_i}{r_i}$. □

In order to shed light on the observed trend of log-likelihood gains for original NPH data, CNPH or DNPH, residing within the lower tail of the bootstrap distribution obtained from the GPH model, the following analysis aims to delve into the underlying causes. This objective will be achieved by inspecting the implications of deviating from the PH assumption through various examples that showcase diverse structures of both PH and NPH data, in particular by comparing the η values derived from Theorem 3.4.3. Through this intensive examination, we hope to uncover any meaningful patterns or explanations for this phenomenon. Examples 3.4.1 and 3.4.2 examine survival data with a binary covariate for balanced groups, while Example A.1.1 in Appendix A.1 examines survival data with imbalanced groups. Given that the summation in η is taken effectively only over event times and to keep these examples simple we assume that censored observations do not exist.

Example 3.4.1 For this example, we consider fitting the GPH model only to PH data with a binary covariate, x , which includes evenly balanced groups that are temporally ordered according to m -blocks. Assume a covariate in which each block consists of an observation from each group in the following order

$$\underbrace{1}_{1,0}, \underbrace{2}_{1,0}, \dots, \underbrace{m}_{1,0}$$

n	$\hat{\beta}$	η	η/n	PH test
4	0.94	1.123	0.281	1
20	0.226	8.87	0.444	0.88
200	0.033	98.35	0.491	0.87
2×10^3	0.004	997.8	0.498	0.92

Table 3.9: The impact on η for evenly balanced PH survival data with a covariate consisting of m -blocks of $(1, 0)$

For the stated type of survival data and a variety of sample sizes, Table 3.9 shows the estimates of the regression coefficients and derivatives of log-profile likelihood with respect to ϵ , denoted by η . Based on the null hypothesis that the corresponding data follow the PH assumption, Table 3.9 also displays p-values of Schoenfeld residuals test. It can be seen that $\hat{\beta}$ converges to zero as $n \rightarrow \infty$, which leads to $\phi_{[0]} = \phi_{[1]}$. Consequently, η can be determined using the following

$$\frac{m}{2m}, \frac{m-1}{2m-1}, \frac{m-1}{2m-2}, \frac{m-2}{2m-3}, \frac{m-2}{2m-4}, \frac{m-2}{2m-5}, \frac{m-3}{2m-6}, \dots, \frac{2}{4}, \frac{1}{3}, \frac{1}{2}, 0$$

Similarly, when the temporally ordered covariate exhibits evenly distributed groups in m -blocks, such that each block consists of two observations from each group and is arranged as the following sequence

$$\overbrace{1, 1, 0, 0}^1, \dots, \overbrace{1, 1, 0, 0}^m$$

As n approaches infinity, $\phi_{[0]} = \phi_{[1]}$ as illustrated in Table 3.10 which indicates that η is comprised of the following

$$\frac{2m}{4m}, \frac{2m}{4m-1}, \frac{2m-2}{4m-2}, \frac{2m-2}{4m-3}, \frac{2m-2}{4m-4}, \frac{2m-2}{4m-5}, \frac{2m-4}{4m-6}, \dots, \frac{2}{4}, \frac{2}{3}, 0, 0$$

We observe the same results for other PH data with equally balanced groups as well. For instance, Table 3.11 indicates that when the temporally ordered binary

n	$\hat{\beta}$	η	η/n	PH test
4	∞	0	0	1
20	0.44	8.51	0.425	0.8
200	0.06	98.1	0.491	0.77
2×10^3	0.008	997.56	0.499	0.86

Table 3.10: The impact on η for evenly balanced PH survival data with a covariate consists of m -blocks of $(1, 1, 0, 0)$

covariate contains m -blocks of $(1, 0, 0, 1)$ or $(1, 0, 0, 0, 1, 1)$, respectively, η can be obtained by

$$\left\{ \frac{2m}{4m}, \frac{2m-1}{4m-1}, \frac{2m-1}{4m-2}, \frac{2m-2}{4m-3}, \frac{2m-2}{4m-4}, \frac{2m-3}{4m-5}, \frac{2m-3}{4m-6}, \dots, \frac{2}{4}, \frac{1}{3}, \frac{1}{2}, 0 \right\}$$

$$\left\{ \frac{3m}{6m}, \frac{3m-1}{6m-1}, \frac{3m-1}{6m-2}, \frac{3m-1}{6m-3}, \frac{3m-3}{6m-4}, \frac{3m-3}{6m-5}, \frac{3m-3}{6m-6}, \frac{3m-4}{6m-7}, \dots, \frac{3}{6}, \frac{2}{5}, \frac{2}{4}, \frac{2}{3}, 0, 0 \right\}$$

In light of the results in Tables 3.9, 3.10, and 3.11, it is reasonable to expect that the maximum derivative of log-profile-likelihood with respect to ϵ^* is of the order $n/2$, with n the number of observations. In fact, we believe that this is true for PH data with no censored observations and the groups are evenly balanced using a considerably large sample, so that $\hat{\beta}$ converges to zero, see Example A.1.1 for imbalance data.

◇

Example 3.4.2 The GPH model is fitted here to NPH data with a binary covariate. The covariate is composed of evenly balanced groups that are temporally ordered according to different scenarios. Suppose that the covariate consists of the following three-blocks

x	n	$\hat{\beta}$	η	η/n	PH test
	4	-0.48	1.236	0.309	0.14
$\overbrace{(1, 0, 0, 1)}^1, \dots, \overbrace{(1, 0, 0, 1)}^m$	20	-0.081	8.915	0.446	0.72
	200	-0.0079	98.3	0.49	0.92
	2×10^3	-0.0007	997.78	0.499	0.97
	6	-1.208	1.54	0.257	0.07
$\overbrace{(1, 0, 0, 0, 1, 1)}^1, \dots, \overbrace{(1, 0, 0, 0, 1, 1)}^m$	24	-0.27	10.58	0.441	0.59
	204	-0.04	100.14	0.491	0.795
	2004	-0.0052	999.58	0.499	0.9

Table 3.11: The impact on η for evenly balanced PH data with different covariates structures

$$\overbrace{1, 1, \dots, 1}^m, \overbrace{0, 0, \dots, 0}^{2m}, \overbrace{1, 1, \dots, 1}^m$$

such that the sample size $n = m + 2m + m = 4m$ for any positive integer m . For this particular covariate structure of NPH data, Table 3.12 indicates that the estimates of the regression coefficients reaches -0.821 when $n = 2000$. Notice, the value of η is smaller than what would be expected for PH data, which is about 37.7% of the sample size in this NPH data. Another possible scenario of non-proportional

n	$\hat{\beta}$	η	η/n	PH test
4	-0.48	1.24	0.31	0.14
20	-0.717	7.209	0.36	6.8×10^{-5}
200	-0.809	75	0.375	4.4×10^{-44}
2×10^3	-0.821	753.29	0.377	0

Table 3.12: The impact on η for evenly balanced 3-blocks NPH survival data using the $(1_{\{m\}}, 0_{\{2m\}}, 1_{\{m\}})$ structure

n	$\hat{\beta}$	η	η/n	PH test
6	0.014	2.186	0.364	0.062
18	-0.0836	7.29	0.405	2×10^{-4}
198	-0.149	84.54	0.427	0
1998	-0.155	857.31	0.429	0

Table 3.13: The impact on η for evenly balanced 3-blocks NPH survival data using the $(1_{\{2m\}}, 0_{\{3m\}}, 1_{\{m\}})$ structure

n	$\hat{\beta}$	η	η/n	PH test
6	0.362	2.158	0.36	0.63
18	0.463	7.216	0.401	0.25
198	0.534	83.94	0.42	3.7×10^{-6}
1998	0.5417	851.5	0.426	8.3×10^{-51}

Table 3.14: The impact on η for evenly balanced 4-blocks NPH survival data using the $(1_{\{m\}}, 0_{\{2m\}}, 1_{\{2m\}}, 0_{\{m\}})$ structure

hazards data is associated with the following 4-blocks of a binary covariate

$$\overbrace{1, 1, \dots, 1}^{2m}, \overbrace{0, 0, \dots, 0}^{3m}, \overbrace{1, 1, \dots, 1}^m$$

The findings presented in Table 3.13 illustrate that applying the GPH model to this structure of covariate values, combined with increasing the sample size, yields a η value approximately 30% of the actual sample size which is obviously less than the log-likelihood increments found for PH data in Example 3.4.1.

Based on Table 3.14, it is evident that the log-likelihood gain is approximately 42% of the sample size for non-proportional hazards data related to the binary covariate comprising of the following blocks

$$\overbrace{1, 1, \dots, 1}^m, \overbrace{0, 0, \dots, 0}^{2m}, \overbrace{1, 1, \dots, 1}^{2m}, \overbrace{0, 0, \dots, 0}^m$$

Various structures of NPH data have been examined, revealing that the approximate log-likelihood increments resulting from using the GPH model is smaller for NPH data compared to PH data for large samples and evenly balanced groups without censored observations.

For extreme cases of survival data, particularly when the regression coefficient approaches infinity or negative infinity, the derivative of the log-profile likelihood with respect to ϵ^* will be zero. An example of that would be a covariate which divides observations into distinct clusters, as can be seen in the following blocks

$$\overbrace{1, 1, \dots, 1}^{n/2}, \overbrace{0, 0, \dots, 0}^{n/2}$$

Clearly, as $\hat{\beta}$ approaches infinity we have $\phi_{[0]} = 1$ and $\phi_{[1]} \rightarrow \infty$. Hence, $\sum_{i=1}^{n/2} \frac{r'_i}{r_i} \rightarrow 0$ for the first $n/2$ components of $\frac{r'_i}{r_i}$. For the second half of events, the value of $r'_i = 0$ since there is no observation left with $x = 1$, so

$$\sum_{i=1}^n \frac{r'_i}{r_i} \rightarrow 0$$

◇

Contrary to our initial proposal, the investigation based on Theorem 3.4.3 and Examples 3.4.1, 3.4.2 and A.1.1, indicate that fitting the GPH model to PH data consistently yielded higher log-likelihood increments compared to NPH data. Further, the disparity between the PH and NPH data diminishes in relation to the log-likelihood gains achieved through the application of the GPH model as the data becomes more imbalanced. A notable aspect highlighted by these examples is that fitting the GPH model to small to moderate PH and NPH data does not result in significant differences in log-likelihood gains. However, the disparity becomes more apparent as the sample size increases. This aspect is consistent with our bootstrap results in Figure 3.19 (left), which show no evidence of different log-likelihood gains for PH versus NPH data when $n = 60$. On the other hand, the disparity in log-likelihood gains resulting from fitting the GPH model for PH and NPH data is clearly evident when using a sample size of $n = 1000$, as in Figure 3.19 (right).

Having acknowledged that our initial proposal was incorrect, our proposal was revised by rejecting the null hypothesis when the test statistic $(\ell_{\epsilon^*} - \ell_0)^*$ related to the original data resides within the lower tail of its bootstrap distribution. Accordingly, the optimal level of imprecision level can be determined by finding the value of ϵ^* at which the null hypothesis cannot be rejected at a particular significance level. Consider DNPH and CNPH data with $N = 1000$, Figure 3.19 (right) suggests that the appropriate level of imprecision can be between 1 – 2 for DNPH and between 0.5–2 for CNPH. Thus, the GPH model can serve as a reliable alternative to the PH model in scenarios where the PH assumption is questionable. This can be achieved by fitting the GPH model and determining the optimal level of imprecision by conducting bootstrap hypothesis testing using $\ell_{\epsilon^*} - \ell_0$ as a test statistic, and gradually increasing the value of ϵ^* until the null hypothesis can no longer be rejected.

3.5 Concluding remarks

In this chapter, two imprecise proportional hazards models are introduced based on Poisson empirical likelihoods, namely an individual-based model, IPH, and a group-based imprecise PH model. The IPH model assumes that each individual in the data set has unique imprecision factors, whereas the GPH model allows groups of individuals to share the same imprecision factors. Hence, the IPH model can be viewed as a generalization of the GPH model, in which each group or category has a unique member. Originally, we attempted to construct the full likelihood of the IPH and the GPH models using the empirical likelihood based on the CDF, as described in Section 2.5.2. Due to the complexity of the problem, however, we have instead considered the Poisson empirical likelihood in light of the fact that proportionality of the hazard can be handled easily.

When the imprecision level is zero, these models produce the MLE of the partial likelihood for the PH model, and the difference of the log-likelihood values equals the number of non-censored observations. As the imprecision level is increased, the the estimate of the parameter converges to zero and the log-likelihood value increase.

The estimation of lower and upper survival functions for specific individuals in these models was approached through restricted and unrestricted survival functions. It was observed that both the IPH and GPH models produced higher upper survival estimates compared to the PH survival estimates. Furthermore, the unrestricted survival estimates for both models exhibited a greater disparity between the lower and upper survival estimates, which can be attributed to the escalating hazard at each event time in contrast to the restricted survival estimates. The restricted survival estimates derived from both the GPH and IPH models are directly influenced by the diversity of covariate values in the survival data. In other words, the GPH and IPH models produce similar results when there is limited repetition of covariate values among individuals. With the exception of GPH unrestricted survival estimates, none of these survival estimations ensure that the PH survival estimates will fall within the upper and lower survival estimates. It has been found that the upper survival functions, both IPH restricted and unrestricted, are less sensitive to imprecise levels than the lower survival functions.

When examining the scenario in which imprecision reaches infinite levels, interesting findings emerged. Specifically, for the GPH restricted hazard estimates, it was shown that the upper hazard estimates for an individual reduce to the Nelson-Aalen estimator, while the lower hazard estimates equal zero. With respect to the GPH unrestricted hazard estimates, it was found that the upper hazard estimates for an individual converge to Breslow PH baseline hazard estimates with the regression coefficient set to zero. The corresponding lower hazard estimates equals zero similar to the restricted type. As the IPH model is simply a special case of the GPH model where each group consists of one member, the upper restricted hazard estimates associated with the individual's event time were reduced to one, while the lower restricted hazard estimates were reduced to zero for event times associated with other individuals, censored times, and lower restricted hazard estimates. The IPH unrestricted hazard estimates, in contrast, converge to 1 for all event times, and zero for right-censored time and for lower unrestricted hazard estimates. In light of these results, our focus has shifted towards the GPH model in the bootstrap assessment in Section 3.4.1.

In order to assess the benefits of using the GPH and the impact of increasing the impression level for both proportional hazards data and non-proportional hazards data, we performed a bootstrap study following Section 2.7.1. Through bootstrapping, we gained insight into when this model can be applied and what imprecision level should be used. Three types of survival data were considered, namely proportional hazards data, non-proportional data with crossing hazard functions and non-crossing hazard functions. A key aspect of the GPH model revealed by the bootstrap results is that for PH data, the GPH model generally results in more significant increase in log-likelihood than for NPH data. In light of this finding, it was decided to modify the null hypothesis in order to fit the purpose of the study. The Log-likelihood increments, which are obtained from fitting the GPH model with and without imprecision, were found to be appropriate statistics for evaluating the validity of the PH assumption. Nevertheless, for small to moderate sample size, there are no significant differences in log-likelihood gains, which suggests that this test statistic can only be reliable for large samples. Further, the bootstrap hypothesis test can also be employed to determine the optimal level of imprecision. This is done by finding the value of ϵ^* at which the null hypothesis cannot be rejected at a particular significance level. Accordingly, the GPH model can be used instead of the PH model safely in scenarios where the validity of the PH assumption is uncertain.

Chapter 4

Robust PH model

4.1 Introduction

In applications, covariates are often subject to measurement errors. A common approach to dealing with this problem is to use the mismeasured version of the covariate directly. However, this naive approach can lead to biased estimates and incorrect conclusions if not properly addressed. In the realm of proportional hazards models, measurement error presents a significant challenge, as evidenced by Prentice [70], Nakamura [64], and Hu *et al.* [43], among others. The literature reveals a diverse range of approaches and solutions to this issue.

Prentice [70], as highlighted by Augustin and Schwarz [8], delves into the limitations of applying simple likelihood-based corrections in the proportional hazards model. Prentice discusses the challenges of maintaining multiplicative form when faced with time-dependent covariates. In response, the focus of analysis has been shifted to the rare disease assumption, which allows analytical solutions to normal measurement errors to be developed, which align with the results of regression calibration [42].

Hu *et al.* [43] contribute to the structural approach by exploring likelihood-based methods within the framework of the classic additive error model, characterized by a known error distribution [19]. They focus on situations where the surrogate covariate is measured on all subjects, with the true covariate ascertained on a validation set [17]. This method allows for the estimation of error distribution and the calculation

of parameter estimators and their variances, offering practical application potential [43].

Nakamura [64] presents a fundamental contribution to the concept of corrected score functions and corrected log-likelihood as a solution to measurement for covariates error in proportional hazards models. Despite the partial likelihood's inherent complexities, Nakamura proposes first and second-order approximations, yielding surprisingly consistent estimators, as noted by Kong and Gu [52]. This method demonstrates effectiveness in simulation studies and extends to non-normal measurement errors, broadening its applicability [15, 80]. The interpretation by Augustin [5] of Nakamura's method offers a different perspective, suggesting that these corrections are exact when applied to the Breslow likelihood. This insight extends Nakamura's method to a wider range of hazard models and measurement error distributions, underscoring its versatility.

The methods in the literature on measurement error in the proportional hazards model are intended to enhance data analysis accuracy and reliability through the reduction of estimation bias. Contrary to these methods, this chapter leverages the incorporation of errors into covariate values as a strategy to mitigate the restrictions imposed by the proportional hazards assumption. As a result, a more robust modeling framework is achieved when the proportional hazards assumption is doubtful. In our approach, we propose a modified version of the observed covariate values by adding error terms for each covariate value. These errors are not restricted to any particular distribution, and are independent of survival times and the covariate, but are assumed to fluctuate within a small interval. The robust PH model is assessed using different likelihoods to determine their impact on the estimation of the regression parameter, the enhancement of the likelihood value, and the estimation of survival functions for a particular individual.

This chapter is structured in the following manner: Section 4.2 provides an overview of the robust PH model formulation. Section 4.3 introduces the model utilizing the Poisson full likelihood, building upon the concepts discussed in Section 2.5.1. Section 4.4 details the implementation of the robust PH model using an empirical likelihood based on the CDF, aligning with the methodologies outlined by

[77], also Section 2.5.2. Detailed simulation studies are provided in Section 4.5 that investigate some aspects of the robust PH model in relation to imprecise survival estimates and the reliability the robust PH model. The chapter concludes with Section 4.6, offering some final observations and remarks.

4.2 Robust PH model

In this section we shall use a slightly different notation to that previously used, in order to distinguish between the standard PH model and the robust PH model. In the robust PH model, instead of using the observed value of the covariate for the i th individual, x_i , we consider a modified version of that value $x_i + \epsilon_i$. Consequently, the function $\phi_i = \exp(\beta x_i)$ in Equation (2.31) will be replaced by $\alpha_i = \exp(\beta(x_i + \epsilon_i))$. The resulting robust PH model can be expressed as follows

$$h_i(t) = h_0(t)\alpha_i \quad (4.1)$$

where ϵ_i is an imprecision term related to the covariate value x_i , for $i = 1, 2, \dots, n$. Furthermore, these imprecision terms are permitted to vary within a typically small interval such that $|\epsilon_i| \leq \epsilon^*$. It could be argued that assuming errors belong to a specific distribution, as is commonly formulated, might be more reasonable than adopting the less familiar unknown bounded errors. However, as emphasized in Belforte *et al.* [10], the information available regarding measurement errors in real-life applications tends to align closely with 'unknown but bounded' errors, including systematic and random errors, which are also often perceived to be bounded, since errors greater than three standard deviations are commonly regarded as impossible.

In the context of applying the robust PH model in Equation (4.1), particularly when the PH assumption is questionable, the task of selecting and validating an appropriate level of imprecision is quite challenging. Therefore, the level of imprecision is currently regarded as subjective. This allows researchers the flexibility to decide on the appropriate level of imprecision for their specific studies to ensure the safe application of a PH-like model. For instance, the level of imprecision, ϵ^* , can be considered as a few percent of the standard deviation or the IQR of the observed covariate values. Consequently, the covariate can be regarded as an imprecise measure

of the covariate value for the individuals (e.g. height may be accurately measured and recorded rounded to the nearest centimeter, so real height can be within half a centimeter either side of the recorded value).

As part of the analysis of the robust PH model, different likelihood representations have been considered and outlined in Sections 4.3 and 4.4 related to the poisson and empirical likelihoods, respectively. The regression coefficients and survival functions will therefore be subscripted by the lower case letters p and e to indicate the corresponding likelihood function.

4.3 Robust-Poisson PH likelihood (RP)

This section illustrates the Poisson full likelihood for the robust PH model. This is essentially a version of the Poisson full likelihood for the PH model, discussed in Section 2.5.1, modified to include imprecision in the values of the covariate. Suppose we observe the following survival data

$$(t_1, \delta_1, x_1), (t_2, \delta_2, x_2), \dots, (t_n, \delta_n, x_n)$$

where t_i is an observed value of $T_i = \min\{V_i, C_i\}$ and $\delta_i = I\{V_i \leq C_i\}$ is the censoring indicator. Here V_i and C_i , for $i = 1, \dots, n$, are independent positive continuous random variables such that C_i is the censoring time associated with the survival time V_i . Additionally, consider x_1, \dots, x_n as values of a one-dimensional covariate that corresponds to the survival times V_1, \dots, V_n . The generalization of this methodology to multidimensional covariates is left for future studies. Based on the robust PH model, the hazard function and cumulative hazard function of the i th individual are

$$h_i(t) = h_0(t)\alpha_i, \quad H_i(t) = H_0(t)\alpha_i \quad (4.2)$$

with $\alpha_i = \exp(\beta(x_i + \epsilon_i))$ as used in Equation (4.1). According to Equation (2.50), the Poisson full likelihood associated with the robust PH model is given by

$$\begin{aligned} L(\beta_p, h_0(t_1), \dots, h_0(t_n), \epsilon_1, \dots, \epsilon_n) &= \prod_{i=1}^n [h_i(t_i)^{\delta_i} \exp(-H_i(t_i))] \\ &= \prod_{i=1}^n [(h_0(t_i)\alpha_i)^{\delta_i} \exp(-H_0(t_i)\alpha_i)] \end{aligned} \quad (4.3)$$

Theorem 4.3.1 Based on Poisson full likelihood function for the robust PH model in Equation (4.3) we have the following:

- (i) The baseline hazard function at time t_i can be estimated, conditional on β_p and $\epsilon_1, \dots, \epsilon_n$, by

$$\hat{h}_0(t_i) = \frac{\delta_i}{r_i} \quad (4.4)$$

where $r_i = \sum_{l=i}^n \alpha_l$, $\alpha_l = \exp(\beta_p(x_l + \epsilon_l))$, and β_p is the regression coefficient.

- (ii) The profiled Poisson likelihood function for the robust PH model becomes

$$L(\beta_p, \epsilon) = \prod_{i=1}^n \left[\frac{\alpha_i}{r_i} \right]^{\delta_i} \quad (4.5)$$

where $\epsilon^T = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ is a vector that represents the amount of imprecision associated with the observed covariate values x_1, x_2, \dots, x_n .

- (iii) The profiled Poisson likelihood function for the robust PH model in Equation (4.5) is equivalent to the partial likelihood for the robust PH model.

Proof.

- (i) (ii) These results can be verified by following the exact same process described in Section 2.5.1 by assuming fixed β and ϵ and profiling out the baseline hazard function which then can be substituted in Equation (4.3) to obtain the profile likelihood function for the robust PH model.
- (iii) Consider $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ distinct ordered failure times where $k \leq n$. As discussed in Section 2.4.2, given that an individual from the risk set R_i has experienced an event at time t_i , the probability that the i^{th} individual with the imprecise covariate value $\tilde{x}_i = x_i + \epsilon_i$ will experience the event at time t_i is

$$L_i(\beta_p, \epsilon_i, \dots, \epsilon_n) = \frac{h_i(t_i)}{\sum_{l \in R_i} h_l(t_i)} \quad (4.6)$$

By substituting the robust PH hazard function from Equation (4.1) and canceling out the baseline hazard, $h_0(t)$, we obtain

$$L_i(\beta_p, \epsilon_i, \dots, \epsilon_n) = \frac{\alpha_i}{\sum_{l \in R_i} \alpha_l} \quad (4.7)$$

Hence, the partial likelihood function for the robust PH model in Equation (4.1) becomes

$$L_p(\beta_p, \boldsymbol{\epsilon}) = \prod_{i=1}^k \frac{\alpha_i}{\sum_{l \in R_i} \alpha_l} \quad (4.8)$$

which is equivalent to the profiled Poisson likelihood function for the robust PH model in Equation (4.5). \square

A constrained optimization technique will be required to maximize the profile likelihood function since the imprecision terms, ϵ_i , are assumed to be restricted by $|\epsilon_i| \leq \epsilon^*$. In the examples that follow, the likelihood function in Equation (4.8) will be maximized using the limited memory quasi-Newton optimization algorithm for bound constrained optimization, L-BFGS-B, to estimate these imprecision terms along with the regression coefficient. The L-BFGS-B optimization was implemented in R version 4.2.2 using the *optim* function applied to the log-likelihood [18, 73].

It should be noted that we assume the absence of ties. Additionally, the covariate values will be re-centered during the optimization process such that $\tilde{x}_i = x_i + \epsilon_i - \mu$ where μ is the mean of $\boldsymbol{x} + \boldsymbol{\epsilon}$, following the common practice of dealing with continuous covariates in the PH model. The impact of the re-centering process disappears as each α_i is multiplied by $\exp(-\beta_p \mu)$, resulting in canceling out their effect according to Equation (4.5). A similar approach can be found in the "survfit" function of the survival package where centering covariate values prevents overflow in the exponential function argument and enhances numerical stability without adversely impacting the estimated parameters, according to Therneau [82, 84].

Theorem 4.3.2 Based on maximizing the likelihood function in Equation (4.5), the imprecision term for the j th observation can be estimated as follows when $\hat{\beta} > 0$:

$$\hat{\epsilon}_j = \begin{cases} \text{Not determined} & ; \text{ for right-censored observations before the first event} \\ \epsilon^* & ; \text{ for the first observed event} \\ -\epsilon^* & ; \text{ for } j = n \text{ and right-censored observations after the first event} \\ \in [-\epsilon^*, \epsilon^*] & ; \text{ for other events} \end{cases}$$

Conversely, ϵ_j can be estimated when the regression estimate is negative as follows

$$\hat{\epsilon}_j = \begin{cases} \text{Not determined} & ; \text{ for right-censored observations before the first event} \\ -\epsilon^* & ; \text{ for the first observed event} \\ \epsilon^* & ; \text{ for } j = n \text{ and right-censored observations after the first event} \\ \in [-\epsilon^*, \epsilon^*] & ; \text{ for other events} \end{cases}$$

Proof.

The following proves the case for positive regression estimates; similar steps can be followed to prove the statement for negative regression estimates. The log-likelihood, ℓ , associated with Equation(4.5) can be expressed as follows

$$\ell(\beta_p, \epsilon) = \sum_{i=1}^n \delta_i [\ln \alpha_i - \ln r_i] \quad (4.9)$$

where the summation is effectively calculated over only failure times. Based upon a particular ϵ_j , it appears that ϵ_j only affects ℓ directly through α_i when $i = j$ and indirectly through r_i when $i \leq j$. Accordingly, the derivative of ℓ with respect to ϵ_j is given by

$$\frac{\partial \ell}{\partial \epsilon_j} = \frac{\partial \ell}{\partial \alpha_j} \frac{\partial \alpha_j}{\partial \epsilon_j}$$

and observe that

$$\frac{\partial \alpha_j}{\partial \epsilon_j} = \beta_p \alpha_j$$

since α_j is always positive, then $\frac{\partial \alpha_j}{\partial \epsilon_j}$ is positive when $\beta_p > 0$ and negative when $\beta_p < 0$. Furthermore,

$$\frac{\partial \ell}{\partial \alpha_j} = \frac{\delta_j}{\alpha_j} - \sum_{i \leq j} \frac{\delta_i}{r_i} \quad (4.10)$$

To maximize the log-likelihood function, the value of ϵ_j is chosen such that the first partial derivative with respect to ϵ_j is positive. This leads to the following consequences:

1. For right-censored observations before the first event, those values of ϵ_j do not affect the log-likelihood and their estimates are therefore not determined.

2. For the first failure time, subscribed by (1) , $\frac{\partial \ell}{\partial \alpha_{(1)}} = \frac{1}{\alpha_{(1)}} - \frac{1}{r_{(1)}}$, which is positive since $r_{(1)} > \alpha_{(1)}$. Hence, $\frac{\partial \ell}{\partial \epsilon_{(1)}} > 0$ only if $\hat{\epsilon}_{(1)} = \epsilon^*$ when $\hat{\beta}_p > 0$ or $\hat{\epsilon}_{(1)} = -\epsilon^*$ when $\hat{\beta}_p < 0$.
3. For right-censored observations after the first event, we have $\frac{\partial \ell}{\partial \alpha_j} = -\sum_{i \leq j} \frac{\delta_i}{r_i} < 0$. Therefore, $\frac{\partial \ell}{\partial \epsilon_j}$ can be positive when $\hat{\epsilon}_j = \epsilon^*$ and $\hat{\beta}_p < 0$ or when $\hat{\epsilon}_j = -\epsilon^*$ and $\hat{\beta}_p > 0$.
4. For the last observation if $\delta_n = 1$, then

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha_n} &= \frac{1}{\alpha_n} - \sum_{i \leq n} \frac{\delta_j}{r_i} \\ &= \frac{1}{\alpha_n} - \sum_{i < n} \frac{\delta_j}{r_i} - \frac{1}{\alpha_n} \\ &= -\sum_{i < n} \frac{\delta_j}{r_i} \\ &< 0 \end{aligned}$$

Consequently, $\frac{\partial \ell}{\partial \epsilon_n}$ is positive only if $\hat{\epsilon}_n = \epsilon^*$ when $\hat{\beta}_p > 0$ or $\hat{\epsilon}_n = -\epsilon^*$ when $\hat{\beta}_p < 0$. Additionally, the third consequence will be followed if $\delta_n = 0$. \square

In light of Theorem 4.3.2, it is evident based on Equation (4.10) that for relatively small α_j and a small j (early time) where there are not many $i \leq j$, then $\partial \ell / \partial \alpha_j$ is likely to be positive. This leads to imprecision terms that match the sign of the regression parameter for early event observations. Conversely, $\partial \ell / \partial \alpha_j$ is likely to be negative for relatively large α_j and j is large (late time) resulting in imprecision terms of the opposite sign to the regression parameter for late event observations. Additionally, it should be noted that $\hat{\epsilon}_i$ values determined by Theorem 4.3.2 will be substituted into the likelihood function. Therefore, the optimization will be performed only over the regression parameter and those $\epsilon_i \in [-\epsilon^*, \epsilon^*]$ for which Theorem 4.3.2 failed to provide clear results. In the subsequent example, we demonstrate the implementation of the robust PH model using simulated data.

ϵ^*	$n = 30$		$n = 60$		$n = 200$	
	$\hat{\beta}_p$	ℓ	$\hat{\beta}_p$	ℓ	$\hat{\beta}_p$	ℓ
0	-0.5252	-51.285	-0.5212	-131.58	-0.5554	-636.33
0.1	-0.5644	-50.252	-0.5520	-129.48	-0.5898	-628.75
0.5	-0.7531	-45.633	-0.6937	-120.49	-0.7510	-596.25
1	-1.1684	-38.811	-0.9625	-107.25	-1.0463	-550.83
1.2	-1.4369	-35.816	-1.1171	-101.28	-1.2044	-531.23

Table 4.1: Estimates of β and values of $\ell(\hat{\beta}_p, \hat{\epsilon})$ for the RP model applied to the datasets in Example 4.3.1 having $\beta = -0.5$. A range of values of ϵ^* is considered: 0, 0.1, 0.5, 1 and 1.2.

Example 4.3.1 This example illustrates the behaviour of the robust PH model and the effect of increasing the level of imprecision on the estimated regression coefficient and imprecision terms, the maximum log-likelihood value, and the estimated baseline hazard function. Three simulated PH datasets are considered with $N = 30, 60$, and 200, each with baseline survival times following the Weibull distribution characterized by $\rho = 1.2$ and $\lambda = 2$. The covariate X was generated using a normal distribution with a mean of 10 and a variance of 9. Further, 20% of the observations are right-censored in accordance with Section 2.6.3, and the regression coefficient β was set to -0.5 . The Robust PH model was then fitted to each dataset, using four different levels of imprecision, namely $\epsilon^* = 0.1, 0.5, 1$, and 1.2. In order to illustrate some characteristics of the model, we additionally generate another dataset with $n = 30$ using the same settings, but assuming a positive regression coefficient $\beta = 0.5$.

For the data set with $\beta = -0.5$, Table 4.1 shows the estimated regression coefficients and the log-likelihood values in the absence of imprecision, i.e. the standard PH model where $\epsilon^* = 0$, and in the presence of imprecision using the robust PH model. In the presence of increasing imprecision, it is observed that the log-likelihood values increase reflecting the fact that the log-likelihood function is maximized over a wide range of possible parameter values and the estimated regression parameters

Data			$\epsilon^* = 0.1 \quad 0.5 \quad 1 \quad 1.2$				
t	$status$	x	$\hat{\beta}$	-0.56	-0.75	-1.17	-1.44
1.50	1	6.59	$\hat{\epsilon}_1$	-0.10	-0.50	-1.00	-1.20
1.71	1	4.10	$\hat{\epsilon}_2$	-0.10	-0.44	0.41	0.80
2.01	1	6.20	$\hat{\epsilon}_3$	-0.10	-0.50	-1.00	-0.73
3.14	1	12.51	$\hat{\epsilon}_4$	-0.10	-0.50	-1.00	-1.20
3.77	1	4.94	$\hat{\epsilon}_5$	0.10	0.50	1.00	1.20
4.40	1	7.81	$\hat{\epsilon}_6$	-0.10	-0.50	-1.00	-1.08
4.51	1	7.94	$\hat{\epsilon}_7$	-0.10	-0.50	-1.00	-0.83
7.13	1	10.46	$\hat{\epsilon}_8$	-0.10	-0.50	-1.00	-1.20
8.51	1	8.32	$\hat{\epsilon}_9$	-0.10	-0.50	-0.74	-0.59
8.97	1	6.92	$\hat{\epsilon}_{10}$	0.10	0.50	0.92	1.05
11.36	1	10.39	$\hat{\epsilon}_{11}$	-0.10	-0.50	-1.00	-1.20
12.70	1	6.80	$\hat{\epsilon}_{12}$	0.10	0.50	1.00	1.20
12.94	1	9.35	$\hat{\epsilon}_{13}$	-0.10	-0.50	-0.77	-0.67
13.21	0	15.36	$\hat{\epsilon}_{14}$	0.10	0.50	1.00	1.20
14.12	1	8.33	$\hat{\epsilon}_{15}$	0.10	0.39	0.52	0.60
17.59	1	11.38	$\hat{\epsilon}_{16}$	-0.10	-0.50	-1.00	-1.20
17.60	1	8.66	$\hat{\epsilon}_{17}$	0.10	0.50	0.67	0.73
24.45	1	11.20	$\hat{\epsilon}_{18}$	-0.10	-0.50	-1.00	-1.20
26.67	0	8.12	$\hat{\epsilon}_{18}$	0.10	0.50	1.00	1.20
29.89	1	11.08	$\hat{\epsilon}_{20}$	-0.10	-0.50	-1.00	-1.20
38.34	0	9.31	$\hat{\epsilon}_{21}$	0.10	0.50	1.00	1.20
38.63	1	8.58	$\hat{\epsilon}_{22}$	0.10	0.50	1.00	1.20
48.98	1	10.21	$\hat{\epsilon}_{23}$	0.10	0.50	0.55	0.60
58.18	1	13.67	$\hat{\epsilon}_{24}$	-0.10	-0.50	-1.00	-1.20
65.54	0	15.15	$\hat{\epsilon}_{25}$	0.10	0.50	1.00	1.20
65.65	1	11.49	$\hat{\epsilon}_{26}$	0.10	0.20	0.23	0.27
71.61	0	13.76	$\hat{\epsilon}_{27}$	0.10	0.50	1.00	1.20
79.96	0	12.10	$\hat{\epsilon}_{28}$	0.10	0.50	1.00	1.20
219.82	1	10.33	$\hat{\epsilon}_{29}$	0.10	0.50	1.00	1.20
244.93	1	14.68	$\hat{\epsilon}_{30}$	0.10	0.50	1.00	1.20

Table 4.2: Estimates of $\beta, \epsilon_1, \dots, \epsilon_{30}$ for the RP model applied to the dataset in Example 4.3.1 with $n = 30$ having $\beta = -0.5$. A range of values of ϵ^* is considered: 0, 0.1, 0.5, 1 and 1.2.

Data			$\epsilon^* = 0.1$	0.5	1	1.2	
t	$status$	x	$\hat{\beta}$	0.54	0.71	1.11	1.38
0.009	1	12.51	$\hat{\epsilon}_1$	0.10	0.50	1.00	1.20
0.066	1	13.67	$\hat{\epsilon}_2$	0.10	0.50	1.00	1.20
0.085	1	15.15	$\hat{\epsilon}_3$	0.10	0.43	-0.37	-0.54
0.117	1	10.46	$\hat{\epsilon}_4$	0.10	0.50	1.00	1.20
0.120	1	14.68	$\hat{\epsilon}_5$	-0.10	-0.19	-0.70	-0.81
0.132	1	15.36	$\hat{\epsilon}_6$	-0.10	-0.50	-1.00	-1.20
0.134	1	11.38	$\hat{\epsilon}_7$	0.10	0.50	1.00	1.20
0.198	1	10.39	$\hat{\epsilon}_8$	0.10	0.50	1.00	1.20
0.216	1	11.20	$\hat{\epsilon}_9$	0.10	0.50	1.00	1.16
0.288	1	13.76	$\hat{\epsilon}_{10}$	-0.10	-0.50	-1.00	-1.20
0.292	1	11.08	$\hat{\epsilon}_{11}$	0.10	0.50	0.61	0.57
0.376	1	12.10	$\hat{\epsilon}_{12}$	-0.10	-0.50	-0.85	-0.89
0.455	1	11.49	$\hat{\epsilon}_{13}$	-0.10	-0.50	-0.66	-0.72
0.536	1	9.35	$\hat{\epsilon}_{14}$	0.10	0.50	1.00	0.99
0.604	1	7.94	$\hat{\epsilon}_{15}$	0.10	0.50	1.00	1.20
0.619	1	6.59	$\hat{\epsilon}_{16}$	0.10	0.50	1.00	1.20
0.654	1	7.81	$\hat{\epsilon}_{17}$	0.10	0.50	1.00	1.20
0.726	0	8.12	$\hat{\epsilon}_{18}$	-0.10	-0.50	-1.00	-1.20
0.830	1	8.32	$\hat{\epsilon}_{18}$	0.10	0.50	0.90	0.84
0.987	1	10.21	$\hat{\epsilon}_{20}$	-0.10	-0.50	-1.00	-1.20
1.074	0	8.33	$\hat{\epsilon}_{21}$	-0.10	-0.50	-1.00	-1.20
1.134	0	9.31	$\hat{\epsilon}_{22}$	-0.10	-0.50	-1.00	-1.20
1.142	1	6.20	$\hat{\epsilon}_{23}$	0.10	0.50	1.00	1.20
1.289	1	8.66	$\hat{\epsilon}_{24}$	-0.10	-0.23	-0.14	-0.13
2.802	1	6.92	$\hat{\epsilon}_{25}$	0.10	0.50	1.00	1.20
3.028	1	8.58	$\hat{\epsilon}_{26}$	-0.10	-0.50	-0.47	-0.40
3.895	0	10.33	$\hat{\epsilon}_{27}$	-0.10	-0.50	-1.00	-1.20
3.973	0	4.10	$\hat{\epsilon}_{28}$	-0.10	-0.50	-1.00	-1.20
4.407	1	6.80	$\hat{\epsilon}_{29}$	-0.10	-0.50	-0.82	-0.85
5.310	0	4.94	$\hat{\epsilon}_{30}$	-0.10	-0.50	-1.00	-1.20

Table 4.3: Estimates of $\beta, \epsilon_1, \dots, \epsilon_{30}$ for the RP model applied to the data set in Example 4.3.1 with $n = 30$ having $\beta = 0.5$. A range of values of ϵ^* is considered: 0, 0.1, 0.5, 1 and 1.2.

diverge from zero. The same pattern was observed for the data set with $\beta = 0.5$, as reported in Table 4.3. For negative β , increasing ϵ^* results in a decrease in the regression coefficient while for positive β it results in an increase. Note that the values of $\hat{\epsilon}_i$ determined by Theorem 4.3.2 will be substituted in the likelihood, and the optimization will only be over the regression parameter and $\epsilon_i \in [-\epsilon^*, \epsilon^*]$ in which Theorem 4.3.2 was not able to clearly figure them out.

In comparison to the GPH/IPH models, where increasing the level of imprecision leads to shrinkage in the estimates of β , increasing the imprecision level in the robust PH model results in an exaggeration of the regression estimate. This occurs because the model tries to account for the additional variability in the covariate values. Therefore, it is crucial to be aware that $\hat{\beta}$ might overstate the true effect of the covariate due to the imprecision introduced in the covariate values. Researchers should be careful not to overinterpret the size of $\hat{\beta}$, particularly when ϵ^* is large. It would be prudent to emphasize that the observed relationship could be inflated by the noise introduced in the covariate.

Similarly, when applying the GPH/IPH models, it is advisable to consider that the effect of the covariate may be underestimated due to the shrinkage effect in the estimates in response to increased imprecision level. Hence, users should be cautious about concluding that a small $\hat{\beta}$ indicates a weak effect of the covariate; instead, it might reflect the imprecision in the model. In both models, the uncertainty increases in response to the increased imprecision level, but the direction of bias in the estimates differs. This introduces complexities in interpreting the regression coefficient, so practitioners should be cautious.

Furthermore, increasing the imprecision level substantially in the robust PH model may result in non-convergence, particularly for small sized PH data. Therefore, the occurrence of non-convergence during the optimization should be interpreted as a diagnostic indicator of an inappropriate selection of the imprecision level, ϵ^* , necessitating a reduction in the imprecision level to successfully achieve an MLE. Although this particular characteristic has not been reported for these data sets, it is worth noting that it will be further highlighted in Section 4.5.

Table 4.2 and 4.3 show the estimates of the regression coefficient and imprecision

terms for the data with $n = 30$. With respect to the estimated imprecision terms, it is observed that a significant proportion of these terms, particularly those associated with early-stage event occurrences, tend to adopt the imprecision limit that matches the sign of the estimated regression parameter, $\hat{\beta}$. This alignment is aimed at maximizing the log-likelihood function of the robust PH model, as illustrated in Theorem 4.3.2. Conversely, for individuals linked to later event and censoring times, the optimization of contributions is often achieved when the majority of imprecision terms assume the imprecision limit with a sign opposite to that of the estimated regression parameter, as evidenced in Table 4.2 and Table 4.3.

However, due to the nature of constrained multivariate optimization, these imprecision terms possess the flexibility to take any value within the predefined imprecision interval. This range includes the potential for a change in their sign with an increase in the imprecision level. The instances highlighted in Table 4.2 and 4.3 display this characteristic. The occurrence of this behaviour appears to be part of the constrained multivariate optimization. This conclusion is drawn by keeping the estimated imprecision terms and the estimated regression parameter constants as shown in Table 4.2, while allowing the value of ϵ_2 to vary within the imprecision interval. It was observed that the log-likelihood function attains its maximum only at the estimated $\hat{\epsilon}_2$ displayed in Table 4.2 and that using other values within the imprecision interval consistently leads to a reduction in the log-likelihood value. Figure 4.1 shows that upon implementing a minimal level of imprecision, the consequent estimations of imprecision terms are observed to conform with the predetermined imprecision limits, either $-\epsilon^*$ or ϵ^* . As the level of imprecision increases, the variability in the estimates of imprecision terms increases accordingly, thereby permitting these values to fluctuate within the defined imprecision limits.

Consider estimating the baseline hazard functions for the simulated survival data with $n = 30$. The estimates of $\hat{\beta}_p$ and the imprecision terms $\hat{\epsilon}$ in Table 4.2 can be substituted in Equation (4.4) to determine the baseline hazard function for the robust PH model. Accordingly, the baseline hazard function is given by

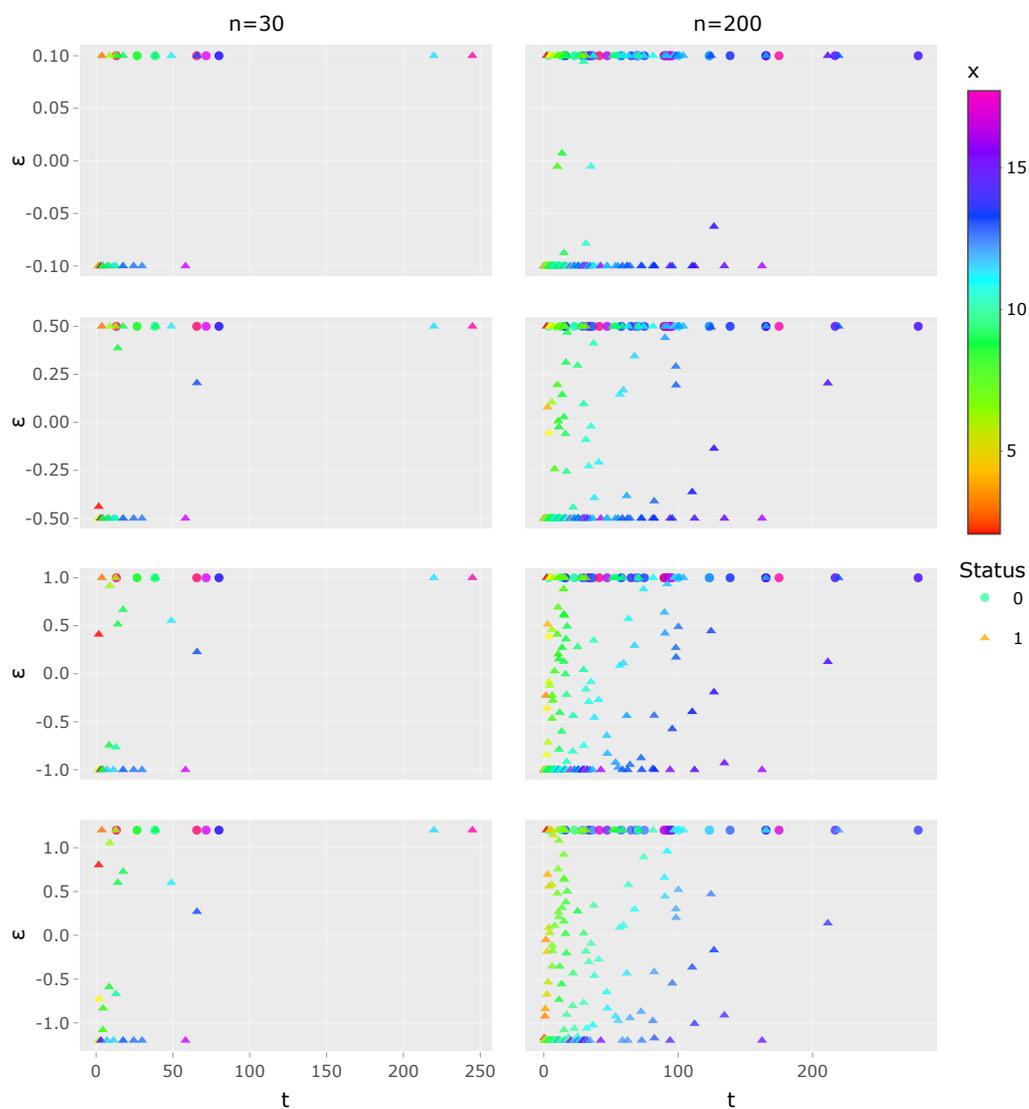


Figure 4.1: Estimates of imprecision terms ϵ_i for the RP model applied to the data set in Example 4.3.1 with $n = 30$ (left) and $n = 200$ (right) having $\beta = -0.5$. A range of values of ϵ^* is considered: 0.1, 0.5, 1 and 1.2 (top-bottom).

$$\begin{aligned}
\hat{h}_0(t_i) &= \frac{\delta_i}{r_i} \\
&= \frac{\delta_i}{\sum_{l=i}^n \alpha_l} \\
&= \frac{\delta_i}{\sum_{l=i}^n \exp\left(\hat{\beta}_p(x_l + \hat{\epsilon}_l)\right)}
\end{aligned} \tag{4.11}$$

This baseline hazard function is intended to be represented as a function of t_i to emphasize that it equals zero except at event times. The construction of the robust PH model utilizes the Poisson empirical likelihood approach, which yielded to an estimation of the baseline hazard function analogous to Breslow's estimate of the baseline hazard function as in Equation (4.11). Consequently, this approach implies that the baseline hazard function related to right-censored times is zero, while the baseline hazard function related to event times can vary between zero and infinity.

Tables A.5 and A.6 in Appendix A.3 present the baseline hazard function estimates derived from applying the robust PH model to the simulated datasets with a sample size $n = 30$ using both negative and positive regression coefficients. These datasets were examined employing different levels of imprecision $\epsilon^* = 0.1, 0.5, 1, 1.2$ along with estimated baseline hazard function results from the standard PH model. Per above, increasing the imprecision level shows that these baseline hazard functions exhibit monotonic behaviour; that is, increasing imprecision leads to a higher baseline hazard for negative estimates of the regression coefficient. This phenomenon occurs due to the fact that an increase in the imprecision level lowers the estimated value of the regression coefficient, as mentioned earlier, resulting in a reduced denominator in Equation (4.11). For a positive regression coefficient, the impact reverses as illustrated in the referenced tables. Figure 4.2 further highlights the relationship between the sign of the estimated regression coefficient and its influence on either augmenting or diminishing the baseline hazard function. This is demonstrated by comparing the baseline hazard estimates obtained from the robust PH model with an imprecision level of $\epsilon^* = 0.1$ and the standard PH model, where ϵ^* is assumed to

be zero.

◇

4.3.1 Imprecise hazard and survival functions

This section describes the estimation of the imprecise hazard and survival functions for individuals with particular covariate values using the robust PH model. A naive approach to determining the imprecise hazard and survival functions can be achieved by directly plugging in the imprecision limits, $-\epsilon^*$, ϵ^* , to Equation (4.2) in accordance with the estimates of the regression coefficient, the imprecision terms, and the baseline hazard function. When the regression coefficient is positive, the upper and lower hazard functions for individuals with $x = x_j$ are as follows:

$$\begin{aligned}\underline{h}_j(t; \epsilon^*) &= \hat{h}_{0p}(t; \epsilon^*) \exp\left(\hat{\beta}(x_j - \epsilon^*)\right) \\ \bar{h}_j(t; \epsilon^*) &= \hat{h}_{0p}(t; \epsilon^*) \exp\left(\hat{\beta}(x_j + \epsilon^*)\right)\end{aligned}\tag{4.12}$$

The lower and upper survival functions for the individual with a covariate value $x = x_j$ at time t given the level of imprecision ϵ^* can be derived as follows

$$\underline{S}_j(t; \epsilon^*) = \exp\left[-\sum_{l:t_l \leq t} \bar{h}_j(t_l; \epsilon^*)\right]\tag{4.13}$$

$$\bar{S}_j(t; \epsilon^*) = \exp\left[-\sum_{l:t_l \leq t} \underline{h}_j(t_l; \epsilon^*)\right]\tag{4.14}$$

The reverse holds true when the regression coefficient is negative, i.e., the lower hazard function is associated with the positive imprecision limit, ϵ^* , while the upper hazard function is associated with the negative imprecision limit, $-\epsilon^*$. The following example illustrates how to estimate the naive hazard and survival functions for individuals using the robust PH model.

Example 4.3.2 In this example, we explore the impact of increasing the level of imprecision on the imprecise hazard and survival functions for different individuals. Our particular implementation of the robust PH model is based on the simulated data set that includes 30 observations with the negative regression coefficient

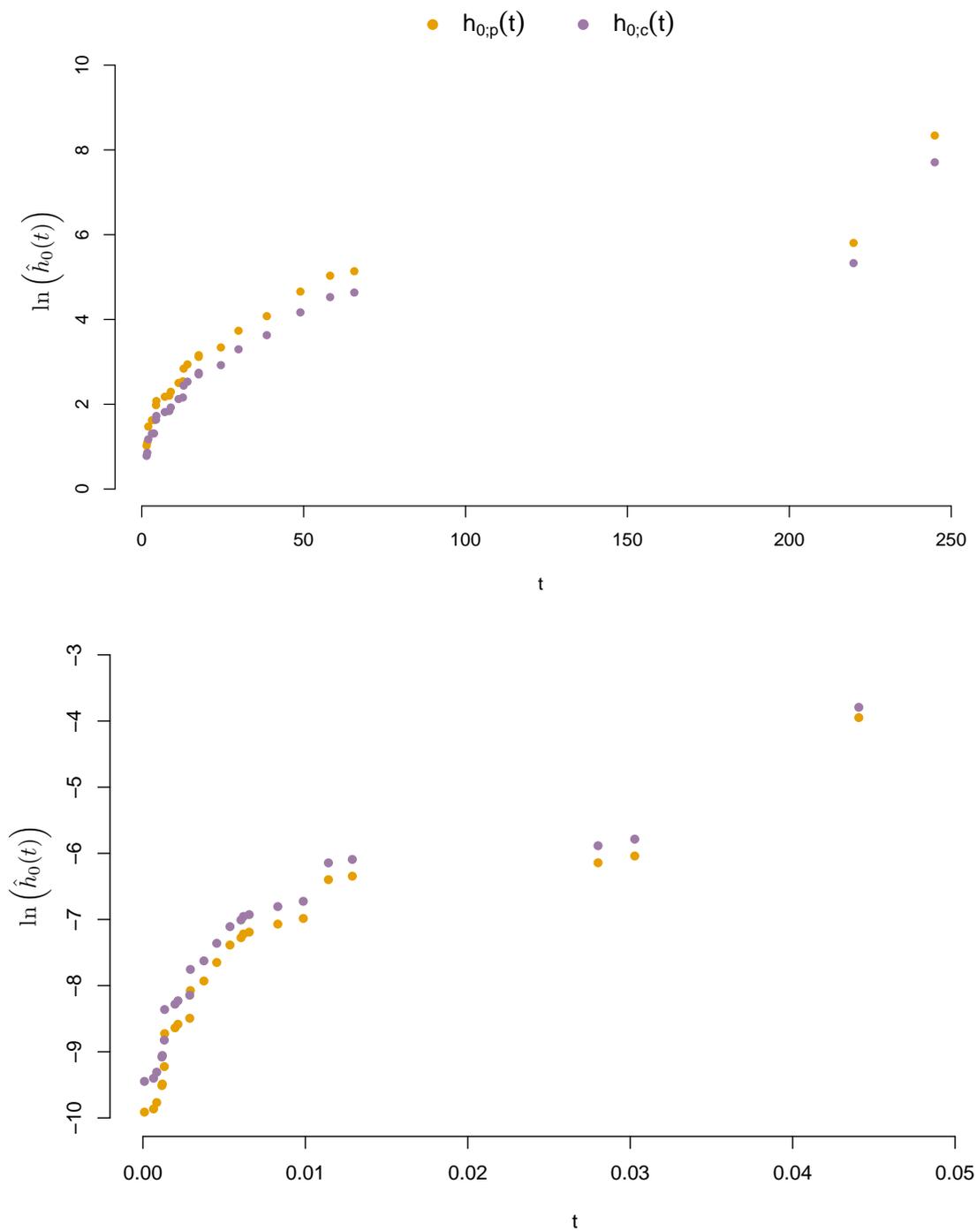


Figure 4.2: Estimates of baseline hazards for the RP model applied to the data set in Example 4.3.1 with $n = 30$ having $\beta = -0.5$ (top) and $\beta = 0.5$ (bottom) using $\epsilon^* = 0.1$, these estimates equal zero at non-event times.

$\hat{\beta} = -0.5644$, as also used in Example 4.3.1. The naive hazard and survival functions will be estimated for three different covariate values, $x = 5, 9, 15$, representing individuals with the minimum, mean, and maximum values of the covariate. Additionally, we consider two levels of imprecision $\epsilon^* = 0.5$, and 1. The resulting lower and upper survival functions as derived in Equations (4.13) and (4.14) together with the survival function corresponding to the standard PH model are presented in Tables 4.4, 4.5, and 4.6 for individuals with $x = 5, 9, 15$, respectively.

Similar to the PH model, the lower and upper survival functions obtained by the robust PH model decrease only at event times, while remain constant between successive event times due to the fact the the baseline hazard function equal zero at right-censored times. As a consequence of constructing the robust PH model based on the Poisson type of empirical likelihood, the survival function for the last observation does not reach zero although the last observation related to an event time. The survival values are therefore expressed as 0.00e+00 instead of 0, representing a number that converges to zero. The tables reveal inconsistent behavior of the imprecise survival estimates when increasing imprecision levels. Notably, the survival function estimates generated by the standard PH model may not necessarily reside within the interval delineated by the lower and upper survival estimates derived from the robust proportional hazards model across all values of x .

Given a negative estimate of the regression coefficient, Table 4.4 shows that for small values of x , the PH survival estimate tends to fall within the imprecise survival estimates primarily at earlier times. On the other hand, the PH survival estimate corresponding to higher values of x may lie between the lower and upper survival estimates mostly at later times, as evidenced in Table 4.6. The duration during which the PH survival estimates are enclosed within the corresponding imprecise survival estimates varies according to changes in x as in Table 4.5, when $x = 9$. Based on epsilon values of 0.5 and 1, Figure 4.3 illustrates the behaviour of the naive imprecise survival estimates in comparison to those derived from the standard PH model for individuals with covariate values of $x = 5, 9, 15$.

Indeed, as evidenced in Example 4.3.1 and other simulations, the baseline hazard estimates exhibit a pronounced monotonic effect with increasing imprecision levels.

time	status	$\hat{S}_{5;c}(t)$	$\epsilon^* = 0.5$		$\epsilon^* = 1$	
			$\hat{S}_{5;p}(t)$	$\bar{\hat{S}}_{5;p}(t)$	$\hat{S}_{5;p}(t)$	$\bar{\hat{S}}_{5;p}(t)$
1.50	1	8.54e-01	7.75e-01	8.87e-01	4.30e-01	9.22e-01
1.71	1	7.20e-01	5.88e-01	7.79e-01	1.63e-01	8.39e-01
2.01	1	5.70e-01	3.31e-01	5.94e-01	2.00e-02	6.85e-01
3.14	1	4.36e-01	1.56e-01	4.17e-01	2.74e-04	4.53e-01
3.77	1	3.33e-01	7.36e-02	2.93e-01	3.74e-06	2.99e-01
4.40	1	2.30e-01	2.23e-02	1.67e-01	1.62e-09	1.41e-01
4.51	1	1.53e-01	5.52e-03	8.64e-02	3.02e-14	4.94e-02
7.13	1	9.83e-02	1.06e-03	3.98e-02	1.59e-21	9.77e-03
8.51	1	6.23e-02	1.96e-04	1.80e-02	5.15e-29	1.85e-03
8.97	1	3.79e-02	2.75e-05	7.12e-03	3.34e-39	1.91e-04
11.36	1	2.07e-02	2.23e-06	2.18e-03	4.64e-53	8.75e-06
12.70	1	1.11e-02	1.62e-07	6.34e-04	8.76e-68	3.30e-07
12.94	1	4.83e-03	3.42e-09	1.03e-04	2.16e-92	1.38e-09
13.21	0	4.83e-03	3.42e-09	1.03e-04	2.16e-92	1.38e-09
14.12	1	1.94e-03	3.71e-11	1.23e-05	4.93e-126	7.75e-13
17.59	1	6.59e-04	1.42e-13	8.91e-07	5.08e-172	2.79e-17
17.60	1	2.14e-04	4.14e-16	5.71e-08	5.31e-221	5.14e-22
24.45	1	5.59e-05	3.53e-19	2.05e-09	5.48e-284	4.20e-28
26.67	0	5.59e-05	3.53e-19	2.05e-09	5.48e-284	4.20e-28
29.89	1	7.91e-06	3.81e-24	9.38e-12	0.00e+00	1.57e-39
38.34	0	7.91e-06	3.81e-24	9.38e-12	0.00e+00	1.57e-39
38.63	1	5.17e-07	1.57e-31	3.13e-15	0.00e+00	1.99e-58
48.98	1	4.85e-09	1.46e-47	8.83e-23	0.00e+00	1.96e-114
58.18	1	6.02e-12	5.72e-72	2.84e-34	0.00e+00	8.36e-225
65.54	0	6.02e-12	5.72e-72	2.84e-34	0.00e+00	8.36e-225
65.65	1	3.46e-15	6.30e-99	5.74e-47	0.00e+00	0.00e+00
71.61	0	3.46e-15	6.30e-99	5.74e-47	0.00e+00	0.00e+00
79.96	0	3.46e-15	6.30e-99	5.74e-47	0.00e+00	0.00e+00
219.82	1	1.14e-21	3.25e-148	3.55e-70	0.00e+00	0.00e+00
244.93	1	1.30e-91	0.00e+00	0.00e+00	0.00e+00	0.00e+00

Table 4.4: Comparison of standard PH survival estimates and naive imprecise survival estimates for individuals with $x = 5$. The estimates were derived from the RP model applied to the data set in Example 4.3.2, with $n = 30$ and $\hat{\beta} = -0.5644$, using ϵ^* values of 0.5 and 1.

time	status	$\hat{S}_{9;c}(t)$	$\epsilon^* = 0.5$		$\epsilon^* = 1$	
			$\hat{S}_{9;p}(t)$	$\bar{S}_{9;p}(t)$	$\hat{S}_{9;p}(t)$	$\bar{S}_{9;p}(t)$
1.50	1	9.81e-01	9.88e-01	9.94e-01	9.92e-01	9.99e-01
1.71	1	9.61e-01	9.74e-01	9.88e-01	9.83e-01	9.98e-01
2.01	1	9.33e-01	9.47e-01	9.75e-01	9.64e-01	9.96e-01
3.14	1	9.03e-01	9.13e-01	9.58e-01	9.26e-01	9.93e-01
3.77	1	8.74e-01	8.80e-01	9.41e-01	8.90e-01	9.89e-01
4.40	1	8.35e-01	8.29e-01	9.16e-01	8.28e-01	9.82e-01
4.51	1	7.95e-01	7.74e-01	8.87e-01	7.48e-01	9.72e-01
7.13	1	7.53e-01	7.14e-01	8.53e-01	6.39e-01	9.58e-01
8.51	1	7.12e-01	6.57e-01	8.21e-01	5.44e-01	9.43e-01
8.97	1	6.70e-01	5.97e-01	7.84e-01	4.37e-01	9.23e-01
11.36	1	6.22e-01	5.27e-01	7.40e-01	3.24e-01	8.97e-01
12.70	1	5.76e-01	4.64e-01	6.96e-01	2.36e-01	8.70e-01
12.94	1	5.21e-01	3.83e-01	6.37e-01	1.39e-01	8.27e-01
13.21	0	5.21e-01	3.83e-01	6.37e-01	1.39e-01	8.27e-01
14.12	1	4.66e-01	3.07e-01	5.73e-01	6.75e-02	7.71e-01
17.59	1	4.08e-01	2.33e-01	5.04e-01	2.51e-02	7.00e-01
17.60	1	3.56e-01	1.75e-01	4.40e-01	8.76e-03	6.33e-01
24.45	1	3.02e-01	1.24e-01	3.74e-01	2.26e-03	5.55e-01
26.67	0	3.02e-01	1.24e-01	3.74e-01	2.26e-03	5.55e-01
29.89	1	2.38e-01	7.06e-02	2.87e-01	1.78e-04	4.34e-01
38.34	0	2.38e-01	7.06e-02	2.87e-01	1.78e-04	4.34e-01
38.63	1	1.70e-01	3.06e-02	1.94e-01	2.65e-06	2.89e-01
48.98	1	9.61e-02	4.98e-03	8.24e-02	1.02e-11	8.67e-02
58.18	1	4.24e-02	3.14e-04	2.24e-02	2.21e-22	8.07e-03
65.54	0	4.24e-02	3.14e-04	2.24e-02	2.21e-22	8.07e-03
65.65	1	1.70e-02	1.49e-05	5.33e-03	2.55e-34	5.67e-04
71.61	0	1.70e-02	1.49e-05	5.33e-03	2.55e-34	5.67e-04
79.96	0	1.70e-02	1.49e-05	5.33e-03	2.55e-34	5.67e-04
219.82	1	2.74e-03	5.61e-08	3.85e-04	1.71e-55	5.09e-06
244.93	1	7.56e-12	2.87e-74	2.35e-35	0.00e+00	0.00e+00

Table 4.5: Comparison of standard PH survival estimates and naive imprecise survival estimates for individuals with $x = 9$. The estimates were derived from the RP model applied to the data set in Example 4.3.2, with $n = 30$ and $\hat{\beta} = -0.5644$, using ϵ^* values of 0.5 and 1.

time	status	$\hat{S}_{15;c}(t)$	$\epsilon^* = 0.5$		$\epsilon^* = 1$	
			$\underline{\hat{S}}_{15;p}(t)$	$\overline{\hat{S}}_{15;p}(t)$	$\underline{\hat{S}}_{15;p}(t)$	$\overline{\hat{S}}_{15;p}(t)$
1.50	1	9.99e-01	1.00e+00	1.00e+00	1.00e+00	1.00e+00
1.71	1	9.98e-01	1.00e+00	1.00e+00	1.00e+00	1.00e+00
2.01	1	9.97e-01	9.99e-01	1.00e+00	1.00e+00	1.00e+00
3.14	1	9.96e-01	9.99e-01	1.00e+00	1.00e+00	1.00e+00
3.77	1	9.94e-01	9.99e-01	9.99e-01	1.00e+00	1.00e+00
4.40	1	9.92e-01	9.98e-01	9.99e-01	1.00e+00	1.00e+00
4.51	1	9.90e-01	9.97e-01	9.99e-01	1.00e+00	1.00e+00
7.13	1	9.88e-01	9.96e-01	9.98e-01	1.00e+00	1.00e+00
8.51	1	9.86e-01	9.95e-01	9.98e-01	9.99e-01	1.00e+00
8.97	1	9.83e-01	9.94e-01	9.97e-01	9.99e-01	1.00e+00
11.36	1	9.80e-01	9.93e-01	9.97e-01	9.99e-01	1.00e+00
12.70	1	9.77e-01	9.92e-01	9.96e-01	9.99e-01	1.00e+00
12.94	1	9.72e-01	9.90e-01	9.95e-01	9.98e-01	1.00e+00
13.21	0	9.72e-01	9.90e-01	9.95e-01	9.98e-01	1.00e+00
14.12	1	9.68e-01	9.87e-01	9.94e-01	9.98e-01	1.00e+00
17.59	1	9.62e-01	9.84e-01	9.93e-01	9.97e-01	1.00e+00
17.60	1	9.57e-01	9.81e-01	9.91e-01	9.96e-01	1.00e+00
24.45	1	9.50e-01	9.77e-01	9.89e-01	9.95e-01	9.99e-01
26.67	0	9.50e-01	9.77e-01	9.89e-01	9.95e-01	9.99e-01
29.89	1	9.40e-01	9.72e-01	9.86e-01	9.92e-01	9.99e-01
38.34	0	9.40e-01	9.72e-01	9.86e-01	9.92e-01	9.99e-01
38.63	1	9.27e-01	9.63e-01	9.82e-01	9.88e-01	9.99e-01
48.98	1	9.05e-01	9.44e-01	9.73e-01	9.77e-01	9.98e-01
58.18	1	8.73e-01	9.16e-01	9.59e-01	9.56e-01	9.96e-01
65.54	0	8.73e-01	9.16e-01	9.59e-01	9.56e-01	9.96e-01
65.65	1	8.40e-01	8.86e-01	9.45e-01	9.33e-01	9.93e-01
71.61	0	8.40e-01	8.86e-01	9.45e-01	9.33e-01	9.93e-01
79.96	0	8.40e-01	8.86e-01	9.45e-01	9.33e-01	9.93e-01
219.82	1	7.77e-01	8.34e-01	9.18e-01	8.92e-01	9.89e-01
244.93	1	3.34e-01	1.58e-01	4.19e-01	7.46e-04	4.99e-01

Table 4.6: Comparison of standard PH survival estimates and naive imprecise survival estimates for individuals with $x = 15$. The estimates were derived from the RP model applied to the data set in Example 4.3.2, with $n = 30$ and $\hat{\beta} = -0.5644$, using ϵ^* values of 0.5 and 1.

Nevertheless, it becomes apparent that this monotonic trend does not uniformly apply to hazard estimates for specific individuals due to variations in the estimates of the regression coefficient and the covariate effect, represented by $\hat{\alpha}_i$, in response to increased imprecision levels. Consequently, this variations gives rise to the observed inconsistencies in behavior of the imprecise survival estimates.

Table 4.7 provides insight into two additional aspects of naive survival estimates by analyzing the disparities between upper and lower survival estimates for individuals denoted as $\overline{\hat{S}}_{x;p}(t) - \underline{\hat{S}}_{x;p}(t)$. Interestingly, Table 4.7 highlights a correlation between the difference in the imprecise survival estimates and the covariate effect. In particular, there is a noticeable tendency for the difference in imprecise survival estimates to decrease over time for maximal covariate effects, related to $x = 5$. However, this trend gradually diminishes and eventually reverses as the covariate effect decreases and reaches its minimal value, i.e. when $x = 15$. Additionally, the table demonstrates another inconsistency in naive survival estimates, which is that increasing the imprecision level does not necessarily result in a wider difference between upper and lower survivals. This phenomenon becomes apparent at later time points for individuals with $x = 5$, namely from time $t = 4.40$, and at earlier time points for individuals with $x = 15$.

◇

Envelope-type of hazard and survival functions

The naive estimation method aims to derive imprecise hazard and survival estimates for an individual based on a specified imprecision level, ϵ^* . However, these imprecise estimates often fail to encompass the hazard and survival estimates derived from lower imprecision levels, including $\epsilon^* = 0$. The inconsistency behavior observed in the naive-type survival estimates presented in Example 4.3.2 is attributed to the failure of monotonic properties in baseline hazard functions to consistently manifest in individual hazard functions. Therefore, an alternative approach known as the envelope method was proposed to overcome this limitation.

The envelope method introduces a key restriction that imprecision terms are allowed to influence only the baseline hazard estimates. This choice is motivated

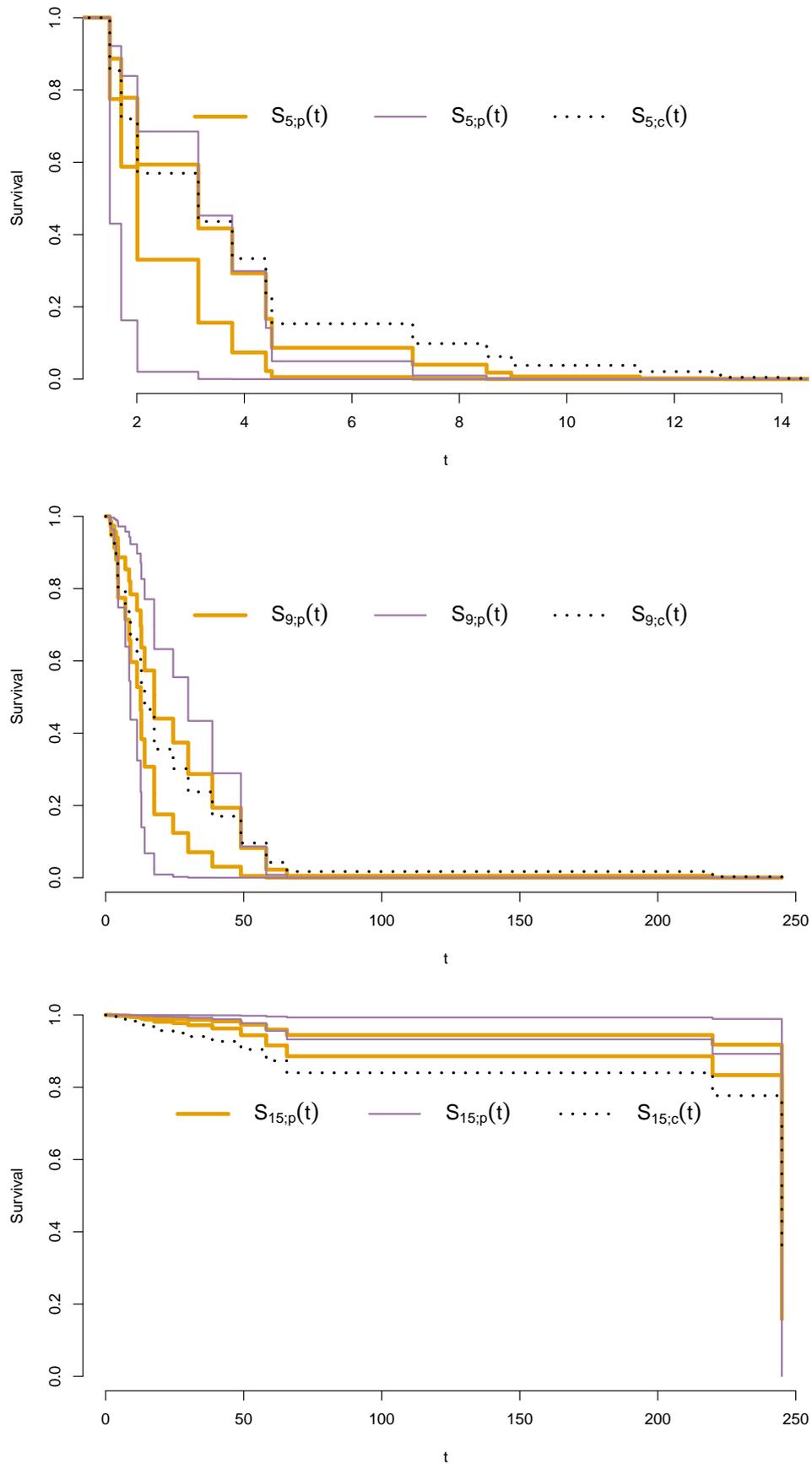


Figure 4.3: Comparison of standard PH survival estimates and naive imprecise survival estimates for $x = 5, 9, 15$ (top-bottom). These estimates were derived from the RP model applied to the data set in Example 4.3.2, using ϵ^* values of 0.5 (amber), and 1 (indigo).

time	status	$\overline{\hat{S}}_{5;p}(t) - \underline{\hat{S}}_{5;p}(t)$		$\overline{\hat{S}}_{9;p}(t) - \underline{\hat{S}}_{9;p}(t)$		$\overline{\hat{S}}_{15;p}(t) - \underline{\hat{S}}_{15;p}(t)$	
		$\epsilon^* = 0.5$	1	0.5	1	0.5	1
1.50	1	1.12e-01	4.92e-01	6.57e-03	7.09e-03	7.23e-05	6.43e-06
1.71	1	1.91e-01	6.76e-01	1.36e-02	1.52e-02	1.51e-04	1.38e-05
2.01	1	2.63e-01	6.65e-01	2.77e-02	3.23e-02	3.14e-04	2.98e-05
3.14	1	2.61e-01	4.52e-01	4.52e-02	6.64e-02	5.26e-04	6.25e-05
3.77	1	2.19e-01	2.99e-01	6.18e-02	9.90e-02	7.39e-04	9.52e-05
4.40	1	1.44e-01	1.41e-01	8.63e-02	1.54e-01	1.08e-03	1.54e-04
4.51	1	8.09e-02	4.94e-02	1.12e-01	2.25e-01	1.47e-03	2.37e-04
7.13	1	3.88e-02	9.77e-03	1.39e-01	3.18e-01	1.94e-03	3.65e-04
8.51	1	1.78e-02	1.85e-03	1.63e-01	3.99e-01	2.41e-03	4.96e-04
8.97	1	7.09e-03	1.91e-04	1.87e-01	4.86e-01	2.97e-03	6.75e-04
11.36	1	2.18e-03	8.75e-06	2.12e-01	5.72e-01	3.67e-03	9.17e-04
12.70	1	6.34e-04	3.30e-07	2.33e-01	6.33e-01	4.41e-03	1.18e-03
12.94	1	1.03e-04	1.38e-09	2.53e-01	6.87e-01	5.49e-03	1.61e-03
13.21	0	1.03e-04	1.38e-09	2.53e-01	6.87e-01	5.49e-03	1.61e-03
14.12	1	1.23e-05	7.75e-13	2.66e-01	7.03e-01	6.75e-03	2.19e-03
17.59	1	8.91e-07	2.79e-17	2.71e-01	6.75e-01	8.29e-03	3.00e-03
17.60	1	5.71e-08	5.14e-22	2.65e-01	6.24e-01	9.91e-03	3.85e-03
24.45	1	2.05e-09	4.20e-28	2.50e-01	5.53e-01	1.18e-02	4.95e-03
26.67	0	2.05e-09	4.20e-28	2.50e-01	5.53e-01	1.18e-02	4.95e-03
29.89	1	9.38e-12	1.57e-39	2.16e-01	4.34e-01	1.50e-02	7.01e-03
38.34	0	9.38e-12	1.57e-39	2.16e-01	4.34e-01	1.50e-02	7.01e-03
38.63	1	3.13e-15	1.99e-58	1.63e-01	2.89e-01	1.96e-02	1.04e-02
48.98	1	8.83e-23	1.96e-114	7.74e-02	8.67e-02	2.93e-02	2.04e-02
58.18	1	2.84e-34	8.36e-225	2.21e-02	8.07e-03	4.36e-02	3.97e-02
65.54	0	2.84e-34	8.36e-225	2.21e-02	8.07e-03	4.36e-02	3.97e-02
65.65	1	5.74e-47	0.00e+0	5.31e-03	5.67e-04	5.87e-02	6.07e-02
71.61	0	5.74e-47	0.00e+0	5.31e-03	5.67e-04	5.87e-02	6.07e-02
79.96	0	5.74e-47	0.00e+0	5.31e-03	5.67e-04	5.87e-02	6.07e-02
219.82	1	3.55e-70	0.00e+0	3.85e-04	5.09e-06	8.43e-02	9.67e-02
244.93	1	0.00e+0	0.00e+0	2.35e-35	0.00e+0	2.61e-01	4.98e-01

Table 4.7: Comparison of the width of naive survival estimates for individuals with $x = 5, 9, 15$ based on the RP model applied to the data set in Example 4.3.2, with $n = 30$ and $\hat{\beta} = -0.5644$, using ϵ^* values of 0.5 and 1.

by the observation that baseline hazard estimates exhibit a monotonic behavior as the level of imprecision increases. By focusing the impact of imprecision on the baseline hazard, the envelope method ensures that the hazard and survival estimates derived from the robust PH model at any lower imprecision level, including the case where $\epsilon^{**} = 0$, are contained within the range of estimates produced at the selected imprecision level, ϵ^{**} . This notation, ϵ^{**} , was specifically introduced to distinguish the envelope method from the naive approach.

To understand how the envelope method operates, consider the process of estimating the hazard function for an individual with a covariate value $x = x_j$ using the standard PH model. This estimation yields a regression estimate, $\hat{\beta}_c$, which is then applied to Breslow's estimator to derive the baseline hazard estimate. This estimated baseline hazard is subsequently used to calculate the individual-specific hazard function, $\hat{h}_{0p}(t) \exp(\hat{\beta}_c x_j)$, as detailed in Section 2.4.4. Extending this methodology, the robust PH model is applied to the same dataset to estimate the hazard function for the same individual. This results in a different estimate of the regression coefficient, $\hat{\beta}_p$, which is then used to determine the corresponding hazard function, $\hat{h}_{0p}(t) \exp(\hat{\beta}_p x_j)$.

Let $\mathfrak{H}(t|x = x_j; \epsilon^{**})$ represent the set of all possible hazard functions for an individual with covariate value x_j at time t , given the imprecision level ϵ^{**} . The lower and upper bounds of the hazard function for this individual can then be defined as:

$$\begin{aligned} \underline{h}_j(t; \epsilon^{**}) &= \min\{\mathfrak{H}(t|x = x_j; \epsilon^{**})\} \\ &= \min\{h_{0c}(t) \exp(\hat{\beta}_c x_j), h_{0p}(t; \epsilon^{**}) \exp(\hat{\beta}_p x_j)\} \end{aligned} \quad (4.15)$$

$$\begin{aligned} \bar{h}_j(t; \epsilon^{**}) &= \max\{\mathfrak{H}(t|x = x_j; \epsilon^{**})\} \\ &= \max\{h_{0c}(t) \exp(\hat{\beta}_c x_j), h_{0p}(t; \epsilon^{**}) \exp(\hat{\beta}_p x_j)\} \end{aligned} \quad (4.16)$$

These formulations highlight the core principle of the envelope method that the imprecision effect directly impacts only the baseline hazard estimates, while it indirectly influences the regression estimate $\hat{\beta}_p$. The lower and upper survival estimates for the individual with a covariate value $x = x_j$ at time t given the level of imprecision ϵ^{**} can be derived as follows

$$\underline{S}_j(t; \epsilon^{**}) = \exp \left[- \sum_{l:t_l \leq t} \bar{h}_j(t_l; \epsilon^{**}) \right] \quad (4.17)$$

$$\bar{S}_j(t; \epsilon^{**}) = \exp \left[- \sum_{l:t_l \leq t} \underline{h}_j(t_l; \epsilon^{**}) \right] \quad (4.18)$$

The envelope method offers a more reliable and encompassing approach to estimating hazard and survival functions under varying levels of imprecision. By confining the imprecision effects to the baseline hazard, it maintains the integrity of the estimates across different imprecision levels, providing a comprehensive view of the potential variability in the data. We demonstrate the envelope estimation of the imprecise hazard and survival functions for individuals in the example presented below.

Example 4.3.3 In this example, we explore the impact of increasing the level of imprecision on the envelope-type of imprecise hazard and survival functions for different individuals. This application of the envelope imprecise estimates in this example is based on the simulated data with $n = 30$ and $\hat{\beta} = -0.5644$, which was also used in Example 4.3.1 and 4.3.2. The envelope imprecise hazard and survival functions will be estimated for three different covariate values, $x = 5, 9, 15$ using two levels of imprecision $\epsilon^{**} = 0.5$, and 1. Tables 4.8, 4.9, and 4.10 present the envelope survival estimates resulting from Equations (4.17) and (4.18), along with standard PH survival estimates for individuals with $x = 5, 9, 15$, respectively.

The survival estimates based on the standard PH model are always between the envelope-type lower and upper survival estimates derived from the robust PH model for all values of x . Further, the tables present consistent results supporting the envelope approach such that survival estimates derived from lower levels of imprecision are integrated within survival estimates obtained from higher levels of imprecision. For small values of x , Table 4.8 shows that the PH survival estimates tend to fall at or near the upper limit of the envelope survival estimates obtained from the robust PH model. On the other hand, the PH survival estimates corresponding to higher values of x lie close or at the lower limit of the envelope survival estimates from the robust PH model, as shown in Table 4.10. As x varies between its minimum and

time	status	$\hat{S}_{5;c}(t)$	$\epsilon^{**} = 0.5$		$\epsilon^{**} = 1$	
			$\hat{\underline{S}}_{5;p}(t)$	$\hat{\bar{S}}_{5;p}(t)$	$\hat{\underline{S}}_{5;p}(t)$	$\hat{\bar{S}}_{5;p}(t)$
1.50	1	8.54e-01	8.39e-01	8.54e-01	7.69e-01	8.54e-01
1.71	1	7.20e-01	6.94e-01	7.20e-01	5.69e-01	7.20e-01
2.01	1	5.70e-01	4.68e-01	5.70e-01	2.97e-01	5.70e-01
3.14	1	4.36e-01	2.80e-01	4.36e-01	7.81e-02	4.36e-01
3.77	1	3.33e-01	1.67e-01	3.33e-01	2.05e-02	3.33e-01
4.40	1	2.30e-01	7.35e-02	2.30e-01	1.85e-03	2.30e-01
4.51	1	1.53e-01	2.82e-02	1.53e-01	6.25e-05	1.53e-01
7.13	1	9.83e-02	9.11e-03	9.83e-02	3.41e-07	9.83e-02
8.51	1	6.23e-02	2.85e-03	6.23e-02	1.60e-09	6.23e-02
8.97	1	3.79e-02	7.41e-04	3.79e-02	1.09e-12	3.79e-02
11.36	1	2.07e-02	1.32e-04	2.07e-02	5.33e-17	2.07e-02
12.70	1	1.11e-02	2.18e-05	1.11e-02	1.40e-21	1.11e-02
12.94	1	4.83e-03	1.55e-06	4.83e-03	3.12e-29	4.83e-03
13.21	0	4.83e-03	1.55e-06	4.83e-03	3.12e-29	4.83e-03
14.12	1	1.94e-03	6.95e-08	1.94e-03	1.08e-39	1.94e-03
17.59	1	6.59e-04	1.52e-09	6.59e-04	5.35e-54	6.59e-04
17.60	1	2.14e-04	2.77e-11	2.14e-04	3.12e-69	2.14e-04
24.45	1	5.59e-05	2.17e-13	5.59e-05	8.00e-89	5.59e-05
26.67	0	5.59e-05	2.17e-13	5.59e-05	8.00e-89	5.59e-05
29.89	1	7.91e-06	8.46e-17	7.91e-06	1.33e-125	7.91e-06
38.34	0	7.91e-06	8.46e-17	7.91e-06	1.33e-125	7.91e-06
38.63	1	5.17e-07	7.24e-22	5.17e-07	2.03e-186	5.17e-07
48.98	1	4.85e-09	7.19e-33	4.85e-09	0.00e+00	4.85e-09
58.18	1	6.02e-12	1.27e-49	6.02e-12	0.00e+00	6.02e-12
65.54	0	6.02e-12	1.27e-49	6.02e-12	0.00e+00	6.02e-12
65.65	1	3.46e-15	4.02e-68	3.46e-15	0.00e+00	3.46e-15
71.61	0	3.46e-15	4.02e-68	3.46e-15	0.00e+00	3.46e-15
79.96	0	3.46e-15	4.02e-68	3.46e-15	0.00e+00	3.46e-15
219.82	1	1.14e-21	5.97e-102	1.14e-21	0.00e+00	1.14e-21
244.93	1	1.30e-91	0.00e+00	1.30e-91	0.00e+00	1.30e-91

Table 4.8: Comparison of standard PH survival estimates and envelope imprecise survival estimates for individuals with $x = 5$. The estimates were derived from the RP model applied to the data set in Example 4.3.2, with $n = 30$ and $\hat{\beta} = -0.5644$, using ϵ^{**} values of 0.5 and 1.

time	status	$\hat{S}_{9;c}(t)$	$\epsilon^{**} = 0.5$		$\epsilon^{**} = 1$	
			$\hat{\underline{S}}_{9;p}(t)$	$\hat{\overline{S}}_{9;p}(t)$	$\hat{\underline{S}}_{9;p}(t)$	$\hat{\overline{S}}_{9;p}(t)$
1.50	1	9.81e-01	9.81e-01	9.91e-01	9.81e-01	9.98e-01
1.71	1	9.61e-01	9.61e-01	9.82e-01	9.61e-01	9.95e-01
2.01	1	9.33e-01	9.33e-01	9.63e-01	9.33e-01	9.89e-01
3.14	1	9.03e-01	9.03e-01	9.39e-01	9.03e-01	9.76e-01
3.77	1	8.74e-01	8.74e-01	9.16e-01	8.74e-01	9.64e-01
4.40	1	8.35e-01	8.35e-01	8.80e-01	8.35e-01	9.43e-01
4.51	1	7.95e-01	7.95e-01	8.39e-01	7.95e-01	9.14e-01
7.13	1	7.53e-01	7.52e-01	7.95e-01	7.53e-01	8.70e-01
8.51	1	7.12e-01	7.10e-01	7.52e-01	7.12e-01	8.28e-01
8.97	1	6.70e-01	6.65e-01	7.07e-01	6.65e-01	7.79e-01
11.36	1	6.22e-01	6.11e-01	6.57e-01	6.06e-01	7.23e-01
12.70	1	5.76e-01	5.59e-01	6.08e-01	5.49e-01	6.70e-01
12.94	1	5.21e-01	4.91e-01	5.50e-01	4.66e-01	6.05e-01
13.21	0	5.21e-01	4.91e-01	5.50e-01	4.66e-01	6.05e-01
14.12	1	4.66e-01	4.21e-01	4.92e-01	3.72e-01	5.42e-01
17.59	1	4.08e-01	3.49e-01	4.31e-01	2.74e-01	4.74e-01
17.60	1	3.56e-01	2.87e-01	3.75e-01	1.97e-01	4.13e-01
24.45	1	3.02e-01	2.26e-01	3.18e-01	1.29e-01	3.51e-01
26.67	0	3.02e-01	2.26e-01	3.18e-01	1.29e-01	3.51e-01
29.89	1	2.38e-01	1.54e-01	2.51e-01	5.87e-02	2.76e-01
38.34	0	2.38e-01	1.54e-01	2.51e-01	5.87e-02	2.76e-01
38.63	1	1.70e-01	8.65e-02	1.80e-01	1.59e-02	1.98e-01
48.98	1	9.61e-02	2.49e-02	1.01e-01	3.29e-04	1.12e-01
58.18	1	4.24e-02	3.74e-03	4.47e-02	1.59e-07	4.92e-02
65.54	0	4.24e-02	3.74e-03	4.47e-02	1.59e-07	4.92e-02
65.65	1	1.70e-02	4.61e-04	1.79e-02	3.08e-11	1.98e-02
71.61	0	1.70e-02	4.61e-04	1.79e-02	3.08e-11	1.98e-02
79.96	0	1.70e-02	4.61e-04	1.79e-02	3.08e-11	1.98e-02
219.82	1	2.74e-03	1.00e-05	2.89e-03	8.03e-18	3.18e-03
244.93	1	7.56e-12	3.18e-51	7.98e-12	0.00e+00	8.79e-12

Table 4.9: Comparison of standard PH survival estimates and envelope imprecise survival estimates for individuals with $x = 9$. The estimates were derived from the RP model applied to the data set in Example 4.3.2, with $n = 30$ and $\hat{\beta} = -0.5644$, using ϵ^{**} values of 0.5 and 1.

time	status	$\hat{S}_{15;c}(t)$	$\epsilon^{**} = 0.5$		$\epsilon^{**} = 1$	
			$\underline{\hat{S}}_{15;p}(t)$	$\overline{\hat{S}}_{15;p}(t)$	$\underline{\hat{S}}_{15;p}(t)$	$\overline{\hat{S}}_{15;p}(t)$
1.50	1	9.99e-01	9.99e-01	1.00e+00	9.99e-01	1.00e+00
1.71	1	9.98e-01	9.98e-01	1.00e+00	9.98e-01	1.00e+00
2.01	1	9.97e-01	9.97e-01	1.00e+00	9.97e-01	1.00e+00
3.14	1	9.96e-01	9.96e-01	9.99e-01	9.96e-01	1.00e+00
3.77	1	9.94e-01	9.94e-01	9.99e-01	9.94e-01	1.00e+00
4.40	1	9.92e-01	9.92e-01	9.99e-01	9.92e-01	1.00e+00
4.51	1	9.90e-01	9.90e-01	9.98e-01	9.90e-01	1.00e+00
7.13	1	9.88e-01	9.88e-01	9.97e-01	9.88e-01	1.00e+00
8.51	1	9.86e-01	9.86e-01	9.97e-01	9.86e-01	1.00e+00
8.97	1	9.83e-01	9.83e-01	9.96e-01	9.83e-01	1.00e+00
11.36	1	9.80e-01	9.80e-01	9.95e-01	9.80e-01	1.00e+00
12.70	1	9.77e-01	9.77e-01	9.94e-01	9.77e-01	1.00e+00
12.94	1	9.72e-01	9.72e-01	9.93e-01	9.72e-01	9.99e-01
13.21	0	9.72e-01	9.72e-01	9.93e-01	9.72e-01	9.99e-01
14.12	1	9.68e-01	9.68e-01	9.91e-01	9.68e-01	9.99e-01
17.59	1	9.62e-01	9.62e-01	9.89e-01	9.62e-01	9.99e-01
17.60	1	9.57e-01	9.57e-01	9.87e-01	9.57e-01	9.99e-01
24.45	1	9.50e-01	9.50e-01	9.84e-01	9.50e-01	9.98e-01
26.67	0	9.50e-01	9.50e-01	9.84e-01	9.50e-01	9.98e-01
29.89	1	9.40e-01	9.40e-01	9.80e-01	9.40e-01	9.98e-01
38.34	0	9.40e-01	9.40e-01	9.80e-01	9.40e-01	9.98e-01
38.63	1	9.27e-01	9.27e-01	9.74e-01	9.27e-01	9.96e-01
48.98	1	9.05e-01	9.05e-01	9.61e-01	9.05e-01	9.93e-01
58.18	1	8.73e-01	8.73e-01	9.41e-01	8.73e-01	9.86e-01
65.54	0	8.73e-01	8.73e-01	9.41e-01	8.73e-01	9.86e-01
65.65	1	8.40e-01	8.40e-01	9.20e-01	8.40e-01	9.79e-01
71.61	0	8.40e-01	8.40e-01	9.20e-01	8.40e-01	9.79e-01
79.96	0	8.40e-01	8.40e-01	9.20e-01	8.40e-01	9.79e-01
219.82	1	7.77e-01	7.77e-01	8.83e-01	7.77e-01	9.65e-01
244.93	1	3.34e-01	2.48e-01	3.80e-01	8.58e-02	4.15e-01

Table 4.10: Comparison of standard PH survival estimates and envelope imprecise survival estimates for individuals with $x = 15$. The estimates were derived from the RP model applied to the data set in Example 4.3.2, with $n = 30$ and $\hat{\beta} = -0.5644$, using ϵ^{**} values of 0.5 and 1.

maximum values in the data set, the PH survival estimates fluctuate between the corresponding imprecise envelope survival estimates derived from the robust model as in Table 4.8, where $x = 9$ representing individuals with covariate value approximately equal the mean, see Figure 4.4. This aspect is naturally consistent with the definition of the lower and upper hazard functions in Equations (4.15) and (4.16). The reverse results are found when the regression coefficient is positive.

Table 4.11 presents the disparities between the envelope upper and lower survival estimates, $\overline{\hat{S}}_{x;p}(t) - \underline{\hat{S}}_{x;p}(t)$. Unlike the naive approach, the table illustrates that the disparity between the envelope-type upper and lower survival functions expands with increasing levels of imprecision. This is an inevitable consequence of the constraints imposed by the envelope estimation methodology. Moreover, although not included in this thesis, the difference between the envelope imprecise survival estimates decreases as the sample size increases.

Regarding the correlation between the disparities of the envelope-type upper and lower survival estimates and the covariate effect, it remains consistent. Specifically, there is a significant tendency for the difference between the imprecise survival estimates to decline over time for maximum covariate effects, associated with $x = 5$ given that the regression coefficient is negative. Conversely, this trend fades away and eventually reverses as the covariate effect decreases and reaches its minimum value, as evidenced by the case of $x = 15$. Accordingly, differences between envelope-type upper and lower survival estimates are not uniform across all individuals, and differ according to the magnitude of the covariate effect, that is, the value of the covariate and the sign of the regression coefficient. It has been observed that this aspect exists in the naive imprecise survival estimates as well, indicating that it is an inherent feature of robust model.

◇

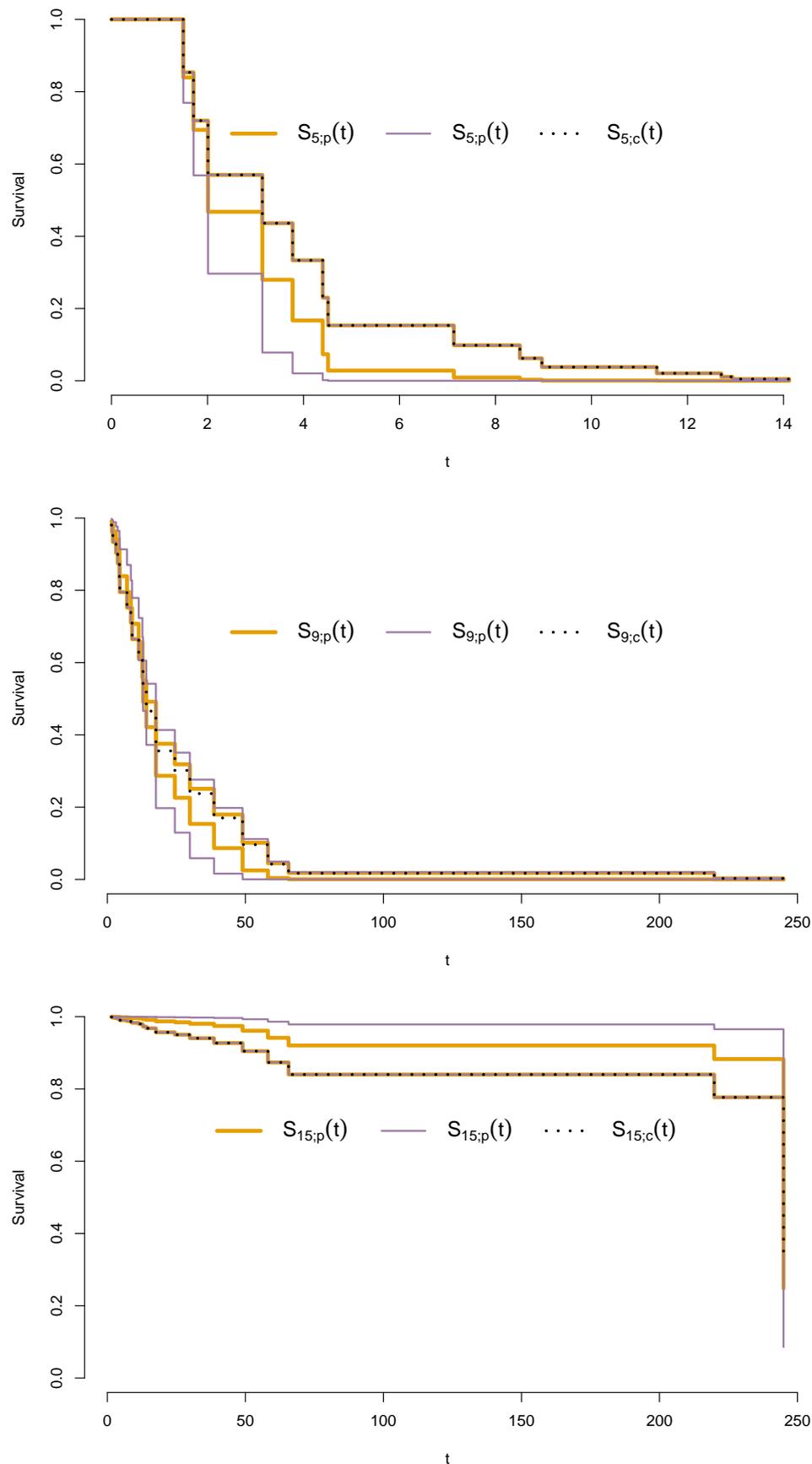


Figure 4.4: Comparison of standard PH survival estimates and envelope imprecise survival estimates for $x = 5, 9, 15$ (top-bottom). These estimates were derived from the RP model applied to the data set in Example 4.3.2, using ϵ^{**} values of 0.5 (amber), and 1 (indigo).

time	status	$\widehat{S}_{5;p}(t) - \underline{\widehat{S}}_{5;p}(t)$		$\widehat{S}_{9;p}(t) - \underline{\widehat{S}}_{9;p}(t)$		$\widehat{S}_{15;p}(t) - \underline{\widehat{S}}_{15;p}(t)$	
		$\epsilon^{**} = 0.5$	1	0.5	1	0.5	1
1.50	1	1.40e-02	8.43e-02	1.06e-02	1.68e-02	7.36e-04	8.27e-04
1.71	1	2.55e-02	1.51e-01	2.16e-02	3.41e-02	1.52e-03	1.71e-03
2.01	1	1.02e-01	2.73e-01	2.99e-02	5.52e-02	2.54e-03	2.93e-03
3.14	1	1.57e-01	3.58e-01	3.58e-02	7.30e-02	3.65e-03	4.31e-03
3.77	1	1.67e-01	3.13e-01	4.15e-02	9.01e-02	4.78e-03	5.70e-03
4.40	1	1.56e-01	2.28e-01	4.42e-02	1.08e-01	6.27e-03	7.62e-03
4.51	1	1.25e-01	1.53e-01	4.42e-02	1.19e-01	7.87e-03	9.70e-03
7.13	1	8.91e-02	9.83e-02	4.27e-02	1.17e-01	9.57e-03	1.20e-02
8.51	1	5.94e-02	6.23e-02	4.13e-02	1.16e-01	1.13e-02	1.43e-02
8.97	1	3.72e-02	3.79e-02	4.26e-02	1.14e-01	1.31e-02	1.68e-02
11.36	1	2.06e-02	2.07e-02	4.62e-02	1.17e-01	1.53e-02	1.98e-02
12.70	1	1.10e-02	1.11e-02	4.94e-02	1.21e-01	1.76e-02	2.29e-02
12.94	1	4.83e-03	4.83e-03	5.89e-02	1.39e-01	2.04e-02	2.70e-02
13.21	0	4.83e-03	4.83e-03	5.89e-02	1.39e-01	2.04e-02	2.70e-02
14.12	1	1.94e-03	1.94e-03	7.04e-02	1.69e-01	2.34e-02	3.14e-02
17.59	1	6.59e-04	6.59e-04	8.16e-02	2.01e-01	2.68e-02	3.66e-02
17.60	1	2.14e-04	2.14e-04	8.87e-02	2.16e-01	3.03e-02	4.20e-02
24.45	1	5.59e-05	5.59e-05	9.26e-02	2.21e-01	3.45e-02	4.83e-02
26.67	0	5.59e-05	5.59e-05	9.26e-02	2.21e-01	3.45e-02	4.83e-02
29.89	1	7.91e-06	7.91e-06	9.71e-02	2.17e-01	4.00e-02	5.73e-02
38.34	0	7.91e-06	7.91e-06	9.71e-02	2.17e-01	4.00e-02	5.73e-02
38.63	1	5.17e-07	5.17e-07	9.30e-02	1.82e-01	4.73e-02	6.94e-02
48.98	1	4.85e-09	4.85e-09	7.65e-02	1.11e-01	5.65e-02	8.83e-02
58.18	1	6.02e-12	6.02e-12	4.10e-02	4.92e-02	6.80e-02	1.13e-01
65.54	0	6.02e-12	6.02e-12	4.10e-02	4.92e-02	6.80e-02	1.13e-01
65.65	1	3.46e-15	3.46e-15	1.75e-02	1.98e-02	8.03e-02	1.39e-01
71.61	0	3.46e-15	3.46e-15	1.75e-02	1.98e-02	8.03e-02	1.39e-01
79.96	0	3.46e-15	3.46e-15	1.75e-02	1.98e-02	8.03e-02	1.39e-01
219.82	1	1.14e-21	1.14e-21	2.88e-03	3.18e-03	1.06e-01	1.88e-01
244.93	1	1.30e-91	1.30e-91	7.98e-12	8.79e-12	1.32e-01	3.29e-01

Table 4.11: Comparison of the width of envelope survival estimates for individuals with $x = 5, 9, 15$ based on the RP model applied to the data set in Example 4.3.2, with $n = 30$ and $\hat{\beta} = -0.5644$, using ϵ^{**} values of 0.5 and 1.

4.4 Robust-Empirical PH likelihood (RE)

Our objective in this section is to construct the empirical likelihood for robust PH models based on the baseline CDF, $F_{0;e}$, as demonstrated in Section 2.5.2. The derivation steps of this likelihood is almost identical to Ren and Zhou [77]. Therefore, we will summarize our finding to avoid extensive repetition. Recall the relation between the survival function and the cumulative hazard function, as in Equation (2.9). This relationship is a crucial tool for deriving the full empirical likelihood function for the robust PH model, where

$$\begin{aligned}
 S_e(t | x_i, \epsilon_i) &= \exp(-H_e(t | x_i, \epsilon_i)) \\
 &= \exp\left(-\int_0^t h_e(t | x_i, \epsilon_i) du\right) \\
 &= \exp\left(-\int_0^t h_{0;e}(t) \alpha_i du\right) \\
 &= S_{0;e}(t)^{\alpha_i}
 \end{aligned} \tag{4.19}$$

By differentiating both sides of Equation (4.19) with respect to t and using $S_e(t) = 1 - F_e(t)$, we obtain the following

$$\begin{aligned}
 \frac{d}{dt} S_e(t | x_i, \epsilon_i) = \frac{d}{dt} S_{0;e}(t)^{\alpha_i} &\Rightarrow -f_e(t | x_i, \epsilon_i) = \alpha_i [-f_{0;e}(t)] [S_{0;e}(t)]^{\alpha_i - 1} \\
 &\Rightarrow f_e(t | x_i, \epsilon_i) = \alpha_i f_{0;e}(t) S_{0;e}(t)^{\alpha_i - 1}
 \end{aligned} \tag{4.20}$$

Then, the likelihood function for the robust PH model can be obtained by substituting Equations (4.19) and (4.20) into the full likelihood for survival data in Equation (2.24), as follows

$$\begin{aligned}
 \prod_{i=1}^n f_e(t_i, \delta_i | x_i, \epsilon_i) &\propto \prod_{i=1}^n f_e(t_i | x_i)^{\delta_i} S_e(t_i | x_i)^{1 - \delta_i} \\
 &\propto \prod_{i=1}^n [\alpha_i f_{0;e}(t) S_{0;e}(t)^{\alpha_i - 1}]^{\delta_i} [S_{0;e}(t_i)^{\alpha_i}]^{1 - \delta_i} \\
 &\propto \prod_{i=1}^n [\alpha_i dF_{0;e}(t_i)]^{\delta_i} [S_{0;e}(t_i)]^{\alpha_i - \delta_i}
 \end{aligned}$$

Accordingly, the full likelihood function for the robust PH model is given by

$$L(\beta, F_{0;e}, \epsilon) = \prod_{i=1}^n [\alpha_i dF_{0;e}(t_i)]^{\delta_i} [S_{0;e}(t_i)]^{\alpha_i - \delta_i} \tag{4.21}$$

In the following theorem, we summarize the results of the optimization of the likelihood function above.

Theorem 4.4.1 Under the full empirical likelihood function for the robust PH model we have:

1. The baseline survival function in Equation (4.19) can be estimated under the constraint that $\alpha_n \geq 1$ for any fixed values of β and ϵ by

$$\hat{S}_{0n;e}(t) = 1 - \hat{F}_{0n;e}(t) = \prod_{i:t_i \leq t} \frac{r_i - \delta_i}{r_i} \quad (4.22)$$

where $r_i = \sum_{j=i}^n \alpha_j$ and $\alpha_n = 1$ to meet the constraint that $r_i \geq 1$ for every event time.

2. The profile empirical likelihood function corresponds to the robust PH model is given by

$$L(\beta_e, \epsilon) = \prod_{i=1}^n \left(\frac{\alpha_i}{r_i} \right)^{\delta_i} \left(\frac{r_i - \delta_i}{r_i} \right)^{r_i - \delta_i} \quad (4.23)$$

Proof.

To avoid unnecessary repetition, this theorem can be proved easily and follows the approach detailed in Appendix A.2 for Theorem 2.5.2, which involves using the definition of the robust PH model outlined in Equation (4.2), where ϕ_i is simply substituted for α_i . \square

It is important to notice that the profile likelihood function for the robust PH model in Equation (4.23) depends on the constraint $r_i \geq 1$ for all event times. By setting $\alpha_n = 1$, this constraint is satisfied since r_i is the sum of positive values α_j with j ranging from i to n , similar to Section 2.5.2. During the optimization process, the values of $x_i + \epsilon_i$ are adjusted by $x_i + \epsilon_i - (x_n + \epsilon_n)$. As a result of this convention, the estimates $\hat{\beta}_e, \hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_n$ can then be substituted into Equation (4.22) to obtain the survival function corresponding to $x = x_n + \epsilon_n$. Thus, the baseline survival function can be determined by

$$\hat{S}_{0;e}(t) = \hat{S}_{0n;e}(t)^{\exp(-\hat{\beta}_e[x_n + \hat{\epsilon}_n])}$$

Imprecise survival functions

The imprecise survival functions can be estimated using two feasible approaches, naive and envelop. The naive imprecise survival estimates may not incorporate survival estimates corresponding to lower levels of imprecision, as seen in Section 4.3.1. The naive survival functions for the robust PH model based on the empirical likelihood can be obtained by directly substitute the imprecision limits, $-\epsilon^*$ and ϵ^* , to Equation (4.19) in accordance with the estimates of the regression coefficient and the baseline survival function. Hence, the resulting naive imprecise survival functions when the estimate of the regression coefficient is positive for an individual with the covariate value $x = x_j$ at time t using the level of imprecision ϵ^* are given by

$$\underline{S}_j(t; \epsilon^*) = \hat{S}_{0;e}(t)^{\exp(\hat{\beta}_e(x_j + \epsilon^*))} \quad (4.24)$$

$$\overline{S}_j(t; \epsilon^*) = \hat{S}_{0;e}(t)^{\exp(\hat{\beta}_e(x_j - \epsilon^*))} \quad (4.25)$$

When the estimate of the regression coefficient is negative, the lower survival estimates correspond to $-\epsilon^*$, whereas the upper survival estimates are related to ϵ^* . The baseline survival estimates obtained from the robust empirical likelihood demonstrate monotonic properties in response to increased imprecision levels, as shown in the simulations and examples. Similar to their counterparts in the robust model based on Poisson likelihood, the inability of the naive approach to preserve this property in the imprecise survival estimates for individuals leads to inconsistent behavior. Consequently, the focus will be on the envelope-type of survival estimates, which possess attractive features that compensate for the shortcomings of the naive approach.

In the envelope approach, we restrict survival estimates derived from lower levels of imprecision to fall within the range of the imprecise survival estimates obtained from the robust PH model at the selected level of imprecision. In this context, the imprecise survival estimates for an individual using a specific imprecision level, ϵ^{**} , are regarded as an envelope encompassing all potential survival estimates associated with imprecision levels less than ϵ^{**} . Let $\mathfrak{S}(t|x = x_j; \epsilon^{**})$ denote the set of all possible survival functions for an individual with a covariate value x_j at the observed time t

given the level of imprecision ϵ^* . It follows that the imprecise survival functions for the individual with a covariate value $x = x_j$ at time t given the level of imprecision ϵ^{**} can be defined by

$$\begin{aligned}\underline{S}_j(t; \epsilon^{**}) &= \min\{\mathfrak{S}(t|x = x_j; \epsilon^{**})\} \\ &= \min\{\hat{S}_0(t)^{\exp(\hat{\beta}_c x_j)}, \hat{S}_{0;e}(t; \epsilon^{**})^{\exp(\hat{\beta}_e x_j)}\}\end{aligned}\quad (4.26)$$

$$\begin{aligned}\overline{S}_j(t; \epsilon^{**}) &= \max\{\mathfrak{S}(t|x = x_j)\} \\ &= \max\{\hat{S}_0(t)^{\exp(\hat{\beta}_c x_j)}, \hat{S}_{0;e}(t; \epsilon^{**})^{\exp(\hat{\beta}_e x_j)}\}\end{aligned}\quad (4.27)$$

In light of this definition, it is apparent that only the baseline survival estimates are directly impacted by the imprecision effect, while the regression coefficient $\hat{\beta}_e$ is influenced indirectly. Furthermore, the subscript c refers to the estimates derived from the PH model based on empirical likelihood, which differs marginally from those derived from the Poisson likelihood.

Theorem 4.4.2 Consider maximizing the likelihood function in Equation (4.23) when the estimates of the regression parameter is positive. Then, the imprecision term for the j -th observation can be estimated as follows:

$$\hat{\epsilon}_j = \begin{cases} \text{Not determined} & ; \text{ for right-censored observations before the first event} \\ \epsilon^* & ; \text{ for the first observed event} \\ -\epsilon^* & ; \text{ for } j = n \text{ and right-censored observations after the first event} \\ \in [-\epsilon^*, \epsilon^*] & ; \text{ for other events} \end{cases}$$

Conversely, ϵ_j can be estimated when the regression estimate is negative as follows

$$\hat{\epsilon}_j = \begin{cases} \text{Not determined} & ; \text{ for right-censored observations before the first event} \\ -\epsilon^* & ; \text{ for the first observed event} \\ \epsilon^* & ; \text{ for } j = n \text{ and right-censored observations after the first event} \\ \in [-\epsilon^*, \epsilon^*] & ; \text{ for other events} \end{cases}$$

Proof.

Although the subsequent proof addresses the case of positive regression estimates, analogous steps can be applied to demonstrate the statement for negative regression

estimates. The log-likelihood associated with Equation(4.23) can be represented as follows

$$\ell(\beta_e, \boldsymbol{\epsilon}) = \sum_{i=1}^n \delta_i \ln \left(\frac{\alpha_i}{r_i} \right) + (r_i - \delta_i) \ln \left(\frac{r_i - \delta_i}{r_i} \right) \quad (4.28)$$

where the summation is effectively calculated over only failure times. Based upon a particular ϵ_j , it appears that ϵ_j only affects ℓ directly through α_i when $i = j$ and indirectly through r_i when $i \leq j$. Accordingly, the derivative of ℓ with respect to ϵ_j is given by

$$\frac{\partial \ell}{\partial \epsilon_j} = \frac{\partial \ell}{\partial \alpha_j} \frac{\partial \alpha_j}{\partial \epsilon_j}$$

and observe that

$$\frac{\partial \alpha_j}{\partial \epsilon_j} = \hat{\beta}_e \alpha_j$$

since α_j is always positive, then $\frac{\partial \alpha_j}{\partial \epsilon_j}$ is positive when $\hat{\beta}_e > 0$ and negative when $\hat{\beta}_e < 0$. Furthermore, the derivative of the log-likelihood with respect to α_j is

$$\frac{\partial \ell}{\partial \alpha_j} = \frac{\partial}{\partial \alpha_j} \sum_{i=1}^n \delta_i \ln \left(\frac{\alpha_i}{r_i} \right) + \frac{\partial}{\partial \alpha_j} \sum_{i=1}^n (r_i - \delta_i) \ln \left(\frac{r_i - \delta_i}{r_i} \right) \quad (4.29)$$

From Equation (4.10), the derivative of the first summation is

$$\frac{\delta_j}{\alpha_j} - \sum_{i \leq j} \frac{\delta_i}{r_i}$$

The derivative of the second summation is

$$\frac{\partial}{\partial \alpha_j} \sum_{i=1}^n (r_i - \delta_i) \ln \left(\frac{r_i - \delta_i}{r_i} \right) = \sum_{i \leq j} \frac{\delta_i}{r_i} + \sum_{i \leq j} \ln \left(\frac{r_i - \delta_i}{r_i} \right)$$

Hence, the derivative of the log-likelihood in Equation (4.29) can be expressed as follows

$$\frac{\partial \ell}{\partial \alpha_j} = \frac{\delta_j}{\alpha_j} + \sum_{i \leq j} \ln \left(\frac{r_i - \delta_i}{r_i} \right) \quad (4.30)$$

In light of these derivatives, it can be determined that ϵ_j has the following properties:

1. These values of ϵ_j for right-censored observations prior to the first event do not affect the log-likelihood, and hence their estimates cannot be determined.
2. For the first failure time, (1), we have

$$\frac{\partial \ell}{\partial \alpha_{(1)}} = \frac{1}{\alpha_{(1)}} + \ln \left(\frac{r_{(1)} - 1}{r_{(1)}} \right)$$

The following demonstrating that $\frac{\partial \ell}{\partial \alpha_1} > 0$:

It is known that $\ln\left(\frac{r_{(1)}-1}{r_{(1)}}\right)$ is an increasing function of $r_{(1)}$. Since $r_{(1)} \geq \alpha_{(1)} + \alpha_{[n]} = \alpha_{(1)} + 1$, then the minimum value that $r_{(1)}$ could be is $r_{(1)} = \alpha_{(1)} + 1$ which follows

$$\ln\left(\frac{r_{(1)}-1}{r_{(1)}}\right) \geq \ln\left(\frac{(\alpha_{(1)}+1)-1}{\alpha_{(1)}+1}\right) = \ln\left(\frac{(\alpha_{(1)})}{\alpha_{(1)}+1}\right)$$

Define a function $g(\alpha_{(1)})$ as follows

$$g(\alpha_{(1)}) = \frac{1}{\alpha_{(1)}} + \ln\left(\frac{(\alpha_{(1)})}{\alpha_{(1)}+1}\right) \leq \frac{\partial \ell}{\partial \alpha_{(1)}}$$

Therefore, proving that $g(\alpha_{(1)}) > 0$ implies that $\frac{\partial \ell}{\partial \alpha_{(1)}} > 0$, so by differentiating $g(\alpha_{(1)})$ with respect to $\alpha_{(1)}$ we have

$$\begin{aligned} \frac{\partial}{\partial \alpha_{(1)}} g(\alpha_{(1)}) &= \frac{-1}{\alpha_{(1)}^2} + \frac{\left(\frac{\alpha_{(1)}+1-\alpha_{(1)}}{(\alpha_{(1)})^2}\right)}{\left(\frac{\alpha_{(1)}}{\alpha_{(1)}+1}\right)} \\ &= \frac{-1}{\alpha_{(1)}^2} + \frac{1}{\alpha_{(1)}^2 + \alpha_{(1)}} \end{aligned} \quad (4.31)$$

The derivative of $g(\alpha_{(1)})$ is always negative for $\alpha_{(1)} > 0$ which indicates that $g(\alpha_{(1)})$ is a decreasing function with respect to $\alpha_{(1)}$. By analyzing the least value of $g(\alpha_{(1)})$ as $\alpha_{(1)}$ approaches infinity, it can be seen that

$$\begin{aligned} \lim_{\alpha_{(1)} \rightarrow \infty} \frac{1}{\alpha_{(1)}} &\rightarrow 0 \\ \lim_{\alpha_{(1)} \rightarrow \infty} \ln\left(\frac{(\alpha_{(1)})}{\alpha_{(1)}+1}\right) &\leq \frac{\partial \ell}{\partial \alpha_{(1)}} \rightarrow 0 \end{aligned}$$

Therefore, $g(\alpha_{(1)}) \rightarrow 0$ from the positive side as $\alpha_{(1)}$ approaches infinity, confirming that $g(\alpha_{(1)}) > 0$ which implies that $\frac{\partial \ell}{\partial \alpha_{(1)}} > 0$. Hence, $\frac{\partial \ell}{\partial \epsilon_{(1)}} > 0$ only if $\hat{\epsilon}_{(1)} = \epsilon^*$ when $\hat{\beta}_e > 0$ or $\hat{\epsilon}_{(1)} = -\epsilon^*$ when $\hat{\beta}_e < 0$.

3. For right-censored observations after the first event, $r_i > 1$ for all $i < n$ since $\alpha_n \geq 1$, leading to $\frac{\partial \ell}{\partial \alpha_j} < 0$. Therefore, α_j needs to be as small as possible, i.e. $\hat{\epsilon}_j = \epsilon^*$ when $\hat{\beta}_e < 0$ and $\hat{\epsilon}_j = -\epsilon^*$ when $\hat{\beta}_e > 0$.

4. For the final observation where $\delta_n = 1$, the analytical proof of this case is complicated, primarily because the likelihood holds only under the constraint that $\alpha_n = 1$. Consequently, this affects all preceding observations where $j < n$. Despite these complexities, extensive simulations reinforce that setting ϵ_n as the limit of imprecision with a sign opposite to that of $\hat{\beta}_e$ consistently yields the highest likelihood. This result aligns with the findings presented in Theorem 4.3.2. \square

Theorem 4.4.2 reveals that imprecision terms related to the robust empirical likelihood are estimated in the same manner as their counterparts in the robust Poisson likelihood. That is for relatively small α_j and a small j (early time) where there are not many $i \leq j$, then $\partial\ell/\partial\alpha_j$ is likely to be positive leading to imprecision terms that match the sign of the regression parameter. Conversely, $\partial\ell/\partial\alpha_j$ is likely to be negative for relatively large α_j and j is large (late time) resulting in imprecision terms with the opposite sign of the regression parameter. As can be seen from Theorem 4.4.2, the estimates of the imprecision terms follow a similar pattern to their counterparts in Poisson likelihood of the robust model. Additionally, it should be noted that the $\hat{\epsilon}_i$ values determined by Theorem 4.4.2 will be substituted into the likelihood function. Consequently, the optimization will be conducted only over the regression parameter and ϵ_i terms within the range $[-\epsilon^*, \epsilon^*]$ for which Theorem 4.4.2 did not yield clear results. The upcoming example illustrates the impact of increasing the imprecision level to the regression coefficient, likelihood value, and the survival estimates for individuals using simulated data, also used in Example 4.3.1.

Example 4.4.1 This example demonstrates how the robust PH model based on the empirical full likelihood behaves when the level of imprecision is increased. The model was fitted on the exact simulated datasets that used in Example 4.3.1. Table 4.12 presents the regression coefficient estimates and log-likelihood values for both the standard proportional hazards model and the robust PH model based on the Empirical likelihood using different values of imprecision levels. The table re-

ϵ^*	$n = 30$		$n = 60$		$n = 200$	
	$\hat{\beta}_e$	ℓ	$\hat{\beta}_e$	ℓ	$\hat{\beta}_e$	ℓ
0	-0.5145	-74.129	-0.5157	-178.20	-0.5531	-799.51
0.1	-0.5546	-73.121	-0.5466	-176.13	-0.5874	-791.98
0.5	-0.7454	-68.577	-0.6878	-167.25	-0.7482	-759.68
1	-1.1659	-61.803	-0.9573	-154.14	-1.0431	-714.51
1.2	-1.4357	-58.814	-1.1120	-148.20	-1.2009	-694.99

Table 4.12: Estimates of β and values of $\ell(\hat{\beta}_e, \hat{\epsilon})$ for the RE model applied to the datasets in Example 4.4.1 having $\beta = -0.5$. A range of values of ϵ^* is considered: 0, 0.1, 0.5, 1 and 1.2.

veals similar characteristics obtained from fitting the robust PH model based on Poisson likelihood. That is increasing the imprecision level for a negative coefficient will decrease the regression coefficient estimates, and the reverse holds true for a positive regression coefficient.

The estimates of the regression coefficient and imprecision terms in Table 4.13 illustrate that there is a tendency for the imprecision terms to align with the sign of the estimated regression parameter, $\hat{\beta}$, particularly for early-stage event occurrences. Conversely, for individuals associated with later event and censoring times, the optimization of contributions is often achieved when the majority of imprecision terms take values with a sign opposite to that of the estimated regression parameter, as demonstrated by Theorem 4.4.2.

These imprecision terms have the flexibility to take any value within the pre-defined imprecision interval, including the potential for a change in their sign in response to changes in the imprecision level, as highlighted in Table 4.13. This behavior has been observed in Example 4.3.1 for the robust model based on the Poisson likelihood which is believed to be an inherent characteristic of constrained multivariate optimization.

In this example, the naive imprecise survival estimates based on the robust RE likelihood have been omitted due to the same inconsistency issue observed in the

Data			$\epsilon^* = 0.1$	0.5	1	1.2	
t	$status$	x	$\hat{\beta}_e$	-0.55	-0.75	-1.17	-1.44
1.50	1	6.59	$\hat{\epsilon}_1$	-0.10	-0.50	-1.00	-1.20
1.71	1	4.10	$\hat{\epsilon}_2$	-0.10	-0.47	0.41	0.80
2.01	1	6.20	$\hat{\epsilon}_3$	-0.10	-0.50	-1.00	-0.73
3.14	1	12.51	$\hat{\epsilon}_4$	-0.10	-0.50	-1.00	-1.20
3.77	1	4.94	$\hat{\epsilon}_5$	0.10	0.50	1.00	1.20
4.40	1	7.81	$\hat{\epsilon}_6$	-0.10	-0.50	-1.00	-1.08
4.51	1	7.94	$\hat{\epsilon}_7$	-0.10	-0.50	-1.00	-0.83
7.13	1	10.46	$\hat{\epsilon}_8$	-0.10	-0.50	-1.00	-1.20
8.51	1	8.32	$\hat{\epsilon}_9$	-0.10	-0.50	-0.74	-0.59
8.97	1	6.92	$\hat{\epsilon}_{10}$	0.10	0.50	0.92	1.05
11.36	1	10.39	$\hat{\epsilon}_{11}$	-0.10	-0.50	-1.00	-1.20
12.70	1	6.80	$\hat{\epsilon}_{12}$	0.10	0.50	1.00	1.20
12.94	1	9.35	$\hat{\epsilon}_{13}$	-0.10	-0.50	-0.76	-0.67
13.21	0	15.36	$\hat{\epsilon}_{14}$	0.10	0.50	1.00	1.20
14.12	1	8.33	$\hat{\epsilon}_{15}$	0.10	0.39	0.52	0.60
17.59	1	11.38	$\hat{\epsilon}_{16}$	-0.10	-0.50	-1.00	-1.20
17.60	1	8.66	$\hat{\epsilon}_{17}$	0.10	0.50	0.67	0.73
24.45	1	11.20	$\hat{\epsilon}_{18}$	-0.10	-0.50	-1.00	-1.20
26.67	0	8.12	$\hat{\epsilon}_{18}$	0.10	0.50	1.00	1.20
29.89	1	11.08	$\hat{\epsilon}_{20}$	-0.10	-0.50	-1.00	-1.20
38.34	0	9.31	$\hat{\epsilon}_{21}$	0.10	0.50	1.00	1.20
38.63	1	8.58	$\hat{\epsilon}_{22}$	0.10	0.50	1.00	1.20
48.98	1	10.21	$\hat{\epsilon}_{23}$	0.10	0.50	0.56	0.60
58.18	1	13.67	$\hat{\epsilon}_{24}$	-0.10	-0.50	-1.00	-1.20
65.54	0	15.15	$\hat{\epsilon}_{25}$	0.10	0.50	1.00	1.20
65.65	1	11.49	$\hat{\epsilon}_{26}$	0.10	0.23	0.23	0.27
71.61	0	13.76	$\hat{\epsilon}_{27}$	0.10	0.50	1.00	1.20
79.96	0	12.10	$\hat{\epsilon}_{28}$	0.10	0.50	1.00	1.20
219.82	1	10.33	$\hat{\epsilon}_{29}$	0.10	0.50	1.00	1.20
244.93	1	14.68	$\hat{\epsilon}_{30}$	0.10	0.50	1.00	1.20

Table 4.13: Estimates of $\beta, \epsilon_1, \dots, \epsilon_{30}$ for the RE model applied to the dataset in Example 4.4.1 with $n = 30$ having $\beta = -0.5$. A range of values of ϵ^* is considered: 0, 0.1, 0.5, 1 and 1.2.

naive imprecise survival estimates based on the robust RP likelihood. The envelope-type of imprecise lower and upper survival estimates, alongside the survival estimates derived from the PH Empirical model, are presented in Tables 4.14, 4.15, and 4.16 for individuals with $x = 5, 9, 15$, respectively. These tables indicate that envelope survival estimates are decreased at event times, but remain unchanged between successive event times similar to the PH model. As compared to their Poisson-based counterparts, envelope survival estimates derived from a robust PH model based on empirical likelihood are always zero for the last observation associated with an event time, as shown in Figure 4.5. Additionally, imprecise survival estimates obtained from either the RP or RE likelihoods are comparable, particularly those based of the naive approach as illustrated in Figure 4.6 (bottom) for individuals with $x = 15$. However, the envelope-type of imprecise survival estimates derived from either the RP or RE likelihood exhibit slight discrepancies, as illustrated in Figure 4.6 (top) for individuals with $x = 15$. This disparity arises because the envelope survival estimates are intended to encompass the PH survival estimates, which inherently vary depending on the likelihood function employed. Further, It is apparent from the tables that the imprecision level is correlated with the disparity between survival estimates for individuals. In spite of the fact that the PH survival estimates have generally fallen within the envelope imprecise survival estimates for all individuals, the sign of the regression coefficient and the values of x are crucial factors in determining how the PH survival estimates fluctuate within their imprecise survival estimates. Similar to the robust model based on the poisson likelihood, the PH survival estimates converge to the upper survival estimates derived from the robust empirical PH model for small x values, as shown in Table 4.14. Conversely, Table 4.16 illustrates that the PH survival estimates align with the lower survival estimates obtained from the RE model. For values of x between the highest and lowest values of the covariate, Table 4.15 demonstrates that the PH survival estimates exhibit fluctuations within the imprecise survival estimates as $x = 9$. Notably, the reverse patterns are found when the regression coefficient is positive.

Table 4.17 displays the disparities between upper and lower survival estimates. The table indicates that as the level of imprecision increases, the differences between

time	status	$\hat{S}_{5;c}(t)$	$\epsilon^{**} = 0.5$		$\epsilon^{**} = 1$	
			$\underline{\hat{S}}_{5;e}(t)$	$\overline{\hat{S}}_{5;e}(t)$	$\underline{\hat{S}}_{5;e}(t)$	$\overline{\hat{S}}_{5;e}(t)$
1.50	1	8.56e-01	8.41e-01	8.56e-01	7.70e-01	8.56e-01
1.71	1	7.24e-01	6.98e-01	7.24e-01	5.69e-01	7.24e-01
2.01	1	5.76e-01	4.73e-01	5.76e-01	2.97e-01	5.76e-01
3.14	1	4.44e-01	2.85e-01	4.44e-01	7.88e-02	4.44e-01
3.77	1	3.42e-01	1.71e-01	3.42e-01	2.09e-02	3.42e-01
4.40	1	2.39e-01	7.70e-02	2.39e-01	1.91e-03	2.39e-01
4.51	1	1.61e-01	3.03e-02	1.61e-01	6.63e-05	1.61e-01
7.13	1	1.05e-01	1.01e-02	1.05e-01	3.78e-07	1.05e-01
8.51	1	6.76e-02	3.25e-03	6.76e-02	1.85e-09	6.76e-02
8.97	1	4.19e-02	8.74e-04	4.19e-02	1.34e-12	4.19e-02
11.36	1	2.34e-02	1.64e-04	2.34e-02	7.23e-17	2.34e-02
12.70	1	1.28e-02	2.84e-05	1.28e-02	2.10e-21	1.28e-02
12.94	1	5.78e-03	2.20e-06	5.78e-03	5.62e-29	5.78e-03
13.21	0	5.78e-03	2.20e-06	5.78e-03	5.62e-29	5.78e-03
14.12	1	2.42e-03	1.09e-07	2.42e-03	2.53e-39	2.42e-03
17.59	1	8.60e-04	2.73e-09	8.60e-04	1.85e-53	8.60e-04
17.60	1	2.94e-04	5.69e-11	2.94e-04	1.62e-68	2.94e-04
24.45	1	8.12e-05	5.27e-13	8.12e-05	7.08e-88	8.12e-05
26.67	0	8.12e-05	5.27e-13	8.12e-05	7.08e-88	8.12e-05
29.89	1	1.26e-05	2.73e-16	1.26e-05	3.41e-124	1.26e-05
38.34	0	1.26e-05	2.73e-16	1.26e-05	3.41e-124	1.26e-05
38.63	1	9.30e-07	3.53e-21	9.30e-07	2.90e-184	9.30e-07
48.98	1	1.10e-08	9.71e-32	1.10e-08	0.00e+00	1.10e-08
58.18	1	1.88e-11	7.58e-48	1.88e-11	0.00e+00	1.88e-11
65.54	0	1.88e-11	7.58e-48	1.88e-11	0.00e+00	1.88e-11
65.65	1	1.49e-14	1.09e-65	1.49e-14	0.00e+00	1.49e-14
71.61	0	1.49e-14	1.09e-65	1.49e-14	0.00e+00	1.49e-14
79.96	0	1.49e-14	1.09e-65	1.49e-14	0.00e+00	1.49e-14
219.82	1	5.77e-21	1.34e-98	5.77e-21	0.00e+00	5.77e-21
244.93	1	0	0	0	0	0

Table 4.14: Comparison of standard PH survival estimates and envelope imprecise survival estimates for individuals with $x = 5$. The estimates were derived from the RE model applied to the data set in Example 4.4.1, with $n = 30$ and $\hat{\beta} = -0.5145$, using ϵ^{**} values of 0.5 and 1.

time	status	$\hat{S}_{9;c}(t)$	$\epsilon^{**} = 0.5$		$\epsilon^{**} = 1$	
			$\underline{\hat{S}}_{9;p}(t)$	$\overline{\hat{S}}_{9;p}(t)$	$\underline{\hat{S}}_{9;p}(t)$	$\overline{\hat{S}}_{9;p}(t)$
1.50	1	9.80e-01	9.80e-01	9.91e-01	9.80e-01	9.98e-01
1.71	1	9.60e-01	9.60e-01	9.82e-01	9.60e-01	9.95e-01
2.01	1	9.32e-01	9.32e-01	9.63e-01	9.32e-01	9.89e-01
3.14	1	9.01e-01	9.01e-01	9.38e-01	9.01e-01	9.76e-01
3.77	1	8.72e-01	8.72e-01	9.14e-01	8.72e-01	9.64e-01
4.40	1	8.33e-01	8.33e-01	8.78e-01	8.33e-01	9.43e-01
4.51	1	7.92e-01	7.92e-01	8.37e-01	7.92e-01	9.13e-01
7.13	1	7.50e-01	7.50e-01	7.92e-01	7.50e-01	8.70e-01
8.51	1	7.09e-01	7.09e-01	7.48e-01	7.09e-01	8.27e-01
8.97	1	6.67e-01	6.67e-01	7.00e-01	6.67e-01	7.73e-01
11.36	1	6.19e-01	6.19e-01	6.43e-01	6.19e-01	7.04e-01
12.70	1	5.73e-01	5.73e-01	5.88e-01	5.73e-01	6.38e-01
12.94	1	5.18e-01	5.16e-01	5.18e-01	5.18e-01	5.41e-01
13.21	0	5.18e-01	5.16e-01	5.18e-01	5.18e-01	5.41e-01
14.12	1	4.63e-01	4.44e-01	4.63e-01	4.32e-01	4.63e-01
17.59	1	4.06e-01	3.68e-01	4.06e-01	3.18e-01	4.06e-01
17.60	1	3.54e-01	3.02e-01	3.54e-01	2.29e-01	3.54e-01
24.45	1	3.00e-01	2.38e-01	3.00e-01	1.51e-01	3.00e-01
26.67	0	3.00e-01	2.38e-01	3.00e-01	1.51e-01	3.00e-01
29.89	1	2.37e-01	1.62e-01	2.37e-01	6.85e-02	2.37e-01
38.34	0	2.37e-01	1.62e-01	2.37e-01	6.85e-02	2.37e-01
38.63	1	1.70e-01	9.18e-02	1.70e-01	1.86e-02	1.70e-01
48.98	1	9.63e-02	2.67e-02	9.63e-02	3.90e-04	9.63e-02
58.18	1	4.27e-02	4.08e-03	4.27e-02	1.94e-07	4.27e-02
65.54	0	4.27e-02	4.08e-03	4.27e-02	1.94e-07	4.27e-02
65.65	1	1.72e-02	5.08e-04	1.72e-02	3.89e-11	1.72e-02
71.61	0	1.72e-02	5.08e-04	1.72e-02	3.89e-11	1.72e-02
79.96	0	1.72e-02	5.08e-04	1.72e-02	3.89e-11	1.72e-02
219.82	1	2.60e-03	1.09e-05	2.60e-03	1.06e-17	2.60e-03
244.93	1	0	0	0	0	0

Table 4.15: Comparison of standard PH survival estimates and envelope imprecise survival estimates for individuals with $x = 9$. The estimates were derived from the RE model applied to the data set in Example 4.4.1, with $n = 30$ and $\hat{\beta} = -0.5145$, using ϵ^{**} values of 0.5 and 1.

time	status	$\hat{S}_{15;c}(t)$	$\epsilon^{**} = 0.5$		$\epsilon^{**} = 1$	
			$\underline{\hat{S}}_{15;p}(t)$	$\bar{\hat{S}}_{15;p}(t)$	$\underline{\hat{S}}_{15;p}(t)$	$\bar{\hat{S}}_{15;p}(t)$
1.50	1	9.99e-01	9.99e-01	1	9.99e-01	1
1.71	1	9.98e-01	9.98e-01	1	9.98e-01	1
2.01	1	9.97e-01	9.97e-01	1	9.97e-01	1
3.14	1	9.95e-01	9.95e-01	9.99e-01	9.95e-01	1
3.77	1	9.94e-01	9.94e-01	9.99e-01	9.94e-01	1
4.40	1	9.92e-01	9.92e-01	9.99e-01	9.92e-01	1
4.51	1	9.89e-01	9.89e-01	9.98e-01	9.89e-01	1
7.13	1	9.87e-01	9.87e-01	9.97e-01	9.87e-01	1
8.51	1	9.84e-01	9.84e-01	9.97e-01	9.84e-01	1
8.97	1	9.82e-01	9.82e-01	9.96e-01	9.82e-01	1
11.36	1	9.78e-01	9.78e-01	9.95e-01	9.78e-01	1
12.70	1	9.75e-01	9.75e-01	9.94e-01	9.75e-01	1
12.94	1	9.70e-01	9.70e-01	9.92e-01	9.70e-01	9.99e-01
13.21	0	9.70e-01	9.70e-01	9.92e-01	9.70e-01	9.99e-01
14.12	1	9.66e-01	9.66e-01	9.91e-01	9.66e-01	9.99e-01
17.59	1	9.60e-01	9.60e-01	9.89e-01	9.60e-01	9.99e-01
17.60	1	9.54e-01	9.54e-01	9.86e-01	9.54e-01	9.99e-01
24.45	1	9.47e-01	9.47e-01	9.84e-01	9.47e-01	9.98e-01
26.67	0	9.47e-01	9.47e-01	9.84e-01	9.47e-01	9.98e-01
29.89	1	9.36e-01	9.36e-01	9.79e-01	9.36e-01	9.98e-01
38.34	0	9.36e-01	9.36e-01	9.79e-01	9.36e-01	9.98e-01
38.63	1	9.22e-01	9.22e-01	9.73e-01	9.22e-01	9.96e-01
48.98	1	8.99e-01	8.99e-01	9.59e-01	8.99e-01	9.93e-01
58.18	1	8.66e-01	8.66e-01	9.39e-01	8.66e-01	9.86e-01
65.54	0	8.66e-01	8.66e-01	9.39e-01	8.66e-01	9.86e-01
65.65	1	8.31e-01	8.31e-01	9.17e-01	8.31e-01	9.78e-01
71.61	0	8.31e-01	8.31e-01	9.17e-01	8.31e-01	9.78e-01
79.96	0	8.31e-01	8.31e-01	9.17e-01	8.31e-01	9.78e-01
219.82	1	7.62e-01	7.62e-01	8.78e-01	7.62e-01	9.65e-01
244.93	1	0	0	0	0	0

Table 4.16: Comparison of standard PH survival estimates and envelope imprecise survival estimates for individuals with $x = 15$. The estimates were derived from the RE model applied to the data set in Example 4.4.1, with $n = 30$ and $\hat{\beta} = -0.5145$, using ϵ^{**} values of 0.5 and 1.

upper and lower envelope-type survival estimates expand, reflecting an inevitable consequence of the constraints embedded within the envelope estimation methodology. Regarding the relationship between these disparities and the covariate effect, which is one of the aspects of the robust model, a consistent pattern is observed. This relationship is determined by the covariate effect, particularly the value of the covariate and the sign regression coefficient. For instance, when $x = 5$ and $\hat{\beta} = -0.5145$, the difference between imprecise survival estimates appears to decrease over time since it is related to the maximum covariate effect. This trend diminishes and then reverses as the covariate effect decreases and reaches a minimum value, as illustrated when $x = 9, 15$.

◇

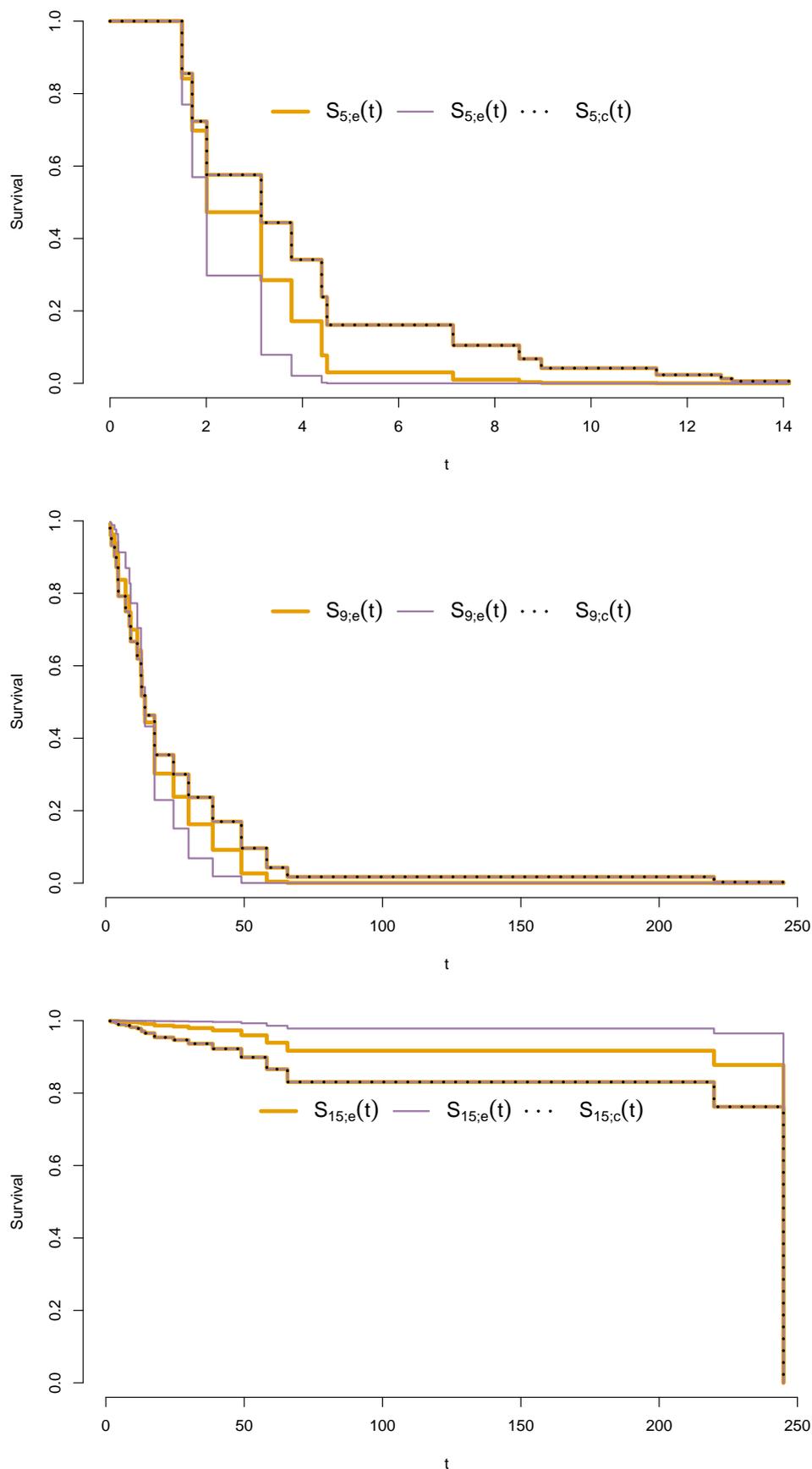


Figure 4.5: Comparison of standard PH survival estimates and envelope imprecise survival estimates for $x = 5, 9, 15$ (top-bottom). These estimates were derived from the RE model applied to the data set in Example 4.4.1, using ϵ^{**} values of 0.5 (amber), and 1 (indigo).

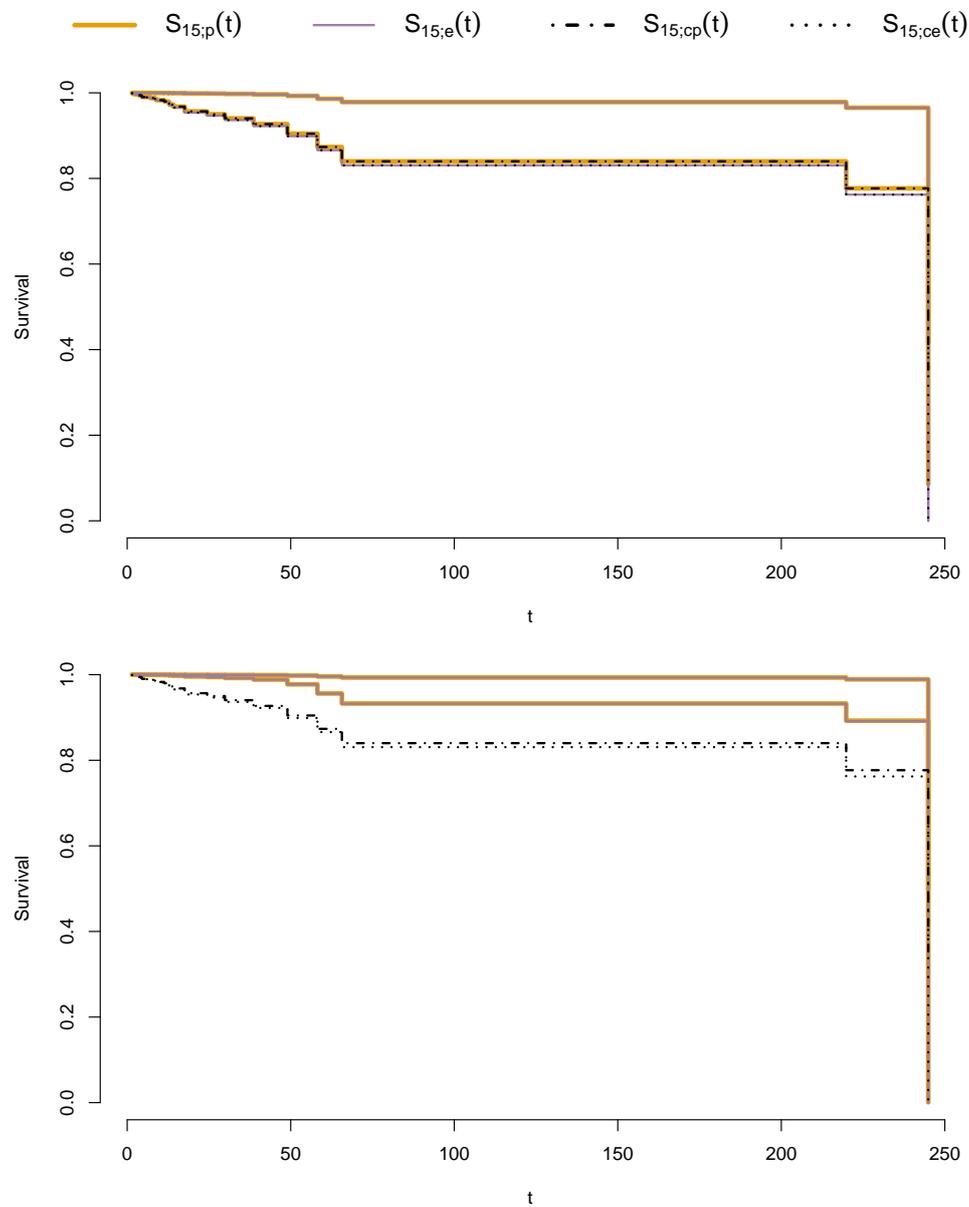


Figure 4.6: Comparison of envelope imprecise survival estimates (top) and naive imprecise survival estimates (bottom) for individuals with $x = 15$. These estimates were derived from both the RP model (amber), and the RE model (indigo) applied to the data set in Examples 4.4.1 and 4.3.2, using $\epsilon^{**} = \epsilon^* = 1$.

time	status	$\overline{\hat{S}}_{5;e}(t) - \underline{\hat{S}}_{5;e}(t)$		$\overline{\hat{S}}_{9;e}(t) - \underline{\hat{S}}_{9;e}(t)$		$\overline{\hat{S}}_{15;e}(t) - \underline{\hat{S}}_{15;e}(t)$	
		$\epsilon^{**} = 0.5$	1	0.5	1	0.5	1
1.50	1	1.41e-02	8.58e-02	1.10e-02	1.73e-02	8.09e-04	9.06e-04
1.71	1	2.56e-02	1.54e-01	2.24e-02	3.52e-02	1.68e-03	1.88e-03
2.01	1	1.03e-01	2.78e-01	3.08e-02	5.67e-02	2.78e-03	3.20e-03
3.14	1	1.59e-01	3.65e-01	3.69e-02	7.49e-02	4.00e-03	4.70e-03
3.77	1	1.70e-01	3.21e-01	4.27e-02	9.23e-02	5.22e-03	6.21e-03
4.40	1	1.62e-01	2.37e-01	4.53e-02	1.10e-01	6.83e-03	8.26e-03
4.51	1	1.31e-01	1.61e-01	4.53e-02	1.21e-01	8.55e-03	1.05e-02
7.13	1	9.49e-02	1.05e-01	4.21e-02	1.20e-01	1.04e-02	1.29e-02
8.51	1	6.44e-02	6.76e-02	3.89e-02	1.18e-01	1.23e-02	1.54e-02
8.97	1	4.10e-02	4.19e-02	3.28e-02	1.06e-01	1.42e-02	1.81e-02
11.36	1	2.33e-02	2.34e-02	2.36e-02	8.51e-02	1.66e-02	2.13e-02
12.70	1	1.28e-02	1.28e-02	1.49e-02	6.50e-02	1.90e-02	2.47e-02
12.94	1	5.78e-03	5.78e-03	1.45e-03	2.35e-02	2.21e-02	2.90e-02
13.21	0	5.78e-03	5.78e-03	1.45e-03	2.35e-02	2.21e-02	2.90e-02
14.12	1	2.42e-03	2.42e-03	1.98e-02	3.09e-02	2.52e-02	3.37e-02
17.59	1	8.60e-04	8.60e-04	3.82e-02	8.80e-02	2.89e-02	3.92e-02
17.60	1	2.94e-04	2.94e-04	5.17e-02	1.25e-01	3.27e-02	4.49e-02
24.45	1	8.12e-05	8.12e-05	6.20e-02	1.50e-01	3.72e-02	5.17e-02
26.67	0	8.12e-05	8.12e-05	6.20e-02	1.50e-01	3.72e-02	5.17e-02
29.89	1	1.26e-05	1.26e-05	7.43e-02	1.68e-01	4.31e-02	6.12e-02
38.34	0	1.26e-05	1.26e-05	7.43e-02	1.68e-01	4.31e-02	6.12e-02
38.63	1	9.30e-07	9.30e-07	7.80e-02	1.51e-01	5.08e-02	7.41e-02
48.98	1	1.10e-08	1.10e-08	6.96e-02	9.60e-02	6.07e-02	9.41e-02
58.18	1	1.88e-11	1.88e-11	3.86e-02	4.27e-02	7.31e-02	1.20e-01
65.54	0	1.88e-11	1.88e-11	3.86e-02	4.27e-02	7.31e-02	1.20e-01
65.65	1	1.49e-14	1.49e-14	1.67e-02	1.72e-02	8.63e-02	1.48e-01
71.61	0	1.49e-14	1.49e-14	1.67e-02	1.72e-02	8.63e-02	1.48e-01
79.96	0	1.49e-14	1.49e-14	1.67e-02	1.72e-02	8.63e-02	1.48e-01
219.82	1	5.77e-21	5.77e-21	2.59e-03	2.60e-03	1.15e-01	2.03e-01
244.93	1	0.00e+0	0.00e+0	0.00e+0	0.00e+0	0.00e+0	0.00e+0

Table 4.17: Comparison of the width of envelope survival estimates for individuals with $x = 5, 9, 15$ based on the RE model applied to the data set in Example 4.4.1, with $n = 30$ and $\hat{\beta} = -0.5145$, using ϵ^{**} values of 0.5 and 1.

4.5 Simulations

In this section, we delve into a comprehensive simulation study aimed at providing valuable insights into the behavior and performance of RP survival models under various scenarios, thereby facilitating their wide adoption and improving the reliability of survival analysis in practical settings. The simulations examined two specific aspects in detail. First, exploring one of the key issues of the RP model, namely the failure to incorporate the PH survival function into the range of naive imprecise survival functions. The second objective is to evaluate the performance of RP models compared to traditional PH models in terms of estimation performance when covariate measurement error is considered.

4.5.1 Inconsistency of the naive imprecise survival functions

This section undertakes an investigation of the inconsistency observed in the behaviour of naive imprecise survival functions. Specifically, it is our intention to determine whether or not the failure to include the PH survival function within the range of imprecise survival functions derived from the robust model is a fundamental characteristic of the naive imprecise survival estimates. While the simulations herein utilize the robust model based on the Poisson likelihood, we anticipate that the conclusions drawn from comparing the results with naive survival estimates based on empirical likelihood will yield similar outcomes, especially for datasets of moderate to large sizes. This simulation could also provide further understanding of how PH survival estimates for individuals fluctuate within their respective envelope-type of imprecise survival estimates according to the sign of the regression coefficient and the values of x . Three distinct scenarios will be considered to gain a comprehensive insight.

Scenario 1

The purpose of the scenario is to examine the impact of the regression coefficient to the occurrence of this phenomenon. We generate $m = 100$ PH survival data of size $n = 600$ with 20% of observations right-censored, as discussed in Section 2.6. The

Weibull distribution is assumed for the baseline survival times with shape parameter $\rho = 1.5$ and scale parameter $\lambda = 0.2$, based on the corresponding functions in Table 2.1. The survival times for individuals are dependent on the pre-specified distribution of a single time-independent covariate $X \sim \mathcal{N}(10, 3^2)$ and different values of the regression coefficient $\beta = \{-1, -0.5, -0.001, 0.001, 0.5, 1\}$. The PH survival function along with the imprecise survival functions based on the level of imprecision $\epsilon^* = 1$ will be estimated for individuals with $x = 5, 10, 15$. For each of the 100 survival data the following steps will be followed:

1. Estimates the survival functions for individuals with $x = 5, 10, 15$
2. Partition the survival time into early, middle, and late intervals, which are denoted by $t\{1\}$, $t\{2\}$, $t\{3\}$, respectively.
3. For each interval, compute the probability in which the PH survival estimates are less than the lower survival estimates, $\underline{S}_x(t\{i\})^- = P(S_{x;c}(t\{i\}) < \underline{S}_{x;p}(t\{i\}))$, and probability that the PH survival estimates are greater than the upper survival estimates, $\overline{S}_x(t\{i\})^+ = P(S_{x;c}(t\{i\}) > \overline{S}_{x;p}(t\{i\}))$, with $i = \{1, 2, 3\}$ and $x = \{5, 10, 15\}$. For instance, the probability that the PH survival estimates for individuals with $x = 5$ are less than the lower imprecise survival estimates at early times is denoted by $\underline{S}_5(t\{1\})^-$. This probability represents the average of the PH survival estimates that are below the lower imprecise survival estimates at each time within the first interval, $t\{1\}$. Similarly, one can determine, $\overline{S}_5(t\{1\})^+$, the PH survival estimates for individuals with $x = 5$ that exceeding the upper imprecise survival estimates at early times.
4. The distributions of these probabilities will be shown in boxplot figures for each value of the regression coefficient $\beta = \{-1, -0.5, -0.001, 0.001, 0.5, 1\}$.

Figure 4.7 illustrates the probabilities in which the naive imprecise survival estimates fail to encompass the PH survival estimates at early, middle, and late times. Figure 4.7 (left) presents the probabilities related to the negative regression coefficients $\beta = \{-0.001, -0.5, -1\}$, respectively from top to bottom. On the other hand, probabilities related to the positive regression coefficients $\beta = \{0.001, 0.5, 1\}$,

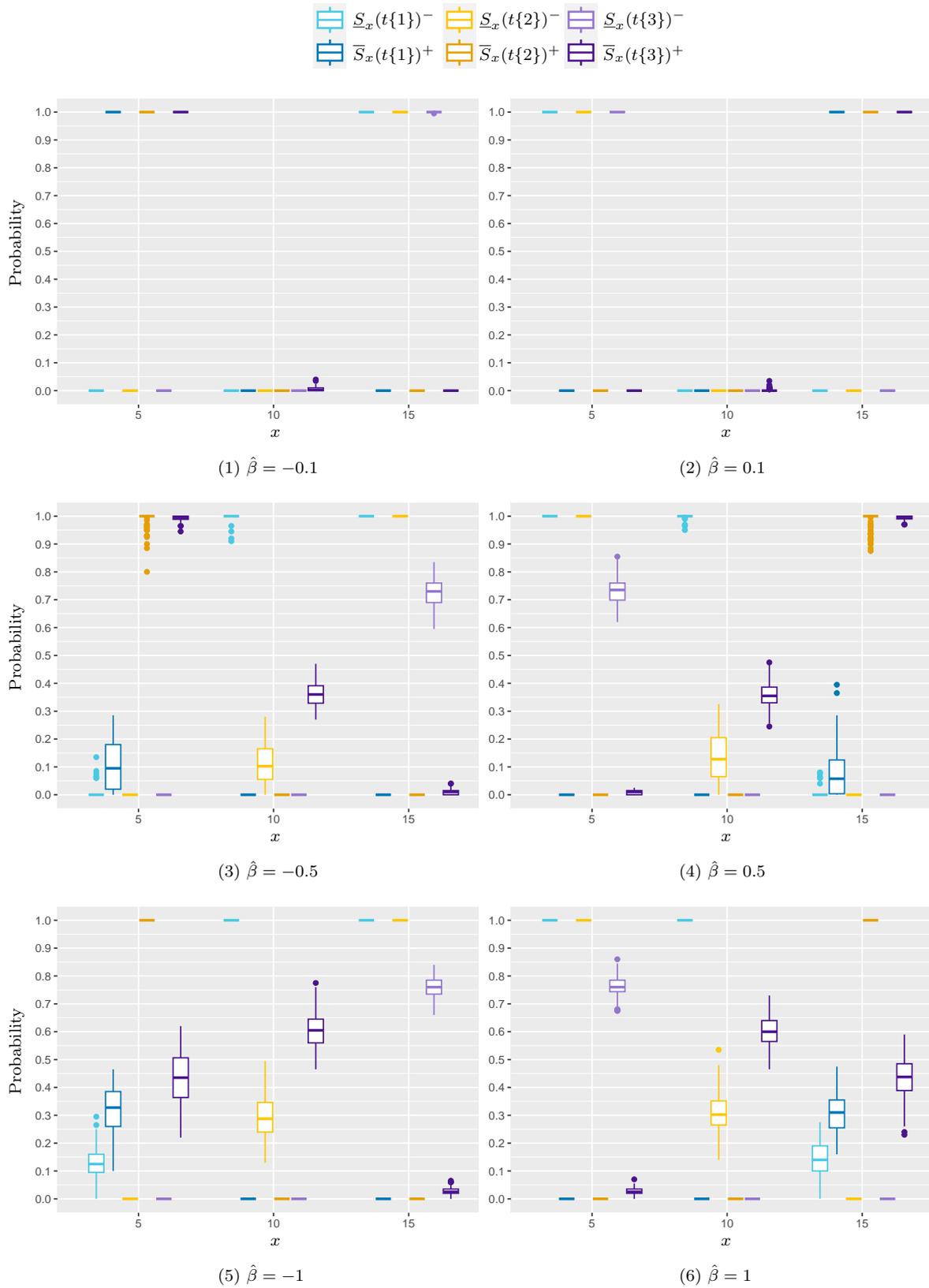


Figure 4.7: Scenario 1: Probabilities of the PH survival estimates deviating from the imprecise survival estimates based on the RP likelihood for $m=100$ datasets each of size 300 using $\epsilon^* = 1$.

are shown in Figure 4.7 (right) from top to bottom, respectively. The simulation results confirm what has been observed in Example 4.3.2.

The duration during which the PH survival estimates are enclosed within the corresponding imprecise survival estimates varies according to the covariate effect, $\hat{\beta}x$. That is the PH survival estimates for individuals with moderate covariate effects tend to fall within their imprecise survival estimates for all values of the regression parameter, as seen for $x = 10$ in Figure 4.7. Clearly, this result is more pronounced at very small regression coefficients as illustrated in Figure 4.7 (1,2), but it is also true at other values of the regression coefficients with some fluctuation where the imprecise survival estimates do not include the PH survival estimates, as in Figure 4.7 (3-6).

The PH survival estimates for individuals with highest covariate effects, i.e. when $x = 5$ for negative regression coefficients Figure 4.7 (1) and when $x = 15$ for positive regression coefficients Figure 4.7 (2), are more likely to lie above the imprecise survival estimates. However, as the regressing coefficients deviate from zero the imprecise survival estimates become able to capture the PH survival estimates, beginning at early time then fluctuate as shown in Figure 4.7 (3, 5) for $x = 5$, and Figure 4.7 (4, 6) for $x = 15$.

For individuals with lowest covariate effects, i.e. when $x = 15$ for negative regression coefficients Figure 4.7 (1) and when $x = 5$ for positive regression coefficients Figure 4.7 (2), the PH survival estimates are more likely to lie below the imprecise survival estimates. The imprecise survival estimates start to capture the PH survival estimates at late times when the regressing coefficients deviate from zero. This occurrence is evident in Figure 4.7 (3,5) for $x = 15$, and Figure 4.7 (4,6) for $x = 5$. Simulating data using other settings and distributions revealed similar patterns.

Scenario 2

This scenario investigates the impact of other factors on enclosing the PH survival estimates with the imprecise survival estimates. These factors include sample size, the proportion of right-censored observations, and the level of imprecision. A total of $m = 1000$ datasets will be generated under different configurations. The simulation

involves two proportions of right-censoring, specifically, specifically 20% and 50%, were examined, with sample sizes of $n = 60$ and 150. Furthermore, the imprecise survival functions will be estimated using two levels of imprecision, $\epsilon^* = 0.05$ and 0.1. Instead of the Weibull distribution, the Gompertz distribution was employed for the baseline survival times, characterized by a shape parameter $\rho = 0.05$ and scale parameter $\lambda = 0.1$, based on the corresponding functions described in Table 2.1. The covariate was assumed to follow the uniform distribution with values ranging between 0 and 1 with the regression coefficient $\beta = -1$.

Similar to the preceding scenario, the probabilities of the PH survival estimates deviating from the imprecise survival estimates were computed over $t\{1\}$, $t\{2\}$, and $t\{3\}$, for $x = 0.2, 0.5, 0.8$ for each of the 1000 data sets in every configuration. Figures 4.8 and 4.9 display the probabilities that the naive imprecise survival estimates fail to encompass the PH survival estimates at early, middle, and late times for the simulated data, with sample sizes $n = 60$ and 150, respectively. In both figures, it is evident that a higher censoring proportion increases the probabilities for the PH survival estimates to be encompassed within the corresponding imprecise survival estimates for individuals with moderate covariate effect, while reducing the inclusion probabilities for individuals with high and low moderate covariate effect, as evidenced by comparisons between the (1,2) and (3,4) in both figures. This phenomenon could be attributed to the fact that the baseline hazard function is only computed at event time and otherwise is zero. Consequently, given that half of the observations are right-censored, it reduces the imprecision of the baseline hazard estimates, leading to narrower imprecision between the upper and lower survival estimates. Regarding the impact of increasing the imprecision level, as anticipated, it gradually diminishes the deviation of PH survival estimates from the imprecise survival estimates. This is observed by comparing the (1,3) and the (2,4) of these figures. Based on a comparison between Figures 4.8 and 4.9, it is apparent that for larger datasets, the imprecision level has less influence on reducing deviation probabilities. Indeed, based on all simulations conducted, although not all are reported, it is observed that the PH survival estimates for individuals with higher covariate effects are more likely to exceed the imprecise survival estimates at very low im-

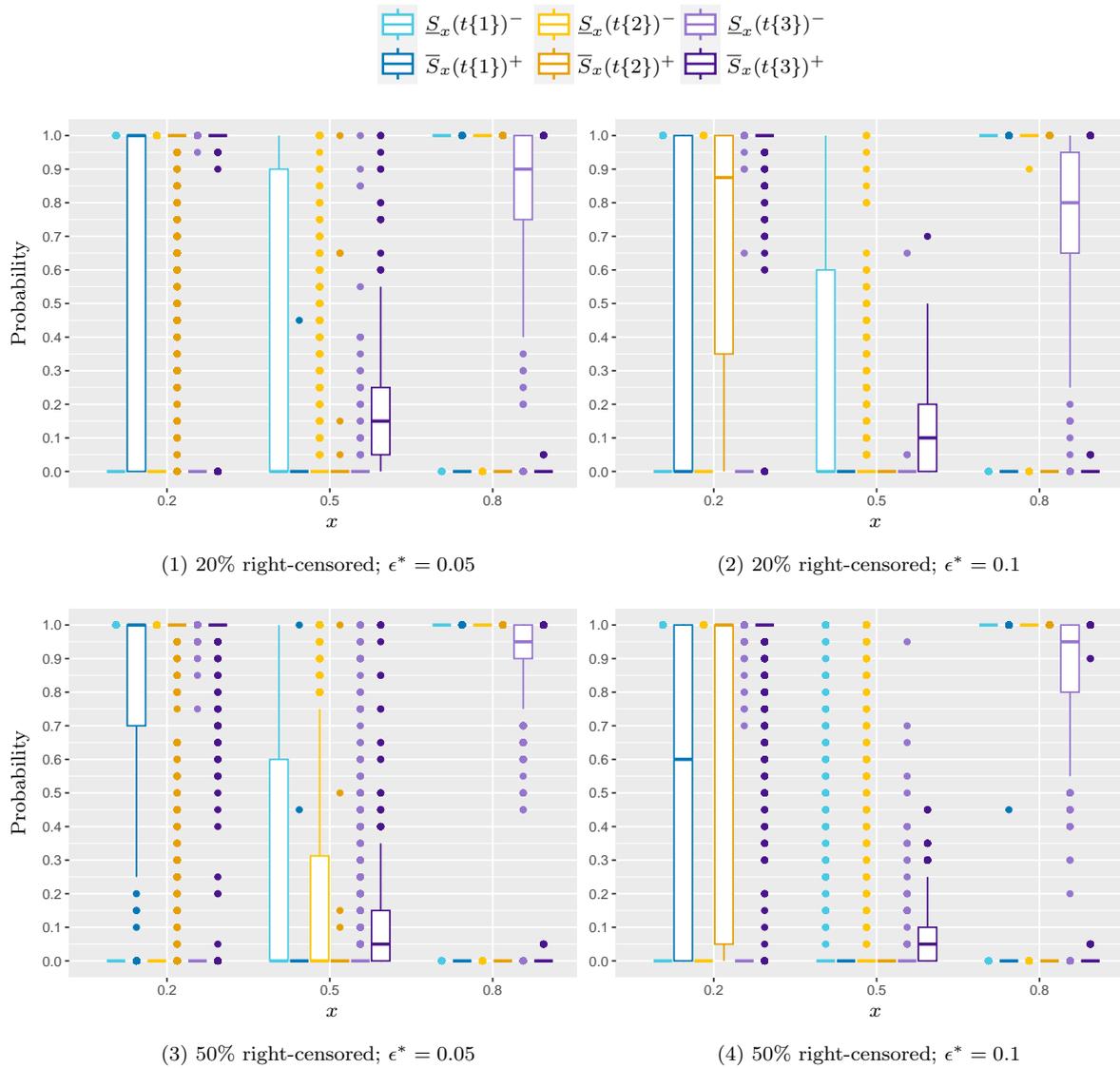


Figure 4.8: Scenario 2: Probabilities of the PH survival estimates deviating from the imprecise survival estimates based on the RP likelihood for $m = 1000$ datasets each of size 60 with $\beta = -1$.

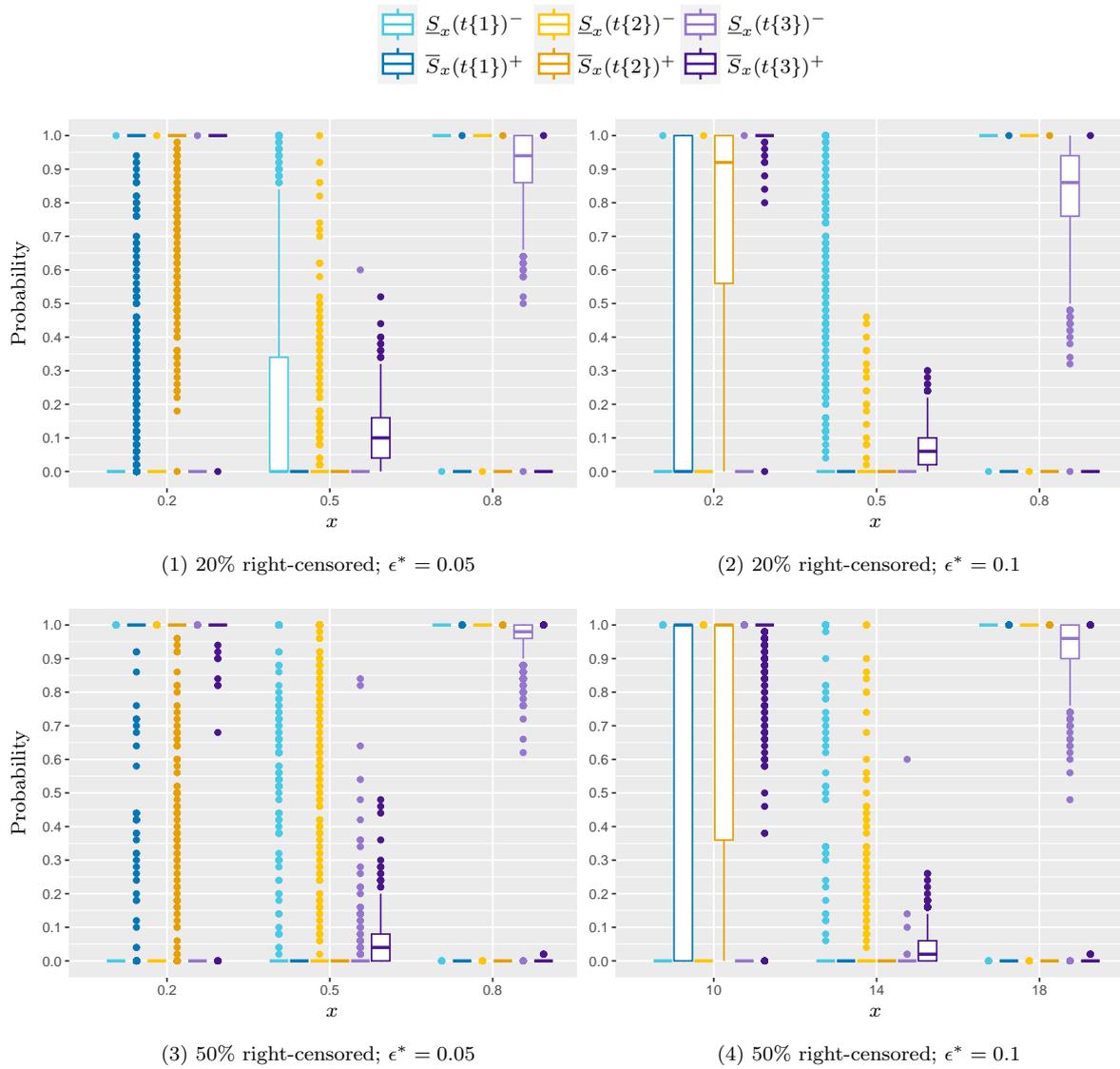


Figure 4.9: Scenario 2: Probabilities of the PH survival estimates deviating from the imprecise survival estimates based on the RP likelihood for $m = 1000$ datasets each of size 150 with $\beta = -1$

precision levels. Conversely, the PH survival estimates for individuals with lower covariate effects typically fall below the imprecise survival estimates. If imprecision is increased, the probability that naive imprecise survival estimates include PH survival estimates increases at early times when covariate effects are higher and later when covariate effects are lower. However, this trend fluctuates in response to the increased level of imprecision.

Scenario 3

In this scenario, we explore whether different behaviour of the naive upper and lower survival estimates can be observed when fitting the robust PH model to non-proportional hazards survival data. Therefore, we generate 100 non-PH survival data, attributed to frailty effects, with each data set containing $n = 600$ observations, of which 20% are subject to right-censoring. The exponential distribution is assumed for baseline survival times, with a fixed hazard rate parameter of $\lambda = 2$, as detailed in Table 2.1. Additionally, a covariate $X \sim \mathcal{N}(14, 3^2)$ is considered. The average treatment effect across all institutions is specified as $\beta = -1.2$. The frailty effect is introduced to reflect variations across 5 distinct institutions or clusters, each consisting of 120 observations. This effect represents institution-specific deviations from the average treatment effect, and is drawn from a normal distribution with a mean of $\mu = 0$ and standard deviation $\sigma = 3$. Subsequently, PH survival functions are estimated alongside imprecise survival functions for individuals with $x = 10, 14, 18$, using the imprecision levels $\epsilon^* = 0.8$ and 1.5.

Figure 4.8 illustrates the probabilities that the naive imprecise survival estimates fail to encompass the PH survival estimates, using $\epsilon^* = 0.8$ (left) and $\epsilon^* = 1.5$ (right). Nevertheless, similar trends were found regarding the deviation of the PH survival estimates from the naive survival estimates. This confirms the influence of the covariate effect on the pattern where the naive survival estimates fail to encompass the PH survival estimates.

In light of the results obtained from the previous three scenarios, these results can be relevant to imprecise survival estimates based on the envelope approach. As opposed to focusing on the time periods during which PH survival estimates

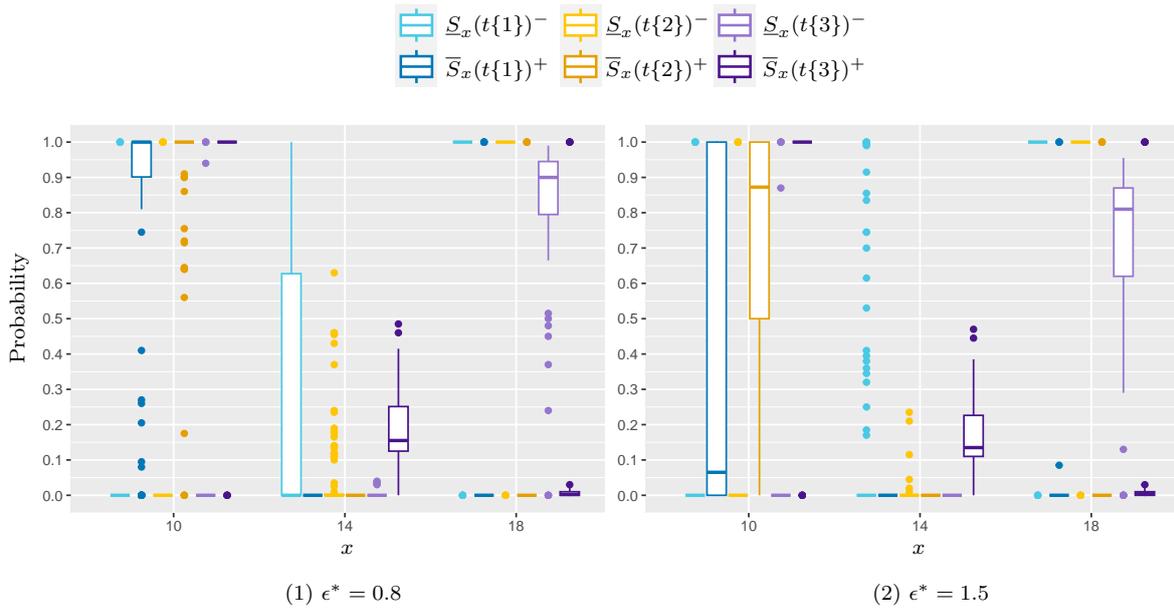


Figure 4.10: Scenario 3: Probabilities of the PH survival estimates deviating from the imprecise survival estimates based on the RP likelihood for $m = 100$ datasets each of size 600 based on NPH data (frailty)

exceed or fall below their corresponding envelope-type of survival estimates, this results should provide insight into the relative positioning of PH survival estimates within their imprecise counterparts. Specifically, PH survival estimates are shown to reside either at their upper survival estimates for individuals characterized by higher covariate effects, at the lower survival estimates for those with lower covariate effects, or fluctuate within the corresponding imprecise survival estimates for individuals with moderate covariate effects.

4.5.2 Robust PH model with covariate measurement error

The introductory section of this chapter provides a comprehensive overview of the diverse range of methodologies and solutions found in the literature addressing covariate measurement errors in PH models. Despite the primary objective of this chapter being to mitigate the PH assumption by incorporating errors into covariate values, a particular focus is directed towards exploring the estimation capabilities of the RP and RE models when compared to the standard PH model in the context of covariate measurement errors. As part of our simulations, we carefully examined two

scenarios: one in which covariate measurement errors are present and one without covariate measurement errors. As a result of these scenarios, we hope to provide insight into the relative effectiveness of the robust PH model at different levels of covariate measurement accuracy.

Scenario 1

This scenario assesses the estimation of the regression coefficient in the robust PH model compared to the standard PH model, using survival data free from covariate measurement errors. A total of 1000 survival data have been generated assuming the Weibull distribution with shape $\rho = 1.2$, and scale $\lambda = 2$ for baseline survival times. To maintain simplicity, we consider a single time-independent covariate following a uniform distribution over the interval $[10, 30]$, with a regression coefficient of $\beta = -0.5$. Additionally, various sample sizes were considered, including $n = 30, 60, 120$, with censorship rates of 20% and 50%. Based on the simulated data, the standard and robust PH models were fitted under each configuration.

Table 4.18 presents the simulation averages of the estimates $\hat{\beta}_p$ and $\hat{\beta}_e$, alongside their respective standard deviations, relative bias, and the number of non-convergence issues denoted as NA. Four levels of imprecision are employed in this simulation: $\epsilon^* = 0.001, 0.5, 1, 1.4$. The relative bias, R.B, is computed as $|(\hat{\beta} - \beta)/\beta|$, where β represents the assumed true value of the regression parameter. Ideally, the relative bias should be close to zero, indicating that the model is on average accurate in estimating the true parameter.

Table 4.19 suggests that the standard PH Model based on either Poisson or Empirical likelihoods, $\epsilon^* = 0$, demonstrates a consistent estimator with lower variability and relative bias aligned with the assumed true parameter. In comparison, the robust PH model exhibits an increase in variability and relative bias of parameter estimates with increasing imprecision, particularly for the RP likelihood. This suggests that the RE method is less affected by the introduced imprecision. Notably, the robust PH model, based on both likelihoods, exhibits non-convergences during optimization due to the infinite summation of log likelihood functions resulting from

Cens %		n=30							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.5322	0.1268	0.0644	0	-0.4951	0.1217	0.0099	0	
0.001	-0.5325	0.1270	0.0651	0	-0.5253	0.1229	0.0507	2	
0.5	-0.7348	0.2469	0.4696	0	-0.7223	0.2687	0.4445	13	
1	-1.1215	0.7773	1.2430	1	-0.9890	1.7303	0.9779	54	
1.4	-1.8588	3.7938	2.7176	9	-1.2385	4.5507	1.4770	155	
		n=60							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.5088	0.0746	0.0176	0	-0.5056	0.0744	0.0111	0	
0.001	-0.5090	0.0747	0.0181	0	-0.5070	0.0906	0.0139	0	
0.5	-0.6726	0.1283	0.3452	0	-0.6668	0.1260	0.3335	1	
1	-0.9296	0.2452	0.8593	0	-0.9111	0.2312	0.8222	17	
1.4	-1.2604	0.4855	1.5209	0	-1.1898	0.4438	1.3796	105	
		n=120							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.5023	0.0483	0.0046	0	-0.5003	0.0482	0.0005	0	
0.001	-0.5025	0.0484	0.0051	0	-0.5005	0.0482	0.0010	0	
0.5	-0.6510	0.0775	0.3021	0	-0.6482	0.0772	0.2964	0	
1	-0.8723	0.1337	0.7446	0	-0.8674	0.1307	0.7349	2	
1.4	-1.1384	0.2220	1.2769	0	-1.1174	0.2283	1.2348	45	
		n=30							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.5624	0.2155	0.1248	0	-0.5516	0.1936	0.1031	0	
0.001	-0.5628	0.2160	0.1255	0	-0.5983	1.0547	0.1965	3	
0.5	-0.8382	1.0287	0.6764	2	-0.7323	2.0666	0.4645	20	
1	-1.5939	3.9721	2.1877	19	-1.2304	2.7029	1.4608	108	
1.4	-3.2603	7.8535	5.5205	41	-1.5656	3.7765	2.1311	208	
		n=60							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.5196	0.1023	0.0393	0	-0.5158	0.1022	0.0316	0	
0.001	-0.5199	0.1024	0.0398	0	-0.5161	0.1023	0.0322	0	
0.5	-0.6902	0.1846	0.3803	0	-0.6805	0.1721	0.3611	6	
1	-0.9694	0.3961	0.9388	0	-0.9511	0.4524	0.9023	33	
1.4	-1.3382	0.7580	1.6764	2	-1.2214	0.7062	1.4429	117	
		n=120							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.5041	0.0604	0.0082	0	-0.5019	0.0603	0.0039	0	
0.001	-0.5043	0.0605	0.0087	0	-0.5022	0.0604	0.0043	0	
0.5	-0.6503	0.0972	0.3006	0	-0.6473	0.0970	0.2945	0	
1	-0.8676	0.1687	0.7353	0	-0.8614	0.1673	0.7228	4	
1.4	-1.1298	0.2840	1.2596	0	-1.1104	0.3247	1.2208	54	

Table 4.18: Scenario 1: Comparison between the robust PH estimates based on the RP and RE likelihoods, using $\epsilon^* = 0.001, 0.5, 1, 1.4$, verses the standard PH estimates, $\epsilon^* = 0$, applied to PH survival data the absence of measurement error.

increasing the imprecision level. The presence of non-convergence in the robust PH model can be perceived as a valuable feature rather than a drawback. This occurrence serves to highlight instances in which the standard PH model offers a superior fit to the data set, thus indicating the need for a potential reduction in level of imprecision or a consideration of the standard PH model. As expected, estimates are more stable when the sample size is larger. Despite its limited effect on the mean of estimates, censoring proportion has a significant impact on the mean of standard deviations and the occurrence of convergence issues. Other simulations were conducted using different distributions for the baseline and the covariate and yielded similar outcomes to those presented.

Scenario 2: In this scenario, we are inspecting the impact of estimating the regression coefficient under the standard PH model in the presence of the covariate measurement errors. Therefore, we consider the same data sets as used in the previous scenario, but standard additive errors, τ_i 's, were added to the covariate value of each individual as measurement errors. These measurement errors are assumed to be normally distributed with mean $\mu_\tau = 0$ and standard deviation $\sigma_\tau = 2$. The simulation mean of the estimators, the standard deviations, the relative biases, and the number of non-convergence issues, are presented in Table 4.19.

The presence of measurement errors in the covariates can result in potential bias when estimating the PH model parameter as discussed in the literature when using the standard PH model [5, 8]. Based on Poisson and Empirical likelihoods, Table 4.19 illustrates that the classical PH model with $\epsilon^* = 0$ underestimates the true parameter $\beta = -0.5$ due to measurement errors in the covariate. Despite the large sample size of a data sets, the bias of estimates can still exist. Other simulations, not presented here, have shown that the bias can become even more pronounced when the true parameter deviates more from zero or when $\sigma_\tau > 2$.

The robust PH model, in contrast, exhibits an attractive attributes in accommodating the covariate measurement errors according on the selected level of imprecision. While the estimates obtained from fitting the robust PH model reveal an increase in variability as epsilon increases, the estimates are progressively improved

Cens %		n=30							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.3476	0.0847	0.3047	0	-0.3431	0.0845	0.3137	0	
0.001	-0.3478	0.0848	0.3044	0	-0.3433	0.0846	0.3134	0	
0.5	-0.4326	0.1244	0.1348	0	-0.4269	0.1237	0.1462	0	
1	-0.5566	0.2098	0.1132	0	-0.5451	0.1877	0.0902	3	
1.4	-0.7040	0.3456	0.4079	0	-0.6435	1.2890	0.2870	11	
		n=60							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.3248	0.0488	0.3504	0	-0.3250	0.0911	0.3499	0	
0.001	-0.3249	0.0489	0.3502	0	-0.3227	0.0488	0.3545	0	
0.5	-0.3927	0.0671	0.2146	0	-0.3899	0.0669	0.2202	0	
1	-0.4834	0.0971	0.0333	0	-0.4797	0.0966	0.0405	0	
1.4	-0.5817	0.1371	0.1634	0	-0.5777	0.1398	0.1553	0	
		n=120							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.3198	0.0343	0.3604	0	-0.3185	0.0343	0.3629	0	
0.001	-0.3199	0.0343	0.3602	0	-0.3186	0.0342	0.3628	0	
0.5	-0.3821	0.0466	0.2359	0	-0.3804	0.0464	0.2392	0	
1	-0.4623	0.0657	0.0754	0	-0.4601	0.0653	0.0797	0	
1.4	-0.5453	0.0886	0.0906	0	-0.5431	0.0888	0.0862	0	
		n=30							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.3718	0.1214	0.2564	0	-0.3668	0.1209	0.2664	0	
0.001	-0.3720	0.1215	0.2561	0	-0.3670	0.1210	0.2660	0	
0.5	-0.4706	0.2001	0.0588	0	-0.4615	0.2000	0.0770	4	
1	-0.6387	0.6130	0.2774	2	-0.5850	0.8212	0.1700	13	
1.4	-1.0533	3.4762	1.1067	4	-0.6831	0.8593	0.3661	43	
		n=60							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.3409	0.0668	0.3181	0	-0.3381	0.0668	0.3237	0	
0.001	-0.3411	0.0668	0.3179	0	-0.3383	0.0668	0.3235	0	
0.5	-0.4129	0.0932	0.1742	0	-0.4095	0.0932	0.1809	0	
1	-0.5110	0.1399	0.0221	0	-0.5059	0.1353	0.0118	1	
1.4	-0.6208	0.2079	0.2415	0	-0.6123	0.1982	0.2246	4	
		n=120							
ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA	
0	-0.3303	0.0443	0.3394	0	-0.3288	0.0443	0.3424	0	
0.001	-0.3304	0.0444	0.3392	0	-0.3290	0.0443	0.3420	0	
0.5	-0.3934	0.0605	0.2132	0	-0.3917	0.0604	0.2166	0	
1	-0.4749	0.0861	0.0502	0	-0.4728	0.0858	0.0543	1	
1.4	-0.5603	0.1188	0.1205	0	-0.5579	0.1185	0.1157	2	

Table 4.19: Scenario 2: Comparison between the robust PH estimates based on the RP and RE likelihoods, using $\epsilon^* = 0.001, 0.5, 1, 1.4$, versus the standard PH estimates, $\epsilon^* = 0$, applied to PH survival data with measurement errors $\tau \sim \mathcal{N}(0, 2^2)$.

ϵ^*	$\hat{\beta}_p$	SD	R.B	NA	$\hat{\beta}_e$	SD	R.B	NA
0	-0.4045	0.0609	0.4944	0	-0.4025	0.0607	0.4969	0
0.001	-0.4047	0.0609	0.4942	0	-0.4046	0.0607	0.4967	0
0.5	-0.5073	0.0920	0.3659	0	-0.5045	0.0916	0.3694	0
1	-0.6587	0.1503	0.1766	0	-0.6545	0.1495	0.1819	1
1.4	-0.8424	0.2400	0.0530	0	-0.8128	0.6802	0.0160	16

Table 4.20: Scenario 2: Comparison between the robust PH estimates based on the RP and RE likelihoods, $\epsilon^* = 0.001, 0.5, 1, 1.4$, verses the standard PH estimates, $\epsilon^* = 0$, applied to PH survival data with measurement errors $\tau \sim \mathcal{N}(0, 2^2)$.

by moving closer to the true value and the bias diminishes considerably. In the context of this specific simulation study, Table 4.19 indicates that for small data sets the optimal value of the imprecision level falls within the range of 0.5 to 1, while an ϵ^* value ranging from approximately 1 to 1.4 seems to be suitable for large data sets as it leads to favorable results. Table 4.19 provides evidence that the incidence of non-convergence is significantly lower than Scenario 1, in which survival data were free of measurement errors. Further analysis of these non-convergence instances revealed that the estimated values exceeded the true parameter value, $\beta = -0.5$, at a lower level of imprecision.

Overall, the robust PH model faces challenges in reliability when confronted with increased variability and potential non-convergence issues at higher levels of imprecision. A comparison between the Poisson likelihood and the Empirical likelihood within the robust PH model suggests that the former yields more stable estimates and encounters fewer non-convergence issues, thereby indicating its potential robustness in handling measurement errors. As a consequence, it is crucial to exercise caution when selecting the imprecision level in the robust PH model. This underscores the importance of developing future methods to assist researchers choose an appropriate level of imprecision.

4.5.3 Bootstrap investigation

In response to the necessity of identifying the optimal level of imprecision in the robust PH model, bootstrap investigations were undertaken. The following

summarizes these investigations. Two types of survival datasets were investigated: survival data with a valid PH assumption, PH, and survival data that violated the PH assumption due to time-dependent covariates, NPH. A total of $M=100$ original survival datasets were studied for both PH and NPH scenarios. Each dataset consisted of $N = 60$ observations, with 20% of them being right-censored. The NPH data were generated with time-dependent effects formed by an interaction with log-time according to the *simsurv* function from the *simsurv* R package [14]. Various levels of imprecision, $\epsilon^* = 0.0001, 0.1, 0.5, 1, 2$, were considered, and two test statistics, $\hat{\ell}_\epsilon$ and $\hat{\ell}_\epsilon - \hat{\ell}_0$, were examined. Here, $\hat{\ell}_{\epsilon^*}$ and $\hat{\ell}_0$ represent the maximum log-likelihood values obtained by fitting the robust model with imprecision levels ϵ^* and zero, respectively. The methodology closely followed the description outlined in Section 3.4.1, where quantiles of the test statistics associated with the original datasets were derived from their corresponding bootstrap distributions.

Figure A.2 (top) shows the quantiles of $\hat{\ell}_\epsilon^*$ (left) and $(\hat{\ell}_\epsilon - \hat{\ell}_0)^*$ (right) for both PH and NPH data. In fact, the figure suggests that increasing the imprecision level does not lead to any significant impact. Further, other types of NPH data were examined by incorporating frailty effects based on distinct clusters, using the same R package. In spite of this, similar results were exhibited as shown in Figure A.2 (bottom).

Our analysis revealed Zelterman's bootstrap method suffers from a negative probability problem caused by Equation (2.70) for continuous covariates, which has been corrected by re-scaling the probability values. Accordingly, another bootstrap method developed by N.L. Hjort [66] was implemented, with the results provided in Figure A.3. Similarly, the bootstrap analysis failed to demonstrate any substantial impact resulting from increased levels of imprecision.

Comparable instances were encountered in Chapter 3 where bootstrap outcomes failed to reveal any substantial influence of increasing imprecision levels in the GPH model, except for larger datasets with $n > 500$. Hence, it is conceivable that exploring survival data with larger sample sizes may yield more noteworthy insights. Nevertheless, the GPH likelihood optimization focuses exclusively on the primary parameter, as the imprecision terms that maximize the likelihood function have

been analytically identified. Despite leveraging parallel computing capabilities for the bootstrap studies related to the robust PH model, computational complexity and time requirements would escalate considerably when dealing with larger data sizes. This is attributed to the necessity of estimating numerous imprecision terms in addition to the primary regression parameter.

4.6 Concluding remarks

In this chapter, the robust PH model was developed to be used instead of the standard PH model in cases where the PH assumption regarding a continuous covariate is in doubt. A robust proportional hazards model incorporates errors into covariate values in order to overcome the limitations of the proportional hazards assumption. As part of the proposed method, the observed covariate values are modified by adding error terms to each covariate value. There is no distribution for these errors, but they are allowed to fluctuate within a small interval determined by the level of imprecision, ϵ^* . The model was constructed based on the Poisson and Empirical full likelihoods, such that when the imprecision level is zero, the robust PH model is reduced to the standard PH model as discussed in Section 2.5. Nonetheless, as the level of imprecision increases, the estimated parameter derived from the robust model deviates from zero.

Theorem 4.3.1 and 4.4.2 indicate that imprecision estimates for right-censored observations prior to the first event time are not determined. Further, imprecision term estimates for early event observations are matched by the regression parameter sign. Contrary, imprecision term estimates corresponding to later event observations and right-censored observations occurring after the first event are likely to have a sign opposite to the regression parameter. For event observations in the middle, imprecision terms may have any value within the predefined imprecision interval due to the nature of constrained multivariate optimization.

The estimation of imprecise hazard and survival functions for specific individuals within the robust PH model has been approached through two methods: the naive approach and the envelope approach. The naive approach directly incorporates the imprecision limits, $-\epsilon^*$, ϵ^* , into the definition of the robust model to derive the imprecise hazard and survival functions. However, it has been observed that the estimates obtained from the standard PH model may not fall within the range of the imprecise naive estimates. To address this limitation, the envelope method has been proposed, which restricts the hazard and survival estimates derived from the robust PH model based on lower imprecision levels, including those obtained from fitting the standard PH model, to fall within the range of imprecise hazard and survival estimates derived from the robust PH model based on the selected ϵ^{**} . Accordingly, the envelope-type of imprecise hazard and survival estimates for a specific imprecision level, ϵ^{**} , is considered to encompass all potential hazard and survival estimates associated with imprecision levels ϵ^* lower than ϵ^{**} .

Several simulation scenarios were conducted to analyze the impact of the covariate effect on the pattern where naive survival estimates fail to include PH survival estimates. Interestingly, these findings can be extended to imprecise survival estimates derived from the envelope approach, as they provide valuable insights into the positioning of standard PH survival estimates within their counterparts, imprecise envelope survival estimates. More precisely, the standard PH survival estimates demonstrate distinct patterns in relation to the corresponding robust imprecise survival estimates according to the covariate effect. The standard PH survival estimates for individuals with higher covariate effects tend to reside at the upper survival estimates. Conversely, the standard PH survival estimates for individuals with lower covariate effects reside at the imprecise lower survival estimates. For individuals with moderate covariate effects, the standard PH survival estimates fluctuate within a range of the corresponding imprecise survival estimates.

Additional simulations were conducted to investigate the estimation capabilities of the robust PH model in comparison to the standard PH model in the context of

measurement errors associated with continuous covariates. Two distinct scenarios were examined: one involving accurately measured covariate and another where the covariate is contaminated with measurement errors. In the absence of measurement errors, the standard PH Model based on either Poisson or Empirical likelihoods provides a stable estimator with lower variability and relative bias aligned with the assumed true parameter. This feature does not hold true when the covariate suffers from measurement errors since the estimates can be significantly biased. In this case, the simulation results illustrate that the estimates derived from the robust PH model gradually approach the true value, and the bias is substantially reduced as Epsilon increases. Nevertheless, the reliability of the robust model is compromised by increased variability and potential non-convergence challenges at higher levels of imprecision.

To identify the optimal level of imprecision in the robust PH model, bootstrap investigations were conducted on survival datasets using various approaches. The studies, based on the selected test statistics, did not yield any valuable conclusions regarding the optimal level of imprecision. Similar results were encountered in Chapter 3 when the GPH model was applied to small to moderate-sized datasets, suggesting that larger datasets with $n > 500$ might provide more significant insights. However, the computational demands would significantly increase due to the considerably escalated number of imprecision terms that need to be estimated.

As a conclusion, this chapter is intended to serve as an introduction to the robust PH model, and to pave the way for more sophisticated methodologies capable of adeptly navigating the trade-off between imprecision and reliability in the context of violations of the PH assumption by providing insights into their advantages. The robust PH model exhibits appealing characteristics that warrant further investigation. As these initial findings suggest, the robust PH model may be a suitable alternative to the conventional PH model in circumstances where the PH assumptions are questionable in relation to a continuous covariate. However, future investigations need to establish a definitive methodology for determining the appropriate level of

imprecision.

Chapter 5

Most Likely Data

5.1 Introduction

A novel imprecise estimation technique referred to as Most Likely Data (MLD) is discussed in this chapter, highlighting some key concepts, possible developments, and potential applications. The objective of introducing the MLD method in this chapter is to pave the way for further investigation and development in the field of statistical inference. This will offer researchers a valuable tool for addressing complex modeling challenges and improving parameter estimates reliability and interpretability.

In statistical inference, the Maximum Likelihood Estimation (MLE) method stands as a standard for parameter estimation, aiming to find parameter values that maximize the likelihood of observing a given dataset under a specified statistical model. By optimizing the likelihood function, MLE provides a point estimate for the parameters, offering a straightforward and widely-used approach in various fields of study. However, if the underlying model assumptions are incorrect, for instance, or the actual data deviates from the model assumptions, MLE can provide biased estimates, e.g. the violation of the PH assumption.

The MLD method, on the other hand, represents a novel approach to parameter estimation that holds significant promise for advancing statistical inference tech-

niques. By shifting the focus from point estimates to interval estimates based on the most likely data configurations, MLD offers a unique perspective on parameter uncertainty and model fitting. The MLD is designed to identify a range of parameter values where the data of interest is the most likely data to be observed compared to other possible data. This objective is achieved by dividing the parameter space into partitions, or intervals, where the data of interest are explicitly the most likely data to be observed compared to any alternative data sets. The resulting intervals, termed MLD estimates or MLD imprecise estimates, which encompass parameter values for which the data exhibit the highest probability. The MLD method provides a nuanced perspective on parameter estimation by considering the most likely data configurations relative to the parameter space. This offers insights into the uncertainty inherent in the estimation process.

This chapter is structured as follows: Section 5.2 discusses the MLD method, its applications to discrete distributions. In Section 5.3, the application of the MLD method to the PH model is investigated, and its limitations as an estimation approach are identified. Following this, Section 5.4 presents a flexible variant of the MLD method and discusses its implications for discrete distributions and the PH model. Lastly, Section 5.5 offers relevant comments and reflections concerning the MLD method.

5.2 MLD Method

As opposed to the MLE, which seeks to determine a point estimate to maximize the likelihood of observing a given dataset, the MLD aims to construct an interval $\langle \underline{\theta}, \bar{\theta} \rangle$ within the parameter space Θ where the data of interest \mathcal{D} are explicitly the most likely observed data compared to any alternative datasets. Suppose \mathbb{D} refers to the set of all potential data. The MLD estimates are constructed based on the following definition

$$\langle \underline{\theta}, \bar{\theta} \rangle = \{ \theta : P(\mathcal{D}|\theta) \geq P(\mathbb{D}|\theta), \forall \theta \in \Theta \} \quad (5.1)$$

where $\underline{\theta}$ and $\bar{\theta}$ represent the upper and lower bounds for the imprecise estimates. As indicated in the definition, the inclusion of values from the parameter space is valid for instances in which the observed data have the same probability as any other data within \mathcal{D} . While the notation of the MLD intervals may suggest closed bounded intervals, this choice was made for convenience as it will be demonstrated later that these intervals do not necessarily have to be closed.

Despite the fact that this thesis examines the application of the MLD concept to discrete data and the PH model, the implications of this concept may extend across a wide range of statistical fields. The next section demonstrates the MLD method in the context of the binomial distribution.

5.2.1 MLD method for Binomial distribution

Recall that the probability mass function for the Binomial distribution of observing exactly x success in n trials

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \end{aligned}$$

Figure 5.1 illustrates the most likely observed value of $x = 0, 1, 2, 3, 4$ over a range of values of p , given $n = 4$. The figure highlights that to apply the MLD method to the Binomial distribution, the following constraints must be satisfied in order to determine the imprecise estimates for observing $X = x$

1. $P(X = x | p, n) \geq P(X = x - 1 | p, n)$
2. $P(X = x | p, n) \geq P(X = x + 1 | p, n)$

Theorem 5.2.1 Assume the above binomial constraints hold, then the MLD imprecise estimates of the parameter p in which $X = x$ is the most likely observed data using the binomial distribution, are determined by

$$\frac{x}{n+1} \leq p \leq \frac{x+1}{n+1} \quad (5.2)$$

Proof.

To begin with, let us consider the first constraint. That is

$$\begin{aligned}
 P(X = x \mid p, n) &\geq P(X = x - 1 \mid p, n) \\
 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} &\geq \frac{n!}{(x-1)!(n-(x-1))!} p^{(x-1)} (1-p)^{n-(x-1)} \\
 \frac{n!}{x(x-1)!(n-x)!} p^x (1-p)^{n-x} &\geq \frac{n!}{(x-1)!(n-x+1)(n-x)!} p^{(x-1)} (1-p)^{n-(x-1)} \\
 p &\geq \frac{x}{n+1}
 \end{aligned}$$

Therefore, in order to fulfill the first constraint, p is set to be greater than or equal to $\frac{x}{n+1}$. The second constraint is determined by the following

$$\begin{aligned}
 P(X = x \mid p, n) &\geq P(X = x + 1 \mid p, n) \\
 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} &\geq \frac{n!}{(x+1)!(n-(x+1))!} p^{(x+1)} (1-p)^{n-(x+1)} \\
 \frac{n!}{x!(n-x)(n-x-1)!} p^x (1-p)^{n-x} &\geq \frac{n!}{(x+1)x!(n-x-1)!} p^{(x+1)} (1-p)^{n-x-1} \\
 x+1 &\geq np - xp + xp + p \\
 p &\leq \frac{x+1}{n+1}
 \end{aligned}$$

Hence, it can be concluded that $X = x$ is most likely to be observed when p lies in the interval $[\frac{x}{n+1}, \frac{x+1}{n+1}]$. \square

Two fundamental properties of the MLD method are noteworthy: firstly, the imprecise estimates for p , making $X = x$ most likely to be observed, are distinct from just implementing the MLE for $x - 1$ and $x + 1$, using $\hat{p}_{x-1} = (x - 1)/n$ and $\hat{p}_{x+1} = (x + 1)/n$, as indicated in Equation (5.2). For instance, the MLD imprecise estimates for $x = 2$ is $p \in [1/(n + 1), 3/(n + 1)]$ which is not equivalent to $\hat{p}_1 = 1/n$ and $\hat{p}_3 = 3/n$.

Secondly, the probability of observing $x - 1$ and x is equivalent at $\underline{p} = x/(n + 1)$, and likewise for x and $x + 1$ at $\bar{p} = (x + 1)/(n + 1)$.

Example 5.2.1 Consider an experiment involving tossing a coin four times, with X representing the number of observed heads. Suppose this experiment involves an unfair coin, where the probability of observing a head, denoted as p , can assume any value in the interval $[0, 1]$, and let $x = 1$. In practical scenarios where one seeks

x	\hat{p}	\underline{p}	\bar{p}
0	0	0	0.2
1	0.25	0.2	0.4
2	0.5	0.4	0.6
3	0.75	0.6	0.8
4	1	0.8	1

Table 5.1: Upper and lower estimates for p where x is most likely to be observed

to estimate p , the MLE method is often employed, yielding $\hat{p} = \frac{x}{n} = \frac{1}{4} = 0.25$. However, our interest lies in an imprecise estimation of p , aiming to identify a range of p values where $x = 1$ is most likely to be observed compared to other values of x . This objective is achieved by dividing the parameter space of p according to the most likely data across various p values.

Instead of solely relying on the MLE of p , these imprecise estimates enhance the robustness of the estimation. The upper and lower estimates for each value of x can be derived by Equation (5.2) as shown in Table 5.1.

5.2.2 MLD method for Poisson distribution

Recall the Poisson distribution in which X is a discrete random variable representing the number of events, x , observed within a given time period. Assume that X follows the Poisson distribution with an average rate of events denoted by λ . Then, the probability mass function for observing x events over specific time period is given by

$$P(X = x) = \frac{\lambda^x \exp(-\lambda)}{x!} \quad (5.3)$$

Figure 5.2 shows the most likely observed number of events for $x = 0, 1, 2, 3, 4, 5, 6, 7, 8$ over range of λ values. The figure indicates that applying the MLD method to the Poisson distribution in which the number of events $X = x$ has the highest probability to be observed are subject to the following constraints

1. $P(X = x | \lambda) \geq P(X = x - 1 | \lambda)$

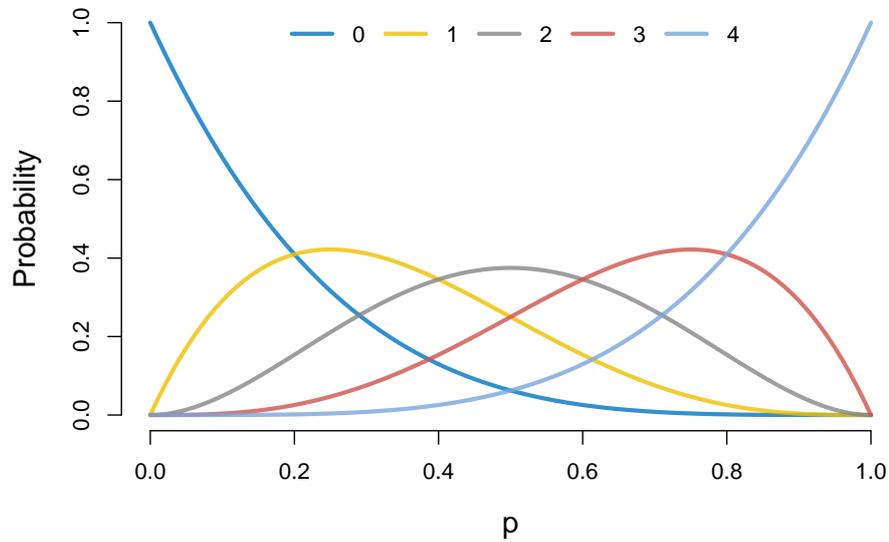


Figure 5.1: MLD estimates for p based on a binomial experiment involving tossing a coin four times

$$2. P(X = x | \lambda) \geq P(X = x + 1 | \lambda)$$

Due to the fact that X are restricted to be Natural numbers, then $x - 1 = 0$ whenever $x = 0$.

Theorem 5.2.2 Assume the Poisson constraints hold, then the MLD imprecise estimates of λ in which $X = x$ is the most likely observed number of event, are determined by

$$x \leq \lambda \leq x + 1 \tag{5.4}$$

Proof.

Consider the first constraint, then

$$\begin{aligned} P(X = x | \lambda) &\geq P(X = x - 1 | \lambda) \\ \frac{\lambda^x \exp(-\lambda)}{x(x-1)!} &\geq \frac{\lambda^{(x-1)} \exp(-\lambda)}{(x-1)!} \\ \lambda &\geq x \end{aligned}$$

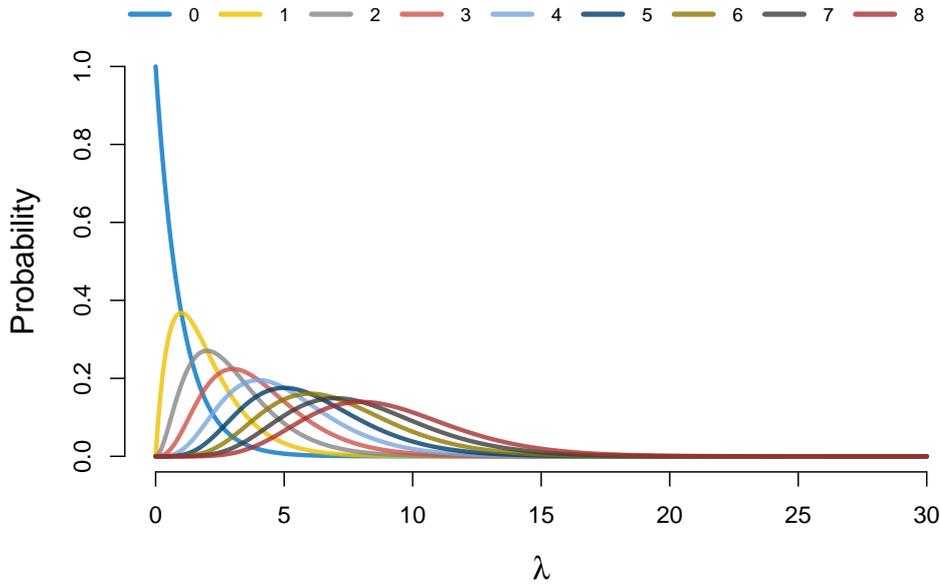


Figure 5.2: MLD estimates for observing $x = 0, 1, 2, 3, 4, 5, 6, 7, 8$ number of events at particular time period based on the Poisson distribution

Similarly, the second constraint leads to $\lambda \leq x + 1$. Hence, the MLD imprecise estimates of λ based on the Poisson distribution are obtained by $x \leq \lambda \leq x + 1$. \square

According to Theorem 5.2.2, the MLD imprecise estimates that makes the number of events $X = 3$ and $X = 6$, for instance, the most likely number of events are $\lambda \in [3, 4]$, and $\lambda \in [6, 7]$, respectively.

5.3 MLD method for the the PH model

A number of estimators have been developed, each grounded in distinct likelihood functions, to estimate the regression parameter β within the PH model. However, as far as our current understanding extends, none of these estimators have exhibited imprecision in estimating the PH model. This section undertakes an exploration of the potential application of the MLD method to the PH model, aiming to relax the PH assumption. The investigation focuses primarily on survival data with binary covariates. The structure of this section will rely on a foundational example to

provide insight into the feasibility of employing the MLD method for estimating a PH model.

Consider the artificial survival data in Table 5.2, comprising two groups, each consisting of two patients. The covariate x represents the gender of those patients such that $x = 0$ for Female patient, and male otherwise.

i	t	δ	x
1	4	1	0
2	6	1	1
3	7	1	0
4	9	1	1

Table 5.2: Data set

To implement the MLD method, all possible orders of events must be considered given the data set in Table 5.2, as in Section 5.2.1. Let β denote the parameter space encompassing all parameter values corresponding to the potential sample orders: MMFF, MFMF, FMMF, MFFM, FMFM, and FFMM. The main objective is to investigate the possibility of dividing the parameter space, β , into subsets according to what is the most likely data. For instance, what is the set of values in the parameter space in which the data FMFM is the most likely order to be observed compares to all others.

5.3.1 MLD for the PH model using marginal probability

One feasible approach to determine the probability of observing these data orders involves examining the conditional probability that the individual with $x = x_j$ experiences the event at time t_i , given that one individual from R_i experiences the event at time t_i , as outlined in Equation (2.32). Consequently, multiplying these conditional probabilities across all event times for all potential orders could represent the probability of observing a specific order as follows

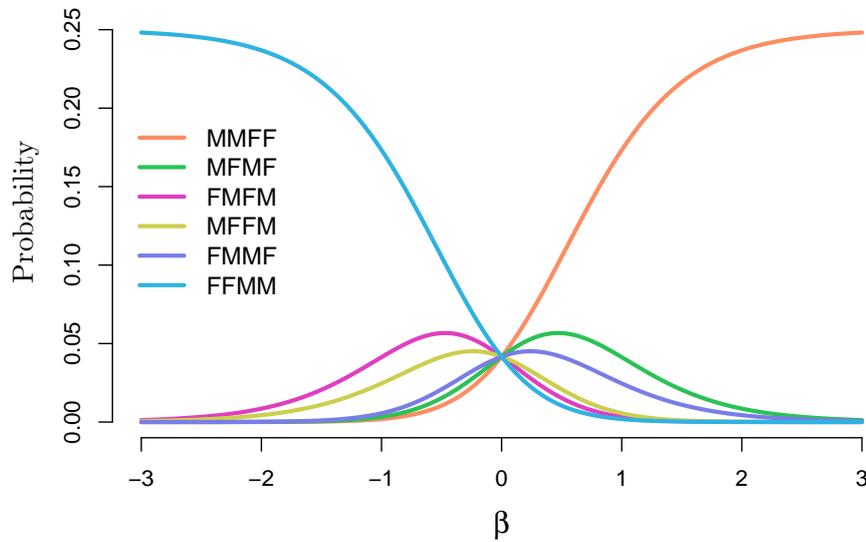


Figure 5.3: MLD for the balanced data in Table 5.2

$$\prod_{i=1}^n \frac{\phi_i}{\sum_{l \in R_i} \phi_l}$$

This represents the marginal distribution of the rank statistic, as highlighted by Kalbfleisch and Prentice [45], in scenarios without ties or censored observations. It also corresponds to the partial likelihood method proposed by Cox [25]. Figure 5.3 illustrates the probabilities corresponding to these orders plotted against potential values from the parameter space, β . Remarkably, according to Figure 5.3, MMFF appears to be the most probable dataset whenever $\beta > 0$, while FFMM dominates when $\beta < 0$.

Figure 5.3 indicates that these probabilities do not adequately account for the proportion of each group in the risk set to adjust the estimate of the parameter accordingly. This limitation becomes evident in Figure 5.4, where an additional female observation is included in the data in Table 5.2. Despite this, MMFFF emerges as the dominant most likely order when $\beta > 0$, illustrating how these probabilities fail to address the impact of unbalanced observations in the risk set.

Indeed, the expectation was that these probabilities would fluctuate across some intervals of β , thereby providing informative insights into the most likely observed

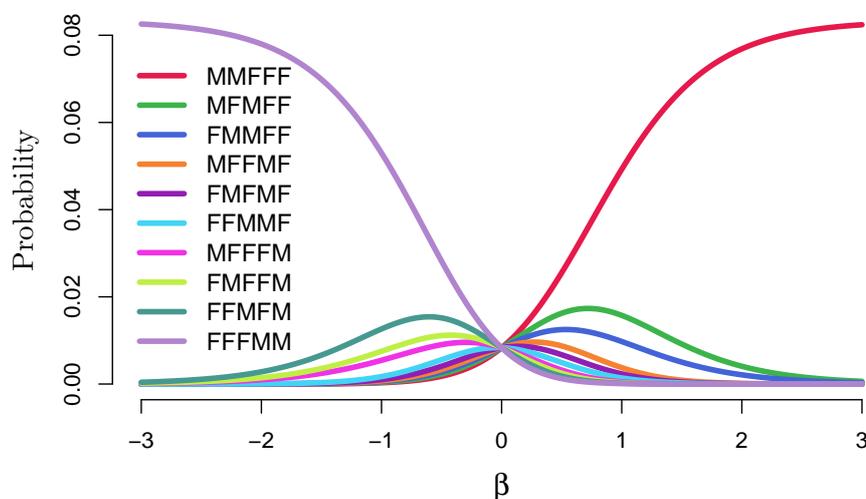


Figure 5.4: MLD for the unbalanced data with one additional female observation order within specific intervals of β .

5.3.2 MLD for the PH model using adjusted probability

In light of the limitations observed in Section 5.3.1 in which the MLD method for the PH model based on marginal probabilities failed to effectively partition the parameter space for different variants in the sample space. This issue arises because marginal probabilities do not address unbalanced observations in the risk set. This section aims to adjust these marginal probabilities to overcome this limitation by considering the number of observations in the risk set at each event time that share the same covariate value with the individual who experienced the event. Consequently, the MLD method will be applied to the PH model according to these adjusted probabilities.

Consider, for instance, if a male have the event at $t_{(i)}$, then the adjusted probability will take into account the number of males in the risk set at $t_{(i)}$, rather than focusing exclusively on that particular male. Let the individual who had the event at $t_{(i)}$ as the i th individual, and R_i be the set of individuals at risk, survive or censored prior to t_i . The number of individuals in the set R_i whose share the same

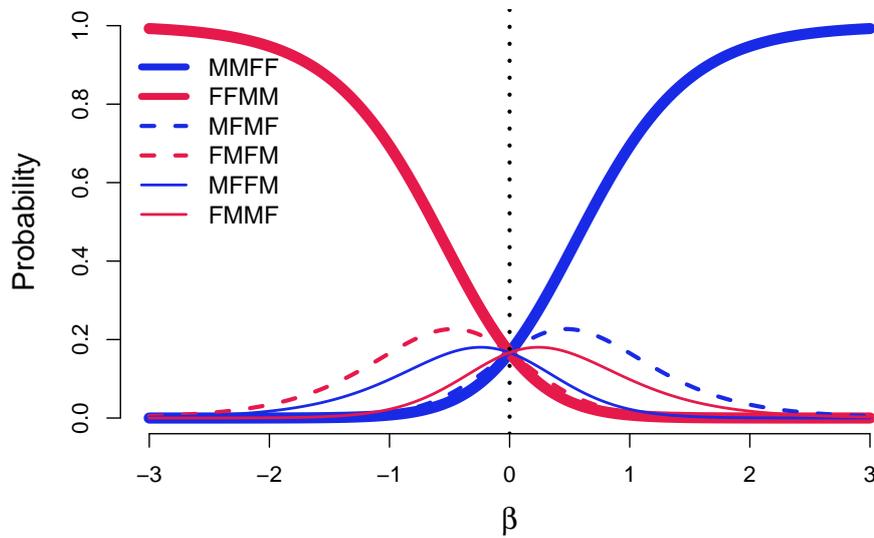


Figure 5.5: MLD for the balanced data in Table 5.2 based on the adjusted probability covariate value as i th, denoted by \tilde{n}_i . Then, the adjusted probability at event time t_i is

$$\frac{\tilde{n}_j \phi_j}{\sum_{l \in R_j} \phi_l} \quad (5.5)$$

The adjusted probabilities for particular data based on the PH model can be written as

$$\prod_{j=1}^n \frac{\tilde{n}_j \phi_j}{\sum_{l \in R_j} \phi_l} \quad (5.6)$$

Generally, the estimate obtained from the adjusted probability is identical to the one obtained by the marginal probabilities for both balance and unbalanced data sets due to the fact that \tilde{n}_i does not involve any information about the parameter, β , and it can be considered as a constant, see Figure (5.5) and Figure (5.6).

Although the suggested probability does not show any advantageous in terms of partitioning the parameter space using the MLD method, the scale to these probabilities, is appealing because at each value in the parameter space the sum of the probabilities for all theses data orders is equal to 1. Therefore, this representation of

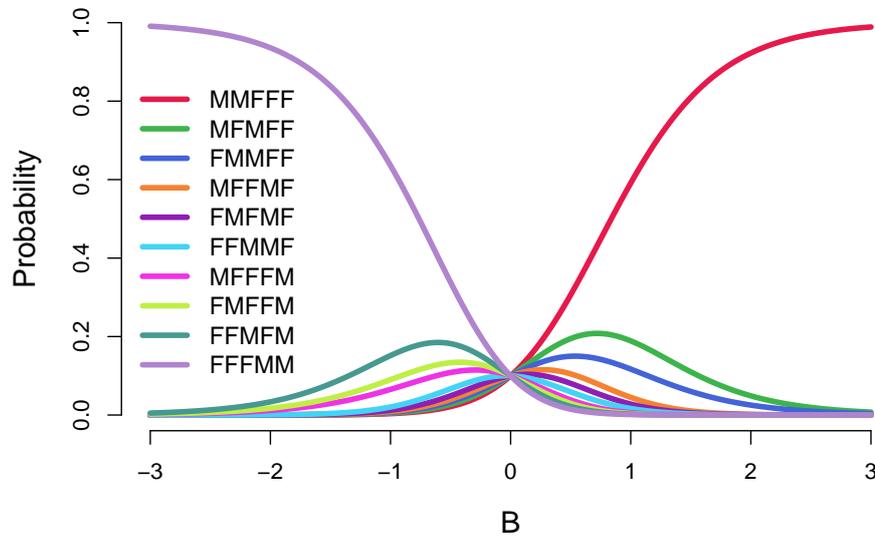


Figure 5.6: MLD for the unbalanced data with additional female observation based on the adjusted probability

probability will adopted for the rest of this chapter in the context of the PH model instead of the original marginal probability.

5.3.3 Time-based MLD method for the PH model

The analyses conducted in Sections 5.3.1 and 5.3.2 yield identical conclusions, namely, FFMM is the most likely observed data when $\beta < 0$, and MMFF whenever $\beta > 0$. This section discusses an alternative approach to applying the MLD method in the context of the PH model, that is a time-based approach. In this approach, the range of parameter values where the observation of interest is most likely to be observed compared to other possible observations will be identified at each event time. The result of this process will be k intervals related to the k distinct event times. Consequently, the intersection of these intervals represents the time-based MLD imprecise estimates. However, an empty intersection indicates $[\underline{\beta}, \bar{\beta}] = [0, 0]$ and will be interpreted as that the given data is never most likely.

Suppose the interval $I_i = [\underline{I}_i, \bar{I}_i]$ represents the range of parameter values in which a particular observation of interest is most likely to be observed at time t_i , with

the lower and upper limits \underline{I}_i and \bar{I}_i , respectively. Let $\mathbf{I} = \{\underline{I}_1, \underline{I}_2, \dots, \underline{I}_k\}$, and $\bar{\mathbf{I}} = \{\bar{I}_1, \bar{I}_2, \dots, \bar{I}_k\}$ denote the sets of lower and upper limits for the intervals I_i for $i = 1, 2, \dots, k$. The time-based MLD estimates for the PH model can be defined as follows

$$[\underline{\beta}, \bar{\beta}] = \begin{cases} \bigcap_{i=1}^k I_i & \max(\mathbf{I}) \leq \min(\bar{\mathbf{I}}) \\ [0, 0] & \max(\mathbf{I}) > \min(\bar{\mathbf{I}}) \end{cases} \quad (5.7)$$

The following theorem illustrate how to identify the lower and upper limits for each interval I_i .

Theorem 5.3.1 Suppose the individual of interest to be most likely observed at t_j is characterized by \tilde{x}_j and \tilde{n}_j which represent the covariate value and the number of individuals from the same group in the risk set R_j . Let the other group represented by x_j^c and n_j^c . The set of β values in which the observation of interest is most likely to experience the event at time t_j can be determined as follows

$$\beta(\tilde{x}_j - x_j^c) \geq \ln \left(\frac{n_j^c}{\tilde{n}_j} \right) \quad (5.8)$$

Note: the direction of this inequality depends on the coding of \tilde{x}_j and x_j^c .

Proof.

Consider the adjusted probability in Equation(5.5) and the definition of the MLD in Equation(5.1), then an individual with \tilde{x}_j is most likely to be observed at t_j is conditional on the following

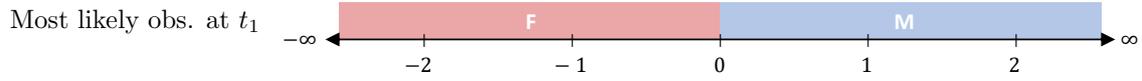
$$\begin{aligned} P(\tilde{x}_j | t_j, \beta) &\geq P(x_j^c | t_j, \beta) \\ \frac{\tilde{n}_j \tilde{\phi}_j}{\sum_{l \in R_j} \phi_l} &\geq \frac{n_j^c \phi_j^c}{\sum_{l \in R_j} \phi_l} \\ \tilde{n}_j \tilde{\phi}_j &\geq n_j^c \phi_j^c \\ \exp(\beta(\tilde{x}_j - x_j^c)) &\geq \frac{n_j^c}{\tilde{n}_j} \\ \beta(\tilde{x}_j - x_j^c) &\geq \ln \left(\frac{n_j^c}{\tilde{n}_j} \right) \end{aligned}$$

□

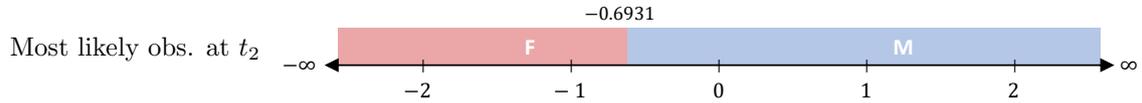
Example 5.3.1 Recall the artificial survival data in Table 5.2. For the sake of simplicity, let $P(F)$ and $P(M)$ denote the probability observing a female and a male at t_1 , respectively. Additionally, $P(M|F)$ represents the probability of observing a male at t_2 given observing a female at t_2 . Using this analogy of notations, $P(F|FM)$ indicates the probability of observing a female at t_3 given observing a female at t_1 and a male at t_2 .

Suppose the data of interest is FMFM as shown in the original data in Table 5.2, then by Equation (5.8) the probability of observing a female at t_1 is attainable with $\beta \in I_1 = (-\infty, 0]$. To observe a male at t_2 given a female was observed at t_1 , the values $\tilde{x}_2 = 1$, $\tilde{n}_2 = 2$, $x_2^c = 0$, and $n_2^c = 1$ will be substituted in Equation (5.8) which suggest that a male is most likely to be observed at $\beta \in I_2 = [-0.6931, \infty)$. As there is only one of each group in the risk set of t_3 , observing a female, given that the first observed individual was female and the second was male, is most likely to experience the event when $\beta \in I_3 = (-\infty, 0]$. There is only one observation left at the last event time, so any value in the parameter space leads to that male being most likely to experience the event. Thus, the last event time can be neglected. By taking the intersection of these intervals according to Equation(5.7), the time-based MLD estimates for observing FMFM are $[-0.69310, 0]$. The following diagram illustrates these steps.

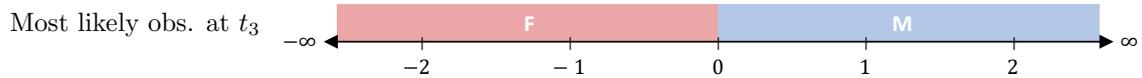
Consider investigating the following two scenarios in which the first observation is female: FFMM and FMMF, as shown in Figure 5.7. The figure indicates that a female is most likely to observe at t_1 , $P(F)$, when β is less than zero. For the FFMM data set, observing female at the second time, $P(F|F)$, is most likely for $\beta < -0.6931$. The remaining are two males, so they are the most likely observed at t_3 and t_4 . By taking the intersection of these two intervals, the time-based MLD estimates for FFMM are $[\underline{\beta}, \overline{\beta}] = (-\infty, -0.6931]$. For the FMMF data set, observing a male at the second event time, $P(M|F)$, is most likely when $\beta > -0.6931$. At t_3 , there is a male and a female, so to observe a male it is most likely when $\beta \in [0, \infty)$. The intersection is equal to zero; therefore, this particular data is never most likely to be observed compared to other.



At t_1 , F is the most likely to be observed when $\beta \in I_1 = (-\infty, 0]$



At t_2 , M is the most likely to be observed when $\beta \in I_2 = [-0.6931, \infty)$



At t_3 , F is the most likely to be observed when $\beta \in I_3 = (-\infty, 0]$

Most likely obs. at t_4 There is only one M left $\Rightarrow \beta \in I_4 = (-\infty, \infty)$

Similarly, one can find that MMFF is most likely when $[\underline{\beta}, \bar{\beta}] = [0.6931, \infty)$, MFMF is the most likely data when $[\underline{\beta}, \bar{\beta}] = [0, 0.6931]$, and MFFM is never most likely data to be observed. Note that the time-based MLD does not hold using the unadjusted marginal probability as non of the data sets in most likely due to the fact that the marginal probability, partial likelihood, does not take into account the number of individual from each group in the risk set.

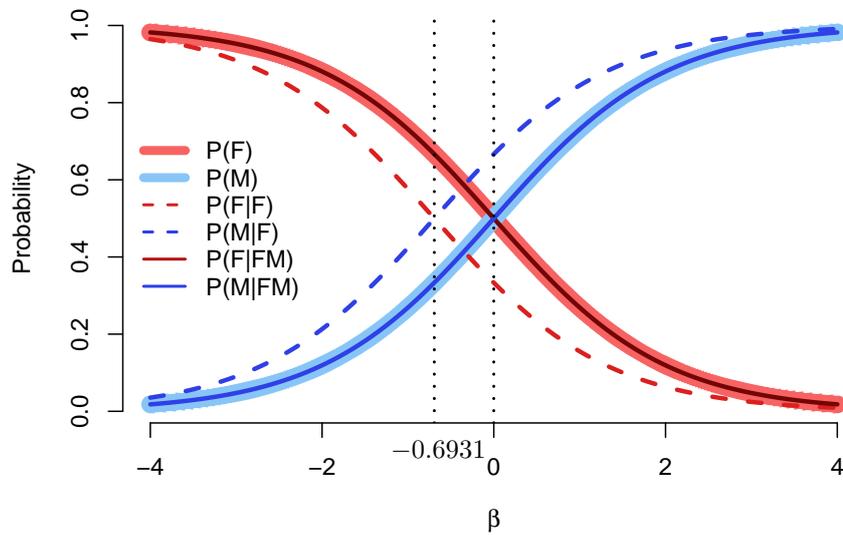


Figure 5.7: Time-based MLD probability for the most likely observations at t_1 , t_2 and t_3 , given female observed at t_1

5.4 Flexible MLD method

Based on the marginal probability or the adjusted variant, it is evident that applying the MLD method to the PH model based on the marginal or adjusted probabilities do not yield any significant insights in terms of partitioning the parameter space into subsets with explicit data are most likely observed data. While the time-based MLD approach for the PH model exhibits appealing characteristics, certain datasets are never most likely data. As a result, this section presents a flexible MLD method designed to mitigate the constraints imposed by the requirement of identifying the most likely data. Essentially, this alternative method widens the scope within the parameter space to accommodate a wider range of potential values, rather than rigidly assuming a dominant probability associated with the most likely data. The flexible MLD approach holds promise, particularly in scenarios like the PH model, where it may offer distinct advantages. Suppose that $\pi \in [0, 1]$ refers to the MLD imprecision term, a key element of the flexible MLD method. The flexible MLD estimates given π are constructed based on the following definition

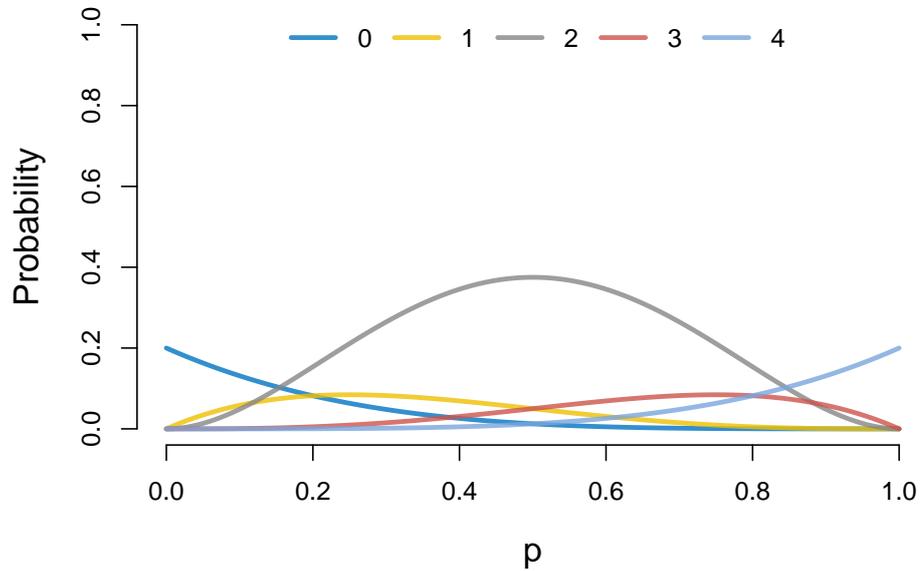


Figure 5.8: Flexible MLD for method Binomial distribution for observing $X = 2$ using imprecision effect $\pi = 0.8$ given $n = 4$

$$\langle \underline{\theta}, \bar{\theta} \rangle = \{ \theta : P(\mathcal{D}|\theta) \geq (1 - \pi)P(\mathbb{D}|\theta), \forall \theta \in \Theta \} \quad (5.9)$$

The following sections implement this method to the binomial distribution, the Poisson, and the PH model.

5.4.1 Flexible MLD for method Binomial distribution

This section describes the application of the flexible MLD for method Binomial distribution. For a given imprecision term π , one may consider applying the flexible MLD method based on the following constraints:

1. $P(X = x | p, n) \geq (1 - \pi)P(X = x - 1 | p, n)$
2. $P(X = x | p, n) \geq (1 - \pi)P(X = x + 1 | p, n)$

In spite of this, a closed-form solution is not feasible for determining flexible imprecise estimates based on the Binomial distribution, since $P(X = x | p, n)$ surpasses the probability for any feasible values of X as the imprecision effect grows.

Figure 5.8 illustrates this challenge considering the flexible imprecise estimates for $X = 2$ using the imprecision term $\pi = 0.8$ with $n = 4$. The figure indicates the difficulty in determining flexible imprecise estimates based on the probabilities of only $X = 1$ and $X = 3$. Therefore, it becomes necessary to conduct calculations encompassing all possible values of X across potential parameter values of p . Then, applying the definition of the flexible MLD method in order to determine the imprecise estimates.

Example 5.4.1 Consider applying the flexible MLD method to a binomial experiment that involving tossing a coin four times as in Example 5.2.1. The imprecise estimates of p in which $X = x$ is the most likely observed based on the flexible method are determined and presented in Table 5.3. the table shows these imprecise estimates for $x = 1, 2, 3, 4$ using various values of the imprecision term π . When $\pi = 0$, the imprecise estimates obtained by flexible method reduces to those obtained from the standard MLD method. With an increase in the MLD imprecision term, the interval of the imprecise estimates of p expanded to accommodate other values from the parameter space in which $X = x$ is more likely to be observe. Additionally, as the imprecision term $\pi = 1$ all potential data have the same probability to be observed using any value from the parameter space.

5.4.2 Relaxed MLD for Poisson distribution

This section presents the application of the flexible MLD methodology within the domain of the Poisson distribution. As with the Binomial distribution, the Poisson distribution lacks a closed-form of deriving imprecise estimates for the parameter λ under the flexible MLD framework. Hence, this challenge highlights the importance of determining the probabilities for all possible values of X , and then identifying the flexible imprecise estimates in which the data of interest is most likely to be observed according to the predetermined imprecision term π . Accordingly, the flexible MLD estimates in which $X = 3$ and $X = 6$ are most likely to be observed based on the data set considered on Section 5.2.2 are given in Table 5.4.

π	x=0		x=1		x=2		x=3		x=4	
	\underline{p}	\bar{p}								
0	0	0.2	0.2	0.4	0.4	0.6	0.6	0.8	0.8	1
0.1	0	0.2174	0.1837	0.4255	0.3750	0.6250	0.5745	0.8163	0.7826	1
0.2	0	0.2381	0.1667	0.4545	0.3478	0.6522	0.5455	0.8333	0.7619	1
0.3	0	0.2632	0.1489	0.4878	0.3182	0.6818	0.5122	0.8511	0.7368	1
0.4	0	0.2941	0.1304	0.5263	0.2857	0.7143	0.4737	0.8696	0.7059	1
0.5	0	0.3333	0.1111	0.5714	0.2500	0.7500	0.4286	0.8889	0.6667	1
0.6	0	0.3846	0.0909	0.6126	0.2105	0.7895	0.3874	0.9091	0.6154	1
0.7	0	0.4271	0.0698	0.6461	0.1828	0.8173	0.3539	0.9302	0.5730	1
0.8	0	0.4772	0.0476	0.6910	0.1544	0.8456	0.3090	0.9524	0.5228	1
0.9	0	0.5635	0.0244	0.5797	0.1144	0.8857	0.2403	0.9756	0.4365	1
1	0	1	0	1	0	1	0	1	0	1

Table 5.3: MLD imprecise estimates for p which maximize the probability of observing x given different values π

π	x=3		x=6	
	$\underline{\lambda}$	$\bar{\lambda}$	$\underline{\lambda}$	$\bar{\lambda}$
0	3	4	6	7
0.1	2.700	4.444	5.400	7.777
0.2	2.401	4.999	4.899	8.366
0.3	2.100	5.345	4.583	8.944
0.4	1.898	5.773	4.243	9.660
0.5	1.733	6.214	3.915	10.58
0.6	1.550	6.694	3.635	11.83
0.7	1.342	7.274	3.302	13.66
0.8	1.096	8.040	2.913	16.73
0.9	0.775	9.235	2.450	23.66
1	0	∞	0	∞

Table 5.4: Flexible MLD imprecise estimates for λ which maximize the probability of observing x number of event during a period of time given different values π

5.4.3 Flexible MLD method for the PH model

The limitations encountered when applying the standard MLD method to the PH model prompted the consideration of the flexible MLD method. Even though the MLD method for the PH model does not offer an advantage in terms of partitioning the parameter space, the flexible MLD method is a more appealing alternative in situations such as those encountered in the PH model. Consider an imprecision term $\pi \in [0, 1]$, the flexible MLD imprecise estimates for the observing the data \mathcal{D} based on the PH model are given by

$$[\underline{\beta}, \bar{\beta}] = \{\beta : P(\mathcal{D}|\beta) \geq (1 - \pi)P(\mathbb{D}|\beta), \forall \beta \in \Theta\} \quad (5.10)$$

In the context where \mathbb{D} represents the dominant data sets, the selection criteria for survival analysis involving binary covariates in the absence of right-censored observations entail the inclusion of only two distinctive data sets within \mathbb{D} . These data sets are distinguished by covariate values demonstrating a monotonic pattern. Specifically, they consist of data instances where covariate values are either $(0s, 1s)$ or $(1s, 0s)$. For instance, within the artificial survival data used in Table 5.2, these extreme data sets correspond to FFMM and MMFF orders. Therefore, the flexible MLD estimates can be determined by the following constraints

$$\begin{aligned} [\underline{\beta}, \bar{\beta}] &= \{\beta : P(\mathcal{D}|\beta) \geq (1 - \pi)P(\text{FFMM}|\beta), \forall \beta \in (-\infty, \infty)\} \\ [\underline{\beta}, \bar{\beta}] &= \{\beta : P(\mathcal{D}|\beta) \geq (1 - \pi)P(\text{MMFF}|\beta), \forall \beta \in (-\infty, \infty)\} \end{aligned} \quad (5.11)$$

Example 5.4.2 Reconsider the artificial survival data in Table 5.2 which related to a binary covariate represent the gender with two individuals in each group. The data sets with the covariate FFMM and MMFF found to be the dominant observations interns of their observing probabilities. On the other hand, the data with the following orders FMFM, FMMF, MFMF, and MFFM are never most likely to be observed. By applying the flexible MLD method, the limitations imposed by the standard MLD method are relaxed, allowing for imprecise estimates for these inferior data sets. Table 5.5 presents the imprecise estimates for the PH model for all potential datasets, employing the flexible MLD method with various values of

	FFMM		FMFM		FMMF	
$\hat{\beta}_c$	Non-Conv.		-0.9406		0.4812	
90% CI	NA		(-2.9871, 1.1059)		(-1.5615, 2.5239)	
π	$\underline{\beta}$	$\bar{\beta}$	$\underline{\beta}$	$\bar{\beta}$	$\underline{\beta}$	$\bar{\beta}$
0	$-\infty$	0	NA	NA	NA	NA
0.1	$-\infty$	0.0225	-0.1003	0.0286	-0.0349	0.0620
0.2	$-\infty$	0.0478	-0.2027	0.0603	-0.0734	0.1291
0.3	$-\infty$	0.0764	-0.3095	0.0960	-0.1166	0.2027
0.4	$-\infty$	0.1095	-0.4236	0.1369	-0.1657	0.2851
0.5	$-\infty$	0.1486	-0.5493	0.1847	-0.2228	0.3798
0.6	$-\infty$	0.1966	-0.6931	0.2426	-0.2914	0.4928
0.7	$-\infty$	0.2587	-0.8673	0.3165	-0.3780	0.6354
0.8	$-\infty$	0.3465	-1.0986	0.4193	-0.4969	0.8333
0.9	$-\infty$	0.4982	-1.4722	0.5926	-0.6931	1.1698
1	$-\infty$	∞	$-\infty$	∞	$-\infty$	∞

	MFMF		MFFM		MMFF	
$\hat{\beta}_c$	0.9406		-0.4812		Non-Conv.	
90% CI	(-1.1059, 2.9871)		(-2.5239, 1.5615)		NA	
π	$\underline{\beta}$	$\bar{\beta}$	$\underline{\beta}$	$\bar{\beta}$	$\underline{\beta}$	$\bar{\beta}$
0	NA	NA	NA	NA	0	∞
0.1	-0.0286	0.1003	-0.0620	0.0349	-0.0225	∞
0.2	-0.0603	0.2027	-0.1291	0.0734	-0.0478	∞
0.3	-0.0960	0.3095	-0.2027	0.1166	-0.0764	∞
0.4	-0.1369	0.4236	-0.2851	0.1657	-0.1095	∞
0.5	-0.1847	0.5493	-0.3798	0.2228	-0.1486	∞
0.6	-0.2426	0.6931	-0.4928	0.2914	-0.1966	∞
0.7	-0.3165	0.8673	-0.6354	0.3780	-0.2587	∞
0.8	-0.4193	1.0986	-0.8333	0.4969	-0.3465	∞
0.9	-0.5926	1.4722	-1.1698	0.6931	-0.4982	∞
1	$-\infty$	∞	$-\infty$	∞	$-\infty$	∞

Table 5.5: Imprecise estimates for β in the PH model based on the flexible method

the imprecision parameter π . When $\pi = 0$, the imprecise estimates derived through the flexible method are equal to those acquired via the standard the MLD method. Furthermore, as the MLD imprecision term increases, the uncertainty surrounding the estimates p becomes larger. This allows for a broader range of feasible values from the parameter space. Additionally, when the imprecision parameter is set to 1, all possible data have an equal probability of being observed, regardless of the specific value chosen from the parameter space.

5.4.4 Flexible time-based MLD method for the PH model

This section illustrates a significant improvement in the imprecise estimates achieved through the time-based MLD approach. Specifically, datasets that were previously unlikely to be observed after partitioning the parameter space according to the most likely data will benefit from the flexible MLD method.

Theorem 5.4.1 Suppose the individual of interest to be most likely observed at t_j is characterized by \tilde{x}_j and \tilde{n}_j which represent the covariate value and the number of individuals from the same group in the risk set R_j . Let the other group represented by x_j^c and n_j^c . The set of β values in which the observation of interest is most likely to experience the event at time t_j can be determined as follows

$$\beta(\tilde{x}_j - x_j^c) \geq \ln \left(\frac{(1 - \pi)n_j^c}{\tilde{n}_j} \right) \quad (5.12)$$

Proof.

The proof can be derived straightforwardly, similar to Theorem 5.3.1. \square

Example 5.4.3 This example illustrates the flexible MLD estimates for one of the never most likely data, namely, FMMF. The estimation will be based on the MLD level of imprecision $\pi = 0.3$.

- At t_1

Based on Theorem 5.4.4 and the adjusted probability in Equation (5.5), the

imprecise estimates for observing a female at t_1 is determined as follows

$$\begin{aligned} P(F) &\geq (1 - \pi)P(M) \\ -\beta &\geq \ln \frac{(1 - 0.3)2}{2} \\ \beta &\leq -\ln 0.7 = 0.3567 \end{aligned}$$

Hence, the imprecise estimates in which a female is more likely to be observed at t_1 are $\beta \in I_1 = (-\infty, 0.3567]$.

- At t_2

To find an interval of β where a male is more likely to be observe can be calculated as follows

$$\begin{aligned} P(M | F) &\geq 0.7P(F | F) \\ \beta &\geq \ln \frac{0.7}{2} \\ \beta &\geq -1.04982 \end{aligned}$$

Hence, the imprecise estimates for β corresponds a male individual to be most likely observed data at the second event time are $\beta \in I_2 = [-1.04982, \infty)$

- At $t_3 = 7$

we have

$$\begin{aligned} P(M | FM) &\geq 0.7P(F | FM) \\ \beta &\geq \ln 0.7 = -0.3567 \end{aligned}$$

A male is the most likely individual to observed at t_3 when $\beta \in I_3 = [-0.3567, \infty)$

Since the last event time has no impact on the time-based imprecise estimates, it will be neglected. Thus, the flexible MLD imprecise estimates for the FMMF data based on the level of imprecision $\pi = 0.3$ are $[\underline{\beta}, \bar{\beta}] = [-0.3567, 0.3567]$. Even though for this particular data the flexible MLD estimates are symmetrical, simulation shows that this cas is not always true.

5.4.5 Simulation study

This section delves into comparing the effectiveness of employing the flexible MLD method, that based on the adjusted probabilities, versus the time-based flexible MLD method within the framework of the PH model. Through a simulation

study, valuable insights are sought into the accuracy of these methodologies in estimating the parameters of the PH model.

One hundred survival data sets were generated under the assumption of a Weibull distribution with shape parameter $\rho = 1.1$ and scale parameter $\lambda = 2$ for baseline survival times related to individuals with $x = 0$. A single covariate was assumed to follow a Binomial distribution with a probability of 0.5, having a regression coefficient of $\beta = -1$. The study considered two sample sizes: small $n = 10$ and moderate $n = 60$. The flexible MLD method and the time-based flexible MLD method were implemented for each dataset at varying levels of imprecision $\pi_i = 0.3, 0.6, 0.8$.

Figure 5.9 illustrates the simulation results for datasets with $n = 10$. Figure 5.9 (top) shows results obtained from applying the time-based flexible MLD method, while (bottom) shows results based on the flexible MLD method. Imprecise estimates are displayed as bars to illustrate the range of the estimates. A teal-colored bar signifies that the corresponding imprecise estimate contained the true parameter value, while a red bar indicates otherwise. Additionally, standard PH estimates are presented as circles, with teal circles indicating inclusion in the associated imprecise estimates and red circles signifying exclusion.

Both methods demonstrate high sensitivity to the imprecision term π . Increasing the level of imprecision raises the probability of imprecise estimates encompassing the true parameter. For the time-based MLD approach, the proportion of imprecise estimates encompassing the true parameter were 22%, 52%, and 85% for $\pi = 0.3, 0.6, 0.8$, respectively. Conversely, the inclusion proportions based the flexible MLD method 3%, 23%, and 50% for the same imprecision levels. Furthermore, it was observed that the time-based estimates fluctuated around the true parameter, while the imprecise estimates derived from the flexible MLD method tended to cluster around 0. Furthermore, it was observed that while the time-based estimates exhibited fluctuations around the true parameter, imprecise estimates derived from the flexible MLD method tended to cluster around 0. As elaborated in Sections 5.3.1 and 5.3.2, this phenomenon appears to be the result of a limitation inherent in the MLD method for the PH model. As discussed in these sections, all datasets are considered as never most likely observed data except the dominant datasets, which

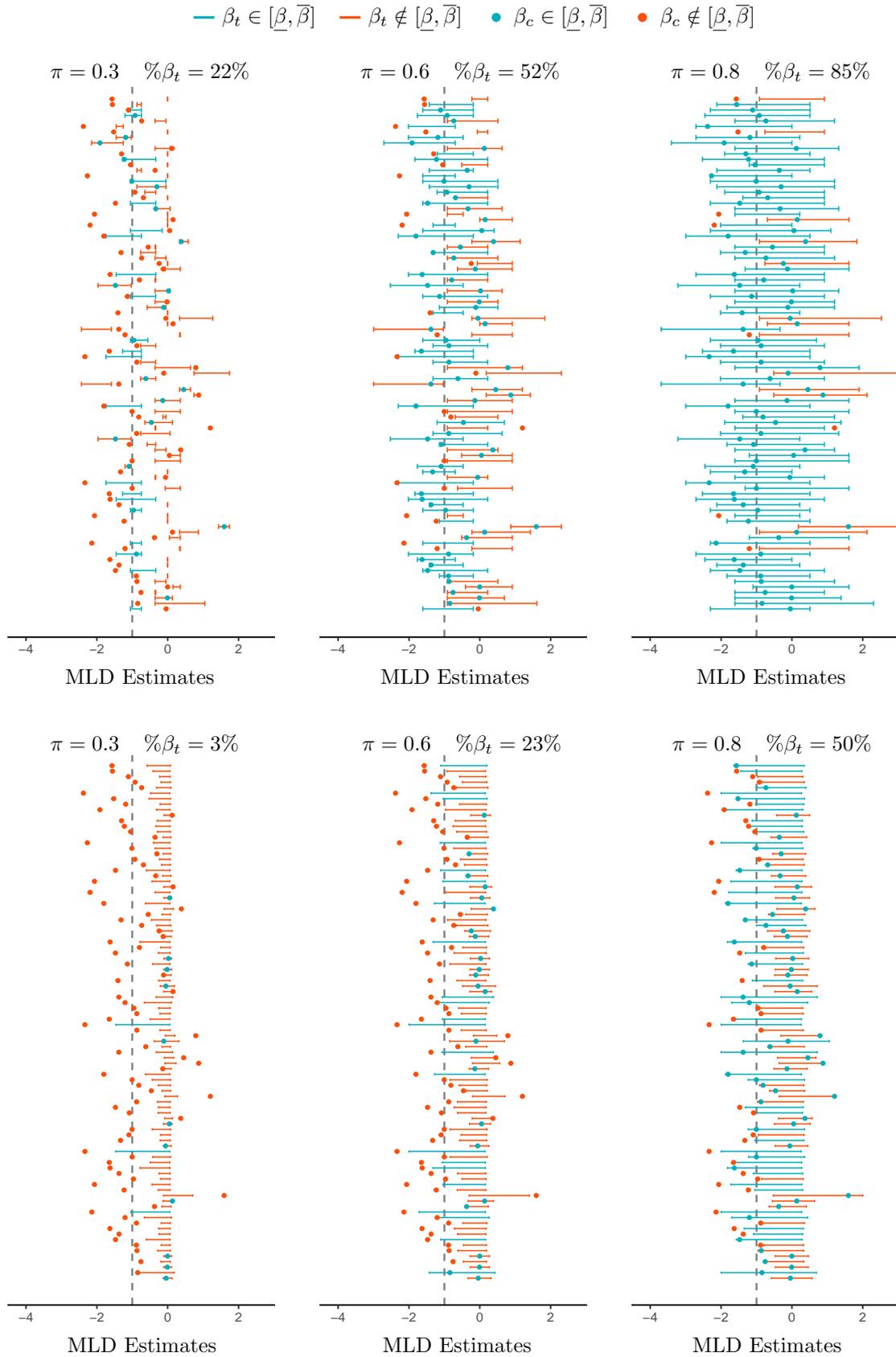


Figure 5.9: MLD estimates for 100 datasets of size 10 using $\pi = 0.3, 0.6, 0.8$ according to the time-based (top) and adjusted probability (bottom) methods, along with the true parameter β_t (dashed line) and the standard PH estimates β_c (circles). Teal color denotes inclusion within the imprecise estimates, while red signifies exclusion.

have monotonous covariate values.

For data sets with $n = 60$, both methods fail to capture the true parameter at lower level of imprecision as illustrated in Figure 5.10. This implies that as the number of observations increases, there is a need for a substantial increase in the level of imprecision to effectively capture the true parameter. As compared to flexible MLD method, the flexible time-based MLD method performs better in terms of robustness and reliability of encompassing both the true parameter and the standard PH estimates across various imprecision levels π . Based on the simulation results, it is important to carefully choose the imprecision level as well as their impact on the trade-offs between estimation inclusion and reliability.

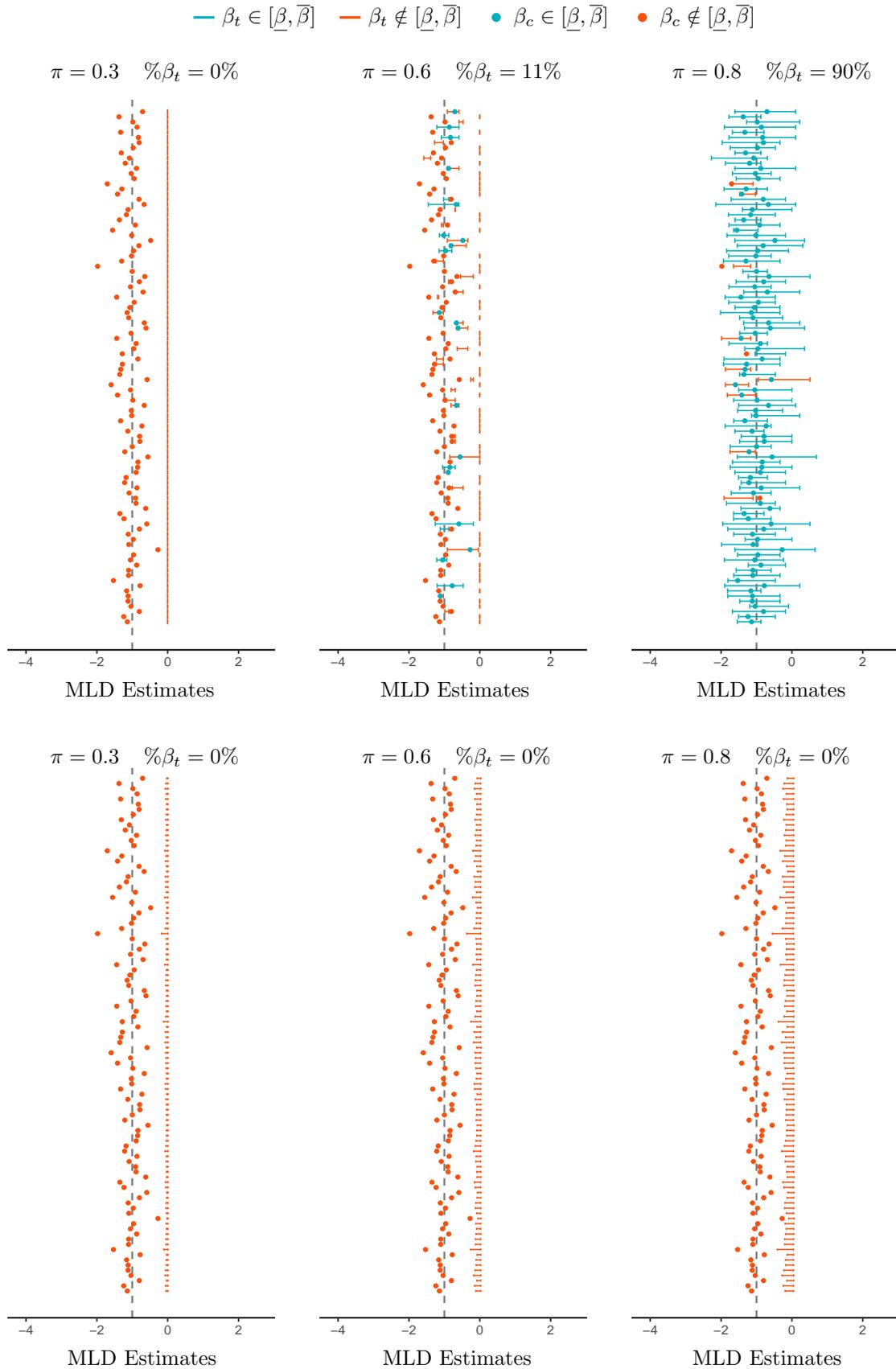


Figure 5.10: MLD estimates for 100 datasets of size 60 using $\pi = 0.3, 0.6, 0.8$ according to the time-based (top) and adjusted probability (bottom) methods, along with the true parameter β_t (dashed line) and the standard PH estimates β_c (circles). Teal color denotes inclusion within the imprecise estimates, while red signifies exclusion.

5.5 Concluding remarks

This chapter has extensively discussed the novel concept of Most Likely Data (MLD) as an imprecise estimation technique, which provides valuable insight into parameter uncertainty and model fitting. Our initial objective was to clarify the inherent limitations of the Maximum Likelihood Estimation (MLE) method, particularly within the Proportional Hazards (PH) model framework, where assumptions such as the Proportional Hazard assumption can lead to inaccurate estimates.

The MLD method has been distinguished by its shift from point estimates to interval estimates, providing a more nuanced understanding of parameter uncertainty when compared to other methods. In our exploration of the application of MLD, particularly with discrete distributions such as Binomial and Poisson and within the PH model, we encountered a number of promising and limiting features.

In the context of discrete distributions, the MLD method applies seamlessly to both the binomial and the Poisson distribution which results in partitioning the parameter space for distinct data and providing a close-form to identifying the imprecise estimates. In spite of this, both faced challenges in deriving closed-form solutions for flexible imprecise estimates, especially when the imprecision effect grew substantial. Similarly, within the PH model, applying the standard MLD method based on marginal or adjusted marginal probabilities failed to effectively partition the parameter space, restricting its utility to only the extreme data in which the covariate values exhibited monotonic characteristics.

Furthermore, our exploration of the time-based MLD approach for the PH model revealed promising improvements; unfortunately, certain datasets were unlikely to be observed, making comprehensive estimations challenging. Nevertheless, while addressing these limitations, the chapter also introduced a flexible MLD method, offering a more adaptive approach to imprecise estimation. This is particularly beneficial in complex models like the PH model. By widening the scope within the parameter space, the flexible MLD method showed promise in mitigating constraints encountered with standard MLD approaches as well as the time-based MLD method.

This chapter aiming to pave the way for advanced methodologies adept at balancing imprecision and reliability in statistical inference as the MLD has some ad-

vantages over conventional models, particularly in scenarios where PH assumptions are questionable. However, to ensure its effectiveness and reliability, future investigations must establish a definitive methodology for determining the appropriate level of imprecision. Despite its potential, further research is needed to address limitations such as effectively partitioning the parameter space. Overall, MLD shows promise for significant contributions to statistical inference techniques, calling for continued exploration and development in this field.

Chapter 6

Conclusions

This chapter summarizes the main findings of this thesis and discusses a few areas for future research. The primary objective of this thesis is to explore opportunities for relaxing the proportional hazards (PH) assumption and to pave the way for further investigation and development in incorporating imprecision into the PH model within the field of statistical inference. This thesis investigates three major contributions: the development of novel imprecise proportional hazards models in two variants, individual-based and group-based, which utilize Poisson empirical likelihoods to handle imprecision. Additionally, a robust PH model has been developed for survival data with continuous covariates; this model incorporates additive errors within covariate values to enhance adaptability by reducing dependency on the proportional hazards assumption. The thesis also introduces the Most Likely Data (MLD) method, a novel imprecise estimation technique that uses interval estimates and prioritizes highly probable data configurations, aiming to refine the statistical inference process in survival analysis. The MLD was implemented in the context of the PH model for survival data with binary covariates. These methodologies have been examined and shown to possess attractive features that enhance the robustness and reliability as alternatives to the standard PH model.

In Chapter 2, we provide a comprehensive overview of key concepts essential to the study's objectives. A description of the PH model is provided, followed by an exploration of parameter estimation and a discussion of baseline survival and

hazard functions. This chapter also provides an overview of various empirical likelihood methods, emphasizing their utility in handling right-censored observations. Furthermore, the chapter provides a detailed overview of methods for generating PH survival data and illustrates how bootstrapping methods are applied. As a result of these concepts, a solid framework is established for the subsequent analysis in the following chapters, framing them as essential tools for the exploration of the thesis.

In Chapter 3, two imprecise proportional hazards models based on Poisson empirical likelihoods are introduced: the individual-based model (IPH) and the group-based imprecise PH model (GPH). The IPH model assigns unique imprecision factors to each individual, while the GPH model allows groups of individuals to share the same imprecision factors, essentially generalizing the IPH model. Initially, attempts were made to construct the full likelihood of these models using empirical likelihood based on the cumulative distribution function (CDF), but due to complexity, the Poisson empirical likelihood was considered instead, particularly advantageous for handling the hazard functions. At zero imprecision levels, these models produce maximum likelihood estimates (MLE) of the partial likelihood for the PH model, with parameter estimates converging to zero and log-likelihood values increasing as imprecision levels rise. Estimation of lower and upper survival functions for specific individuals in these models was approached through restricted and unrestricted survival functions, revealing disparities between lower and upper survival estimates. When imprecision levels reach infinity, distinctive patterns emerge in hazard estimates for both models. Specifically, the upper the GPH restricted and unrestricted hazard estimates. The GPH upper restricted hazard estimates reduce to Nelson-Aalen estimates, while the GPH upper unrestricted hazard estimates converge to Breslow estimates for the baseline hazard function of the PH model. Bootstrap studies were conducted to assess the benefits of using the GPH model and the impact of increasing imprecision levels for both proportional and non-proportional hazards data, with findings suggesting the GPH model can be a safe alternative to the PH model in scenarios where the validity of the PH assumption is uncertain. Surprisingly, the likelihood values evident to has higher benifets when fitting the GPH

model for survival data with a valid PH assumption.

In Chapter 4, we introduce the robust PH model as an alternative to the standard PH model, particularly in cases where the PH assumption regarding a continuous covariate is doubtful. By incorporating errors into covariate values, the robust PH model aims to overcome limitations associated with the proportional hazards assumption. The method involves modifying observed covariate values with error terms, which fluctuate within a small interval determined by the level of imprecision, ϵ^* . The robust model is constructed based on Poisson and Empirical full likelihoods, and when the imprecision level is zero, the robust model is reduced to the standard PH model. However, as imprecision increases, estimated parameters deviate from zero. The chapter explores the estimation of imprecise hazard and survival functions through two methods: the naive approach and the envelope approach. Simulation studies were conducted to examine the impact of covariate effects on survival estimates and compare the robust PH model's estimation capabilities versus a standard PH model in the context of measurement errors. Moreover, bootstrap investigations aim to identify the optimal level of imprecision, indicating that no feasible results from increasing imprecision levels. This appears to be linked to the size of the samples used in these investigations. Overall, the chapter provides a basis for further exploration of methodologies to determine appropriate levels of uncertainty.

In Chapter 5, we delved into the Most Likely Data method as an imprecise estimation technique, offering valuable insights into parameter uncertainty and model fitting. It begins by addressing the MLE limitations, particularly within the PH model framework. The MLD method stands out for its shift from point to interval estimates, providing a nuanced understanding of parameter uncertainty. While exploring its application in various contexts, including discrete distributions like Binomial and Poisson and within the PH model, both promising and limiting features are uncovered. Challenges arise in deriving closed-form solutions for flexible imprecise estimates, especially with substantial imprecision effects. Despite promising improvements with the time-based MLD approach for the PH model, certain datasets

pose challenges for comprehensive estimations. However, the chapter introduces a flexible MLD method, offering an adaptive approach to imprecise estimation, particularly beneficial in complex models like the PH model. Future investigations must establish methodologies for determining appropriate imprecision levels to ensure effectiveness and reliability. Despite limitations, MLD holds promise for significant contributions to statistical inference techniques, prompting further exploration and development in this field.

In future research, conducting simulation studies could offer valuable insights into comparing the accuracy of various methods such as the Individual-Based Proportional Hazards (IPH), Group-Based Proportional Hazards (GPH), robust PH, and Most Likely Data (MLD) for PH approaches in estimating true survival functions, both under proportional hazards (PH) and non-proportional hazards (NPH) scenarios. Moreover, employing a double bootstrap technique may help mitigate biases arising from small to moderate-sized datasets. Addressing a key limitation of this thesis, the Kolmogorov-Smirnov test could serve as an alternative to bootstrap methods for evaluating the efficacy of the introduced models and determining optimal levels of imprecision. This iterative process, gradually increasing the degree of imprecision until minimal deviation is observed in survival estimates from Nelson-Aalen estimates, holds promise for any of the proposed methods. Additionally, survival and hazard estimates for imprecise estimates generated by flexible MLD or flexible time-based MLD methods for the PH model could be straightforwardly derived using Kalbfleisch-Prentice or Breslow estimates. Lastly, investigating the impact of right-censored observations on MLD methods for the PH model could further enhance the applicability of MLD techniques to real-world datasets.

Appendix A

Additional materials

A.1 η for PH and NPH data

In this appendix, we continue the our examples of Section 3.4.1 related to the impact of including small ϵ^* to the the value of $\ell_{\epsilon^*} - \ell_0$ for PH and NPH data. Unlike Examples 3.4.1 and 3.4.2, the following example examines different covariate patterns for imbalanced groups.

Example A.1.1 In order to gain a deeper understanding of the disparity in log-likelihood improvement between the original NPH data sets and their bootstrap samples due to imprecision, this example examines the effect of small ϵ^* values on the incremental changes in log-likelihood values when fitting the GPH model to both PH and NPH data that exhibit an imbalance in group sizes. Specifically, we focus on scenarios where the number of individuals in group one is twice and four times larger than the number of individuals in group zero.

Consider fitting the GPH model to PH survival data that follow the same pattern of covariate as in Example 3.4.1, but with two times as many 1's as 0's. Although this analysis can be conducted using various PH data patterns, we are considering using similar pattern of covariates to maintain consistency in these examples. For PH data with the following patterns of covariates

$$\underbrace{1, 1, 0}_{1}, \dots, \underbrace{1, 1, 0}_{m}$$

x	n	$\hat{\beta}$	η	η/n	PH test
$\overbrace{1, 1, 0, \dots, 1, 1, 0}^1$	21	0.34	8.021	0.382	0.80
$\overbrace{1, 1, 1, 1, 0, 0, \dots, 1, 1, 1, 1, 0, 0}^m$	201	0.05	87.57	0.436	0.81
$\{\frac{m}{3m}, \frac{m}{3m-1}, \frac{2m-2}{3m-2}, \frac{m-1}{3m-3}, \frac{m-1}{3m-4}, \frac{2m-4}{3m-5}, \dots, \frac{1}{3}, \frac{1}{2}, 0\}$	1998	0.006	885.7	0.443	0.89
$\overbrace{1, 1, 1, 1, 0, 0, \dots, 1, 1, 1, 1, 0, 0}^1$	12	1.17	3.151	0.263	0.72
$\overbrace{1, 1, 1, 1, 0, 0, \dots, 1, 1, 1, 1, 0, 0}^m$	204	0.09	88.3	0.433	0.67
$\{\frac{2m}{6m}, \frac{2m}{6m-1}, \frac{2m}{6m-2}, \frac{2m}{6m-3}, \frac{4m-4}{6m-4}, \dots, \frac{2}{6}, \frac{2}{5}, \frac{2}{4}, \frac{2}{3}, 0, 0\}$	1998	0.01	885.2	0.443	0.79
$\overbrace{1, 1, 0, 0, 1, 1, \dots, 1, 1, 0, 0, 1, 1}^1$	18	-0.132	7.07	0.393	0.57
$\overbrace{1, 1, 0, 0, 1, 1, \dots, 1, 1, 0, 0, 1, 1}^m$	198	-0.011	86.59	0.437	0.89
$\{\frac{2m}{6m}, \frac{2m}{6m-1}, \frac{4m-2}{6m-2}, \frac{4m-2}{6m-3}, \frac{2m-2}{6m-4}, \dots, \frac{2}{6}, \frac{2}{5}, \frac{2}{4}, \frac{2}{3}, 0, 0\}$	1998	-0.001	886.08	0.443	0.97

Table A.1: The impact on η for PH data using different covariates patterns with two times as many 1's as 0's

x	n	$\hat{\beta}$	η	η/n	PH test
$\overbrace{1, \dots, 1}^{2m}, \overbrace{0, \dots, 0}^{2m}, \overbrace{1, \dots, 1}^{2m}$	18	-0.56	6.34	0.352	5×10^{-4}
$\overbrace{1, \dots, 1}^{2m}, \overbrace{0, \dots, 0}^{2m}, \overbrace{1, \dots, 1}^{2m}$	198	-0.63	73.7	0.372	2.1×10^{-37}
$\overbrace{1, \dots, 1}^{2m}, \overbrace{0, \dots, 0}^{2m}, \overbrace{1, \dots, 1}^{2m}$	1998	-0.64	747.57	0.374	0
$\overbrace{1, \dots, 1}^{4m}, \overbrace{0, \dots, 0}^{3m}, \overbrace{1, \dots, 1}^{2m}$	18	0.004	6.78	0.377	7e-04
$\overbrace{1, \dots, 1}^{4m}, \overbrace{0, \dots, 0}^{3m}, \overbrace{1, \dots, 1}^{2m}$	198	-0.039	78.95	0.399	0
$\overbrace{1, \dots, 1}^{4m}, \overbrace{0, \dots, 0}^{3m}, \overbrace{1, \dots, 1}^{2m}$	1998	-0.043	801.01	0.401	0
$\overbrace{1, \dots, 1}^{2m}, \overbrace{0, \dots, 0}^{2m}, \overbrace{1, \dots, 1}^{4m}, \overbrace{0, \dots, 0}^m$	18	0.545	6.056	0.336	0.22
$\overbrace{1, \dots, 1}^{2m}, \overbrace{0, \dots, 0}^{2m}, \overbrace{1, \dots, 1}^{4m}, \overbrace{0, \dots, 0}^m$	198	0.649	69.85	0.353	0
$\overbrace{1, \dots, 1}^{2m}, \overbrace{0, \dots, 0}^{2m}, \overbrace{1, \dots, 1}^{4m}, \overbrace{0, \dots, 0}^m$	1998	0.661	708.07	0.354	0

Table A.2: The impact on η for NPH data using different covariates patterns with two times as many 1's as 0's

x	n	$\hat{\beta}$	η	η/n	PH test
	20	0.62	4.94	0.247	0.71
$\overbrace{1, 1, 1, 1, 0}^1, \dots, \overbrace{1, 1, 1, 1, 0}^m$	200	0.09	62.25	0.311	0.72
	2000	0.01	637.88	0.319	0.84
	20	1.32	3.58	0.179	0.55
$\overbrace{1, 1, 1, 1, 1, 1, 1, 1, 0, 0}^1, \dots, \overbrace{1, 1, 1, 1, 1, 1, 1, 1, 0, 0}^m$	200	0.17	61.28	0.306	0.52
	2000	0.02	636.96	0.318	0.7
	20	-0.19	5.74	0.287	0.41
$\overbrace{1, 1, 1, 1, 0, 0, 1, 1, 1, 1}^1, \dots, \overbrace{1, 1, 1, 1, 0, 0, 1, 1, 1, 1}^m$	200	-0.02	63.03	0.315	0.85
	2000	-0.002	638.67	0.319	0.95

Table A.3: The impact on η for PH data using different covariates patterns with four times as many 1's as 0's.

$$\overbrace{1, 1, 1, 1, 0, 0}^1, \dots, \overbrace{1, 1, 1, 1, 0, 0}^m$$

$$\overbrace{1, 1, 0, 0, 1, 1}^1, \dots, \overbrace{1, 1, 0, 0, 1, 1}^m$$

Table A.1 indicates that as the sample size n increases, the log-likelihood increment for these non-balanced PH data converges to approximately 44% of the sample size. Similarly, when examining PH data with covariates consisting of m -blocks following specific patterns, such as $(1, 1, 0, 0, 0, 1, 1, 1, 1)$, $(1, 1, 1, 1, 0, 0, 0, 1, 1)$, and $(1, 0, 0, 0, 1, 1, 1, 1, 1)$, consistent results are observed. Table A.2 reveals that the log-likelihood increments for a range of non-proportional hazards (NPH) data, with double the number of 1s as 0s, converge to 35 – 40% as the sample size increases. These results illustrates that the log-likelihood gains achieved by fitting the GPH model with a small epsilon to NPH survival data are smaller compared to their PH data counterparts.

The same technique was employed to examine the effects of fitting the GPH model to PH and NPH data, where one group has four times the number of members as the other. In line with previous results, a significant decline in the log-profile

x	n	$\hat{\beta}$	η	η/n	PH test
	20	-0.462	5.47	0.274	0.003
$\overbrace{1, \dots, 1}^{4m}, \overbrace{0, \dots, 0}^{2m}, \overbrace{1, \dots, 1}^{4m}$	200	-0.508	57.78	0.289	0
	2000	-0.513	581.1	0.291	0
	15	0.064	4.16	0.277	0.016
$\overbrace{1, \dots, 1}^{8m}, \overbrace{0, \dots, 0}^{3m}, \overbrace{1, \dots, 1}^{4m}$	195	0.0262	58.14	0.298	0
	1995	0.023	598.2	0.299	0
	15	0.602	3.42	0.228	0.31
$\overbrace{1, \dots, 1}^{4m}, \overbrace{0, \dots, 0}^{2m}, \overbrace{1, \dots, 1}^{8m}, \overbrace{0, \dots, 0}^m$	195	0.777	46.59	0.239	0
	1995	0.797	478.45	0.24	0

Table A.4: The impact on η for NPH data using different covariates patterns with four times as many 1's as 0's.

likelihood increments of the GPH model with small Epsilon for PH data, reaching approximately 32% of the sample size when n increases, as shown in Table A.3. Moreover, Table A.4 illustrated that the log-likelihood gain for NPH data was reduced to 24 – 30% of the sample size. Despite the higher log-profile likelihood increments for PH data compared to NPH data, the disparity decreases as the data become more imbalanced.

A.2 Proof of Theorem 2.5.2

Here we profile out F_0 , the nuisance parameter. For simplicity, assume the absence of ties in t_i 's such that $t_1 < t_2 < \dots < t_n$. Following Owen [67], the empirical likelihood approach is used to parametrize Equation (2.58) by restricting all possible MLE of F_0 to the distribution functions which places probability masses p_i only on the observed time points t_i and the interval (t_n, ∞) , see Figure (A.1). Now, let $p_i = F_0(t_i) - F_0(t_i^-)$ for $i = 1, 2, \dots, n$; p_{n+1} represents the probability mass of F_0 in the interval (t_n, ∞) , and the distribution function $F(t) = \sum_{i=1}^n p_i 1_{(t_i \leq t)}$

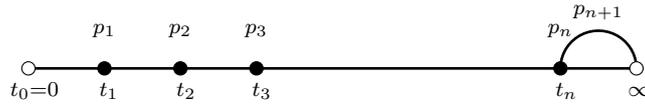


Figure A.1: Probability masses

such that $\sum_{i=1}^{n+1} p_i = 1$, since the empirical cumulative distribution function, ECDF, has been proven to be the non-parametric maximum likelihood estimator, NPMLE [68, 77, 92]. The baseline survival function is given by $S_0(t_i) = 1 - F_0(t_i) = 1 - \sum_{j=1}^n p_j 1_{(t_j \leq t_i)}$ and since $\sum_{i=1}^{n+1} p_i = 1$, then $S_0(t_i) = \sum_{j=i+1}^{n+1} p_j$. Consequently, the full likelihood in Equation (2.58) can be represented as follows

$$L(\beta, F_0) = \prod_{i=1}^n (\phi_i p_i)^{\delta_i} \left(\sum_{j=i+1}^{n+1} p_j \right)^{\phi_i - \delta_i} \quad (1.1)$$

Let $b_{i+1} = \sum_{j=i+1}^{n+1} p_j$, then the properties hold:

1. $b_i = \sum_{j=i}^{n+1} p_j$
2. $b_1 = \sum_{i=1}^{n+1} p_i = 1$
3. $b_{n+1} = p_{n+1}$
4. $b_{i+1} = \sum_{j=i+1}^{n+1} p_j = \sum_{j=i}^{n+1} p_j - p_i = b_i - p_i$
5. let $a_i = p_i / b_i$, then $b_{i+1} = (1 - (p_i / b_i))b_i = (1 - a_i)b_i$

Employing Property (5), p_i and b_{i+1} can be expressed as $p_i = a_i b_i$ and $b_{i+1} = (1 - a_i)b_i$. Substituting these expressions into Equation (1.1) leads to the following likelihood function

$$\begin{aligned} L(\beta, F_0) &= \prod_{i=1}^n \left[(\phi_i a_i b_i)^{\delta_i} ((1 - a_i)b_i)^{\phi_i - \delta_i} \right] \\ &= \prod_{i=1}^n \left[(\phi_i a_i)^{\delta_i} b_i^{\delta_i + \phi_i - \delta_i} (1 - a_i)^{\phi_i - \delta_i} \right] \\ &= \prod_{i=1}^n \left[\left(\frac{\phi_i a_i}{1 - a_i} \right)^{\delta_i} \right] \prod_{i=1}^n \left[b_i^{\phi_i} (1 - a_i)^{\phi_i} \right] \end{aligned} \quad (1.2)$$

In accordance with properties (2) and (5), the second product in Equation (1.2) can be determined by

$$\begin{aligned}
\prod_{i=1}^n (b_i(1-a_i))^{\phi_i} &= (b_1(1-a_1))^{\phi_1} \times (b_2(1-a_2))^{\phi_2} \times \cdots \times (b_n(1-a_n))^{\phi_n} \\
&= (1-a_1)^{\phi_1} \times ((1-a_1)(1-a_2))^{\phi_2} \times ((1-a_1)(1-a_2)\cdots(1-a_n))^{\phi_n} \\
&= (1-a_1)^{\sum_{i=1}^n \phi_i} (1-a_2)^{\sum_{i=2}^n \phi_i} \cdots (1-a_n)^{\phi_n} \\
&= \prod_{i=1}^n (1-a_i)^{r_i}
\end{aligned} \tag{1.3}$$

This result obtained by multiplying the contribution of all i 's and rearranging the likelihood function with reference to $1-a_i$, for $i=1, 2, \dots, n$, using $r_i = \sum_{i=1}^n \phi_i$. The empirical full likelihood function for the PH model can be rewritten by substituting Equation (1.3) into Equation (1.2) as follows

$$L(\beta, F_0) = \prod_{i=1}^n [\phi_i a_i]^{\delta_i} [1-a_i]^{r_i-\delta_i} \tag{1.4}$$

Let the log full likelihood of (β, F_0) denoted by $\ell = \ln L(\beta, F_0)$, then we have

$$\begin{aligned}
\ell &= \sum_{i=1}^n \ln [(\phi_i a_i)^{\delta_i} (1-a_i)^{r_i-\delta_i}] \\
&= \sum_{i=1}^n [\delta_i \ln(\phi_i a_i) + (r_i - \delta_i) \ln(1-a_i)]
\end{aligned} \tag{1.5}$$

The gradient of ℓ is determined with respect to each fixed a_i where $1 \leq i \leq n$ to identify the critical values. Consequently, the gradient is set zero and solve for a_i as follows

$$\begin{aligned}
\frac{\partial \ell}{\partial a_i} &= \delta_i \frac{\phi_i}{\phi_i a_i} + (r_i - \delta_i) \frac{-1}{1-a_i} \\
&= \frac{\delta_i}{a_i} - \frac{r_i - \delta_i}{1-a_i} \\
\frac{\partial \ell}{\partial a_i} = 0 &\Rightarrow \frac{\delta_i}{a_i} - \frac{r_i - \delta_i}{1-a_i} = 0 \\
&\Rightarrow \delta_i(1-a_i) - (r_i - \delta_i)a_i = 0 \\
&\Rightarrow \delta_i - \delta_i a_i - r_i a_i + \delta_i a_i = 0 \\
&\Rightarrow \delta_i - r_i a_i = 0 \\
&\Rightarrow \hat{a}_i = \frac{\delta_i}{r_i}; \forall 1 \leq i \leq n
\end{aligned} \tag{1.6}$$

The estimates \hat{a}_i 's are evidently valid provided that $r_i \neq 0$. To consider the estimation $\hat{a}_i = \frac{\delta_i}{r_i}$ as the maximum estimation of a_i 's, the Hessian matrix, \mathbf{H} , is examined to determine if it is positive or negative semidefinite. **Note:**

- The $n \times n$ Hessian matrix called positive semidefinite \Rightarrow for any nonzero vector \mathbf{z} with n real numbers $\mathbf{z}^T \mathbf{H} \mathbf{z} > 0$; Convex.
- The $n \times n$ Hessian matrix called negative semidefinite \Rightarrow for any nonzero vector \mathbf{z} with n real numbers $\mathbf{z}^T \mathbf{H} \mathbf{z} < 0$; Concave.

The second partial derivative of the log full likelihood for any i is given by

$$\begin{aligned} \frac{\partial^2 \ell}{\partial a_i^2} &= \frac{-\delta_i}{a_i^2} - \frac{r_i - \delta_i}{(1 - a_i)^2} \\ &= \left[\frac{\delta_i}{(1 - a_i)^2} - \frac{\delta_i}{a_i^2} \right] - \frac{r_i}{(1 - a_i)^2} \end{aligned} \quad (1.7)$$

Therefore, the second partial derivatives $\frac{\partial^2 \ell}{\partial a_l a_i} = 0$ for any $l \neq i$. Hence, Hessian matrix can be represented by

$$\begin{aligned} \mathbf{H} &= \begin{bmatrix} \frac{\partial^2 \ell}{\partial a_1^2} & \frac{\partial^2 \ell}{\partial a_2 a_1} & \cdots & \frac{\partial^2 \ell}{\partial a_n a_1} \\ \frac{\partial^2 \ell}{\partial a_1 a_2} & \frac{\partial^2 \ell}{\partial a_2^2} & \cdots & \frac{\partial^2 \ell}{\partial a_n a_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell}{\partial a_1 a_n} & \frac{\partial^2 \ell}{\partial a_2 a_n} & \cdots & \frac{\partial^2 \ell}{\partial a_n^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{-\delta_1}{a_1^2} - \frac{r_1 - \delta_1}{(1 - a_1)^2} & 0 & \cdots & 0 \\ 0 & \frac{-\delta_2}{a_2^2} - \frac{r_2 - \delta_2}{(1 - a_2)^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{-\delta_n}{a_n^2} - \frac{r_n - \delta_n}{(1 - a_n)^2} \end{bmatrix} \end{aligned}$$

It is necessary to confirm that for any nonzero vector $\mathbf{z}^T = (z_1, z_2, \dots, z_n)$ with $z_i \in \mathbb{R}$ for $1 \leq i \leq n$, the condition $\mathbf{z}^T \mathbf{H} \mathbf{z} < 0$ holds, as follows

$$\begin{aligned} \mathbf{z}^T \mathbf{H} \mathbf{z} &= \left(z_1 \left[\frac{-\delta_1}{a_1^2} - \frac{r_1 - \delta_1}{(1 - a_1)^2} \right], \dots, z_n \left[\frac{-\delta_n}{a_n^2} - \frac{r_n - \delta_n}{(1 - a_n)^2} \right] \right) \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \\ &= z_1^2 \left[\frac{-\delta_1}{a_1^2} - \frac{r_1 - \delta_1}{(1 - a_1)^2} \right] + \cdots + z_n^2 \left[\frac{-\delta_n}{a_n^2} - \frac{r_n - \delta_n}{(1 - a_n)^2} \right] \end{aligned} \quad (1.8)$$

To satisfy the maximization requirement we need $\frac{-\delta_i}{a_i^2} - \frac{r_i - \delta_i}{(1-a_i)^2} < 0$ for all $1 \leq i \leq n$. However, two cases will be discussed depend on the value of $\delta_i = 0$ or 1 .

Case 1: $\delta_i = 0$

$$\begin{aligned} \frac{-\delta_i}{a_i^2} - \frac{r_i - \delta_i}{(1-a_i)^2} &= \frac{-0}{\hat{a}_i^2} - \frac{r_i - 0}{(1-\hat{a}_i)^2} \\ &= \frac{-r_i}{(1-\hat{a}_i)^2} < 0; \quad \text{only when } r_i > 0 \end{aligned}$$

This is satisfied since $r_i = \sum_{j=i}^n \phi_j$ and it is known that ϕ_j always positive, so $\mathbf{z}^T \mathbf{H} \mathbf{z} < 0$ in this case.

Case 2: $\delta_i = 1 \Rightarrow \hat{a}_i = \frac{\delta_i}{r_i} = \frac{1}{r_i}$

$$\begin{aligned} \frac{-\delta_i}{a_i^2} - \frac{r_i - \delta_i}{(1-a_i)^2} &= \frac{-1}{\left(\frac{1}{r_i}\right)^2} - \frac{r_i - 1}{\left(1 - \frac{1}{r_i}\right)^2} \\ &= -r_i^2 + \frac{1 - r_i}{\left(1 - \frac{1}{r_i}\right)^2} \\ &= \frac{-r_i^2 \left(1 - \frac{1}{r_i}\right)^2 + 1 - r_i}{\left(1 - \frac{1}{r_i}\right)^2} \\ &= \frac{-r_i^2 \left(1 - \frac{2}{r_i} + \frac{1}{r_i^2}\right) + 1 - r_i}{\left(1 - \frac{1}{r_i}\right)^2} \\ &= \frac{-r_i^2 + \frac{2r_i^2}{r_i} - \frac{r_i^2}{r_i^2} + 1 - r_i}{\left(1 - \frac{1}{r_i}\right)^2} \\ &= \frac{-r_i^2 + 2r_i - 1 + 1 - r_i}{\left(1 - \frac{1}{r_i}\right)^2} \\ &= \frac{r_i - r_i^2}{\left(1 - \frac{1}{r_i}\right)^2} < 0; \quad \text{only when } r_i - r_i^2 = r_i(1 - r_i) < 0 \end{aligned}$$

Either $r_i < 0$ which contradicted the fact that r_i always positive. Thus, $1 - r_i$ must be negative which implies $r_i > 1$; that is, $r_i = \sum_{j=i}^n \phi_j = \phi_i + \phi_{i+1} + \dots + \phi_n > 1$, so this constraint can be satisfied when $\phi_n \geq 1$ since $r_i > \phi_n \geq 1$ for all $1 \leq i \leq n$. We can conclude that $\mathbf{z}^T \mathbf{H} \mathbf{z} < 0$ under the condition that $\phi_n \geq 1$. Consequently, $\hat{a}_i = \frac{\delta_i}{r_i}$ for all $1 \leq i \leq n$ maximizes the full likelihood function in Equation (1.4) under the constraint that $\phi_n \geq 1$. As a result, the full profile likelihood function of β based on the PH model, L_p , can be presented as follows by

$$\begin{aligned}
 L_p(\beta) &= \prod_{i=1}^n [\phi_i \hat{a}_i]^{\delta_i} [1 - \hat{a}_i]^{r_i - \delta_i} \\
 &= \prod_{i=1}^n \left[\phi_i \frac{\delta_i}{r_i} \right]^{\delta_i} \left[1 - \frac{\delta_i}{r_i} \right]^{r_i - \delta_i} \\
 &= \prod_{i=1}^n L_{pl}(\beta) \left[\frac{r_i - \delta_i}{r_i} \right]^{r_i - \delta_i}
 \end{aligned} \tag{1.9}$$

such that $L_{pl}(\beta) = \prod_{i=1}^n \left[\frac{\phi_i}{r_i} \right]^{\delta_i}$ is the partial likelihood function of β driven by Cox [25].

Before optimizing the log-likelihood function of Equation (1.9), it is essential to ensure the constraint $\phi_n \geq 1$ is valid. Ren and Zhou [77] suggests that extensive simulation studies indicate that setting $x_n = 0$, $\phi_n = 1$, results in stable performance. This can be attained by re-coding the covariate x for each individual such that $\tilde{x}_i = x_i - x_n$. Consequently, the MLE of the full profile log-likelihood function, $\hat{\beta}_e$, can be utilized to determine the survival estimates for individuals with $x = x_n$, as follows

$$\begin{aligned}
 \hat{S}_n(t_i) &= \hat{b}_{i+1} \\
 &= (1 - \hat{a}_1)(1 - \hat{a}_2) \dots (1 - \hat{a}_i) \\
 &= \prod_{l \leq i} 1 - \hat{a}_l \\
 &= \prod_{l \leq i} 1 - \frac{\delta_l}{r_l} \\
 &= \prod_{l \leq i} \frac{r_l - \delta_l}{r_l}
 \end{aligned} \tag{1.10}$$

The baseline survival function corresponds to $x = 0$, $S_0(t_i)$, can be estimated by $\hat{S}_n^{exp[-\hat{\beta}x_n]}$. If we are interested in the survival function corresponds to the mean of x , then it can be estimated by $\hat{S}_n^{exp[\hat{\beta}(\bar{x} - x_n)]}$.

A.3 Robust PH model (Poisson)

<i>time</i>	<i>status</i>	$h_{0;c}(t)$	$h_{0;p}^{\epsilon^*=0.1}(t)$	$h_{0;p}^{\epsilon^*=0.5}(t)$	$h_{0;p}^{\epsilon^*=1}(t)$	$h_{0;p}^{\epsilon^*=1.2}(t)$
1.496	1	2.189	2.781	7.560	90.395	502.441
1.709	1	2.351	2.996	8.193	104.184	643.284
2.010	1	3.233	4.363	17.071	224.244	1466.894
3.143	1	3.692	5.068	22.239	459.799	3344.994
3.774	1	3.711	5.091	22.297	460.103	3345.970
4.401	1	5.133	7.234	35.424	829.427	6602.518
4.511	1	5.609	7.976	41.363	1167.048	11288.112
7.132	1	6.142	8.818	48.805	1795.379	19298.865
8.508	1	6.301	9.049	50.156	1847.893	19939.613
8.966	1	6.847	9.917	58.250	2513.504	28491.089
11.360	1	8.357	12.219	74.447	3418.904	40709.971
12.703	1	8.666	12.686	77.825	3632.817	44020.387
12.936	1	11.465	17.114	114.360	6071.674	80096.517
13.208	0	0	0	0	0	0
14.120	1	12.575	18.920	134.047	8300.937	115809.730
17.588	1	14.937	22.582	165.134	11347.746	167443.174
17.601	1	15.525	23.492	173.008	12086.431	180835.103
24.446	1	18.573	28.206	209.483	15542.930	241057.943
26.669	0	0	0	0	0	0
29.893	1	27.003	41.812	339.153	29180.386	485892.780
38.340	0	0	0	0	0	0
38.628	1	37.688	59.025	504.106	48249.129	911048.042
48.983	1	64.511	105.323	1094.517	143016.336	3219747.784
58.177	1	92.473	153.224	1666.332	281857.432	7609017.234
65.542	0	0	0	0	0	0
65.648	1	103.079	170.304	1840.406	315415.415	8704825.458
71.607	0	0	0	0	0	0
79.959	0	0	0	0	0	0
219.825	1	206.218	332.016	3364.963	559459.487	15684171.752
244.929	1	2225.212	4186.390	92064.839	90133044.247	8074690681.285

Table A.5: Estimates of the baseline hazard functions resulted by fitting the PH model and the robust PH model with $\epsilon^* = 0.1, 0.5, 1, 1.2$ using the survival data with $n=30$ and negative β as in Example (4.3.1)

<i>time</i>	<i>status</i>	$h_{0;c}(t)$	$h_{0;p}^{\epsilon^*=0.1}(t)$	$h_{0;p}^{\epsilon^*=0.5}(t)$	$h_{0;p}^{\epsilon^*=1}(t)$	$h_{0;p}^{\epsilon^*=1.2}(t)$
9.30×10^{-3}	1	7.88×10^{-5}	4.95×10^{-5}	4.67×10^{-6}	2.01×10^{-8}	4.67×10^{-10}
6.55×10^{-2}	1	8.27×10^{-5}	5.20×10^{-5}	4.91×10^{-6}	2.16×10^{-8}	5.05×10^{-10}
8.48×10^{-2}	1	9.06×10^{-5}	5.72×10^{-5}	5.58×10^{-6}	2.96×10^{-8}	8.48×10^{-10}
1.16×10^{-1}	1	1.14×10^{-4}	7.40×10^{-5}	8.82×10^{-6}	5.07×10^{-8}	1.58×10^{-9}
1.19×10^{-1}	1	1.16×10^{-4}	7.57×10^{-5}	9.02×10^{-6}	5.16×10^{-8}	1.61×10^{-9}
1.31×10^{-1}	1	1.47×10^{-4}	9.86×10^{-5}	1.24×10^{-5}	7.34×10^{-8}	2.37×10^{-9}
1.33×10^{-1}	1	2.33×10^{-4}	1.62×10^{-4}	2.43×10^{-5}	2.09×10^{-7}	8.22×10^{-9}
1.97×10^{-1}	1	2.53×10^{-4}	1.77×10^{-4}	2.75×10^{-5}	2.63×10^{-7}	1.14×10^{-8}
2.15×10^{-1}	1	2.66×10^{-4}	1.87×10^{-4}	2.94×10^{-5}	2.87×10^{-7}	1.26×10^{-8}
2.88×10^{-1}	1	2.90×10^{-4}	2.04×10^{-4}	3.36×10^{-5}	3.73×10^{-7}	1.86×10^{-8}
2.92×10^{-1}	1	4.28×10^{-4}	3.11×10^{-4}	5.84×10^{-5}	8.43×10^{-7}	4.83×10^{-8}
3.75×10^{-1}	1	4.87×10^{-4}	3.59×10^{-4}	7.52×10^{-5}	1.35×10^{-6}	8.84×10^{-8}
4.54×10^{-1}	1	6.35×10^{-4}	4.75×10^{-4}	1.06×10^{-4}	2.18×10^{-6}	1.61×10^{-7}
5.36×10^{-1}	1	8.18×10^{-4}	6.19×10^{-4}	1.45×10^{-4}	3.51×10^{-6}	2.95×10^{-7}
6.03×10^{-1}	1	9.04×10^{-4}	6.92×10^{-4}	1.73×10^{-4}	5.45×10^{-6}	5.40×10^{-7}
6.18×10^{-1}	1	9.54×10^{-4}	7.31×10^{-4}	1.86×10^{-4}	6.16×10^{-6}	6.43×10^{-7}
6.54×10^{-1}	1	9.81×10^{-4}	7.52×10^{-4}	1.92×10^{-4}	6.34×10^{-6}	6.63×10^{-7}
7.26×10^{-1}	0	0	0	0	0	0
8.30×10^{-1}	1	1.10×10^{-3}	8.49×10^{-4}	2.17×10^{-4}	7.32×10^{-6}	8.03×10^{-7}
9.87×10^{-1}	1	1.20×10^{-3}	9.25×10^{-4}	2.45×10^{-4}	9.30×10^{-6}	1.06×10^{-6}
1.07	0	0	0	0	0	0
1.13	0	0	0	0	0	0
1.14	1	2.14×10^{-3}	1.66×10^{-3}	4.41×10^{-4}	1.53×10^{-5}	1.65×10^{-6}
1.28	1	2.26×10^{-3}	1.75×10^{-3}	4.66×10^{-4}	1.61×10^{-5}	1.73×10^{-6}
2.80	1	2.78×10^{-3}	2.14×10^{-3}	5.74×10^{-4}	2.05×10^{-5}	2.23×10^{-6}
3.02	1	3.07×10^{-3}	2.38×10^{-3}	6.48×10^{-4}	2.38×10^{-5}	2.66×10^{-6}
3.89	0	0	0	0	0	0
3.97	0	0	0	0	0	0
4.40	1	2.25×10^{-2}	1.92×10^{-2}	8.90×10^{-3}	1.16×10^{-3}	2.63×10^{-4}
5.31	0	0	0	0	0	0

Table A.6: Estimates of the baseline hazard functions resulted by fitting the PH model and the robust PH model with $\epsilon^* = 0.1, 0.5, 1, 1.2$ using the survival data with $n=30$ and positive β as in Example (4.3.1)

A.4 Bootstrap investigation (Chapter 4)

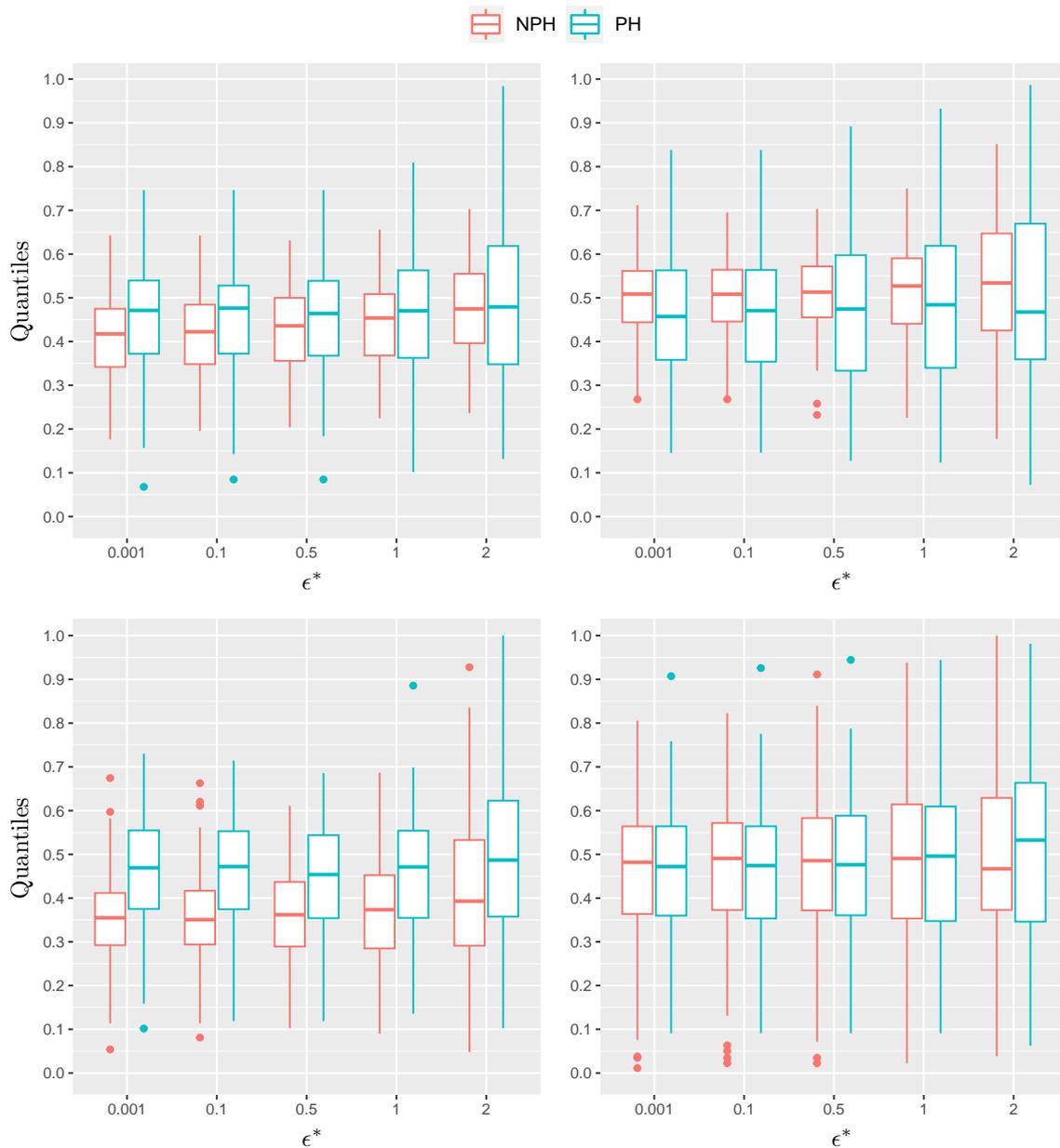


Figure A.2: Quantiles of $\hat{\ell}_{\epsilon^*}^*$ (left) and $(\hat{\ell}_{\epsilon} - \hat{\ell}_0)^*$ (right) based on Zelterman's bootstrap for PH, time-dependent-NPH (top) and frailty-NPH (bottom) survival data with 20% of right-censored observations using the robust Poisson PH model on fitted to $m = 100$ data sets each with $N = 60$ observations.

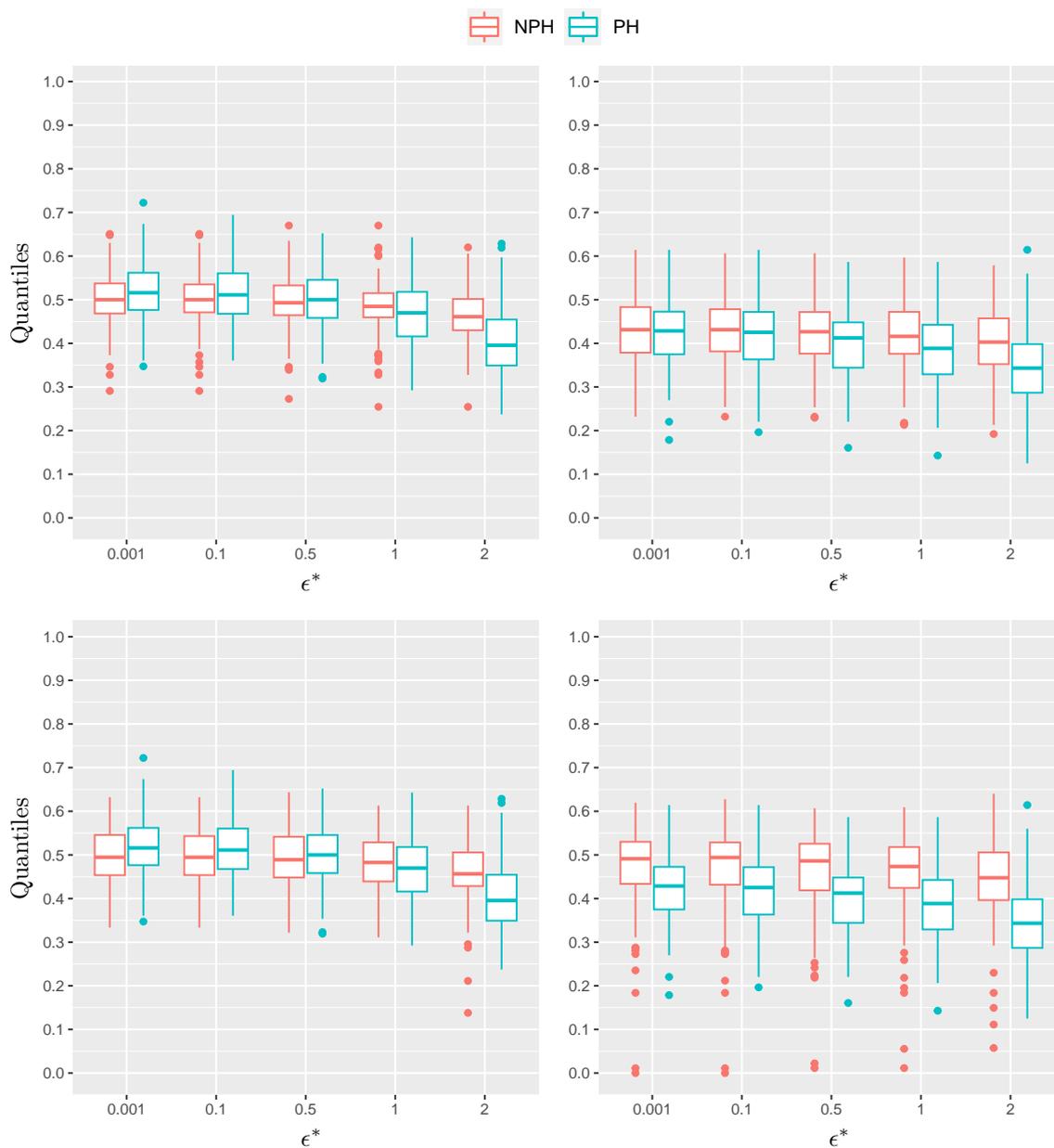


Figure A.3: Quantiles of $\hat{\ell}_{\epsilon^*}^*$ (left) and $(\hat{\ell}_{\epsilon} - \hat{\ell}_0)^*$ (right) based on Hjort's bootstrap for PH, time-dependent-NPH (top) and frailty-NPH (bottom) survival data with 20% of right-censored observations using the robust Poisson PH model on fitted to $m = 100$ data sets each with $N = 60$ observations.

Appendix B

R codes

B.1 Simulating PH and NPH survival data

```
## simulate Weib.PH right-censored data (continuous covariate)
simWeib.PH.Cont = function(N, beta, lambda, rho, Cens_par)
{ # covariate
  x = runif(N,10, 30)
  # Event times (Weibull/Exponential)
  v = runif(n=N)
  V = (-log(v) / (lambda * exp(x * beta)))^(1/rho)
  Fv=ecdf(V)
  C=unname(quantile(Fv, rbeta(N,Cens_par[1],Cens_par[2])))
  status=as.numeric(V <= C)
  time = pmin(V, C)
  Data<-data.frame("time"=time, "status"=status, x)
  Data<-Data[order(Data[,1],-Data[,2]),]
  i1 <- duplicated(Data[,1])
  Data[i1,1]<-Data[i1,1]+cumsum(rep(min(abs(diff(Data[,1]))[diff(
    Data[,1])!=0]))/(length(Data[,1])^2),sum(i1)))
  return(Data)
}

# simulate Weib crossing hazards NPH right-censored data
simWeib.NPH.Bi <- function(N, lambda, rho, Cens_par)
{ v1 <- runif(n=as.integer(N/2))
  V1 <- (-log(v1) / (lambda[1]))^(1/rho[1])
```

```

Fv1<-ecdf(V1)
C1<-unname(quantile(Fv1, rbeta(as.integer(N/2), Cens_par[1], Cens_
  par[2])))
status1<-as.numeric(V1 <= C1)
time1 <- pmin(V1, C1)
v2 <- runif(n=as.integer(N/2))
V2 <- (-log(v2) / (lambda[2]))^(1/rho[2])
Fv2<-ecdf(V2)
C2<-unname(quantile(Fv2, rbeta(as.integer(N/2), Cens_par[1], Cens_
  par[2])))
status2<-as.numeric(V2 <= C2)
time2 <- pmin(V2, C2)
Data<-data.frame("time"=c(time1, time2), "status"=c(status1,
  status2), x=c(rep(0, as.integer(N/2)), rep(1, as.integer(N/2))))
Data<-Data[order(Data[,1], -Data[,2]),]
i1 <- duplicated(Data[,1])
Data[i1,1]<-Data[i1,1]+cumsum(rep(min(abs(diff(Data[,1]))[diff(
  Data[,1])!=0]))/(length(Data[,1])^2), sum(i1)))
return(Data)
}
# Gompertz Obs. required "VGAM" Package.
#simulate NPH right-censored data with non crossing hazards (
  Weibull&Gompertz) .
simWeib.NPH.Bi.W_G <- function(N, lambda, rho, Cens_par)
{ ## Generate Weibull
  V1=rweibull(N/2, shape=rho[2], scale =lambda[2])
  Fv1<-ecdf(V1)
  C1<-unname(quantile(Fv1, rbeta(as.integer(N/2), Cens_par[1], Cens_
    par[2])))
  status1<-as.numeric(V1 <= C1)
  time1 <- pmin(V1, C1)
  #Generate2 Gompertz
  V2=rgompertz(N/2, scale =lambda[1], shape=rho[1])
  Fv2<-ecdf(V2)
  C2<-unname(quantile(Fv2, rbeta(as.integer(N/2), Cens_par[1], Cens_
    par[2])))
  status2<-as.numeric(V2 <= C2)

```

```

time2 <- pmin(V2, C2)
Data<-data.frame("time"=c(time1,time2), "status"=c(status1,
  status2), x=c(rep(0,as.integer(N/2)),rep(1,as.integer(N/2))))
Data<-Data[order(Data[,1],-Data[,2]),]
i1 <- duplicated(Data[,1])
Data[i1,1]<-Data[i1,1]+cumsum(rep(min(abs(diff(Data[,1]))[diff(
  Data[,1])!=0]))/(length(Data[,1])^2),sum(i1)))
return(Data)
}

```

B.2 IPH model

```

# Log-lik function for the IPH
logL_IPH<-function(initial,s,x,eps){
  epsi<-eps[1]-eps[2]
  v<-exp(initial*x)
  vv<-v*exp(epsi)
  r<-c(rev(cumsum(rev(vv)))[-1],0)
  L<-ifelse(s==0,0,log((v*exp(-1))/(v+(r*s))))
  logl<- sum(L)
  return(-logl)
}

# IPH: fit, imprecise hazard for x_i, imprecise survival for x_i
# x_i is a covariate value for particular individual
# type= 1 for restricted, and 2 for Unrestricted # eps*>=0
IPH<-function(data,eps,x_i,type=1) {
  eps=c(-eps,eps)
  data<-data[order(data[,1],-data[,2]),]
  row.names(data)=NULL
  s=data[,2]
  x=data[,3]
  Fit<-tryCatch({optim(0, fn=logL_IPH, s=s, x=x, eps=eps, method
= "BFGS", control = list(maxit=1000))$par}, error = function(e)
{0})
  if(Fit==0) {eps=c(0,-Inf)}
  phi_all=exp(Fit*x)
  phi_x=exp(Fit*x_i)
}

```

```

r=c(rev(cumsum(rev(phi_all)))[-1],0)
h.denominator=phi_all*exp(max(eps))+exp(min(eps))*r
h.t_j=ifelse(s==0,0,1/h.denominator)
if (type==1) {
  u.eps_i.tj=ifelse(x==x_i,max(eps),min(eps))
  u.h_x_t_j=h.t_j*exp(u.eps_i.tj)*phi_x
  l.eps_i.tj=rep(min(eps),nrow(data))
  l.h_x_t_j=h.t_j*exp(l.eps_i.tj)*phi_x
  Sx=exp(-apply(cbind(u.h_x_t_j,l.h_x_t_j),2,cumsum))
  return(list(Fit,data.frame(l.h_x=l.h_x_t_j,u.h_x=u.h_x_t_j),
data.frame(LS.R=Sx[,1],SU.R=Sx[,2])))
} else {
  u.eps_i.tj=rep(max(eps),nrow(data))
  u.h_x_t_j=h.t_j*exp(u.eps_i.tj)*phi_x
  l.eps_i.tj=rep(min(eps),nrow(data))
  l.h_x_t_j=h.t_j*exp(l.eps_i.tj)*phi_x
  Sx=exp(-apply(cbind(u.h_x_t_j,l.h_x_t_j),2,cumsum))
  return(list(Fit,data.frame(l.h_x=l.h_x_t_j,u.h_x=u.h_x_t_j),
data.frame(LS.R=Sx[,1],SU.R=Sx[,2])))
}
}

```

B.3 GPH model

```

require(fastDummies,spatstat.utils)
# fastDummies for dummy_cols and spatstat.utils for revcumsum
# Log-lik function for the GPH
logL_GPH<-function(initial,s,x,eps){
  v=exp(initial*x)
  x.dum <- dummy_cols(x,remove_selected_columns = TRUE)[-1]
  r<-if (sum(dim(x.dum)[2])==0) { data.frame(r1=c(rev(cumsum(rev
((1-x.dum)*v))), r2=c(rev(cumsum(rev(x.dum*v))))))
} else {data.frame(X0=rev(cumsum(rev((1-rowSums(x.dum))*v))),
sapply(seq.int(sum(dim(x.dum)[2])), function(X,xv) rev(cumsum(
rev(xv[,X]))), xv=x.dum*v))}
  eps.r<-exp(sapply(seq_len(ncol(r)), function (X,x,eps) ifelse(x
==X-1,max(eps),min(eps)),x,eps))

```

```

L.denominator<-exp(min(eps))*rowSums(eps.r * r)
L=v/L.denominator
logl<-sum(ifelse(s==0,0,log(L)))
return(-logl)
}
# GPH: fit, imprecise hazard for x_i, imprecise survival for x_i
# x_i is a covariate value for particular individual
# type= 1 for restricted, and 2 for Unrestricted # eps*>=0
GPH<-function(data,eps,x_i,type=1) {
  eps=c(-eps,eps)
  data<-data[order(data[,1],-data[,2]),]
  row.names(data)=NULL
  s=data[,2]
  x=data[,3]
  Fit<-tryCatch({optim(0, fn=logL_GPH, s=s, x=x, eps=eps, method
= "BFGS", control = list(maxit=1000))$par}, error = function(e)
{0})
  phi_all=exp(Fit*x)
  phi_x=exp(Fit*x_i)
  v=exp(rowSums(t(Fit*t(x))))
  x.dum <- dummy_cols(x,remove_selected_columns = TRUE)[,-1]
  r<-if (sum(dim(x.dum)[2])==0) { data.frame(r1=c(rev(cumsum(rev
((1-x.dum)*v))), r2=c(rev(cumsum(rev(x.dum*v))))))
} else {data.frame(X0=rev(cumsum(rev((1-rowSums(x.dum))*v))),
sapply(seq.int(sum(dim(x.dum)[2])), function(X,xv) rev(cumsum(
rev(xv[,X]))), xv=x.dum*v))}
  if(Fit==0) {eps=c(0,-Inf)}
  eps.r<-exp(sapply(sort(unique(x)), function (X,x,eps) ifelse(x
==X,max(eps),min(eps)),x,eps))
  h.denominator<-rowSums(eps.r * r)
  h.t_j=ifelse(s==0,0,1/h.denominator)
  if (type==1) {
    u.eps_i.tj=ifelse(x==x_i,max(eps),min(eps))
    u.h_x_t_j=h.t_j*exp(u.eps_i.tj)*phi_x
    l.eps_i.tj=rep(min(eps),nrow(data))
    l.h_x_t_j=h.t_j*exp(l.eps_i.tj)*phi_x
    hAll=data.frame(Lh.R=l.h_x_t_j,Uh.R=u.h_x_t_j)
  }
}

```

```

h_x_t_j=data.frame(apply(hAll,1,min),apply(hAll,1,max))
S_x_t_j=exp(-apply(h_x_t_j,2,cumsum))
return(list(Fit,h_x_t_j,data.frame(LS.R=S_x_t_j[,2],SU.R=S_x_
t_j[,1])))} else {
u.eps_i.tj=rep(max(eps),nrow(data))
u.h_x_t_j=h.t_j*exp(u.eps_i.tj)*phi_x
l.eps_i.tj=rep(min(eps),nrow(data))
l.h_x_t_j=h.t_j*exp(l.eps_i.tj)*phi_x
hAll=data.frame(Lh.R=l.h_x_t_j,Uh.R=u.h_x_t_j)
h_x_t_j=data.frame(apply(hAll,1,min),apply(hAll,1,max))
S_x_t_j=exp(-apply(h_x_t_j,2,cumsum))
return(list(Fit,h_x_t_j,data.frame(LS.UnR=S_x_t_j[,2],SU.UnR=
S_x_t_j[,1])))}
}

```

B.4 Zelterman *et al.* [91]'s Bootstrap

```

##### Functions #####
# The logL_IPH and logL_GPH functions are required.
# Our test statistic can be Le, Le-L0, or (Le-L0)/L0
Le<-function(A,epsilons){
  A<-data.frame(A[order(A[,1],-A[,2]),])
  i1 <- duplicated(A[,1])
  if (sum(i1)>0) A[i1,1]<-A[i1,1]+cumsum(rep(min(abs(diff(A[,1]))[
diff(A[,1])!=0]))/(length(A[,1])^2),sum(i1)))
  tryCatch({
    PL<-optim(0, fn=model, s=A[,2], x=A[,3], eps=c(0,0), method
= "BFGS", control = list(maxit=1000))
    return(sapply(epsilons, FUN = function(X,c)(-optim(c, fn=
model, s=A[,2], x=A[,3], eps=c(-X,X), method = "BFGS", control =
list(maxit=1000))$value),c=PL$par))),
  error = function(e) rep(NA,length(epsilons)),
  warning = function(w) rep(NA,length(epsilons)))
}
Le_L0<-function(A,epsilons){
  A<-data.frame(A[order(A[,1],-A[,2]),])
  i1 <- duplicated(A[,1])

```

```

  if (sum(i1)>0) A[i1,1]<-A[i1,1]+cumsum(rep(min(abs(diff(A[,1]))[
diff(A[,1])!=0]))/(length(A[,1])^2),sum(i1)))
  tryCatch({
    PL<-optim(0, fn=model, s=A[,2], x=A[,3], eps=c(0,0), method
= "BFGS", control = list(maxit=1000))
    return(sapply(epsilons, FUN = function(X,c,b)(-optim(c, fn=
model, s=A[,2], x=A[,3], eps=c(-X,X), method = "BFGS", control =
list(maxit=1000))$value-b),c=PL$par,b=-PL$value)),
    error = function(e) rep(NA,length(epsilons)),
    warning = function(w) rep(NA,length(epsilons)))
  }
Le_L0_adj<-function(A,epsilons){
  A<-data.frame(A[order(A[,1],-A[,2]),])
  i1 <- duplicated(A[,1])
  if (sum(i1)>0) A[i1,1]<-A[i1,1]+cumsum(rep(min(abs(diff(A[,1]))[
diff(A[,1])!=0]))/(length(A[,1])^2),sum(i1)))
  tryCatch({
    PL<-optim(0, fn=model, s=A[,2], x=A[,3], eps=c(0,0), method
= "BFGS", control = list(maxit=1000))
    return(sapply(epsilons, FUN = function(X,c,b)(-optim(c, fn=
model, s=A[,2], x=A[,3], eps=c(-X,X), method = "BFGS", control =
list(maxit=1000))$value-b)/b,c=PL$par,b=-PL$value)),
    error = function(e) rep(NA,length(epsilons)),
    warning = function(w) rep(NA,length(epsilons)))
  }
# Zelterman for complete data
Zel.boot.complete<- function (d,R,epsilons) {
  fit<-coxph(Surv(d[,1L],d[,2L])~d[,3L],data=d, ties = "breslow")
  ex <- exp(fit$coef*c(0,1))
  t<-apply(as.matrix.noquote(data.frame(table(d[d[,2L]>0,1L])))
,2,as.numeric)
  n.e<-sum(t[,2L])
  t.e<-nrow(t)
  x.probs<-c(1-mean(d[,3]),mean(d[,3]))
  t.probs<-t[,2L]/n.e
  epi<-matrix(NA,t.e,2)
  for (i in seq_len(t.e)){

```

```

    d_pi<-colSums(t[,2] * epi, na.rm=TRUE)
    unscaled <- ex*(n.e*x.probs-d_pi)
    #unscaled = pmax(unscaled, 0)
    epi[i,]<-unscaled/sum(unscaled)
  }
  epi<-apply(epi, 2L, function(X) {k<-X[!is.na(.bincode(X, c(0,
1), include.lowest =TRUE))]);
  X[is.na(.bincode(X, c(0, 1), include.lowest =TRUE))]<-k[
length(k)];
  return(X)})
  b.se<-matrix(sample(t[,1L], n.e * R,replace=T), nrow = R)
  x<-matrix(NA,R,n.e)
  for (j in seq_len(t.e)) {
    x[,1:n.e][b.se==t[j,1L]]<-sample(c(0, 1), sum(b.se==t[j,1L
]),replace=TRUE, prob = epi[j,])
  }

  boot.all<-list(b.se,c(rep(1,n.e)),x)
  return(sapply(seq_len(R),FUN=function(X,AA,epsilons) delta(A=
data.frame(AA[[1]][X,],AA[[2]],AA[[3]][X,]),epsilons),AA=boot.
all,epsilons))
}
# Zelterman for only unique censored obs
Zel.boot.unique.cens<- function (d,R,epsilons) {
  fit<-coxph(Surv(d[,1L],d[,2L])~d[,3L],data=d, ties = "breslow")
  ex <- exp(fit$coef*c(0,1))
  t<-apply(as.matrix.noquote(data.frame(table(d[d[,2L]>0,1L])))
,2,as.numeric)
  n.e<-sum(t[,2L])
  t.e<-nrow(t)
  x.probs<-c(1-mean(d[,3]),mean(d[,3]))
  t.probs<-t[,2L]/n.e
  epi<-matrix(NA,t.e,2)
  for (i in seq_len(t.e)){
    d_pi<-colSums(t[,2] * epi, na.rm=TRUE)
    unscaled <- ex*(n.e*x.probs-d_pi)
    #unscaled = pmax(unscaled, 0) # needed to fix negative

```

```

probabilities for contentious covariate
  epi[i,]<-unscaled/sum(unscaled)
}
epi<-apply(epi, 2L, function(X) {k<-X[!is.na(.bincode(X, c(0,
1), include.lowest =TRUE))]);
  X[is.na(.bincode(X, c(0, 1), include.lowest =TRUE))]<-k[
length(k)];
  return(X)})
epin<-epi[t.e,2] # needed to fix NA in censoring probabilities
c<-apply(as.matrix.noquote(data.frame(table(d[d[,2L]==0,1L])))
,2,as.numeric)
n.c<-sum(c["Freq"])
deno.allcpi<-sum(t[t[,1]>=c[1],2]/n.e)
x1.num<-(t[,2]/n.e)*epi[,2]
x1.cpi<-sum(x1.num[t[,1]>=c[1]])/deno.allcpi
x1.cpi[is.na(x1.cpi)]<-epin
cpi<-cbind(1-x1.cpi,x1.cpi)
b.se<-matrix(sample(t[,1L], n.e * R,replace=T), nrow = R)
b.sc<-matrix(c[1L], n.c * R)
stat<-c(rep(1,n.e),rep(0,n.c))
x<-matrix(NA,R,n.e+n.c)
for (j in seq_len(t.e)) {
  x[,1:n.e][b.se==t[j,1L]]<-sample(c(0, 1), sum(b.se==t[j,1L
]),replace=TRUE, prob = epi[j,])
}
x[,n.e+1]<-sample(c(0, 1), R,replace=TRUE, prob = cpi)
boot.all<-list(b.se,b.sc,stat,x)
return(sapply(seq_len(R),FUN=function(X,AA,epsilons) delta(A=
data.frame(c(AA[[1]][X,],AA[[2]][X,]),AA[[3]],AA[[4]][X,]),
epsilons),AA=boot.all,epsilons))
}
# Zelterman for right-censored data
Zel.boot.cens<- function (d,R,epsilons) {
  fit<-coxph(Surv(d[,1L],d[,2L])~d[,3L],data=d, ties = "breslow")
  ex <- exp(fit$coef*c(0,1))
  t<-apply(as.matrix.noquote(data.frame(table(d[d[,2L]>0,1L])))
,2,as.numeric)

```

```

n.e<-sum(t[,2L])
t.e<-nrow(t)
x.probs<-c(1-mean(d[,3]),mean(d[,3]))
t.probs<-t[,2L]/n.e
epi<-matrix(NA,t.e,2)
for (i in seq_len(t.e)){
  d_pi<-colSums(t[,2] * epi, na.rm=TRUE)
  unscaled <- ex*(n.e*x.probs-d_pi)
  #unscaled = pmax(unscaled, 0)
  epi[i,]<-unscaled/sum(unscaled)
}
epi<-apply(epi, 2L, function(X) {k<-X[!is.na(.bincode(X, c(0,
1), include.lowest =TRUE))]});
X[is.na(.bincode(X, c(0, 1), include.lowest =TRUE))]<-k[
length(k)];
return(X)})
epin<-epi[t.e,2] # needed to fix NA in censoring probabilities
c<-apply(as.matrix.noquote(data.frame(table(d[d[,2L]==0,1L])))
,2,as.numeric)
n.c<-sum(c[,2L])
deno.allcpi<-sapply(seq_len(n.c),function(X,a,b,n.e) sum(a[a
[,1]>=b[X,1],2]/n.e), a=t,b=c,n.e)
x1.num<-(t[,2]/n.e)*epi[,2]
x1.cpi<-sapply(seq_len(n.c),function(X,a,b,w,k) sum(w[a[,1]>=b[
X,1]])/k[X], a=t,b=c, w=x1.num,k=deno.allcpi)
x1.cpi[is.na(x1.cpi)]<-epin
cpi<-cbind(1-x1.cpi,x1.cpi)
b.se<-matrix(sample(t[,1L], n.e * R,replace=T), nrow = R)
b.sc<-matrix(sample(c[,1L], n.c * R,replace=T), nrow = R)
stat<-c(rep(1,n.e),rep(0,n.c))
x<-matrix(NA,R,n.e+n.c)
for (j in seq_len(t.e)) {
  x[,1:n.e][b.se==t[j,1L]]<-sample(c(0, 1), sum(b.se==t[j,1L
]),replace=TRUE, prob = epi[j,])
}
for (j in seq_len(n.c)) {
  x[, (n.e+1):(n.e+n.c)][b.sc==c[j,1L]]<-sample(c(0, 1), sum(b

```

```

.sc==c[j,1L]),replace=TRUE, prob = cpi[j,])
  }
  boot.all<-list(b.se,b.sc,stat,x)
  return(sapply(seq_len(R),FUN=function(X,AA,epsilons) delta(A=
data.frame(c(AA[[1]][X,],AA[[2]][X,]),AA[[3]],AA[[4]][X,]),
epsilons),AA=boot.all,epsilons))
}
# Bootstrap main function is to select the appropriate bootstrap
function based on the censoring scheme
Zel.boot.all<- function (d,R,epsilons) {
  if (sum(d[,2])<nrow(d)-1) Zel.boot.cens(d,R,epsilons)
  else {if (sum(d[,2])==nrow(d)-1) Zel.boot.unique.cens(d,R,
epsilons)
      else Zel.boot.complete(d,R,epsilons)}
}
# imp_args contains arguments associated with generated data,
including:
# Cens: F for complete data or T for right censored data given Cens
_par
# Cens_par: vector of two parameters controlling right censoring
proportion and position.
# model: either 'IPH' or 'GPH' which indicates the fitted model.
Consequently, it will call logL_IPH or logL_GPH
# delta: 1,2,3 refer to the evaluated measures Le, Le-L0, or (Le-L0
)/L0, which will call L0, Le, Le_L0, or Le_L0_adj, respectively.
imp_args<-function(Cens,model,delta,Cens_par){
  Cens<-Cens
  model<-ifelse(model=="CIPL",logL_CIPL,logL_CIPI)
  if (delta==1) {
    delta<-Le
    delta_type<-"Le"} else {
  if (delta==2) {
    delta<-Le_L0
    delta_type<-"Le_L0"} else {
    delta<-Le_L0_adj
    delta_type<-"(Le-L0)/L0"}
}

```

```

    Cens_par<-Cens_par
  }

# Example
imp_args(Cens=T,model="GPH",delta=3,Cens_par=c(.4,.1))
epsilons=c(.1,.5,1,2)
set.seed(123)
# Simulate survival data M data sets each with N observations
Data=replicate(M,ifelse(Cens==F,list(simWeib.PH.com(N, beta=-.5,
  lambda=2, rho=3)),list(simWeib.PH(N, beta=-.5, lambda=2, rho=3,
  Cens_par))))
##### Zelterman et al.'s bootstrap #####
Zel.Boot=sapply(seq_len(M), FUN = function (X, Data,epsilons) {
  original<-delta(Data[[X]],epsilons) # calculate gamma* for
  original data sets use different lvls of epsilon*
  Zel.Results=Zel.boot.all(Data[[X]],R,epsilons) # calculating
  gamma for bootstrap samples
  sapply(seq_len(epsilons), function(X,a,b) ecdf(a[X,])(b[X]),a=
  Zel.Results,b=original)}, Data,epsilons)

```

B.5 Robust PH model

```

# required in both RP and RE to determine imprecision terms
New_eps_all<-function(Data, eps, Beta.and.eps) {
  eps.all=rep(NA,nrow(Data))
  eps.all[Data[,2]==0]==-sign(Beta.and.eps[1])*eps
  eps.all[1:(which(Data[,2]==1)[1]-1)]=0
  eps.all[length(eps.all)]==-sign(Beta.and.eps[1])*eps
  eps.all[which(Data[,2]==1)[1]]=sign(Beta.and.eps[1])*eps
  eps.all[is.na(eps.all)]=Beta.and.eps[-1]
  return(c(Beta.and.eps[1],eps.all))
}

# Log-lik function for the PH model
logL_pl=function(par,s,x){
  x=x-mean(x)
  v=exp(par*x)
  r=rev(cumsum(rev(v)))

```

```

L=(v/r)^s
  return(-sum(log(L)))
}
# Log-lik function for the RP model
logL_RP=function(par,s,x,eps){ # This function required pars=sum(
  Data[-nrow(Data),2])-1 as input
  eps.all=rep(NA,length(x))
  eps.all[s==0]=-sign(par[1])*eps
  eps.all[1:(which(s==1)[1]-1)]=0
  eps.all[length(eps.all)]=-sign(par[1])*eps
  eps.all[which(s==1)[1]]=sign(par[1])*eps
  eps.all[is.na(eps.all)]=par[-1]
  eps.all[is.na(eps.all)]=0
  x=x+eps.all
  x=x-mean(x)
  v=exp(par[1]*x)
  r=rev(cumsum(rev(v)))
  L=(v/r)^s
  return(-sum(log(L)))
}

# RP: fit, imprecise hazard for x_i, imprecise survival for x_i
# x_i is a covariate value for particular individual
# type= 1 for naive, and 2 for envelop # eps*>=0
RP<-function(data,eps,x_i,type=1) {
  if(eps==0) {
    BetaPH=optim(0, fn=logL_pl, s=data[,2], x=data[,3], method =
"BFGS", control = list(maxit=1000))$par
    vPH=exp(BetaPH*data[,3])
    rPH=rev(cumsum(rev(vPH)))
    hx=(data[,2]/rPH)*exp(BetaPH*x_i)
    Sx=exp(-cumsum(hx))
    return(list(BetaPH,hx,Sx))
  } else {
    pars=sum(data[-nrow(data),2])-1
    Beta_imp=optim(c(0,rep(0,pars)), fn=logL_RP, s=data[,2], x=
data[,3], method = "L-BFGS-B", eps=eps, lower = c(-Inf,rep(-eps,

```

```

pars)), upper=c(Inf,rep(eps,pars)), control = list(maxit=1000))$
par
  Beta_imp=New_eps_all(data, eps=eps, Beta.and.eps=Beta_imp)
  v=exp(Beta_imp[1]*(data[,3]+Beta_imp[-1]))
  r=rev(cumsum(rev(v)))
  h.imp=(data[,2]/r)
  if (type==1) {
    hx_imp=cbind(h.imp*exp(Beta_imp[1]*(x_i-eps)),h.imp*exp(
Beta_imp[1]*(x_i+eps)))
    if(sign(Beta_imp[1])==1) {lu=c(1,2)} else {lu=c(2,1)}
    hx_imp=hx_imp[,lu]
    colnames(hx_imp)=c("Lhx.naive","Uhx.naive")
    Sx=exp(-apply(hx_imp[,c(2,1)], 2,cumsum))
    colnames(Sx)=c("LSx.naive","USx.naive")
    return(list(Beta_imp,hx_imp,Sx))
  } else {
    BetaPH=optim(0, fn=logL_pl, s=data[,2], x=data[,3], method
= "BFGS", control = list(maxit=1000))$par
    vPH=exp(BetaPH*data[,3])
    rPH=rev(cumsum(rev(vPH)))
    h.=(data[,2]/rPH)
    hx_imp=cbind(h.*exp(BetaPH*x_i),h.imp*exp(Beta_imp[1]*x_i))
    hx_imp=cbind(apply(hx_imp,1,min),apply(hx_imp,1,max))
    colnames(hx_imp)=c("Lhx.envelop","Uhx.envelop")
    Sx=exp(-apply(hx_imp[,c(2,1)], 2,cumsum))
    colnames(Sx)=c("LSx.envelop","USx.envelop")
    return(list(Beta_imp,hx_imp,Sx))
  } }
}

# Log-lik function for the empirical PH model
logL_E=function(par,s,x){
  x=x-x[length(x)]
  v=exp(par*x)
  r=rev(cumsum(rev(v)))
  L=ifelse(r==1,(s*par*x),(s*par*x)-(r*log(r))+((r-s)*(log(r-s))))
  logl<- sum(subset(L,s==1))
}

```

```

    return(-logl)
}

# Log-lik function for the RE model
logL_RE=function(par,s,x,eps){ # This function required pars=sum(
  Data[-nrow(Data),2])-1 as input
  eps.all=rep(NA,length(x))
  eps.all[s==0]=-sign(par[1])*eps
  eps.all[1:(which(s==1)[1]-1)]=0
  eps.all[length(eps.all)]=-sign(par[1])*eps
  eps.all[which(s==1)[1]]=sign(par[1])*eps
  eps.all[is.na(eps.all)]=par[-1]
  eps.all[is.na(eps.all)]=0
  x=x+eps.all
  x=x-x[length(x)]
  v=exp(par[1]*x)
  r=rev(cumsum(rev(v)))
  L=ifelse(r==1,1,((v/r)^s)*(((r-s)/r)^(r-s)))
  return(-sum(log(L)))
}

# RE: fit, imprecise survival for x_i
# x_i is a covariate value for particular individual
# type= 1 for naive, and 2 for envelop # eps*>=0
RE<-function(data,eps,x_i,type=1) {
  BetaPH=optim(0, fn=logL_pl, s=data[,2], x=data[,3], method = "
    BFGS", control = list(maxit=1000))$par
  BetaE=optim(BetaPH, fn=logL_E, s=data[,2], x=data[,3], method = "
    BFGS", control = list(maxit=1000))$par
  if(eps==0) {
    xn=data[nrow(data),3]
    v=exp(BetaE*(data[,3]-xn))
    r=rev(cumsum(rev(v)))
    Sn=cumprod((r-data[,2])/r)
    Sx=Sn^exp((-BetaE*xn)+(BetaE*x_i))
    return(list(BetaE,Sx))
  } else {

```

```

pars=sum(data[-nrow(data),2])-1
Beta_imp=optim(c(BetaE,rep(0,pars)), fn=logL_RE, s=data[,2], x=
data[,3], method = "L-BFGS-B", eps=eps, lower = c(-Inf,rep(-eps,
pars)), upper=c(Inf,rep(eps,pars)), control = list(maxit=1000))$
par
Beta_imp=New_eps_all(data, eps=eps, Beta.and.eps=Beta_imp)
xn_imp=data[nrow(data),3]+Beta_imp[length(Beta_imp)]
v.=exp(Beta_imp[1]*(data[,3]+Beta_imp[-1]-xn_imp))
r.=rev(cumsum(rev(v.)))
Sn_imp=cumprod((r.-data[,2])/r.)
if (type==1) {
  Sx=cbind(Sn_imp^exp((-Beta_imp[1]*xn_imp)+(Beta_imp[1]*(x_i-
eps))),Sn_imp^exp((-Beta_imp[1]*xn_imp)+(Beta_imp[1]*(x_i+eps)))
)
  if(sign(Beta_imp[1])==1) {Sx=Sx[,c(2,1)]}
  colnames(Sx)=c("LSx.naive", "USx.naive")
  return(list(Beta_imp,Sx))
} else {
  xn=data[nrow(data),3]
  v=exp(BetaE*(data[,3]-xn))
  r=rev(cumsum(rev(v)))
  Sn=cumprod((r-data[,2])/r)
  Sx=Sn^exp((-BetaE*xn)+(BetaE*x_i))
  Sx_imp=Sn_imp^exp((-Beta_imp[1]*xn_imp)+(Beta_imp[1]*x_i))
  LU_Sx=cbind(apply(cbind(Sx,Sx_imp),1,min),apply(cbind(Sx,Sx_
imp),1,max))
  colnames(LU_Sx)=c("LSx.envelop", "USx.envelop")
  return(list(Beta_imp,LU_Sx))
}
}
}

```

B.6 MLD Method

```
## MLD for Binomial Distribution
```

```

#eps level of imprecision, n number of trails, x number of
  success, digits precision (opt)
BinMLD=function(eps,n,x,digits=.001){
  p=seq(0,1,digits)
  All=data.frame(p,sapply(X=0:n, function(X,p,n) choose(n, X)* p^
X * (1-p)^(n-X), p, n))
  if(any(eps==0) & length(eps)==1L) {return(c(x/(n+1),(x+1)/(n+1)
))} else {
  if(any(eps==0)) {eps=eps[-which(eps==0)]; eps0=c(0,x/(n+1),(x
+1)/(n+1))}
  min_max=sapply(eps, function(X,All,x) { All[,-c(1,x+2)]=All
[, -c(1,x+2)]*(1-X)
  PCond=p[which(rowSums(All[,x+2])>=All[,-1])==(ncol(All)-1)
][ -1]
  return(c(X,min(PCond),max(PCond)))}, All,x)
  min_max[,eps==1]=c(1,0,1)
  if(exists("eps0")) {min_max=cbind(eps0[1:3],min_max)}
  return(t(min_max))}
}
BinMLD(eps=seq(0,1,by=.1),n=4,x=2)

## MLD for Poisson Distribution
#eps level of imprecision, x number of events, digits precision (
opt), tolerance range of x values to compare with default 10 (i.
e. 1:[x*10] wills be used as alternatives ti x) (opt)
PoisMLD=function(eps,x,digits=0.01, tolerance=10){
  lambda_range <- seq(0, x/2*100, by = digits)[-1] # Range of
lambda values
  if(any(eps==0) & length(eps)==1L) {return(c(x,x+1))} else {
  if(any(eps==0)) {eps=eps[-which(eps==0)]; eps0=c(0,x,x+1)}
  min_max=sapply(eps, function(X,x,lambda_range,tolerance) {
  results <- numeric(length(lambda_range))
  for (i in seq.int(length(results))) {
  # Calculate probabilities for X=x and other values of X
  p_k <- dpois(x,lambda_range[i])
  max_p_m <- max(sapply(seq.int(x*tolerance), dpois, lambda
= lambda_range[i])*(1 - X))

```

```

        # Store result where  $P(X = x)$  is greater than or equal to
         $(1-\pi) * P(X = m)$  for all  $m \neq x$ 
        results[i] <- p_k >= max_p_m
    }
    return(c(X, range(lambda_range[results==1])))
}, x, lambda_range, tolerance)
min_max[, eps==1]=c(1, 0, Inf)
if(exists("eps0")) {min_max=cbind(eps0[1:3], min_max)}
return(t(min_max))
}
}
PoisMLD(eps=c(0, 0.1), x=3)

## MLD for the PH model (Time-based)
# x is a binary covariate vector, obtained after sorting the
# survival observations by time
# This function takes a unique imprecision level only
TimeBasedPHMLD=function(x, eps) {
  r<-data.frame(r1=c(rev(cumsum(rev((1-x))))), r2=c(rev(cumsum(
  rev(x))))))
  Beta_limits=sapply(1:length(x[-1]), function(X,x,r,eps) {
    sign=(x[X]-(1-x[X]))
    limit1=Inf*sign
    limit2=sign*log((1-eps)*(r[X,(1-x[X])+1]/r[X,(x[X]-1)+2]))
    return(c(min(limit1, limit2), max(limit1, limit2)))}
  , x=x, r=r, eps=eps)
  return(c(max(Beta_limits[1,]), min(Beta_limits[2,])))}
TimeBasedPHMLD(x=c(0, 1, 0, 1), eps=0.2)

## MLD for the PH model (marginal probabilities)
# Thesis for only one data x must be ordered according to
# survival time
# The code takes unique level of imprecision, default is eps=0
# The range of the parameter can be control with the precision,
# default is Beta=seq(-2, 2, 0.001)
PHMLD=function(x, Beta=seq(-2, 2, 0.001), eps=0){
  Dominant.x1=c(rep(1, sum(x)), rep(0, length(x)-sum(x)))

```

```
prob.Dominant.Xs1=sapply(Beta,logL_pl,x=Dominant.x)
Max.Prob.Dominants=apply((1-eps)*cbind(rev(prob.Dominant.Xs1),
prob.Dominant.Xs1),1,max)
Prob.x=sapply(Beta,logL_pl,x=x)
MLD=Prob.x>=Max.Prob.Dominants
range(Beta[MLD])
}
PHMLD(x=c(0,1,0,1),eps=0.1)
```

Bibliography

- [1] Aalen O. (1978). Nonparametric Inference for a Family of Counting Processes. *The Annals of Statistics*, 6(4), 701–726.
- [2] Allison P.D. (2010). *Survival Analysis Using SAS: A practical Guide*. SAS Institute, second edi edition.
- [3] Andersen P.K., Borgan Ø., Gill R.D. and Keiding N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- [4] Aslett L., Coolen F.P.A. and Bock J. (2022). *Uncertainty in Engineering Introduction to Methods and Applications: Introduction to Methods and Applications*. Springer, 1st edition.
- [5] Augustin T. (2004). An Exact Corrected Log-Likelihood Function for Cox’s Proportional Hazards Model under Measurement Error and Some Extensions. *Scandinavian Journal of Statistics*.
- [6] Augustin T. and Coolen F.P.A. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, 124(2), 251–272.
- [7] Augustin T., Coolen F.P.A., De Cooman G. and Troffaes M.C. (2014). *Introduction to Imprecise Probabilities*. John Wiley & Sons.
- [8] Augustin T. and Schwarz R. (2002). Cox’s proportional hazards model under covariate measurement error - A review and comparison of methods. In *Total least squares and errors-in-variables modeling : analysis, algorithms and appli-*

- cations*, (Editors) S. Van Huffel and P. Lemmerling, pp. 175–184. Dordrecht: Springer Netherlands.
- [9] Bardo M., Huber C., Benda N., Brugger J., Fellingner T., Galaune V., Heinz J., Heinzl H., Hooker A.C., Klinglmüller F., König F., Mathes T., Mittlböck M., Posch M., Ristl R. and Friede T. (2023). Methods for non-proportional hazards in clinical trials: A systematic review.
- [10] Belforte G., Bona B. and Cerone V. (1987). Bounded measurement error estimates: Their properties and their use for small sets of data. *Measurement*.
- [11] Bender R., Augustin T. and Blettner M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*.
- [12] Bernardini A. and Tonon F. (2010). *Bounding uncertainty in civil engineering: Theoretical background*. Springer Berlin, Heidelberg.
- [13] Breslow (1972). Discussion on Professor Cox’s Paper. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34, 216–217.
- [14] Brilleman S.L., Wolfe R., Moreno-Betancur M. and Crowther M.J. (2020). Simulating Survival Data Using the *simsurv* R Package. *Journal of Statistical Software*, 97(3), 1–27.
- [15] Buonaccorsi J.P. (1996). A modified estimating equation approach to correcting for measurement error in regression. *Biometrika*, 83(2), 433–440.
- [16] Burr D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model. *Journal of the American Statistical Association*, 89(428).
- [17] Buzas J.S. (2021). Covariate Measurement Error in Survival Data. In *Handbook of Measurement Error Models*. Taylor & Francis, first edition.
- [18] Byrd R.H., Lu P., Nocedal J. and Zhu C. (1995). A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16, 1190–1208.

- [19] Carroll R.J., Ruppert D., Stefanski L.A. and Crainiceanu C.M. (2006). *Measurement error in nonlinear models: A modern perspective*. Taylor & Francis, second edition.
- [20] Castañon E., Sanchez-Arreaez A., Alvarez-Manceñido F., Jimenez-Fonseca P. and Carmona-Bayonas A. (2020). Critical reappraisal of phase III trials with immune checkpoint inhibitors in non-proportional hazards settings. *European Journal of Cancer*, 136, 159–168.
- [21] Collett D. (2014). *Modelling survival data in medical research*. CRC Press, 3 edition.
- [22] Coolen F.P.A. and Augustin T. (2009). A nonparametric predictive alternative to the Imprecise Dirichlet Model: The case of a known number of categories. *International Journal of Approximate Reasoning*, 50(2).
- [23] Coolen F.P.A. and Yan K.J. (2004). Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference*, 126(1).
- [24] Coolen-Schrijner P., Coolen F.P.A., Troffaes M.M. and Augustin T. (2009). Imprecision in statistical theory and practice. *Journal of Statistical Theory and Practice*, 3(1).
- [25] Cox D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society. Series B:*, 34(2), 187–220.
- [26] Cox D.R. (1975). Partial Likelihood. *Biometrika*, 62(2), 269–276.
- [27] Cox D.R. and Oakes D. (1984). *Analysis of survival data*, volume 5. Chapman and Hall.
- [28] Dencœur T. (2016). 40 years of Dempster–Shafer theory. *International Journal of Approximate Reasoning*, 79.
- [29] Efron B. (1977). The efficiency of cox’s likelihood function for censored data. *Journal of the American Statistical Association*, 72(359).

- [30] Efron B. (1981). Censored Data and the Bootstrap. *Journal of the American Statistical Association*, 76(374), 312.
- [31] Efron B. and Gong G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician*, 37(1).
- [32] Efron B. and Tibshirani R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1).
- [33] E.L. Kaplan P.M. (1958). Nonparametric Estimation from Incomplete Observations Author (s): E . L . Kaplan and Paul Meier Source : Journal of the American Statistical Association , Vol . 53 , No . 282 (Jun . , 1958), pp . 457- Published by : American Statistical Association Sta. *American statistical Association*, 53(282), 457–481.
- [34] Escobar L.A. and Meeker W.Q. (1992). Assessing Influence in Regression Analysis with Censored Data. *Biometrics*, 48(2), 507–528.
- [35] Ferson S., Kreinovich V., Ginzburg L., Myers D.S. and Sentz K. (2003). Constructing Probability Boxes and Dempster-Shafer Structures. Technical Report January, Sandia National Laboratories.
- [36] Fleming T.R. and Harrington D.P. (2005). *Counting Processes and Survival Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- [37] Grambsch P.M. and Therneau T.M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*.
- [38] Hall P. and Scala B.L. (1990). Methodology and Algorithms of Empirical Likelihood. *International Statistical Review / Revue Internationale de Statistique*, 58(2), 109–127.
- [39] Hess K.R. (1995). Graphical methods for assessing violations of the proportional hazards assumption in cox regression. *Statistics in Medicine*.

- [40] Hosmer D.W., Lemeshow S. and May S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. John Wiley & Sons, second edition.
- [41] Hougaard P. (1995). Frailty models for survival data. *Lifetime Data Analysis*.
- [42] Hu C. and Lin D.Y. (2002). Cox regression with covariate measurement error. *Scandinavian Journal of Statistics*.
- [43] Hu P., Tsiatis A.A. and Davidian M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics*, 54(4), 1407–1419.
- [44] Johansen S. (1978). The Product Limit Estimator as Maximum Likelihood Estimator. *Statistics*, 6(2).
- [45] Kalbfleisch A.J.D. and Prentice R.L. (1973). Marginal Likelihoods Based on Cox's Regression and Life Model. *Biometrika*, 60(2), 267–278.
- [46] Kalbfleisch J.D. and Prentice R.L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- [47] Kiefer J. and Wolfowitz J. (1956). Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. *The Annals of Mathematical Statistics*, 27(4).
- [48] Klein J., van Houwelingen H., Ibrahim J. and Scheike T. (2013). *Handbook of Survival Analysis*. Chapman and Hall/CRC, first edit edition.
- [49] Klein J.P. and Moeschberger M.L. (2003). *Survival Analysis: Techniques for censored and Truncated Data*. Springer, second edition.
- [50] Kleinbaum D.G.v. and Klein M. (2012). *Survival analysis : a self-learning text*. New York : Springer, 3rd ed. edition.
- [51] Kofler E. (2001). Linear partial information with applications. *Fuzzy Sets and Systems*, 118(1).

- [52] Kong F.H. and Gu M. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statistica Sinica*.
- [53] Kuitunen I., Ponkilainen V.T., Uimonen M.M., Eskelinen A. and Reito A. (2021). Testing the proportional hazards assumption in cox regression and dealing with possible non-proportionality in total joint arthroplasty research: methodological perspectives and review. *BMC Musculoskeletal Disorders*, 22(1), 489.
- [54] Lee E.T. and Go O.T. (1997). Survival analysis in public health research.
- [55] Lee E.T. and Wang J.W. (2003). *Statistical methods for survival data analysis*. New York: Wiley.
- [56] Levi I. (1980). *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. 2. Cambridge: MIT Press.
- [57] Li G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statistics and Probability Letters*.
- [58] Li J. and Ma S. (2013). *Survival Analysis in Medicine and Genetics*. Chapman & Hall, 1st edition edition.
- [59] Lin D.Y. (2003). Survival Analysis.
- [60] Lin D.Y. and Ying Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*.
- [61] Lu Y., Fang J., Tian L. and Jin H. (2015). *Advanced Medical Statistics*. World Scientific, 2 edition.
- [62] Murphy S.A. (1995). Likelihood ratio-based confidence intervals in survival analysis. *Journal of the American Statistical Association*.
- [63] Murphy S.A. and Van Der Vaart A.W. (2000). On profile likelihood. *Journal of the American Statistical Association*.

- [64] Nakamura T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics*, 48(3), 829–838.
- [65] Nelson W. (1972). Theory and Applications of Hazard Plotting for Censored Failure Data. *Technometrics*, 14(4), 945–966.
- [66] N.L. Hjort (1985). Bootstrapping Cox’s Regression Model.
- [67] Owen A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2).
- [68] Owen A.B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC.
- [69] Pan X.R. (1997). *Empirical likelihood ratio Method for censored data*. Ph.d. dissertation, University of Kentucky.
- [70] Prentice R.L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2), 331–342.
- [71] Qin J. and Lawless J. (1994). Empirical Likelihood and General Estimating Equations. *The Annals of Statistics*, 22(1), 300 – 325.
- [72] R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [73] R Core Team (2020) (2020). *R: A language and environment for statistical computing*.
- [74] Rahman R., Fell G., Ventz S., Arfé A., Vanderbeek A.M., Trippa L. and Alexander B.M. (2019). Deviation from the proportional hazards assumption in randomized phase 3 clinical trials in oncology: Prevalence, associated factors, and implications. In *Clinical Cancer Research*.
- [75] Rao P.V., Hosmer D.W. and Lemeshow S. (2000). Applied Survival Analysis: Regression Modeling of Time to Event Data. *Journal of the American Statistical Association*.
- [76] Reid N. (1981). Estimating the median survival time. *Biometrika*, 68(3).

- [77] Ren J.J. and Zhou M. (2011). Full likelihood inferences in the Cox model: an empirical likelihood approach. *Annals of the Institute of Statistical Mathematics*, 63(5), 1005–1018.
- [78] Royston P. and Parmar M.K.B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21(15), 2175–2197.
- [79] Ruggeri F., Ríos Insua D. and Martín J. (2005). Robust Bayesian analysis. In *In D. Dey and C. Rao, editors, Handbook of Statistics. Bayesian Thinking: Modeling and Computation*, volume 25. Elsevier,.
- [80] Stefanski L.A. (1989). Unbiased Estimation of a Nonlinear Function of a Normal Mean with Application to Measurement Error Models. *Communications in Statistics - Theory and Methods*.
- [81] Therneau T.M. (1999). A Package for Survival Analysis in S. *Mayo Clinic Foundation*, pp. 70–73.
- [82] Therneau T.M. (2022). A Package for Survival Analysis in R.
- [83] Therneau T.M. (2024). *A Package for Survival Analysis in R*. R package version 3.7-0.
- [84] Therneau T.M. and Grambsch P.M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer.
- [85] Thomas D.R. and Grunkemeier G.L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, 70(352).
- [86] Variyath A.M., Chen J. and Abraham B. (2010). Empirical likelihood based variable selection. *Journal of Statistical Planning and Inference*, 140(4), 971–981.

-
- [87] Walley P. (1991). Statistical Reasoning with Imprecise Probabilities. In *Monographs on Statistics & Applied Probability*, volume 42. Chapman & Hall.
- [88] Wei L.J. (1992). The accelerated failure time model: A useful alternative to the cox regression model in survival analysis. *Statistics in Medicine*.
- [89] Weichselberger K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2-3).
- [90] Xu R. and Adak S. (2002). Survival analysis with time-varying regression effects using a tree-based approach. *Biometrics*.
- [91] Zelterman D., Le C.T. and Louis T.A. (1996). Bootstrap techniques for proportional hazards models with censored observations. *Statistics and Computing*, 6(3).
- [92] Zhou M. (2015). *Empirical Likelihood Method in Survival Analysis*. New York: Chapman and Hall/CRC, 1st edition.