Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Estimation of the Hazard in Patients with Myocardial Infarction Using A Shape-Constrained Estimator

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

IVONNE MARTIN

Delft, the Netherlands
December 2010

TU Delft

MSc THESIS APPLIED MATHEMATICS

"Estimation of the Hazard in Patients with Myocardial Infarction Using A Shape-Constrained Estimator"

IVONNE MARTIN

Delft University of Technology

Daily supervisor

Dr. H. P Lopuhaä

Responsible professor

Prof. Dr. ir. G. Jongbloed

Other thesis committee members

MSc. ir. N. van Geloven

Dr. D. Kurowicka

Prof. J.G.P. Tijssen

December 2010

Delft, the Netherlands

MSc THESIS APPLIED MATHEMATICS

"Estimation of the Hazard in Patients with Myocardial Infarction Using A Shape-Constrained Estimator"

IVONNE MARTIN

Delft University of Technology

AMC Clinical Research Unit

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In general, this thesis is aiming at predicting the survival probability of patients with heart disease. Heart disease is considered to be one of the mortal diseases that is caused by several factors. Among them, the most common cause is atherosclerosis, which is an accumulation of a plaque in the artery wall as a result of excessive cholesterol in the blood stream. Details about this disease will be explained in Section 1.1. Since this disease is one of the leading causes of death, people have been trying to improve the medication. A lot of research in different areas has been conducted for this purpose.

Typically, research conducted from a mathematical point of view is to investigate the survival probability of a patient having this disease. For instance, by comparing the survival probabilities of patients with different medications. We would like to estimate the curve of the survival probability function. This would describe the survivability of patient with a certain medication over time. Several estimators for describing the estimate of the survival probability will be described in Chapter 2. In this chapter an overview about the disease itself and the purpose of writing this thesis will be given. At the end of this chapter, an outline of this thesis is given.

1.1 Acute Coronary Syndrome (ACS)

In general, a human's heart consists of heart muscles and arteries. It pumps oxygenated blood to the whole body through blood vessels. In normal arteries, blood flows without any obstruction. In some cases, we may find a narrowing of artery wall which results in the obstruction of blood flow. For instance, in the case of atherosclerosis. The obstruction of blood flow in the cardiac artery wall will raise symptoms called ACS, which is a life-threatening manifestation of atherosclerosis characterized by the symptoms chest pain (radiating to the left arm or jaw),nausea and vomiting. It is associated with an increased risk of cardiac death.

In Figure 1.1s, an illustration of the accumulation process in the wall artery until it results in ACS is given. In this picture, we see 6 cross sections of the artery. The first cross section represents a normal artery. In cross section 2, there is a lesion initiation and extracellular lipid in the artery wall as a result of a high amount of cholesterol in the blood stream. This will grow into a plaque which clogs and hardens the artery, which makes an obstacle to the blood flow (see the third and fourth cross section). The situation can worsen if the plaque ruptures (see the fifth cross section), stimulating thrombogenesis and thus forming a thrombus (blood clot). Thrombus formation can completely or partially occlude the blood vessel and reduce blood flow. Blood flow reduction (and consequently oxygen reduction) can cause ischemic discomfort which

is known as angina pectoris. The sudden process of plaque disruption and subsequent thrombus formation is known as ACS.

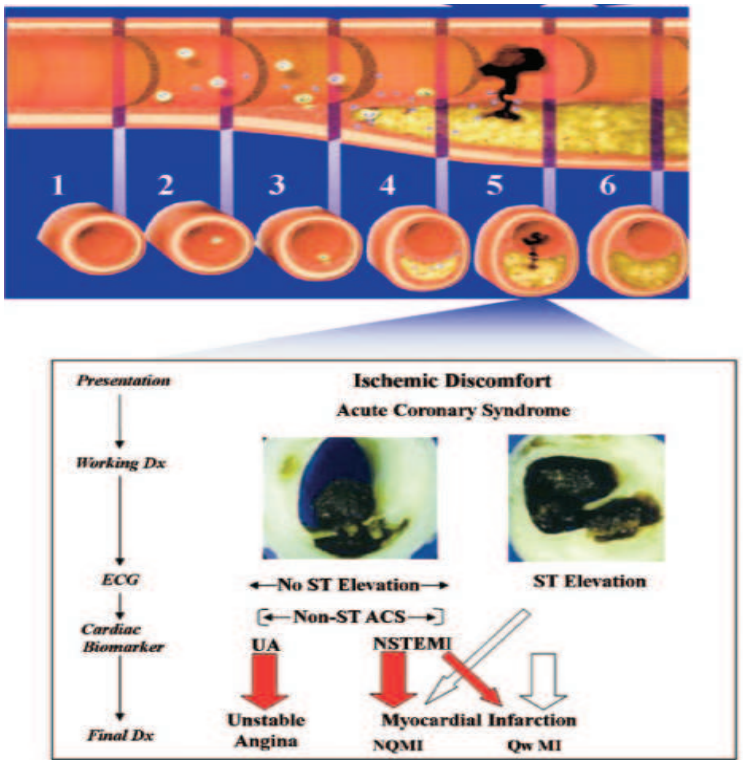

Figure 1.1: An Illustration about the process in human coronary artery until becoming an ACS. *Source:* Circulation, Journal of American Heart Association.

Patients with ACS are classified into two types based on the graphical interpretation of the Electrocardiogram. If it shows an ST-segment elevation then it can be classified as ST-Elevation Myocardial Infarction (STEMI). In the absence of such an elevation, it can be categorized as NSTEMI or Unstable Angina. While STEMI is associated with a total occlusion of the coronary artery, NSTEMI and angina pectoris is associated with partial occlusion.

1.2 Objective of the Thesis

The research work presented in this thesis is about predicting the survival probability of patients with heart disease. In particular, the work is focused on a method to improve the approach to estimate the survival probability and studying the effects of applying the new method. This work is done in collaboration with Department of Cardiology, Academisch Medisch Centrum (AMC) Amsterdam, the Netherlands. In this section, it will be described why such a prediction of the survival probability is important in medical applications as well as the reason why it is needed to be improved.

When a person is diagnosed with a certain life-threatening disease, it is natural to question how large is the probability that he/she will survive. There are different interests which motivate the calculation of this probability. One interest comes from medical applications. For instance, to compare the efficacy of treatments given to patients with this disease. This will be explained shortly.

It is not common in practice to assume that the survival probability follows a certain parametric distribution. Rather, we deal with a non-parametric estimator and the most well known non-parametric

estimator is the product limit estimator (also known as Kaplan-Meier estimator). The distribution function obtained in this way is merely an estimate of the true survival distribution function.

It may be the case that certain characteristics of the patients, for instance the age of patient, gender type, smoking habits and so on, affect the survival probability. If this is the case, one has to formulate a regression survival model that describes the survival probability and its relationship with several covariates. The most well-known regression survival model is the proportional hazard model introduced by Cox(1972). This model is somewhere in between a parametric and nonparametric model and is an example of a so-called semiparametric model. It still assumes a specific structure on the survival distribution, namely proportionality of the hazards of different individuals, which can be described by a finite number of parameters, but it also contains an unknown baseline hazard as an infinite dimensional component.

In the Cox model, the proportional hazard assumption can be used extensively to compare the efficacy of several treatments given to patients with certain disease. This assumption means that the hazard between patients with different covariates (for instance, treatment) are just a constant multiplication of one another. By this, we can draw conclusions about which treatment gives the worst hazard values. Take for example two patients with heart disease. One patient is given treatment A and the other is given treatment B. In this case, the Cox model will result in different hazard functions for those individuals. Based on this information, a conclusion about the efficacy of treatment A and treatment B can be compared.

In this thesis, survival regression is applied to patients from the ICTUS trial[1]. It consists of 1200 patients with heart disease in the AMC hospital in Amsterdam, The Netherlands. These patients are randomized into two treatment groups and the event of interest is the death of a patient or development of a myocardial infarction. In this particular situation, there is empirical evidence and biological reasoning that suggest the hazard of patients to be decreasing over time. This additional information is not incorporated in the traditional estimators, such as the Kaplan-Meier estimator or the estimators for the parameters in the Cox model. Thus, for that purpose a new estimator that incorporates the shape information will be introduced.

Having applied the new estimator, we would like to investigate the behavior of the new estimator and compare it with the traditional estimators. The investigation will include the differences in the survival curves between the traditional and the new estimator.

1.3 Outline of the Thesis

The thesis is organized as follows.

In Chapter 2, the problem is introduced. Some theoretical background, which is needed in survival analysis and survival regression, will be given. In this chapter, an illustration of the cumulative hazard and the hazard itself is given, such that reader can see the shortcomings of the traditional estimator.

In Chapter 3, a shape constrained estimator is introduced and explained in more detail. Applying this estimator will raise the question whether this estimator is reasonable for the dataset. For this, a statistical test will be conducted and the bootstrap method is used to compute the p-value.

In Chapter 4, we will see the effect on applying a shape constrained estimator in the Cox model. Observations in this chapter are intended to investigate whether the shape constrained estimator is a better estimator compared to the traditional one especially in estimating the parameters in the Cox model.

In Chapter 5, another method of estimating a shape constrained estimator is introduced. We do a similar investigation as in Chapter 4 to study the performance of this estimator compared to the traditional one.

In Chapter 6, we give a conclusion about the performance of this estimators compared to the traditional estimator.

Chapter 2

Traditional Estimator for Survival Distribution

In this chapter, we will start with an introduction of the problem in survival analysis. Some theoretical background which support the understanding in survival analysis are given in Section 2.1. It consists of several formula which illustrate different aspects in a data so as to capture the survival pattern in a population. Later on, we use this knowledge to analyze the survivability of a population which is based on the ICTUS dataset. Information regarding the background and characteristics of patients in the ICTUS dataset is given in Section 2.1.

In order to draw inference regarding the survivability pattern, we used several approaches. As a non-parametric approach, the Kaplan-Meier estimator is used. See Section 2.2 for more details. In the semi-parametric approach, survival regression is used. For this purpose we used the Cox model which will be explained in Section 2.3. This involves a brief introduction to the Cox model and the procedure to construct the model. In Section 2.4, we will illustrate how the Cox model is used in estimating the survival distribution for the ICTUS trial dataset. At the end of this chapter, we give a summary regarding the use of the Kaplan-Meier and the Cox model to estimate the survival distribution.

2.1 Introduction of the Problem

In survival analysis, the random variable of interest is the time to an event or simply the follow-up time. It means that a subject is observed from a certain time origin until it shows an event. In general, the term "event" refers to a failure of a mechanism, for instance death, development of new disease, etc. The time origin should be defined unambiguously for each subject. In a randomized clinical trial, it is meaningful to state the time origin as the first time when a patient enters the randomization process.

The probability distribution of the follow-up time can be characterized by several quantities, such as the cumulative distribution function, the survival function, and the cumulative hazard function, or by their derivatives such as the probability density or hazard function. One of the main purposes is to estimate these quantities on the basis of the observed follow-up times. In observing the follow-up times, there is a possibility that for some subject the full time to failure cannot be observed due to the following causes.

1. During the study period, a subject decided to leave the study or can no longer be contacted.

2. At the end of the study, a subject has not shown a failure yet.

In both of these circumstances, the actual survival time is larger than the observed follow-up time. An incomplete observation like this is (right) censored and the presence of censoring may lead to difficulties in estimating the probability distribution.

In this thesis, we want to estimate the probability distribution of patients with non-ST-elevation myocardial infarction (NSTEMI) in the AMC hospital, Amsterdam, The Netherlands. For that purpose, the dataset provided by the Department of Cardiology at the AMC Hospital is used. The dataset is derived from the ICTUS (Invasive versus Conservative in Treatment of the Unstable coronary Syndromes) trial. The study has this name because of comparing two treatment strategies (namely the Invasive and the Conservative treatment) in patients with NSTEMI. It consists of 1200 patients who had a myocardial infarction and met the criteria to be included in the study[1]. These 1200 patients were randomized into two types of treatment. The time origin is the time when these patients were randomized into two types of treatment. The event of interest in this study is the subsequent MI or cardiovascular death. The follow-up time for each patient was recorded during the 5 years. If at the end of the study a patient did not show an event, then its observed follow-up time is considered to be right censored. In this dataset, several characteristics which are considered to affect the survival distribution are also recorded. Up to 14 characteristics were recorded, which are listed below.

Treat	: the randomized treatment
Age	: Age of the patient at the time it enters the study
BMI	: Body Mass Index (in unit kg/m^2)
Gen	: Gender type
Smo	: smoking habit
Hyp	: hypertension
Hypl	: hyperlipidemia
Dbet	: diabetes
Fam.H	: Family history of having MI
Mi.H	: History of MI
PCI	: Having a PCI history
St. D	: Having ST Depression in electrocardiogram
CRP	: Having an increment of C-reactive protein $\geq 10mg/L$
Risc	: FRISC score, the sum of 7 factors presents at admission [1]

Right-Censored Data and Ties. Along this manuscript we deal with a right censored dataset which is in the form of triplets $(t_i, \delta_i, \boldsymbol{z}_i)$. The subscript i represents the ith patient. The variables t_i, δ_i and $\boldsymbol{z}_i$ represent the follow-up time, the censoring indicator and the vector of characteristics of the ith patient, respectively. The censoring indicator can have either two values, $\delta_i = 1$, when we actually observe the event (t_i is the exact survival time) or $\delta_i = 0$ when the subject is lost to follow-up (the exact survival time is larger than t_i). The censoring indicators are modelled as the consequence of the random censoring mechanism

$$T_i = \min\{X_i, C_i\}, \tag{2.1}$$

where X_i is the survival time and C_i is the censoring time. The censoring time in right-censored data is defined as the time when a patient was lost to follow-up. The vector of the characteristics $\boldsymbol{z}_i = (z_{i1}, z_{i2}, ..., z_{ip})$ consists of p values of characteristic of the ith patient. For example, as in the ICTUS dataset, each patient has 14 characteristics as listed earlier.

Another important term that we may find in survival analysis are tied follow-up times. Tied follow-up times refer to the situation in which different patients have the same follow-up time. This could be the case

when the follow-up time is measured in month or weeks. Consider the following simple survival dataset consists of 5 individuals: $(4,1),(4,1),(2,1),(2,1),(2,0),(8,1)$. In this dataset, the follow-up times 2 and 4 occur more than once. If this is the case then 2 and 4 are tied follow-up times.

2.2 Theoretical Background

In this section, several quantities which will be used later in this manuscript will be given. These quantities characterize the survival distribution, such as the cumulative distribution function and the cumulative hazard function as well as their derivatives. These characterizations are possible since all of these quantities are equivalent, i.e., if one form is known then the others can be derived. For completeness, the relations between these quantities are provided. Let T be the survival time.

1. The survival function, as a function of t, is defined as the probability that an individual survives longer than time t:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t),$$

where $F(t)$ is the cumulative distribution function. Therefore, the probability density function can be defined as

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}(1 - S(t)). \tag{2.2}$$

2. The hazard function $h(t)$ is defined as the conditional failure rate, i.e., the infinitesimal probability of failure during a small time interval, given that a subject has survived to the beginning of the interval:

$$h(t) = \lim_{\epsilon \downarrow 0} \frac{P(t < T \leq t + \epsilon \mid T > t)}{\epsilon},$$

which is equal to

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\log[S(t)]. \tag{2.3}$$

The last term of (2.3) is obtained by substituting the form $f(t)$ from equation (2.2).

3. The following relation can be obtained by integrating (2.3) from zero to t and using $S(0) = 1$:

$$-\int_0^t h(x)\,dx = \log[S(t)],$$

which leads to the relation between the cumulative hazard H and the survival function:

$$H(t) = -\log[S(t)], \tag{2.4}$$

or equivalently

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(x)\,dx\right]. \tag{2.5}$$

2.3 The Kaplan-Meier Estimator

A nonparametric approach to estimate the survival distribution is based on the maximum likelihood principle. The idea is to form the likelihood function. This is a function of the parameters that we want to estimate by maximizing this function over all possible parameters. Suppose that we have right censored observations (t_i, δ_i), $i = 1, 2, ..., n$ as explained in the previous subsection. If the distribution of the survival time has survival function $S(t)$ and probability density $f(t)$, the likelihood for the observations can be written as

$$\ell(F) = [f(t_1)]^{\delta_1} [S(t_1)]^{1-\delta_1} \times \cdots \times [f(t_n)]^{\delta_i} [S(t_i)]^{1-\delta_n},$$

or written more compactly

$$\ell(F) = \prod_{i=1}^{n} [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i}. \tag{2.6}$$

This represents the "probability" that the observations occur if they would be generated from a distribution with density f and survival function S. The principle of maximum likelihood (ML) prescribes to find the probability distribution for which the observations are most likely to occur. Maximizing $\ell(F)$ over all possible F (or equivalently all f and S) is not possible, but it can be shown that when one maximizes over all survival functions S that are constant between observed survival times, the ML estimator exists and is given by the Kaplan-Meier estimator

$$\widehat{S}(t) = \prod_{t_{(i)} \le t} \left(1 - \frac{d_i}{n_i}\right), \tag{2.7}$$

where $t_{(1)} < \cdots < t_{(m)}$ are the ordered observed survival times and

$$d_i = \text{number of observed events at } t_{(i)};$$
$$n_i = \text{number of individuals at risk at } t_{(i)},$$

with the convention that $\widehat{S}(t) = 1$ for $t < t_{(1)}$. Individuals at risk at $t_{(i)}$ are those whose follow-up time (either censored or uncensored) is at least $t_{(i)}$.

From its definition it can be seen that the Kaplan-Meier estimator is piecewise constant between successive survival times

$$S(t) = S(t_{(i)}), \qquad \text{for } t_{(i)} \le t < t_{(i+1)}.$$

This is in contrast with a parametric approach, in which the survival function will usually be a smooth function following a pre-specified form. In many applications, such a fixed pre-specified form may not be suitable. The difference is illustrated in Figure 2.1, which shows the Kaplan-Meier estimator for ICTUS dataset and the estimated survival function for the exponential model for a particular dataset.

We will concentrate on estimating the cumulative hazard function $H(t)$, rather than estimating the survival function $S(t)$. Relation (2.4) suggest a natural estimator

$$\widehat{H}(t) = -\ln\left[\widehat{S}(t)\right] = \sum_{t_{(i)} \le t} -\ln\left(1 - \frac{d_i}{n_i}\right).$$

This estimator should not be confused with the Nelson-Aalen estimator for the cumulative hazard,

$$\widehat{H}_{\mathrm{NA}}(t) = \sum_{t_{(i)} \le t} \frac{d_i}{n_i}. \tag{2.8}$$

Although, the estimators are different and $\widehat{H}_{\mathrm{NA}}(t) \ge \widehat{H}(t)$, the difference is usually very small, especially when d_i is small compared to n_i.

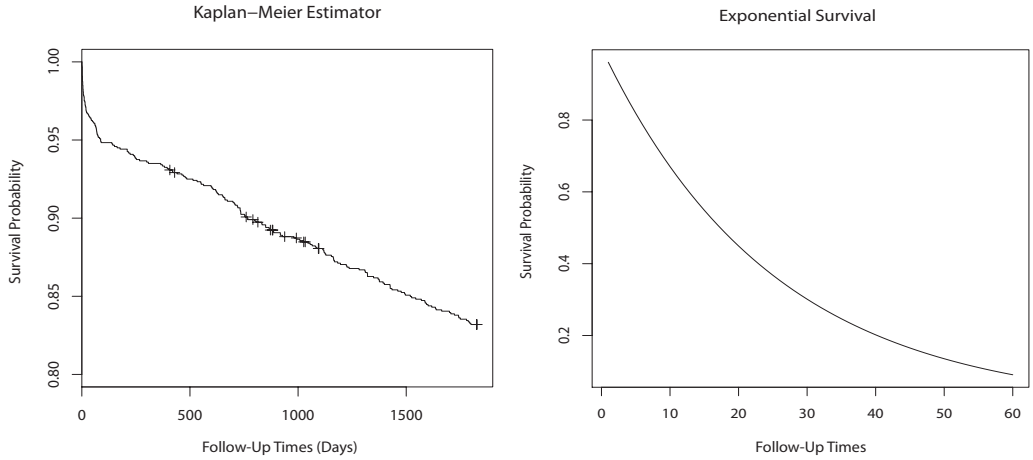

Figure 2.1: Kaplan - Meier (left) and Exponential (right) Survival Curve.

2.4 The Proportional Hazard Model

In dealing with survival data without any prejudice about the underlying distribution, a nonparametric or a semiparametric approach might be appealing to describe the relationship between several variables and the survival probability. When incorporating explanatory variables, the most popular method is the Cox[2] Proportional Hazard Model (or simply the Cox Model). Let $\boldsymbol{z} = (z_1, \ldots, z_p)'$ be the vector of explanatory variables, also referred to as covariates. For instance, in the ICTUS dataset the covariates are the patients' characteristics. Cox proposed the hazard function of the survival time distribution to be of the form

$$h(t) = h(t, \boldsymbol{z}, \boldsymbol{\beta}) = h_0(t) \exp(\boldsymbol{z'\beta}), \tag{2.9}$$

where h_0 is the unknown baseline hazard function and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$. In this model, the hazard function $h(t)$ is characterized by a function of survival time separated from a function of the explanatory variables. Note that the hazard function is equal to $h_0(t)$ when $\exp(\boldsymbol{z'\beta}) = 1$, i.e., when $\boldsymbol{z'\beta} = 0$.

The cumulative hazard function is given by

$$H(t) = H(t, \boldsymbol{z}, \boldsymbol{\beta}) = \int_0^t h_0(u) \exp(\boldsymbol{z'\beta})\, du = H_0(t) \exp(\boldsymbol{z'\beta}), \tag{2.10}$$

where H_0 is the baseline cumulative hazard function. For completeness, we also specify the survival function in the Cox model. According to equation (2.5), the survival function for model (2.9) can be formulated as follows,

$$S(t, \boldsymbol{z}, \boldsymbol{\beta}) = \exp[-H(t, \boldsymbol{z}, \boldsymbol{\beta})] = \left[\mathrm{e}^{-H_0(t)}\right]^{\exp(\boldsymbol{z'\beta})} = [S_0(t)]^{\exp(\boldsymbol{z'\beta})}, \tag{2.11}$$

where $S_0(t)$ is the baseline survival function. Having specified the cumulative hazard function of the Cox model, the problem is to find the estimate of baseline cumulative hazard function $H_0(t)$ and the regression coefficients $\boldsymbol{\beta}$.

Assume that the survival data for the i-th individual consist of a follow-up time t_i, a censoring indicator δ_i, and covariate values $\boldsymbol{z}_i = (z_{i1}, z_{i2}, ..., z_{ip})'$. The hazard ratio gives the relative hazard between two subjects having different covariate values. For instance, consider two subjects having covariate values $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$.

$$\mathrm{HR}(t, \boldsymbol{z}_1, \boldsymbol{z}_2) = \frac{h(t, \boldsymbol{z}_1, \boldsymbol{\beta})}{h(t, \boldsymbol{z}_2, \boldsymbol{\beta})} = \frac{h_0(t) \exp(\boldsymbol{z}_1'\boldsymbol{\beta})}{h_0(t) \exp(\boldsymbol{z}_2'\boldsymbol{\beta})} = \frac{\exp(\boldsymbol{z}_1'\boldsymbol{\beta})}{\exp(\boldsymbol{z}_2'\boldsymbol{\beta})} = \mathrm{e}^{(\boldsymbol{z}_1 - \boldsymbol{z}_2)'\boldsymbol{\beta}}. \tag{2.12}$$

Notice that this ratio is independent of time and can be described by a linear combination of the coefficients $\beta_1, \ldots, \beta_p$ and the differences of the covariate values. Obviously, the hazards of both individuals are proportional to another, which explains why the Cox model is often referred to as the proportional hazards model.

2.4.1 Estimating the Coefficients in the Cox Model

The likelihood of the observations $(t_i, \delta_i, \boldsymbol{z}_i)$, $i = 1, 2, \ldots, n$ can be obtained similar to equation (2.6). Because from (2.3) it follows that $f(t) = h(t)S(t)$, the likelihood for the Cox model becomes

$$\ell(h_0, \boldsymbol{\beta}) = \prod_{i=1}^{n} \left[h(t_i, \boldsymbol{z}_i, \boldsymbol{\beta}) S(t_i, \boldsymbol{z}_i, \boldsymbol{\beta})\right]^{\delta_i} \left[S(t_i, \boldsymbol{z}_1, \boldsymbol{\beta})\right]^{1-\delta_i} \tag{2.13}$$

By inserting (2.9) and (2.11), the right hand side will be a function of h_0 and $\boldsymbol{\beta}$. Theoretically, one could maximize this over all possible hazard functions h_0 and coefficients $\boldsymbol{\beta}$. However, it is more convenient to consider the log-likelihood $L(h_0, \boldsymbol{\beta}) = \ln \ell(h_0, \boldsymbol{\beta})$ and maximize this, which would yield the same maximizer (if it exists). The log-likelihood for the Cox model becomes

$$L(h_0, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \delta_i \ln h_0(t_i) + \delta_i \boldsymbol{z}_i' \boldsymbol{\beta} + \mathrm{e}^{\boldsymbol{z}_i' \boldsymbol{\beta}} \ln S_0(t_i) \right\}. \tag{2.14}$$

In practice, maximizing $L(h_0, \boldsymbol{\beta})$ over all baseline distributions with hazard function h_0 and survival function S_0 and over all coefficients $\boldsymbol{\beta} \in \mathbb{R}^p$ is not possible. While the estimation of $\boldsymbol{\beta}$ is usually of primary interest, the baseline hazard h_0 is seen as a nuisance parameter. In order to circumvent the dependence of the full likelihood on h_0, Cox [2] proposed to use the "partial likelihood function" of the data to estimate the regression coefficients $\boldsymbol{\beta}$. For simplicity in deriving the partial likelihood, assume that there is only one covariate z. Suppose there are n observations, with follow-up times $t_1, \ldots, t_n$. If we denote the ordered follow-up times by $t_{(1)} \leq \cdots \leq t_{(n)}$, then the partial likelihood formula is given by the following expression

$$\ell_p(\beta) = \prod_{i=1}^{n} \left[\frac{e^{z_i' \beta}}{\sum_{j \in R(t_{(i)})} e^{z_j' \beta}} \right]^{\delta_i}, \tag{2.15}$$

with $R(t_{(i)})$ denotes the risk set, consisting of all subjects that are at risk at $t_{(i)}$, i.e., with a follow-up time greater than or equal to $t_{(i)}$. The derivation is explained in [3]. The formula (2.15) above assumes there are no ties in the dataset. If there are ties in the dataset, approximations for the partial likelihood proposed by Breslow [4] or Efron [5] can be used.

Formula (2.15) can be simplified. For a censored observation one has $\delta_i = 0$, so that the partial likelihood can be reduced to

$$\ell_p(\beta) = \prod_{i=1}^{m} \frac{e^{z_{(i)} \beta}}{\sum_{j \in R(t_{(i)})} e^{z_j \beta}} \tag{2.16}$$

where the product is over m distinct ordered observed survival times, $z_{(i)}$ denotes the value of the covariate for the subject with ordered survival time $t_{(i)}$. The log partial likelihood function is then

$$\ell_p(\beta) = \sum_{i=1}^{m} \left\{ z_{(i)} \beta - \ln \left(\sum_{j \in R(t_{(i)})} e^{z_j \beta} \right) \right\}. \tag{2.17}$$

The maximum partial likelihood estimator can be calculated by first differentiating the right hand side of (2.17) with respect to β, setting the derivative equal to zero and solving for the unknown parameter. The derivative of (2.17) with respect to β is

$$\frac{\partial \ell_p(\beta)}{\partial \beta} = \sum_{i=1}^{m} \left\{ z_{(i)} - \frac{\sum\limits_{j \in R(t_{(i)})} z_j e^{z_j \beta}}{\sum\limits_{j \in R(t_{(i)})} e^{z_j \beta}} \right\}. \tag{2.18}$$

The solution of setting (2.18) equal to zero, is the estimated regression coefficient $\widehat{\beta}$.

This method can also be applied to a multicovariate model. Assume there are p covariate values for subject i denoted by the vector $\boldsymbol{z}_i = (z_{i1}, z_{i2}, \ldots, z_{ip})'$. This vector may be any collection of covariates: continuous covariates, categorial covariates, design variables for nominal scale covariates, interactions between covariates, and other higher order terms. The data are now described in the triplet form $(t_i, \delta_i, \boldsymbol{z}_i)$, for $i = 1, 2, \ldots, n$. The partial likelihood for the multivariable model is obtained by replacing the single covariate, z in (2.16) with the vector of covariates $\boldsymbol{z}$ and the its coefficient β with $\boldsymbol{\beta}$.

When putting the derivative equal to zero, this yields p equations, one for each covariate. The derivative for the kth covariate is

$$\frac{\partial \ell_p(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{i=1}^{m} \left\{ z_{(ik)} - \frac{\sum\limits_{j \in R(t_{(i)})} z_{jk} e^{\boldsymbol{z}_j' \boldsymbol{\beta}}}{\sum\limits_{j \in R(t_{(i)})} e^{\boldsymbol{z}_j' \boldsymbol{\beta}}} \right\}. \tag{2.19}$$

A similar procedure as with the previous single covariate model can be followed. The derivatives of the log partial likelihood are set equal to zero and solved simultaneously, yielding the maximum partial likelihood estimator $\widehat{\boldsymbol{\beta}}$.

Tied Follow-up Times. In the previous subsection, the partial likelihood function is derived for the case when there are no tied follow-up times. In most applied settings, tied follow-up times may occur and modifications in the partial likelihood will be needed. Basically, there are three types of approaches to handle tied follow-up times which are used in software packages, i.e., an exact expression and approximations due to Breslow and Effron. The expression for exact approximation will not be presented here.

The approximations proposed by Breslow [4] and Efron [5] are designed to provide an easier form to be used in most software programs, and yet still account for ties which occur in the observed values of follow-up times. The following is the approximation to the partial likelihood that is used by Breslow:

$$l_{p1}(\beta) = \prod_{i=1}^{m} \frac{z_{(i)+}\beta}{\left[\sum\limits_{j \in R(t_{(i)})} e^{z_j \beta} \right]^{d_i}},$$

where d_i denotes the number of subjects with follow-up times $t_{(i)}$ and $z_{(i)+}$ is equal to the sum of the covariate values over the d_i subjects or it can be formulated as

$$z_{(i)+} = \sum_{j \in D(t_{(i)})} z_j$$

where $D(t_{(i)})$ represents the uncensored subjects with survival time equals $t_{(i)}$.

The Effron approximation is a bit more complicated and yields a slightly better approximation to the exact partial likelihood than the Breslow approximation. The following expression is the approximation of the partial likelihood proposed by Effron:

$$l_{p2}(\beta) = \prod_{i=1}^{m} \frac{z_{(i)+}\beta}{\prod_{k=1}^{d_i} \left[\sum_{j \in R(t_{(i)})} e^{z_j \beta} - \frac{k-1}{d_i} \sum_{j \in D(t_{(i)})} e^{z_j \beta} \right]}.$$

The maximum partial likelihood estimator for β in the tied case is obtained in the same manner as for the non tied case.

2.4.2 Estimating the Baseline Survival Function

Once the regression coefficients have been estimated, the baseline survival function, or related quantities that characterize the baseline distribution, can be estimated. The principal idea in deriving the estimate for the baseline survival function is to insert the partial likelihood estimate $\widehat{\boldsymbol{\beta}}$ into the log-likelihood $L(h_0, \boldsymbol{\beta})$ in (2.14) and then maximize over baseline distributions with hazard function h_0. However, as with the Kaplan-Meier likelihood (2.6) it is not possible to maximize over all baseline distributions, and we restrict to maximize over all distributions for which the survival function is constant between successive survival times. The essential idea of the likelihood approach is to write

$$S_0(t_{(i)}) = \frac{S_0(t_{(i)})}{S_0(t_{(i-1)})} \times \frac{S_0(t_{(i-1)})}{S_0(t_{(i-2)})} \times \cdots \times \frac{S_0(t_{(1)})}{S_0(0)} \times S_0(0),$$

where $S_0(0) = 1$. The key point is the use of the quantity

$$\alpha_i = \frac{S_0(t_{(i)})}{S_0(t_{(i-1)})}.$$

Then with (2.11), for the survival function we find

$$\frac{S(t_{(i)}, \boldsymbol{z}, \boldsymbol{\beta})}{S(t_{(i-1)}, \boldsymbol{z}, \boldsymbol{\beta})} = \left\{ \frac{S_0(t_{(i)})}{S_0(t_{(i-1)})} \right\}^{\exp(\boldsymbol{z}'\boldsymbol{\beta})} = \alpha_i^{\exp(\boldsymbol{z}'\boldsymbol{\beta})}.$$

After inserting this in the log-likelihood (2.14), it can be shown that the maximizing baseline survival function is a product of α_i's that satisfy

$$\sum_{l \in D_i} \frac{\widehat{\theta}_l}{1 - \alpha_i \widehat{\theta}_l} = \sum_{l \in R_i} \widehat{\theta}_l, \tag{2.20}$$

with $\widehat{\theta}_l = \exp(\boldsymbol{z}_l' \widehat{\boldsymbol{\beta}})$ and

$$R_i = \text{all subjects that are at risk at } t_{(i)};$$
$$D_i = \text{uncensored subjects that are at risk at } t_{(i)}.$$

If there are no tied follow-up times, D_i contains one subject and the solution to (2.20) is given by explicitly by

$$\widehat{\alpha}_i = \left[1 - \frac{\widehat{\theta}_l}{\sum_{l \in R_i} \widehat{\theta}_l} \right]^{1/\widehat{\theta}_i}.$$

If there are tied follow-up times, the solution to (2.20) is obtained using iterative methods. The estimator of the baseline survival function is the product of the individual estimators of the conditional baseline survival probabilities

$$\widehat{S}_0(t) = \prod_{t_{(i)} \leq t} \widehat{\alpha}_i, \tag{2.21}$$

where $\widehat{\alpha}_i$ is the solution to (2.20). To obtain this solution, the expression $\alpha_i \widehat{\theta}_l$ on the left hand side of (2.20) is often replaced by the approximation $\alpha_i \widehat{\theta}_l \sim 1 + \widehat{\theta}_l \ln(\alpha_i)$. The solution for the baseline survival function in (2.11) can then be computed explicitly:

$$\widetilde{\alpha}_i = \exp\left[\frac{-d_i}{\sum\limits_{l \in R_i} \widehat{\theta}_l}\right] \tag{2.22}$$

which is again the product of the individual conditional survival probabilities.

2.4.3 The Cumulative Baseline Hazard and Baseline Hazard

Since in this manuscript we will deal mostly with the baseline hazard (or the cumulative baseline hazard), a derivation of its formula is given in this subsection. We used the suggestion by Breslow in Cox's paper [2]. He noted that in order to obtain the estimate for the baseline hazard, the attention is restricted in to a hazard that is constant between the subsequent ordered survival times:

$$h_0(t) = h_i, \quad t_{(i-1)} < t \leq t_{(i)}, i = 1, \ldots, k$$

We adopt the convention that every censored time is censored in the preceding uncensored survival time. This means, whenever we have the subsequent failure times and we have a censored follow-up times in between, these censored follow-up times are shifted back to the previous uncensored failure time. It turns out that the partial likelihood estimate of $\boldsymbol{\beta}$ and the following naive estimate of the cumulative baseline hazard

$$\widehat{H}_0(t) = \int_0^t h_0(u)du = \sum_{t_{(i)} \leq t} \frac{d_i}{\sum\limits_{l \in R_i} \exp\left(\boldsymbol{z}_l' \boldsymbol{\beta}\right)}, \tag{2.23}$$

is maximizing the Cox's full likelihood. In this equation, R_i consists of all individuals who are at risk at time $t_{(i)}$. Thus, the estimate for the baseline hazard is obtained by taking the numerical derivative from the estimate of the cumulative baseline hazard above. Its formula is given by the following

$$\widehat{h}_0(X_{(i)}) = \frac{d_i}{\left(t_{(i+1)} - t_{(i)}\right) \sum\limits_{l \in R(t_i)} \mathrm{e}^{\boldsymbol{z}'\boldsymbol{\beta}}}, \quad i = 1, \ldots, k. \tag{2.24}$$

2.5 Cumulative Hazard of ICTUS dataset

In estimating the cumulative hazard from a given dataset, one can use the Nelson-Aalen estimator (as a nonparametric approach) or the Breslow estimator in the Cox model (as a semiparametric approach). In most statistical software, there is already an available package for this purpose. In a nonparametric approach, the cumulative hazard of the ICTUS dataset is estimated by the Nelson-Aalen estimator (equation (2.8)). The estimate of the cumulative hazard by using this estimator is presented in Figure 2.2. In this

section, we will estimate the cumulative hazard using the Cox model. This means that an estimate of the cumulative hazard might depend on several predictors in this dataset. The first step in building the Cox model is choosing significant covariates to be involved in the model. Despite having 14 covariates for each patient in the ICTUS dataset, it might be that only several covariates influence the survival probability. Thus, it is more efficient to include only the covariates which have an influence on the survival distribution. For this purpose, we first do an analysis for choosing the covariates which give a significant contribution to the survival distribution. This will be done in Sections 2.5.1 and 2.5.2. A detailed explanation about the procedure of fitting the Cox model can be found in [6].

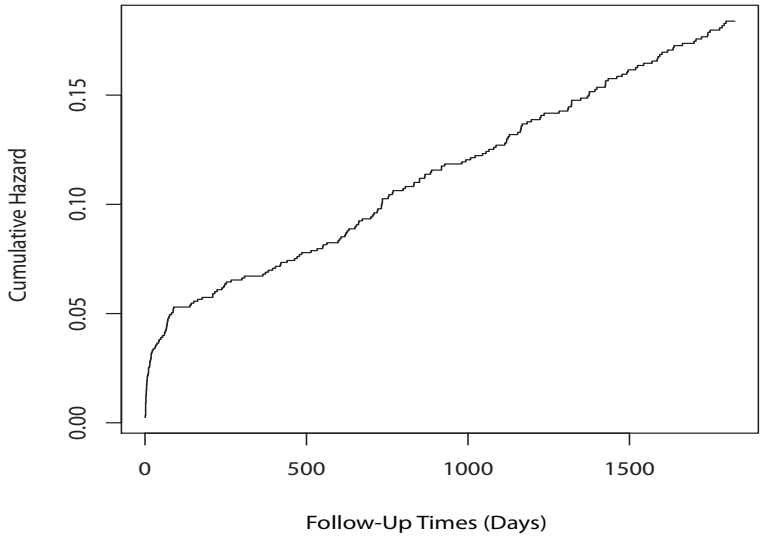

Figure 2.2: The Nelson-Aalen Cumulative Hazard of the ICTUS dataset

2.5.1 Bivariate Analysis

The bivariate analysis is done for all covariates separately. The analysis is based on the procedure explained in [6]. This is done by using the Kaplan-Meier estimator for the specific covariates. In Table 2.1, all $p-$values for the partial likelihood and log rank test are listed. It is common that a p-value between the range 20 - 25$\%$ still indicates the significance of the covariates. On the basis of this suggestion, we can see from Table 2.2, there are three covariates which can be considered insignificant with the survival distribution namely Treat, St.D, and CRP.

Table 2.2 provides the discrete form of the bivariate analysis for continuous covariates. If by either a discrete or a continuous form, the covariate is significant, it should be included in the multivariate model. Notice that the covariate BMI is actually not a continuous covariate. The values of this covariate fall into 7 categories [7]. The range of BMI in this dataset falls between category underweight to obese class 3. Further consideration of these categories lead to the decision of dividing the covariate into two groups, which are normal and overweight or obesity. The covariate Risc is divided into 3 categories, i.e., category 1 which includes all patients having FRISC Score from 1 and 2, category 2 which includes patients with FRISC Score 3 and 4, and the remaining FRISC Score belong to the third category. These categories are based on [1].

No	Covariates	Category	Log-Rank $p-$value	Part.Lik $p-$value
1	Age	Polytomous	< 0.0001	5.64E-09
2	Treat	Dichotomous	0.515	0.515
3	BMI	Polytomous	0.0459	0.0513
4	Gen	Dichotomous	0.0826	0.0890
5	Smo	Dichotomous	0.0461	0.0440
6	Hyp	Dichotomous	0.0040	0.0045
7	Hypl	Dichotomous	< 0.0001	< 0.0001
8	Dbet	Dichotomous	< 0.0001	< 0.0001
9	Fam.H	Dichotomous	0.127	0.1243
10	Mi.H	Dichotomous	< 0.0001	< 0.0001
11	St.D	Dichotomous	0.742	0.7423
12	CRP	Dichotomous	0.774	0.7735
13	Risc	Polytomous	0	< 0.0001
14	PCI	Dichotomous	0.0021	0.0044

Table 2.1: Bivariate Analysis for All Covariates

Variable	Change	Haz.Ratio for Change (.95 CI)	Wald Test p-value	ParLik Ratio p-value
Age	11 years	1.7258 (1.465, 2.033)	7.03E-11	7.78E-12

Table 2.2: Estimated Hazard Ratios for Continuous Covariate

	coef	exp(coef)	lower .95	upper .95	se(coef)	z	Pr($>\|z\|$)
Treat	0.0960	1.1008	0.8281	1.4634	0.1452	0.661	0.5083
Age	0.0318	1.0324	1.0134	1.0517	0.0095	3.367	0.0007
BMIhi	-0.3425	0.7100	0.5247	0.9608	0.1543	-2.219	0.0265
Gen	-0.3252	0.7223	0.5022	1.0391	0.1855	-1.753	0.0795
Smo	-0.0263	0.9740	0.7082	1.3396	0.1626	-0.162	0.8713
Hyp	0.1819	1.1995	0.8946	1.6082	0.1496	1.216	0.2241
Hypl	0.3439	1.4104	1.036	1.9201	0.1574	2.185	0.0289
Dbet	0.3803	1.4627	0.9891	2.1629	0.1996	1.905	0.0567
Fam.H	-0.1488	0.8617	0.6399	1.1604	0.1518	-0.98	0.3271
Mi.H	0.2914	1.3383	0.9386	1.9083	0.1810	1.61	0.1075
PCI	0.2072	1.2303	0.8291	1.8256	0.2014	1.029	0.3034
St.D	-0.0111	0.9889	0.7423	1.3175	0.1463	-0.076	0.9394
CRP	-0.0869	0.9168	0.6502	1.2928	0.1753	-0.495	0.6203
Risc2	0.4313	1.5393	1.0321	2.2957	0.2039	2.115	0.0344
Risc3	1.0125	2.7524	1.3631	5.5577	0.3585	2.824	0.0047

Rsquare= 0.092		
Likelihood ratio test	= 112.1 on 15 df,	p=1.110e$-$16
Wald test	= 118.3 on 15 df,	p=0
Score (logrank) test	= 137.6 on 15 df,	p=0

Table 2.3: Multivariate Model

After applying the bivariate analysis, the fitting of a Cox model is done for all covariates simultaneously. We do not eliminate the insignificant estimates which are obtained by bivariate analysis. Some practice shows that it could be the case that a covariate with a high p-value became highly significant in the multivariable model [6]. The resulting parameters estimates of the model are given in the Table 2.3.

The covariate BMIhi represents patients having a body mass index higher than 25 $\mathrm{kg/m^2}$, which can be considered as overweight or obesed. In the last column of Table 2.4, the $p-$values of the Wald test for each covariate are given. There are 8 covariates which seem to be insignificant. In the following subsection, we decide whether to include or exclude the insignificant covariates.

2.5.2 Inclusion/Deletion of Covariates

Inclusion or exclusion of a covariate which is not found to be significant in the bivariate analysis, will include two important steps, i.e., verification of the partial likelihood test and verification of confounding. A covariate is called a confounder if it is known that this covariate has a correlation with other significant covariates in the model. Suppose for example, that after the separated bivariate analysis smoking habit is not significant while diabetes is. When we put them together in the model and inspect any correlation between these two, we consider smoking habit as a confounder. Possible interaction is inspected by means of a statistical test.

The verification of the partial likelihood test should be done per covariate and cannot be done simultaneously. The procedure for inclusion or exclusion is explained as follows. The partial likelihood of the fitted Cox model with and without the insignificant covariate are compared. By this comparison, we would like to confirm whether there is any significant change from the model with the covariate and the model without the covariate. If it is not significant, then verification whether that covariate is a confounder for another variable should also be done. If it is also not a confounder for another covariate, then it can be excluded from the model. This is done for all insignificant covariates.

The verification of being a confounder is done in the following way. We first fit the Cox model without the insignificant covariate and obtain the estimate of the covariate coefficient. We then fit the Cox model including the insignificant covariate. If the covariate coefficient is changing considerably (a suggestion is a change more than 20%) this means that the insignificant covariate is a confounder. Details about this can be read in [6]. When applying the procedure of inclusion/deletion mentioned above, it is found that all the insignificant covariates are not confounding for any of the other covariates. Therefore they can be removed from the model. The last step in fitting the Cox model is the verification for the possible significant interactions between covariates. For all five 5 significant covariates, we tested all possible interactions between two covariates. As a result, in this dataset no interactions were found to be significant.

After applying the procedure mentioned above, the subset of covariates which are considered to be significant can be found in Table 2.4. If we look on the final model, the covariate Treat is still included even though it is not significant. This is done because of the purpose of this project which is to analyze the effect of treatment.

2.5.3 The Cumulative Baseline Hazard and Baseline Hazard Function of ICTUS dataset

As the regression coefficients have now been estimated, the baseline cumulative hazard $H_0(t)$ can also be estimated using the Breslow estimator (2.23). Figure 2.3 shows the Breslow estimate $\widehat{H}_0(t)$ for the

	coef	exp(coef)	lower .95	upper .95	se(coef)	Pr($>\|z\|$)
Treat	0.0588	1.0606	0.8033	1.4002	0.1417	0.6782
Age	0.0425	1.0434	1.0278	1.0593	0.0077	3.49E-08
BMIhi	−0.3445	0.7086	0.5289	0.9493	0.1492	0.0210
Hypl	0.3643	1.4395	1.0698	1.937	0.1514	0.0161
Dbet	0.7155	2.0451	1.4843	2.8179	0.1635	1.21E-05
Mi.H	0.5238	1.6884	1.2485	2.2834	0.1540	0.0007

Rsquare	0.078	
Likelihood ratio test	= 98 on 6 df,	p=0
Wald test	= 101.2 on 6 df,	p=0
Score (logrank) test	= 108.8 on 6 df,	p=0

Table 2.4: Final Model

cumulative baseline hazard function. In spite of the differences in their values, the cumulative baseline

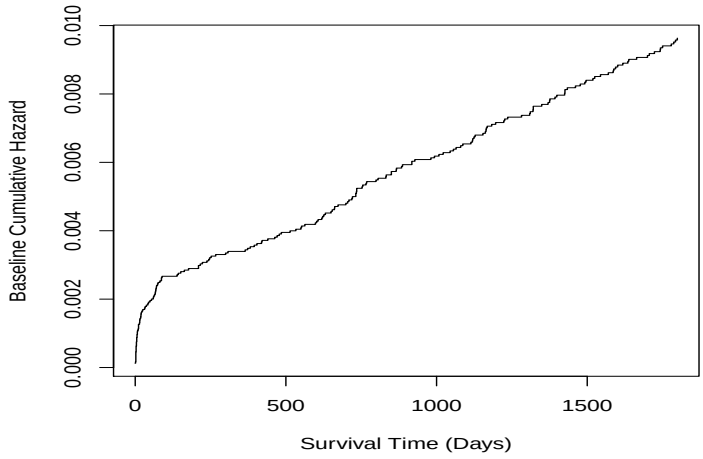

Figure 2.3: The Baseline Cumulative Hazard from Cox Model

hazard shapes in Figure 2.2 and Figure 2.3 are quite similar. As can be seen in both figures, the curve is increasing fastly with decaying steepness in the first 100 days meaning that in this study many patients develop a new MI or die in the first 100 days. In the following days until the end of study, the curve shows more or less a linear trend. This phenomenon suggests that the hazard is decreasing in the first 100 days and remains fairly constant afterwards. This empirical evidence suggests that the hazard is decreasing. Using formula (2.24), we can estimate the baseline hazard at every ordered distinct follow-up times as illustrated in Figure 2.4. The two pictures in Figure 2.4 represent the baseline hazard. The left figure is the hazard derived from the Cox model using the procedure explained in Subsection 2.4.3. This hazard function is not a smooth function. In medical applications, it is desirable to use a smooth function. Therefore, a smooth version of the hazard function is calculated by using a kernel estimator and is depicted in Figure 2.4 in the right picture.

It can be seen from the picture that in the beginning of the study, the risk is quite high. We see that in general the curve is decreasing with a slightly increasing pattern in certain follow-up times interval. Several suspicious jumps occur between day 500 and 1500. This jump pattern means that after the risk of getting subsequent MI decreases, it might increase in the future. This jump pattern in the MI hazard function cannot be explained from a medical perspective. It is believed that in the beginning of the study,

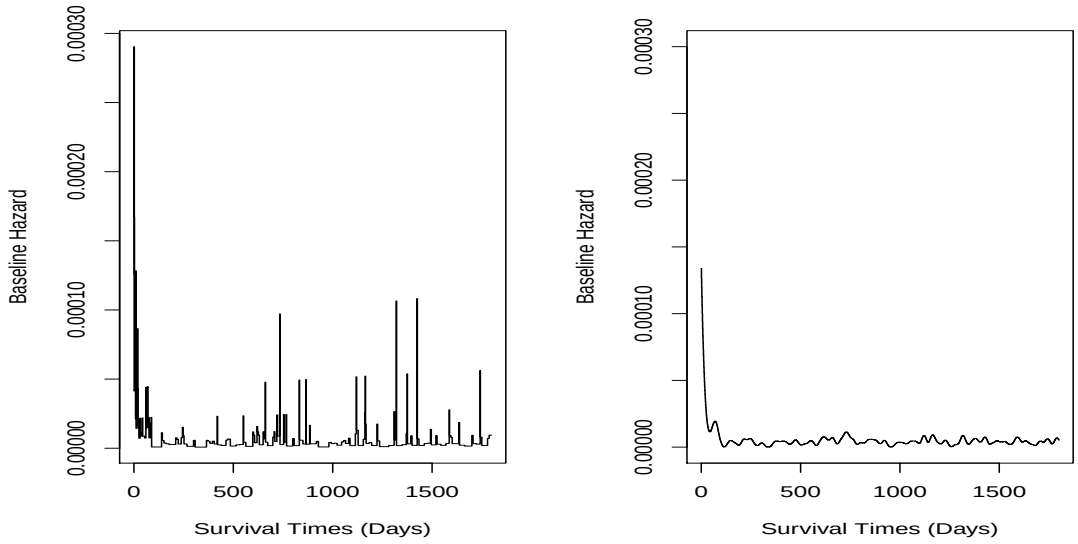

Figure 2.4: The Baseline Hazard from Cox Model

a lot of events happen but once it passes some constant level, it stays there or even decreases. Thus, the baseline hazard curve is expected to decrease over time until a certain base level. In other words, the hazard function should be monotone decreasing. We see that this information is not incorporated by either the Cox model or the Kaplan-Meier estimator.

2.6 Summary

In this chapter, we estimated the survival distribution (in term of cumulative hazard function) of patients having an MI using the traditional estimators. The Nelson-Aalen estimator is used in the nonparametric approach to right-censored data and the Breslow estimator is applied in the semiparametric approach. Using both estimators, we see that the cumulative hazard does not assume a monotone function. Their cumulative hazard curves show a similar pattern even though the values are different. As a consequence, the hazard function is not monotone decreasing. From the medical point of view, this is not reasonable. A patient who undergoes treatment and survives at a certain time would not be expected to have an increase hazard from this specific time onwards. This assumption is not incorporated when using the traditional estimator. In a later chapter, a new estimator which overcomes this drawback will be introduced.

Chapter 3

Shape Constrained Estimator for Right Censored Data

In this chapter, we will investigate a new estimator, which incorporates the information that we have about the shape of the probability distribution. We will assume that the hazard function $h(t)$ of the survival time distribution is nonincreasing. We concentrate on estimating the cumulative hazard function $H(t)$, which then necessarily must be concave. In the construction of this new estimator, the nonparametric estimator $\widehat{H}(t)$ (or alternatively $\widehat{H}_{\mathrm{NA}}$), will be adjusted so that it will result in a continuous concave curve, say $\widetilde{H}(t)$. To obtain a concave estimate for $H(t)$, we construct the so-called least concave majorant of the nonparametric estimator $\widehat{H}(t)$. The new cumulative hazard estimate $\widetilde{H}(t)$ turns out to be piecewise linear and its (left) derivative is piecewise constant. Due to the concavity of $\widetilde{H}(t)$, the estimate $\widetilde{h}(t)$ for the hazard is nonincreasing and therefore incorporates the assumptions of the model. Such a construction leads to sensible estimators and this will be explained in Section 3.1 .

Afterwards, we need to investigate whether it is reasonable to apply this new estimator in the ICTUS dataset. For that purpose, a statistical hypothesis test will be performed in Section 3.3. One step in statistical tests is to built the probability distribution of the test statistic. A bootstrap method is one way to determine the test statistic distribution. More about the bootstrapping technique will be explained in Section 3.4. In Section 3.6, the bootstrap method is performed to construct the distribution of test statistics in ICTUS data and in Section 3.7, we validate the test statistic.

3.1 The Least Concave Majorant

Let G be some function on $[0, \infty)$. The smallest concave function $\widetilde{G}$ on $[0, \infty)$ that lies above G is called the least concave majorant (LCM) of G on $[0, \infty)$, i.e.,

$$\widetilde{G}(t) \geq G(t) \qquad \text{for } t \in [0, \infty).$$

The idea of constructing the LCM originates from the problem of estimating a nonincreasing probability density function on $[0, \infty)$. This problem is the simplest one in a wide range of statistical problems that fall under the name "shape constrained nonparametric estimation". As shown by Grenander [8], the maximum likelihood estimator for a nonincreasing density exists within the class of all nonincreasing densities on $[0, \infty)$. This means there exists a nonincreasing density $\widehat{f}$ that maximizes the likelihood of the observations $y_1, \ldots, y_n$, i.e.,

$$\ell(f) = \prod_{i=1}^{n} f(y_i)$$

over all nonincreasing densities on $[0, \infty)$. Moreover, the maximum likelihood estimator $\widehat{f}$ can be characterized as the left-derivative of the least concave majorant of the empirical distribution function of the observations. The empirical distribution function can be seen as a naive estimator for the cumulative distribution function (CDF) that does not incorporate any information about the probability distribution. The LCM "monotonizes" the naive empirical CDF, which in this case results in a piecewise linear density estimate whose derivative is piecewise constant and nonincreasing, and therefore incorporates the assumption on the density.

Because the empirical CDF is piecewise constant, the construction of the LCM is straightforward. Illustratively, the LCM can be obtained by fixing a rubber band at the origin and stretching it over the empirical CDF. In this way, the LCM will be piecewise linear and touches the empirical CDF at some of the observations. Suppose there are n observations and $0 < y_1 < y_2 < \ldots < y_n$ are the corresponding order statistics. The construction of the LCM is illustrated in Figure 3.1 for the case $n = 5$.

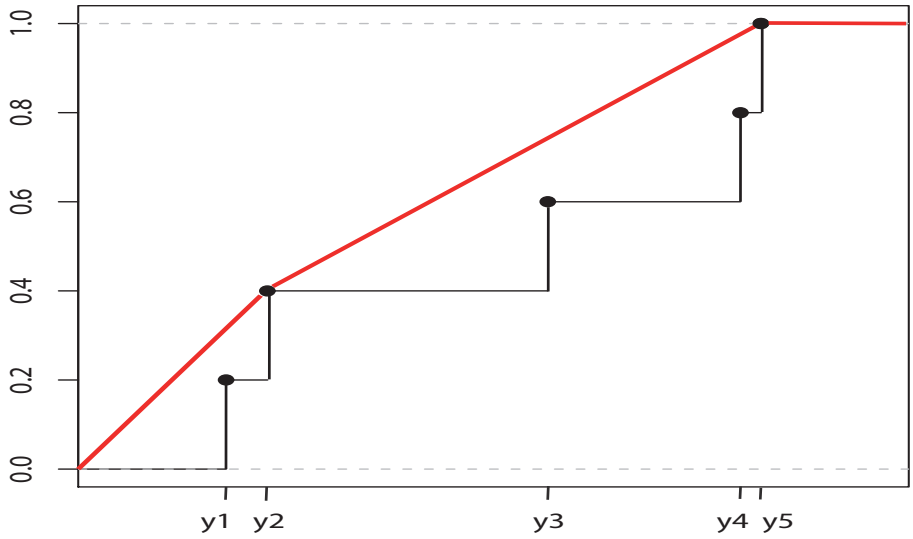

Figure 3.1: Empirical CDF and LCM

The emprical CDF is the staircase starting at the origin having an upward jump at each observation of size $1/n$ and remaining constant equal to 1 after the last observation. The LCM starts at the origin and is obtained by stretching a rubber band over the staircase. In this particular example it touches the empirical CDF at the second order statistic and at the last observation after which it remains constant equal to 1.

The LCM of the empirical CDF can be computed by the following algorithm. Suppose there are n observations, then the empirical distribution function F_n is represented by the points:

$$(0,0),\ (y_1, F_n(y_1)),\ (y_2, F_n(y_2)),\ \ldots\ , (y_n, F_n(y_n)).$$

The algorithm consists of the following steps:

1. Start with the first point $(0,0)$ and calculate the slopes of all lines connecting the first point with all succeeding points. Thus, there will be n slopes:

$$\frac{F_n(y_i)}{y_i}, \quad i = 1, 2, \ldots, n.$$

2. Find the line with the maximum slope and denote the corresponding point by $(y_{(e_1)}, F_n(y_{(e_1)}))$. Connect the points $(0,0)$ and $(y_{(e_1)}, F_n(y_{(e_1)}))$ with a straight line. This will be the first segment of the LCM.

3. Use the endpoint of the first segment $(y_{(e_1)}, F_n(y_{(e_1)}))$ as the new starting point and calculate all slopes between this point with the succeeding points:

$$\frac{F_n(y_{(e_1)}) - F_n(y_i)}{y_{(e_1)} - y_i}, \quad y_i > y_{(e_1)}.$$

Find the line with the maximum slope and denote the endpoint by $(y_{(e_2)}, F_n(y_{(e_2)}))$. Connect the points $(y_{(e_1)}, F_n(y_{(e_1)}))$ and $(y_{(e_2)}, F_n(y_{(e_2)}))$ with a straight line. This will be the second segment of the LCM.

4. Repeat step 3 by taking the endpoint of the last segment as the new starting point and continue until the LCM reaches the last point $(y_n, F_n(y_n) = (y_n, 1)$.

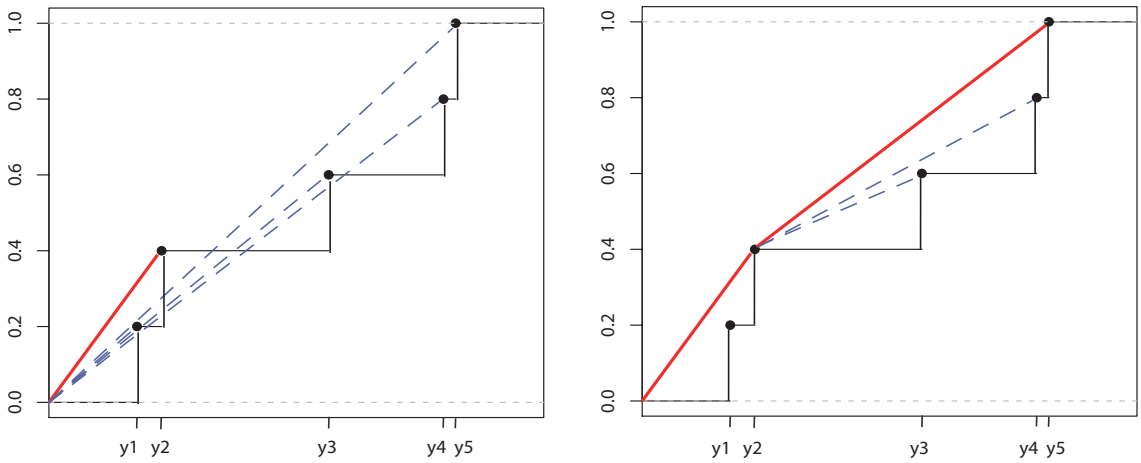

Figure 3.2: Illustration of Algorithm in Constructing LCM

One may feel uncomfortable with the fact that "monotonizing" the empirical CDF leads to a LCM which lies completely above the empirical CDF. This typically is the consequence of misinterpreting the LCM as being an "adjustment" of the empirical CDF, in which case one expects a concave curve that runs through the points of the empirical CDF instead of dominating it completely. However, the LCM should not be seen as an adjustment for the empirical CDF to incorporate concavity of the underlying CDF. It is the probability density for which we try to find the maximum likelihood estimator and it can be shown that this estimator is obtained as the left-derivative of the LCM. If one looks at the density and draws a naive estimate for the density, e.g., a histogram or some smooth alternative like a kernel estimator, then the ML estimator (which is the derivative of the LCM) indeed would be a nonincreasing curve that runs through the naive estimate.

The idea of "monotonicing" a naive estimator has been extended to several statistical models, such as the estimation of a decreasing regression curve or a decreasing hazard function of regular data (see Robertson *et al.* [9]). In the context of survival analysis, this idea has been investigated for the estimation of a decreasing density or a decreasing hazard function of right censored observations. The naive way to estimate the CDF of right censored observations is to use the Kaplan-Meier estimator. Incorporating the constraint that the corresponding density is nonincreasing, suggests to "monotonize" the Kaplan-Meier estimator in the same way as in the density model: determine the LCM of the Kaplan-Meier estimator and compute its left-derivative. Once again this results in a piecewise constant nonincreasing estimate. It can be shown [10] that the true maximum likelihood estimator in this model is slightly different from the monotonized Kaplan-Meier, but the difference tends to zero as the sample size n tends to infinity. A shape

constrained estimator of a nonincreasing hazard function can be obtained by taking the left-derivative of the LCM of the Nelson-Aalen estimator, which is the naive estimate for the cumulative hazard function [11].

These examples suggest to employ the same idea in the Cox model. The naive estimator for the baseline cumulative hazard function H_0 is the Breslow estimator, which is Nelson-Aalen equivalent for the Cox model. To incorporate the assumption that baseline hazard h_0 is nonincreasing, in view of the previous examples, one could "monotonize" the Breslow estimator, i.e., construct its LCM $\widetilde{H}_0$ and use the left-derivative $\widetilde{h}_0$ as an estimate for h_0.

3.2 The LCM for the Cumulative Hazard in the ICTUS dataset

We are now going to apply the procedure of adjusting the traditional estimator of the cumulative hazard using the LCM explained in the previous subsection. Both estimators for the cumulative hazard (the Nelson-Aalen and the Breslow estimator) explained in Section 2.5.3 provide a similar estimate in their shapes. The application of the LCM will result in similar shapes of the baseline hazard. In R software, there is a package ("fdrtool") that supplies the command to produce the LCM given the set of points from a staircase function. Figure 3.3 provides the graph of the Nelson-Aalen cumulative hazard with the corresponding LCM.

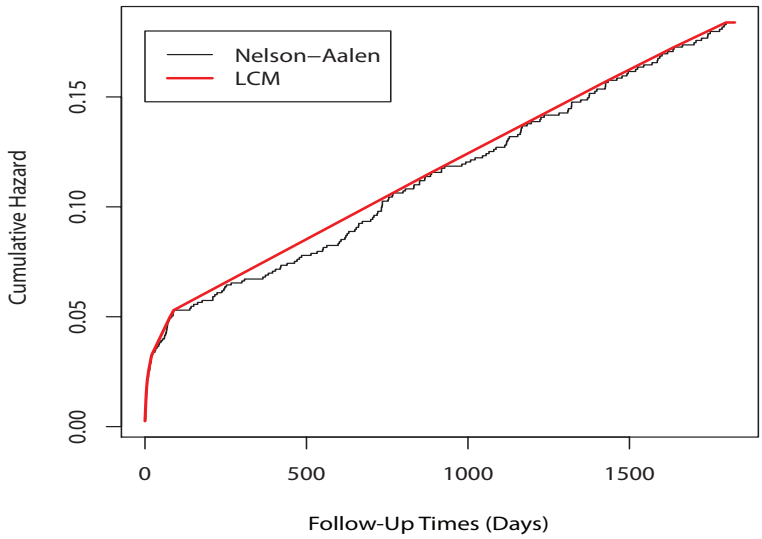

Figure 3.3: Least Concave Majorant of The Nelson-Aalen Cumulative Hazard

In Figure 3.3, we can see that the resulting curve slightly differs from the traditional estimator. This red line connects the points which were produced by the command to produce the LCM in R software. After day 89, the line is linear until day 1434. From this, it can be said that after day 89, the risk of having an event is getting constant. This situation can be explained by the picture of the hazard function in Figure 3.4. The graph of the hazard function derived from the LCM is obtained by taking the slope of the line segment in the LCM of the cumulative hazard. It is confirmed by this figure that the hazard from day 89 is constant at certain level until some certain point and then drops again close to 0.

The hazard curve obtained by applying the LCM is guaranteed to be monotone decreasing over time while using the traditional estimator we still see some ups and downs (the numerical derivative from the Nelson-Aalen estimator in Figure 3.5). However, the hazard curves seem comparable except in some points

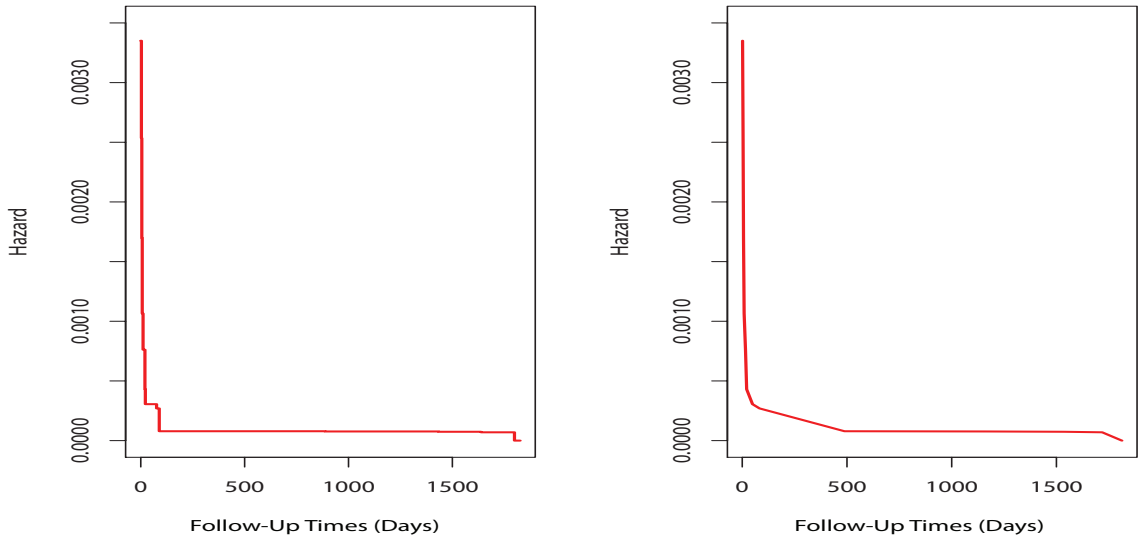

Figure 3.4: The Baseline Hazard from the Monotonized Cumulative Baseline Hazard using LCM (Left Figure) Slope of LCM (Right Figure) The Smoothed version of slope from LCM.

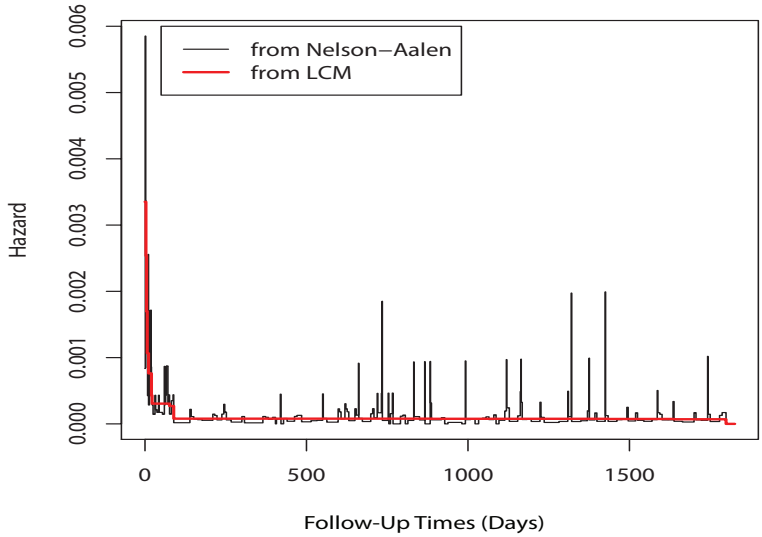

Figure 3.5: Baseline Hazard Comparison

where the naive estimator has jumps (see Figure 3.5). This suggests that the LCM is reasonable to estimate the cumulative hazard for the ICTUS dataset. To investigate this, we perform a test in Section 3.6.

3.3 Testing Monotonicity

It is seen that the LCM incorporates the assumption of monotonicity. If we look at Figure 3.3, the LCM gives a cumulative hazard function which is slightly different from the traditional estimator. Even if their values are only slightly different, it still has to be verified whether this difference is still acceptable to conclude that the cumulative hazard for the ICTUS dataset is concave. Therefore, in this section we describe the procedure to statistically test whether a monotone hazard is a reasonable assumption for this dataset.

The procedure for the statistical test is started with stating the null hypothesis. In this case, the null hypothesis is that the hazard function is monotone decreasing. Thus, the corresponding cumulative hazard function is concave. Assuming that the null hypothesis is true, one tries to disprove the alternative hypothesis, which says the hazard function is not monotone decreasing. For this purpose, an appropriate

test statistic is needed. Based on the distribution of the test statistic under the assumption that the null hypothesis is true, one can compute the p-value. The smaller the p-value (usually smaller than 0.05) the more one is inclined to reject the null hypothesis.

The next step is to describe a relevant test statistic. Consider the curves resulting from the LCM and the Nelson-Aalen cumulative hazard in Figure 3.3. One could think about a possible test statistic as the total area between the LCM curve and the Nelson-Aalen estimator for the baseline cumulative hazard, which can be formulated as follows

$$A = \int (\widetilde{H}(t) - \widehat{H}(t))dt \tag{3.1}$$

where $\widehat{H}$ is the cumulative hazard curve derived using the Nelson-Aalen estimator and $\widetilde{H}$ is the LCM of the Nelson-Aalen cumulative hazard curve. This total area would take larger values if the true cumulative hazard is not concave. If this is the case, then the null hypothesis will be rejected. If it is small, then the null hypothesis is not rejected. However, the term "small" should be made clear. This term can be quantified if we know the distribution of the test statistic. Having this distribution, we can then compute the probability that the test statistic is larger than the observed value, known as the p-value.

The p-value is the probability that the test statistic is at least as large as the observed test statistic, assuming that the null hypothesis is true. This value can be calculated if the distribution of the test statistic is known. However, in our situation, there is no knowledge about the distribution of the test statistic under the null hypothesis. Therefore, the distribution of the test statistic will be approximated and for this purpose, the bootstrapping procedure is used.

3.4 Bootstrap Method

A sampling distribution is the distribution of a given statistic based on a random sample of certain size, say n. It could be possible for a statistic to depend on information about the population from which samples are drawn. If the population distribution is unknown, the problem of estimating the distribution of the statistic maybe difficult. However, this problem can be solved if we can estimate the population distribution. The empirical distribution function is known to be a consistent estimator for the true cumulative distribution function. After having this estimate of the true distribution, a sample can be drawn and one value of the statistic can be obtained. If this process is repeated several times, we have an approximation of the true distribution of the statistic. This method is known as the bootstrap.

The bootstrap idea is based on the fact that original sample represents the population from which it was drawn. Thus, resamples from an estimate based on this random sample, represent random samples that one would obtain from the population. Bootstrapping right-censored data in the form (t_i, δ_i) is not straightforward. It is not done by randomly resampling from the sample since the variable t_i comes from the random censoring mechanism (2.1). We should sample both the survival time X_i and censoring time C_i from their distributions instead. The bootstrap procedure for the right censored data will be explained in Section 3.5.

3.5 Bootstrap Method for General Right Censored Data

In testing monotonicity of the hazard, the bootstrap method is needed to approximate the distribution of the test statistic under the assumption that the null hypothesis is true. To this end, we resample the survival times from the LCM and censoring times from a naive estimator of the cumulative hazard function.

Consider the right censored data in the form (t_i, δ_i), $i = 1, 2, ..., n$. Here n represents the resamples size which is the same size as the original sample. For convenience, the follow-up times are ordered as $t_{(1)} < t_{(2)} < ... < t_{(n)}$. As mentioned above, bootstrapping the right censored data is done by bootstrapping the survival times X_i and the censoring times C_i. This means by randomly sample those variables from their estimated distribution. Their actual distribution is unknown but we do have one random sample. Thus, under the null hypothesis, we can estimate their distribution by the LCM of the Kaplan-Meier estimator for the survival function or the LCM of the Nelson-Aalen estimator for the cumulative hazard. We denote the estimated survival function by $\widehat{S}_X(t)$.

The estimated survival function for the censoring variable C_i can be obtained similarly. Note that the uncensored observation of X_i corresponds to a censored observation of C_i and vice versa. The new censoring indicator $\Gamma_i = 1 - \Delta_i$ is introduced and γ_i represents the censoring indicator for the ith subject. Consider the dataset of the form (t_i, γ_i). When $\gamma_i = 1$, then t_i is the exact censoring time. We can estimate the survival function for C_i, say $\widehat{S}_C(t)$ by the Kaplan-Meier estimate

$$\widehat{S}_C(t) = \prod_{t_{(j)} \leq t} \frac{n_j^o - d_j^o}{n_j^o},$$

where d_j^o is the number of observed censoring events at $t_{(j)}$ and n_j^o is the number of individuals at risk at $t_{(j)}$. Individuals at risk here are those whose observed follow-up time is at least $t_{(j)}$.

In general, the procedure for bootstrapping the survival and censored times is described as follows.

1. Independently sample the survival times X_i^* from the estimated survival function $\widehat{S}_X(t)$ and the censoring times C_i^* from the estimated survival function $\widehat{S}_C(t)$. To do this, notice that if a random variable W has a continuous survival distribution S, then the random variable $U = S(W)$ is uniformly distributed on [0,1]. In order to sample a survival time X_i^* from the estimated survival function $\widehat{S}_X(t)$, one take a random number U which is uniformly distributed on [0,1] and compute $X_i^* = \widehat{S}_X^{-1}(U)$. A similar procedure is applied to sample a censoring time C_i^* from the estimated survival function $\widehat{S}_C(t)$.

2. Define the follow-up time $T_i^* = \min\{X_i^*, C_i^*\}$ for the ith patient. The indicator Δ_i^* can take on two values,

$$\Delta_i^* = \begin{cases} 1, & \text{if } X_i^* \leq C_i^*; \\ 0, & \text{if } X_i^* > C_i^*. \end{cases}$$

   The kth bootstrap sample will be of the form (T_i^*, Δ_i^*), $i = 1, 2, ..., n$ where n is the same as for the original sample. We take many bootstrap samples and for each bootstrap sample we compute the value of the test statistic. By doing this, we expect to approximate the true sampling distribution.

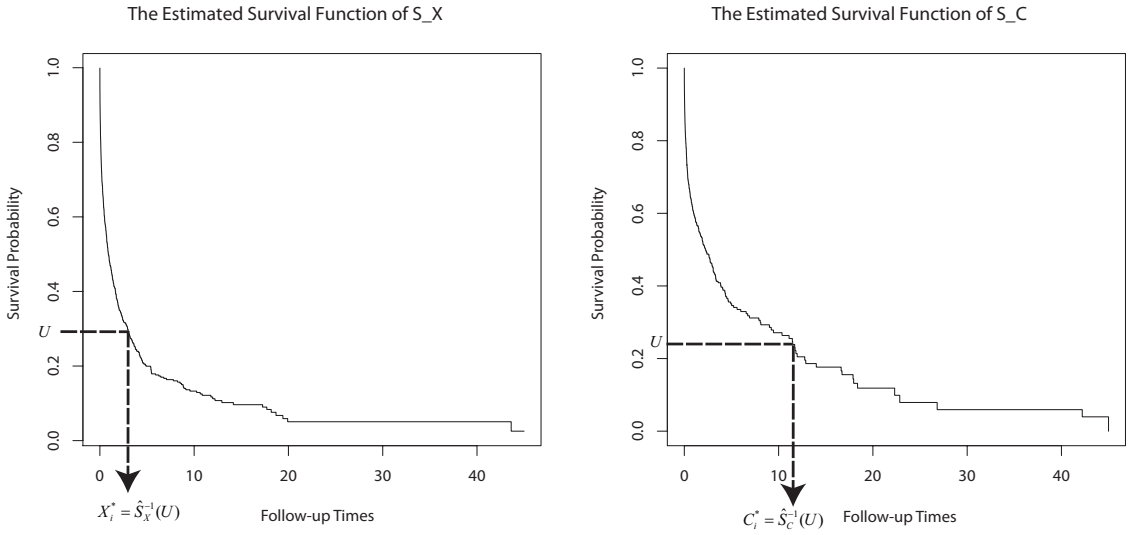

Figure 3.6: Illustration of Sampling Procedure from Survival and Censoring Distribution

3.6 Procedure of Testing Monotonicity in the ICTUS Dataset

As explained in Section 3.5, a bootstrap sample of survival and censoring times, respectively denoted as X_i^* and C_i^* are generated for $i = 1, ..., n$. Part of the procedure for testing monotonicity of the baseline hazard in the ICTUS trial has been explained in Section 3.3 including the definition of the test statistic. In this subsection the procedure is continued, especially by approximating the distribution of the test statistic using the bootstrap method.

The bootstrap technique consist of drawing a lot of resamples. Assume there are B resamples. Each resample is of the form $(T_i^*, \Delta_i^*), i = 1, ..., n = 1200$. For each resample, we can compute the Nelson-Aalen estimator for the cumulative hazard $\widehat{H}^*(t)$ as well as its LCM $\widetilde{H}^*(t)$. Having these two estimators, one can then calculate the test statistic A^* for each resample k. Thus, the p-value can be approximated as

$$p\text{-value} \approx P(A^* \geq A) \simeq \frac{\#(A^* \geq A)}{B} \tag{3.2}$$

where A is the value of the test statistic for the observed data.

Now, we are going to bootstrap the survival and censoring times from their corresponding estimated distributions. It should be noted that bootstrapping the survival times X_i^* is done under the assumption that the null hypothesis is true. Therefore, the X_i^* are obtained from the LCM which will be explained in more detail as follows. Whilst bootstrapping the censoring distribution is done on the basis of its naively estimated distribution. We first explain the bootstrap process for the survival times. The cumulative hazard from LCM is transformed into the survival curve to make sure that it only takes value between [0,1]. By randomly sample U from uniform distribution on $[0, 1]$ and then transform it back into the cumulative hazard by using the inverse of LCM, one obtains the bootstrap sample of X_i^*'s.

Bootstrapping the censoring times C_i^* is done as follows. A new event indicator should be assigned, i.e., $\Gamma_i = \mathbf{1}\{C_i \leq X_i\}$. The Nelson-Aalen estimator applied to the ICTUS results in Figure 3.7. The curve shows a slight increasing pattern between 0 and 1825. At day 1826, the curve shows an extreme increases. Since this is the cumulative hazard for the censoring distribution, it can be seen that before day 1000, several patients were lost to be contacted. The study is then terminated at day 1826, and no

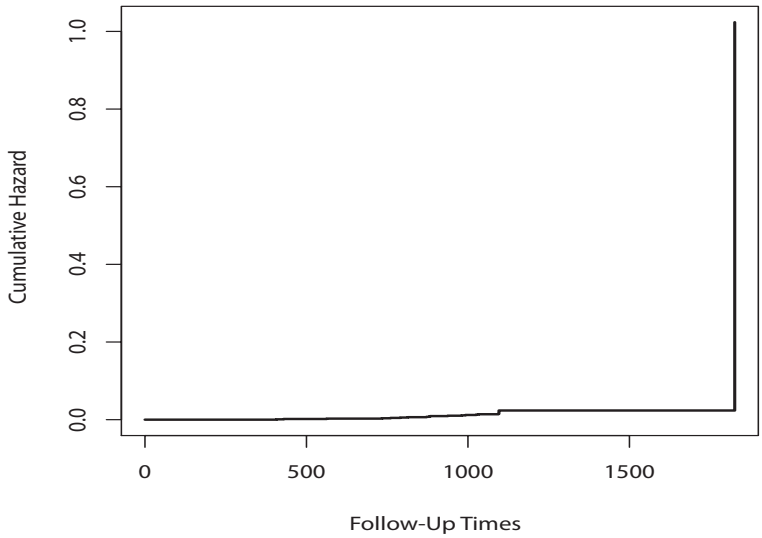

Figure 3.7: The Cumulative Hazard Distribution for the Censoring Times.

patients showed an event at this time. Having the censoring distribution, we now use the similar procedure as generating the survival times. First, transform this curve into the survival curve and randomly sample U from uniform distribution on $[0, 1]$. Substitute this random variable into the inverse of this Nelson-Aalen curve will give a bootstrap sample of C_i^*'s.

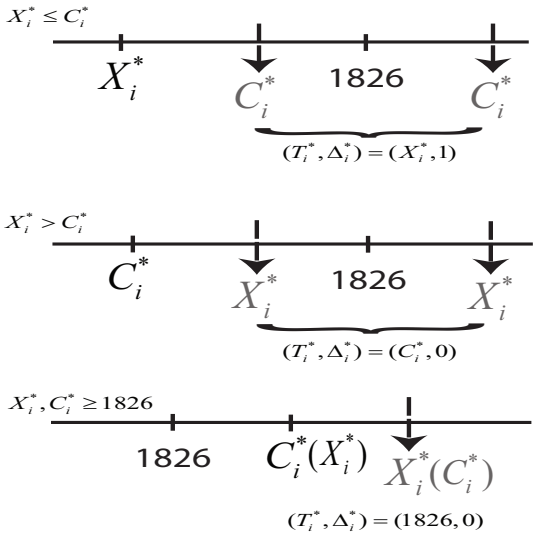

Figure 3.8: The Scheme for the Follow-up Times

The bootstrap dataset then will be in the form (T_i^*, Δ_i^*), $i = 1, \ldots, 1200$. There are three possibilities to assign the values of T_i^*'s and Δ_i^*'s based on the scheme illustrated in Figure 3.8. This scheme is based on the duration of the study which is 1826 days. First, consider $X_i^* \le C_i^*$ and $X_i^* \le 1826$, then for each possible C_i^* satisfying that condition the data is assigned as $(T_i^*, \Delta_i^*) = (X_i^*, 1)$. In the case $X_i^* > C_i^*$, when $X_i^* \le 1826$ the corresponding individual left the study and his/her true actual event times will not be known. For this case, the corresponding data will be $(T_i^*, \Delta_i^*) = (C_i^*, 0)$. If both X_i^* and C_i^* are larger than 1826 (whichever X_i^* or C_i^* comes first) then the data will be of the form $(T_i^*, \Delta_i^*) = (1826, 0)$.

Applying this bootstrap procedure to the ICTUS dataset with $B = 10000$ bootstrap samples, gives an approximation of the distribution of the test statistic under the null hypothesis and can be seen in Figure 3.9. The approximation of the p-value is 0.701. This value suggests that there is not enough evidence to

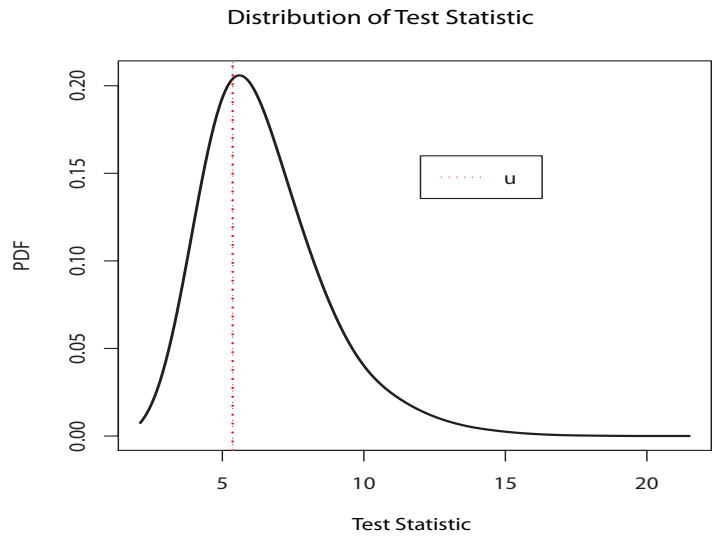

Figure 3.9: The Distribution of the Test Statistic

reject the null hypothesis of a monotone decreasing hazard. Thus, it can be concluded that it is reasonable to assume that the cumulative hazard function for the ICTUS dataset is concave and can be estimated by the LCM method.

3.7 Validating the Test Statistic

The above sampling distribution is obtained by bootstrapping from the estimated distribution. This is done because the true distribution is unknown. We only have one random sample, which is the ICTUS dataset. We want to investigate whether bootstrapping from the true distribution would give a different sampling distribution compared to bootstrapping from this one random sample.

In order to validate our test statistic, we use a parametric distribution such that we know the distribution of the population. Further, we fix one random sample from this parametric distribution such that it resembles the ICTUS dataset. The idea is to bootstrap the test statistic from the true distribution (scheme 1) and compare it to the bootstrapped test statistic from the fixed random sample from this true distribution (scheme 2). The comparison of the distribution of test statistic resulting from these two schemes is based on their shape and the p-value.

The next question that might occur is regarding the parametric distribution that we are going to use for this validation. Since we are dealing with a concave cumulative hazard, we are interested in a distribution for which the cumulative hazard function is concave. One possible choice is the Weibull distribution, in which the concavity of the cumulative hazard is controlled by one parameter. Details about this distribution are given in Section 3.8.

3.8 Weibull Distribution

The Weibull dstribution is a continuous probability distribution. If a random variable T has a Weibull distribution then its probability density function is defined as follows.

$$f(t) = \begin{cases} \frac{\kappa}{\lambda}\left(\frac{t}{\lambda}\right)^{(\kappa-1)} \exp\left(-\left(\frac{t}{\lambda}\right)^{\kappa}\right), & t \geq 0; \\ 0, & t < 0. \end{cases} \tag{3.3}$$

where $\kappa > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. Based on the definition of the probability density function, other quantities such as the hazard and the cumulative hazard function can also be derived. The hazard function of Weibull distribution is

$$h(t) = \frac{\kappa}{\lambda}\left(\frac{t}{\lambda}\right)^{\kappa-1}. \tag{3.4}$$

and the corresponding cumulative hazard function

$$H(t) = \left(\frac{t}{\lambda}\right)^{\kappa} \tag{3.5}$$

As can be seen from (3.4) and (3.5), the shape of the cumulative distribution function (and correspondingly the hazard function) is completely determined by the value of the shape parameter κ. Since κ assumes a value on $(0, \infty)$, then for $0 < \kappa < 1$, the hazard function is monotone decreasing. Correspondingly, the cumulative hazard function will be concave. In the case when $\kappa = 1$, the Weibull distribution becomes a specific distribution, namely the exponential distribution with parameter $\frac{1}{\lambda}$ and the hazard for this distribution is constant over time. If κ assumes a value between $(1, \infty)$ then the hazard function is monotone increasing and correspondingly the cumulative hazard function is convex. The differences in the shape of the hazard function regarding the variation in the shape parameter value are given in Figure 3.10.

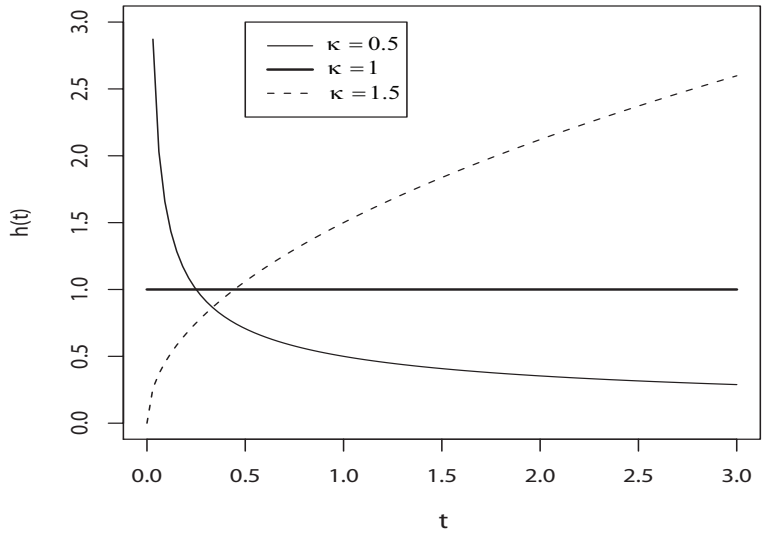

Figure 3.10: The Weibull Hazard Function Depending on the Shape Parameters.

We will use the Weibull distribution to validate the test statistic. For this purpose, the parameters of the Weibull distribution are based on the ICTUS dataset. We will use equation (2.6) to estimate these parameters which will be explained in detail later on. It should be noted that the Weibull distribution based on the ICTUS data involves two derivations. We choose a Weibull distribution for the survival time distribution and also for the distribution of the censoring times. The reason to choose Weibull distribution

as the censoring distribution is based on the shape of its estimated cumulative hazard. The curve is convex, thus the Weibull distribution with shape parameter larger than 1 is suitable.

We first determine the Weibull distribution for the survival times. From the right censored data in ICTUS $(t_i, \delta_i), i = 1, \ldots, 1200$, the likelihood function as (2.6) is formed with the probability density and survival function of the Weibull distribution. We substitute all values of the data into this likelihood and maximize this likelihood over all possible values of parameters of the Weibull distribution. It turns out that the estimated Weibull distribution for the survival times has scale parameter $\lambda_X = 50771.97$ and shape parameter $\kappa_X = 0.5155445$.

The derivation of the Weibull distribution for the censoring times is quite similar to the derivation for X_i. It differs in the sort of right censored data that is used in the likelihood function. Instead of using right censored data of the form (t_i, δ_i), we use right censored data (t_i, γ_i) where $\gamma_i = 1 - \delta_i$. Maximizing the likelihood that is formed for the censoring times, we got scale parameter $\lambda_C = 1825.224$ and shape parameter $\kappa_C = 56.00124$. Both Weibull approximations for X and C are given in the following figure.

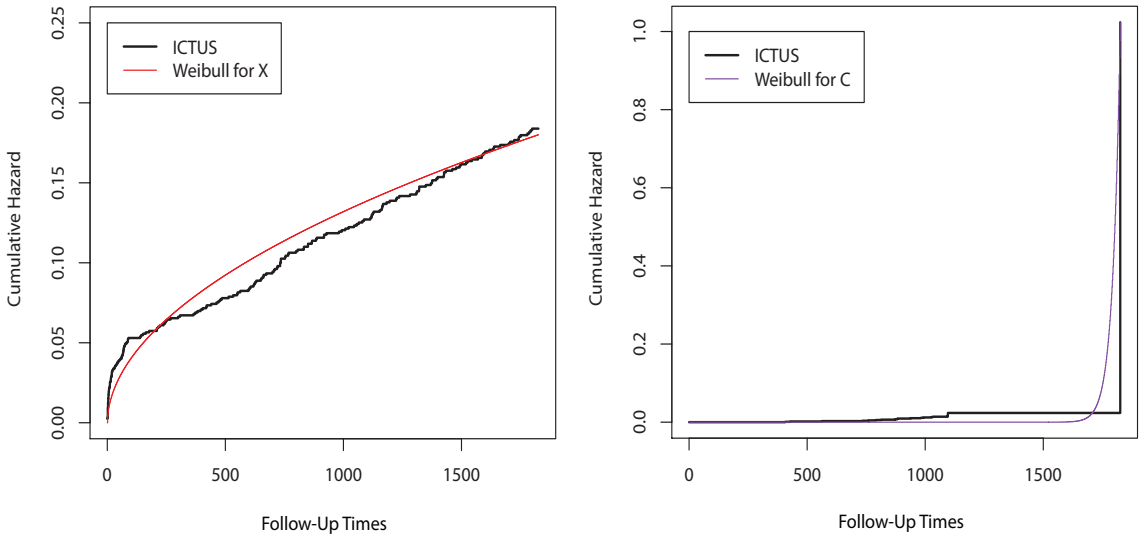

Figure 3.11: The Weibull Hazard Function Depending on the shape parameters

Having estimated the Weibull distributions for X and C, we are now going to compare the distribution of the test statistic using scheme 1 and scheme 2 explained above. The details of scheme 1 and scheme 2 are as follows.

Scheme 1: Sampling from the true Weibull distribution. The procedure for bootstrapping the test statistic in this scheme is given by the following procedure. Let W_X and W_C be the Weibull distributions for the survival times X and censoring times C respectively.

1. Randomly sample $X_i^{(1)}$ from W_X and $C_i^{(1)}$ from W_C, $i = 1, \ldots, 1200$.
2. Define $(T_i^{(1)}, \Delta_i^{(1)})$ as the follow up time and censoring indicator, where $T_i^{(1)} = \min\{X_i^{(1)}, C_i^{(1)}\}$ and $\Delta_i^{(1)} = \mathbf{1}\{X_i^{(1)} \leq C_i^{(1)}\}$.
3. Construct the Nelson-Aalen estimator (2.8) for the cumulative hazard $(\widehat{H}^{(1)}(t))$ based on the right censored data $(T_i^{(1)}, \Delta_i^{(1)})$.
4. Construct the LCM of the Nelson-Aalen estimator; denote this estimator as $\widetilde{H}^{(1)}(t)$.

5. Calculate the test statistic A, defined in (3.1).

6. Repeat step 1 to step 5 for B times, here $B = 1000$ times.

Scheme 2: Sampling from a sample from the true Weibull distribution. For this scheme we first do step 1 to step 4 from scheme 1. By doing these steps, we are given with one fixed sample from the Weibull distribution $(T_i^{(2)}, \Delta_i^{(2)}), i = 1, \ldots, 1200$ as well as the estimated Weibull cumulative hazard, which is obtained from step 3 and step 4. From this estimate of the Weibull cumulative hazard, we can derive the estimates of the distribution for X and C. Let the estimate distributions for X and C be $\hat{W}_S$ and $\hat{W}_C$, respectively.

1. Randomly sample $X_i^{(2)}$ from $\hat{W}_S$ and $C_i^{(2)}$ from $\hat{W}_C$, $i = 1, \ldots, 1200$.

2. Define $(T_i^{(2*)}, \Delta_i^{(2*)})$ as the follow up time and censoring indicator, where $T_i^{(2*)} = \min\{X_i^{(2)}, C_i^{(2)}\}$ and $\Delta_i^{(2*)} = \mathbf{1}\{X_i^{(2)} \leq C_i^{(2)}\}$.

3. Construct the Nelson-Aalen estimator for the cumulative hazard $(\widehat{H}^{(2*)}(t))$ based on the right censored data $(T_i^{(2*)}, \Delta_i^{(2*)})$.

4. Construct the LCM of the Nelson-Aalen estimator (2.8); denote this estimator as $\widetilde{H}^{(2*)}(t)$.

5. Calculate the test statistic A, defined in Section (3.1).

6. Repeat step 1 to step 5 for B times, here $B = 1000$ times.

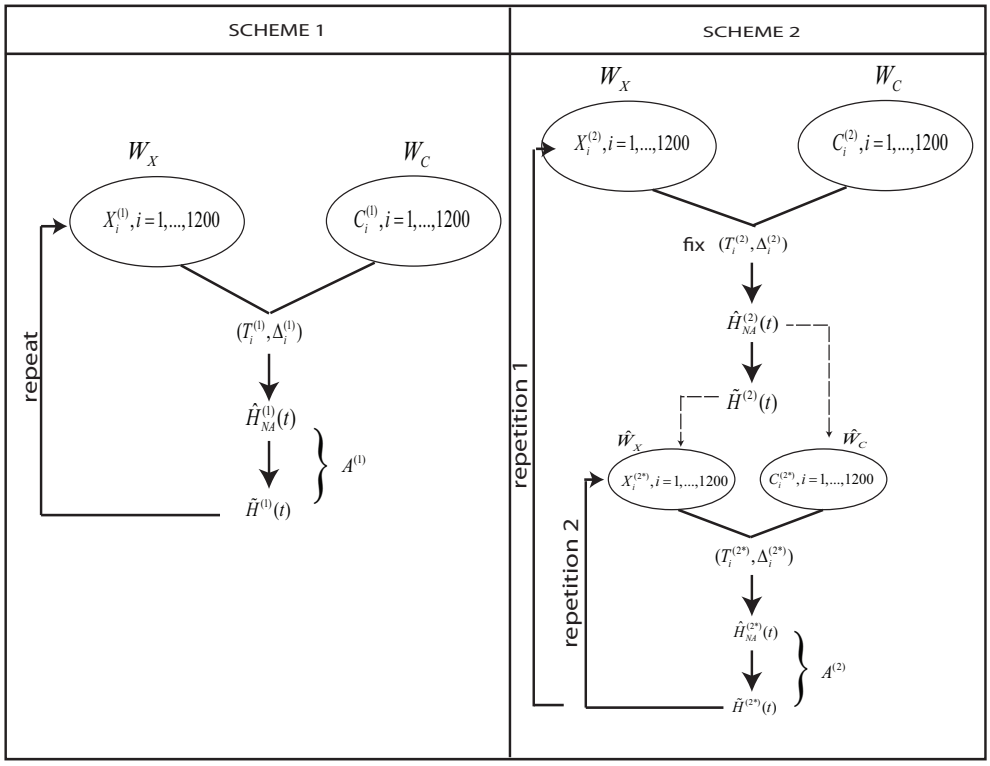

Figure 3.12: The Illustration of the Schemes to Validate the Test Statistic

Figure 3.13 gives an illustration of the resulting distribution of the test statistic given by both schemes. The grey dashed lines are the distribution of test statistic resulted from scheme 2 for several fixed Weibull samples. While the solid bold blue line resulted from scheme 1. From this illustration, we can see that

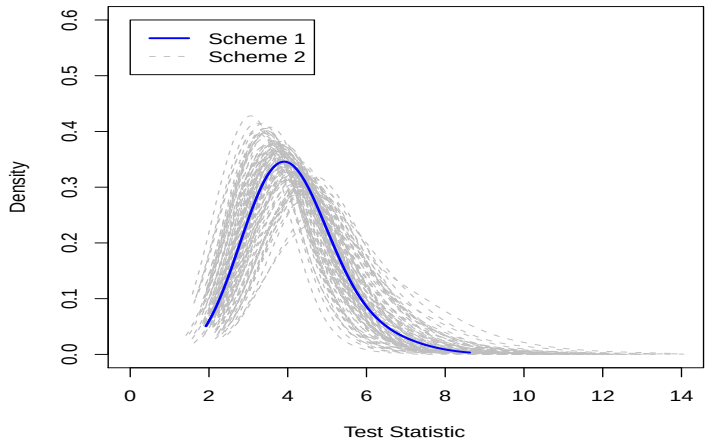

Figure 3.13: The Distribution Comparison between Schemes

actually the test statistic distribution which are calculated from the sample will have a distribution that approximates the distribution of test statistic drawn from the known distribution. We also consider the comparison in the $p-$values between schemes. In scheme 2, we have one $p-$values in each iteration of "repetition 2". This is calculated on the basis of the observed areas between the traditional cumulative hazard and LCM. In order to compute the $p-$values on scheme 1, we use these observed area. The mean of the $p-$value obtained by the true distribution (scheme 1) is 0.51 with standard deviation 0.28, while using the scheme 2 is 0.48 with standard deviation 0.26. Thus, the shapes of the test statistic distributions which were given by scheme 1 and 2 and their corresponding p-values mean supports the validity of the test statistic procedure which is based on an estimated distribution.

3.9 Summary

In this chapter, a new estimator that incorporates the assumption of a monotone hazard function was introduced. Graphically, it has a comparable hazard to the traditional estimator (hazard by the Nelson-Aalen estimator). Hypothesis testing is then performed to quantitatively investigate whether assuming a monotone hazard function is reasonable in the ICTUS dataset. In performing the test, we dealt with the problem of generating the distribution of the test statistic. The distribution of the test statistic is unknown and bootstrap method is needed for this purpose. The p-value which was computed using the bootstrapped distribution of the test statistic did not give enough evidence to reject the null hypothesis.

The problem that arises later on is the validity of the test itself. The distribution of the test statistic is obtained by bootstrapping. We investigated whether there is a large difference in the distribution of the test statistic if it were simulated from the real distribution. Using the Weibull distribution, in which all parameters are obtained based on the ICTUS data, we see that the distribution of the test statistic based on one sample closely approximates the distribution of the test statistic simulated from the true distribution.

Chapter 4

Effects of applying the LCM in the Cox Model

In Chapter 3, it is confirmed by hypothesis testing that the LCM is reasonable to estimate the cumulative hazard and the corresponding hazard in the ICTUS dataset. It is also interesting to investigate the effect of LCM on estimating Cox's baseline hazard. By applying the LCM to the estimated cumulative baseline hazard, an estimate for the baseline hazard is obtained. We are interested on how this estimate behaves, whether it is a better estimator for the baseline hazard compared to Breslow's. The term "better estimate" refers to the estimator having smaller standard deviation and bias than the existing estimator. This will be explained in more detail in Section 4.3.

Recall that in the Cox model, there are two parameters which have to be estimated, i.e., the regression coefficients and the baseline hazard. The effect of LCM on the estimate of the regression coefficients is also an interesting issue to be investigated. The partial likelihood method which is proposed by Cox does not incorporate any assumption on the baseline hazard. This method is proposed due to infeasibility of maximizing the full log-likelihood (2.14). In brief, we use the estimate of the baseline hazard, which is obtained from the LCM, to maximize the full likelihood over all possible regression coefficients. We investigate the distribution of these regression coefficients compared to those which are obtained by partial likelihood in Section 4.4.

In order to assess the quality of the estimates, we need to form a Cox dataset from a certain distribution. By doing this, we know the true value of the regression coefficients and the baseline hazard so that we can measure the distance between the true value and their estimators. Details on generating a dataset based on a Cox model is given in Section 4.1.

4.1 Generating the Cox Data

As defined in equation (2.9), the two parameters in Cox model are the baseline hazard and the regression coefficients. In order to generate a dataset based on a fixed Cox model, we need information about the baseline distribution and the covariates values. As in Chapter 3, we choose the Weibull distribution because its concavity is controlled by one parameter. We use Weibull distribution with shape parameter between 0 and 1 as the baseline hazard. For simplicity, we use only one covariate which is a dichotomous variable (its value is either 0 or 1) and fix the regression coefficient (β) equal to 1.

On the basis of a fixed Cox model, we need to generate the dataset which consists of n subjects in the form (T_i, Δ_i, Z_i) with T_i is the follow-up time, Δ_i is the censoring indicator and Z_i is the covariate assigned

to the ith subject. The values of Z_i are randomly assigned to the ith subject. The problem remains on how to construct the dataset in the form (T_i, Δ_i). As in Chapter 3, the follow-up time and censoring indicator for each subject depends on its survival time and censoring time, whichever occurs first. For this purpose, we need to generate survival times X_i and censoring times C_i from the survival and censoring distributions. Note that in this chapter, we deal with the Cox model, so there is a slight difference in generating the survival and the censoring times. Details on generating the Cox dataset can be read in [12].

First, the survival times X_i are randomly generated from the survival distribution. Since X has a hazard function of equation (2.9) or a cumulative hazard formulated in equation (2.10), we use this to generate the X_i's. From the cumulative hazard of the Cox model, we can derive the survival distribution (2.11). Therefore, if X has the following survival distribution

$$S(x, \boldsymbol{z}, \boldsymbol{\beta}) = [\exp(-H_0(x))]^{\exp(\boldsymbol{z}'\boldsymbol{\beta})}$$

then a random variable $U = S(X, \boldsymbol{z}, \boldsymbol{\beta})$ has a Uniform distribution on [0,1]. Thus, by randomly sampling U from Uniform[0,1] and setting

$$X = H_0^{-1}\left(-\log(U)\exp(-\boldsymbol{z}'\boldsymbol{\beta})\right), \tag{4.1}$$

we can generate a survival time X_i. Furthermore, since the baseline hazard function is already assumed to be from a Weibull distribution, we can set the baseline cumulative hazard H_0 as in the equation (3.5). The inverse of the baseline cumulative hazard is thus the inverse of Weibull hazard as follows.

$$H_0^{-1}(u) = \lambda u^{\frac{1}{\kappa}} \tag{4.2}$$

Generating censoring times C_i in the Cox model is similar to generating censoring times based on the Kaplan-Meier estimator. Suppose that the censoring distribution has a cumulative hazard H_c. Then, we begin by drawing a random number U from Uniform[0,1] and setting

$$C = H_c^{-1}(U) \tag{4.3}$$

with H_c^{-1} is the inverse of the cumulative hazard function.

Several questions related with censoring distribution might raise up, such as the type of the distribution and the censoring proportion. In [13], the exponential, the uniform and truncation are several common distributions which were chosen for the censoring distribution in the Cox model. Thus, we need to determine the parameters for these distribution. Another important factors in determining these parameters are the censoring proportion. Given the censoring proportion and the survival distribution, Halabi and Singh [14] provides a formula to compute the parameters for several censoring distribution. Having generated the survival and censoring times X_i and C_i respectively, we could define T_i and Δ_i as before. As a result, we have a Cox dataset in the form $(T_i, \Delta_i, \boldsymbol{Z}_i)$, say the Weibull-Cox dataset.

4.2 The Effect on ICTUS's Baseline Hazard

In this section, the performance of the LCM based estimator will be compared to the Breslow estimator. To do this, we need a true value of the baseline hazard. To this end, a Weibull-Cox dataset based on the ICTUS data is generated. We estimate the Weibull parameters for the survival and censoring using

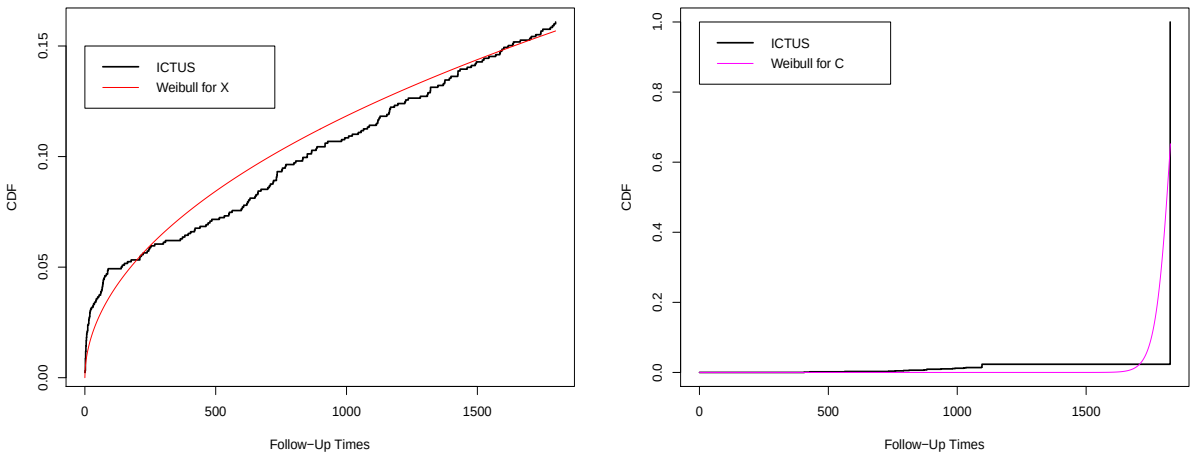

Figure 4.1: The Weibull CDF for the Survival and Censoring Distribution Based on the ICTUS Dataset

maximum likelihood as we did in Chapter 3. We use Weibull distributions as before for the survival and censoring distributions. It turns out that the Weibull parameters for the survival distribution are (shape = 0. 5156, scale 55557.67) while for the censoring distribution are (shape =56.00124, scale = 1825.224). To assess the performance, we use a pointwise measure, such as the Mean Squared Error (MSE) to quantify the differences between these estimators and the true hazard. We will also calculate the bias and the variance for each estimators so as to measure the variability of its value over several points in the domain of the baseline hazard. The Mean Integrated Squared Error (MISE) will also be calculated for both estimators.

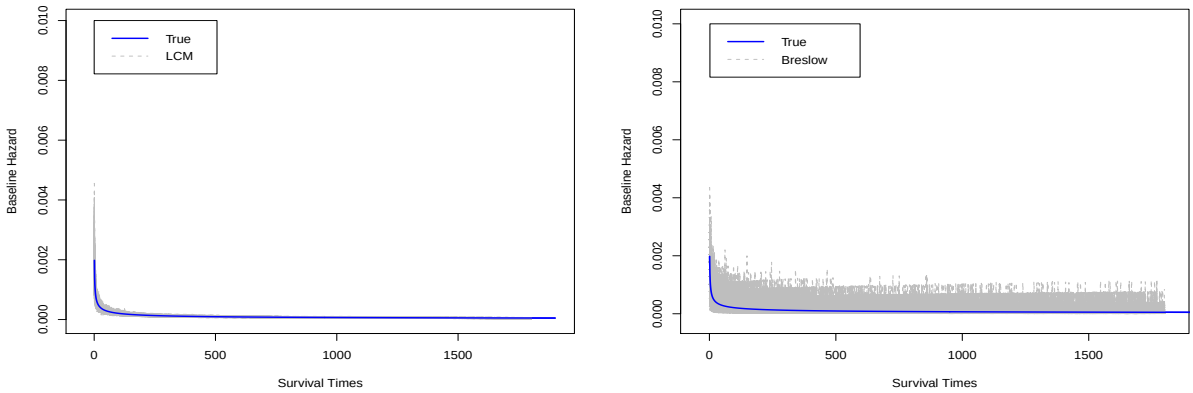

Figure 4.2: The LCM (Left) and Breslow (Right) Performance on ICTUS-Based Weibull Dataset

Figure 4.1 provides the cumulative distribution function for the survival and censoring distribution based on the ICTUS dataset. We choose the size of the dataset to be 2000. In 1000 repetitions, we obtain the estimate of the baseline hazard from the LCM and the Breslow estimator as depicted in Figure 4.2. We see that the overall performance of LCM in estimating the baseline hazard is better than the Breslow estimators. Figure 4.3 also provides the distribution of the baseline hazard over certain points. In each time point we see that the variance of LCM is smaller than the Breslow estimator. The global measure of distance (MISE) also confirms that LCM performs better. Table 4.1 provides numerical measure between these two estimators. Here the term "hat" refers to the Breslow estimator while the term "tilde" refers to the LCM based estimator.

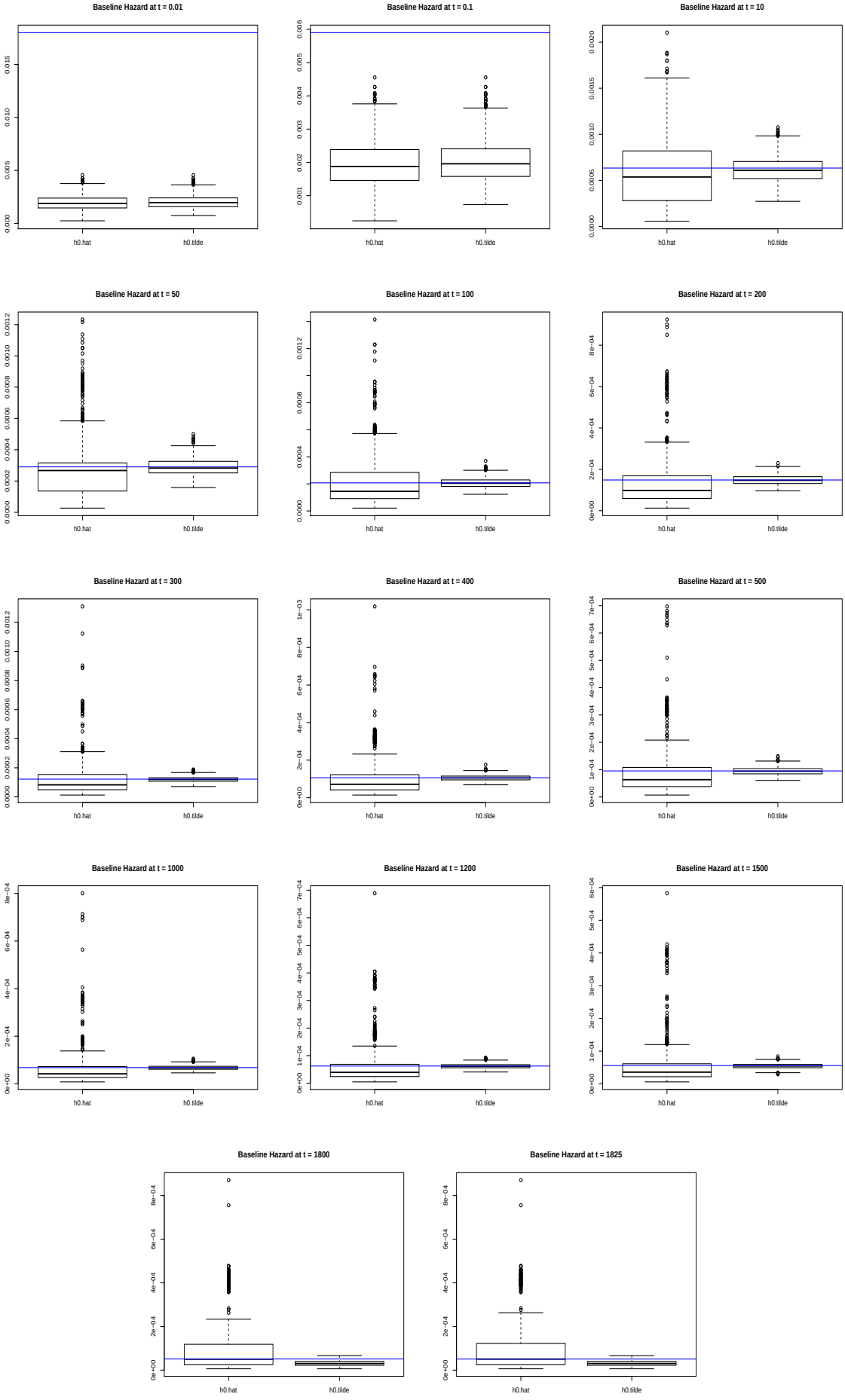

Figure 4.3: Boxplots of the Hazard Values in Each Observed Times of ICTUS-Based Weibull

Points	Bias.hat	Var.hat	MSE.hat	Bias.tilde	Var.tilde	MSE.tilde
0.01	-1.61E-02	5.20E-07	2.58E-04	-1.60E-02	4.07E-07	2.55E-04
0.1	-3.96E-03	5.20E-07	1.62E-05	-3.86E-03	4.07E-07	1.53E-05
10	-3.36E-05	1.28E-07	1.29E-07	-1.27E-05	1.93E-08	1.94E-08
50	-1.39E-07	4.40E-08	4.40E-08	-3.12E-07	3.01E-09	3.00E-09
100	2.47E-06	3.38E-08	3.37E-08	-6.16E-07	1.31E-09	1.31E-09
200	6.60E-07	1.91E-08	1.91E-08	3.72E-07	5.61E-10	5.60E-10
300	5.37E-06	1.70E-08	1.70E-08	-1.11E-06	3.39E-10	3.40E-10
400	-2.60E-06	1.01E-08	1.01E-08	-3.79E-07	2.26E-10	2.26E-10
500	-2.46E-06	8.86E-09	8.85E-09	-5.14E-07	1.83E-10	1.83E-10
1000	-1.79E-06	6.36E-09	6.36E-09	-5.42E-07	8.47E-11	8.50E-11
1200	-1.15E-06	4.76E-09	4.76E-09	-2.89E-07	6.78E-11	6.79E-11
1500	4.71E-07	4.74E-09	4.73E-09	-1.43E-06	6.30E-11	6.50E-11
1800	5.00E-05	1.50E-08	1.74E-08	-1.97E-05	1.48E-10	5.38E-10
1825	5.21E-05	1.55E-08	1.82E-08	-1.94E-05	1.48E-10	5.25E-10

MISE.hat	MISE.tilde
2.34E-05	3.16E-06

Table 4.1: The Measures of LCM Performance on ICTUS-Based Weibull

4.3 Effect on the Baseline Hazard

The application of LCM to the cumulative baseline hazard gives an estimate of a decreasing baseline hazard. As we have seen for the Weibull-Cox baseline hazard based on ICTUS dataset, the LCM performs better in estimating the baseline hazard compared to the Breslow's estimator. In this section, we will investigate the LCM performance in other decreasing shapes of the Weibull baseline hazard. For this purpose we consider 3 shapes for the Weibull distribution which we are going to use in the experiment, namely strong concave, less concave and linear. In this investigation, we choose the uniform distribution $U(0, \theta_c)$ as a censoring distribution. By using the formula provided in [14], we compute the parameter θ_c for the censoring distribution such that the dataset has a censoring proportion p_c. For our investigation, we set the proportion censoring similar with the ICTUS case, i.e., 20%. We also truncate the follow-up time in certain position. In this investigation, it is set at 210. This means all follow-up times which are larger than 210 will equal to 210 and the corresponding event indicator will be 0. This mechanism is designed to represents the situation in real application where a study is terminated at some time. The parameters pairs are listed in Table 4.2 and their corresponding cumulative distribution function curves are depicted in Figure 4.4.

	Strong Concave	Less Concave	Linear
Survival distribution : Weibull(shape, scale)	(0.28,60000)	(0.76,1200)	(1,700)
Censoring distribution: Uniform$(0, \theta_c)$	224	237	234

Table 4.2: The Parameters for the Survival and Censoring Distribution

Using these pairs of survival and censoring distribution, we have a range of follow-up times from 0 to 210. Moreover, we also round-up the generated survival times and the censoring times, such that the tied follow-up times present in the dataset. We will investigate the performance of both estimators in 15 time points on this interval for each case. As before, the size of the data is 2000 and we do 1000 iterations. The visual comparison between the performance of the Breslow estimator and the LCM in estimating the

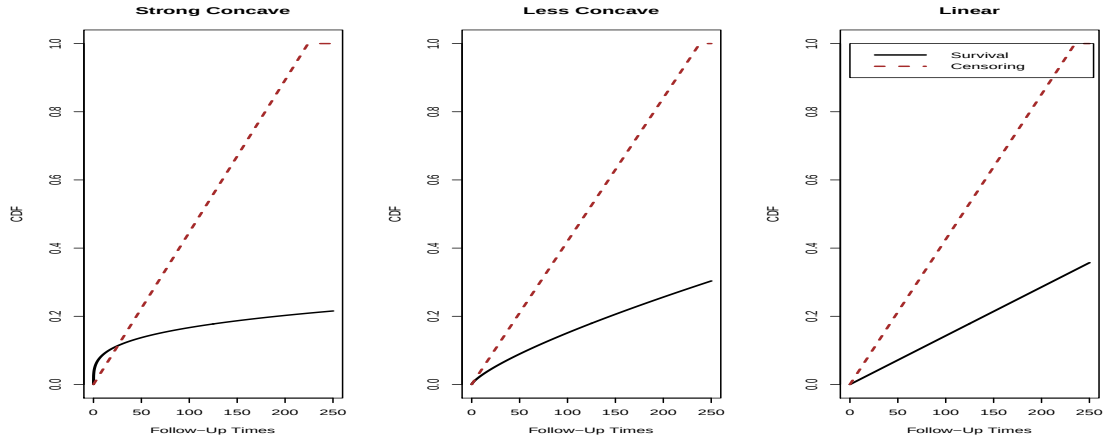

Figure 4.4: The Pair of Weibull CDF in Survival and Censoring Distribution

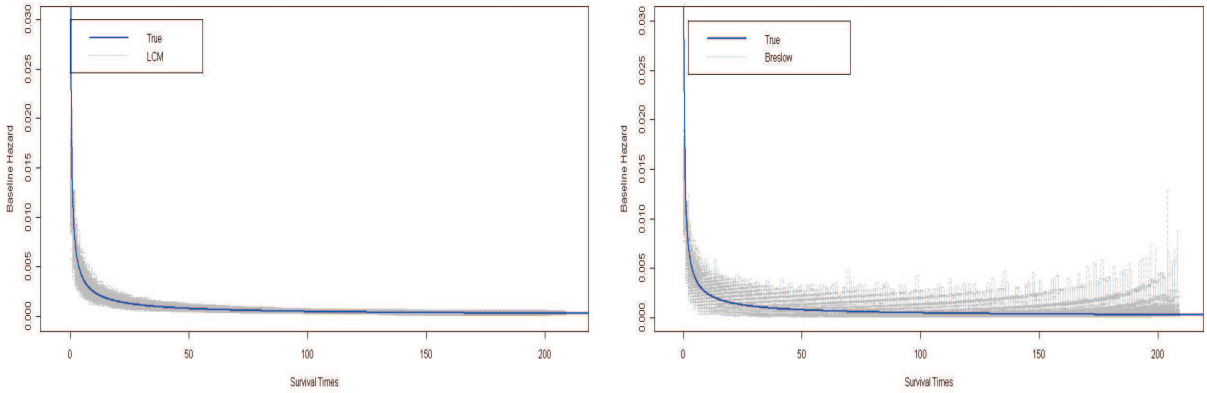

Figure 4.5: The LCM (Left) and the Breslow (Right) Estimate of the Baseline Hazard in the Strong Concave Cumulative Hazard

baseline hazard for the three cases can be seen in Appendix A, Figure A.3 to Figure A.5. The numerical quantities to measure the pointwise differences are listed in Table 4.3 to Table 4.5.

In strong concave case, we see that the bias between the Breslow estimator and the LCM are slightly different. The MISE for this case are also in the same order. These suggest that the Breslow estimator has relatively similar shape to the LCM. However, on the basis of their variance, the LCM is still preferable since it gives smaller values for each time points. In other cases, as we might have expected, the LCM performs better than the Breslow estimator. This is shown by the bias and variance for the LCM's baseline hazard in each observation point which is smaller than the bias and variance for the Breslow estimator. The MISE suggests that the distance between LCM and the true hazard is closer than the Breslow estimator.

If we look on the boxplots representation for visual comparison between the Breslow and the LCM (Appendix A), we might think that the performance of the LCM near zero is as bad as the Breslow. We also notice that it has a large distance to the true value. This will lead into a conclusion that using the LCM is only good in latter time points. This is not true. This large distance is caused by rounding up the follow-up time to the nearest integer. A rigorous study in which we round up the follow-up times to the nearest third digit decimals shows that near zero, the LCM has a closer distance to the true hazard.

Points	Bias.hat	Var.hat	MSE.hat	Bias.tilde	Var.tilde	MSE.tilde
0.01	-3.41E-01	4.88E-06	1.16E-01	-3.41E-01	4.86E-06	1.16E-01
0.1	-5.39E-02	4.88E-06	2.92E-03	-5.39E-02	4.86E-06	2.91E-03
1	-5.00E-03	2.83E-06	2.78E-05	-4.92E-03	2.50E-06	2.67E-05
5	-5.34E-04	1.21E-06	1.50E-06	-4.43E-04	4.53E-07	6.48E-07
10	-1.78E-04	7.63E-07	7.94E-07	-1.40E-04	1.88E-07	2.08E-07
20	-7.09E-05	5.23E-07	5.28E-07	-3.65E-05	6.49E-08	6.62E-08
50	1.11E-05	2.87E-07	2.87E-07	7.02E-06	2.01E-08	2.02E-08
75	-1.28E-06	1.89E-07	1.89E-07	1.50E-05	1.15E-08	1.17E-08
100	-5.07E-06	1.45E-07	1.45E-07	1.58E-05	7.58E-09	7.82E-09
125	1.02E-06	1.37E-07	1.37E-07	2.08E-05	6.68E-09	7.11E-09
150	1.52E-05	1.40E-07	1.40E-07	1.98E-05	8.03E-09	8.42E-09
175	5.44E-05	1.49E-07	1.52E-07	6.23E-06	1.05E-08	1.05E-08
190	2.00E-04	3.17E-07	3.57E-07	6.72E-06	9.99E-09	1.00E-08
200	3.80E-04	5.04E-07	6.48E-07	1.67E-05	9.92E-09	1.02E-08
210	5.62E-04	9.93E-07	1.31E-06	2.65E-05	9.92E-09	1.06E-08

MISE.hat	MISE.tilde
4.02E-04	3.55E-04

Table 4.3: The Bias, Variance, MSE and MISE for both Estimators Applied to The Strong Concave Cumulative Baseline Hazard

Points	Bias.hat	Var.hat	MSE.hat	Bias.tilde	Var.tilde	MSE.tilde
0.01	-7.03E-03	1.03E-06	5.05E-05	-6.81E-03	7.21E-07	4.71E-05
0.1	-2.58E-03	1.03E-06	7.70E-06	-2.36E-03	7.21E-07	6.29E-06
1	-5.05E-04	8.72E-07	1.13E-06	-3.91E-04	3.56E-07	5.09E-07
5	-8.07E-05	6.13E-07	6.19E-07	-3.81E-05	1.37E-07	1.38E-07
10	-1.79E-05	5.49E-07	5.49E-07	3.71E-06	7.55E-08	7.55E-08
20	-1.56E-05	5.26E-07	5.26E-07	5.46E-06	4.64E-08	4.64E-08
50	-2.39E-05	5.05E-07	5.05E-07	2.20E-05	2.52E-08	2.57E-08
75	-1.17E-05	5.12E-07	5.11E-07	2.16E-05	1.87E-08	1.92E-08
100	-6.15E-05	5.24E-07	5.28E-07	3.03E-05	1.75E-08	1.84E-08
125	2.70E-05	6.55E-07	6.55E-07	3.37E-05	1.77E-08	1.89E-08
150	6.05E-05	7.69E-07	7.71E-07	2.51E-05	2.15E-08	2.22E-08
175	-1.40E-06	6.56E-07	6.55E-07	-1.62E-05	3.25E-08	3.28E-08
190	5.83E-05	9.32E-07	9.35E-07	-7.63E-05	4.78E-08	5.36E-08
200	1.59E-04	1.27E-06	1.30E-06	-1.62E-04	6.29E-08	8.89E-08
210	9.77E-04	2.89E-06	3.84E-06	-1.76E-04	5.97E-08	9.05E-08

MISE.hat	MISE.tilde
1.36E-04	8.70E-06

Table 4.4: The Bias, Variance, MSE, and MISE for Both Estimator Applied to The Less Concave Cumulative Baseline Hazard

Points	Bias.hat	Var.hat	MSE.hat	Bias.tilde	Var.tilde	MSE.tilde
0.01	-2.30E-05	3.97E-07	3.97E-07	3.83E-04	1.33E-07	2.79E-07
0.1	-2.30E-05	3.97E-07	3.97E-07	3.83E-04	1.33E-07	2.79E-07
1	-2.00E-05	3.80E-07	3.80E-07	2.60E-04	6.29E-08	1.30E-07
5	-8.52E-07	3.98E-07	3.97E-07	1.53E-04	2.82E-08	5.18E-08
10	-1.18E-05	3.87E-07	3.87E-07	1.15E-04	2.09E-08	3.41E-08
20	1.53E-06	4.70E-07	4.69E-07	9.01E-05	1.73E-08	2.54E-08
50	-1.66E-05	5.91E-07	5.91E-07	6.34E-05	1.55E-08	1.95E-08
75	1.93E-05	6.86E-07	6.86E-07	5.02E-05	1.63E-08	1.88E-08
100	-8.30E-07	6.87E-07	6.86E-07	3.69E-05	1.68E-08	1.82E-08
125	-1.50E-05	8.86E-07	8.85E-07	1.98E-05	1.81E-08	1.84E-08
150	-3.54E-05	1.03E-06	1.03E-06	-3.43E-06	2.19E-08	2.19E-08
175	1.87E-05	1.31E-06	1.31E-06	-5.89E-05	3.37E-08	3.72E-08
190	6.32E-05	1.56E-06	1.56E-06	-1.44E-04	6.45E-08	8.50E-08
200	1.04E-04	1.95E-06	1.96E-06	-2.77E-04	1.11E-07	1.87E-07
210	1.16E-03	3.93E-06	5.27E-06	-3.26E-04	1.11E-07	2.17E-07

MISE.hat	MISE.tilde
1.86E-04	7.69E-06

Table 4.5: The Bias, Variance, MSE and MISE for Both Estimator Applied to the Linear Cumulative Hazard

4.4 The Change in ICTUS's Regression Coefficient

As it is already explained in Section 2.4.1, Cox introduced the partial likelihood to estimate the covariate coefficient because maximizing the full likelihood function (2.14) over all possible $\boldsymbol{\beta}$ and baseline distributions is not feasible. The partial likelihood that he proposed will have a resulting estimate with similar distributional properties as the full maximum likelihood estimator. In fitting the Cox model to a right-censored dataset, we first estimate the covariate coefficient by maximizing the partial likelihood which is then followed by estimating the baseline function. We see that changing the estimate of the baseline distribution will not alter the values of the estimated covariate coefficients. However, if we substitute the baseline distribution, which were estimated by the LCM, into the full likelihood function, we can maximize it over all possible $\boldsymbol{\beta}$. As a result we obtain a new estimated $\boldsymbol{\beta}$ with regard to the LCM, say $\widetilde{\boldsymbol{\beta}}$.

The idea to see a change in the estimate of the regression coefficients can be explained as follows. We first fit the ICTUS dataset with the Cox model. Using the partial likelihood, we estimate the coefficients of the six covariates which are considered to be significant. Having estimated these coefficients, the baseline cumulative hazard (say, $\widehat{H}_0$) can be estimated. The LCM estimator is then applied to the estimated baseline cumulative hazard, which results in the new estimator of baseline cumulative hazard (say, $\widetilde{H}_0$). Based on the function $\widetilde{H}_0$, we can use the formulas explained in Section 2.2 to derive the formulas for the baseline survival $\widetilde{S}_0$ and the baseline hazard $\widetilde{h}_0$. By substituting these estimates of $\widetilde{h}_0$ and $\widetilde{S}_0$ into equation (2.14), we have an expression which only depends on $\boldsymbol{\beta}$. Thus maximizing this log-likelihood over all possible parameter values will give us an estimate of $\widetilde{\boldsymbol{\beta}}$.

Suppose that this likelihood maximization process is repeated, we hope that in the end, we will have an estimate of the baseline hazard and the regression coefficients that maximizes the total likelihood. It would be interesting to see whether this would produce a better estimator. In the above procedure, we only estimate the $\widetilde{\boldsymbol{\beta}}$ after only one iteration. It would be interesting to numerically iterate this process to see whether the maximum likelihood value converges. First, we will investigate the outcome of this process

when we apply it into the ICTUS dataset. The following figure gives the illustration of the procedure to investigate the change of the covariates values.

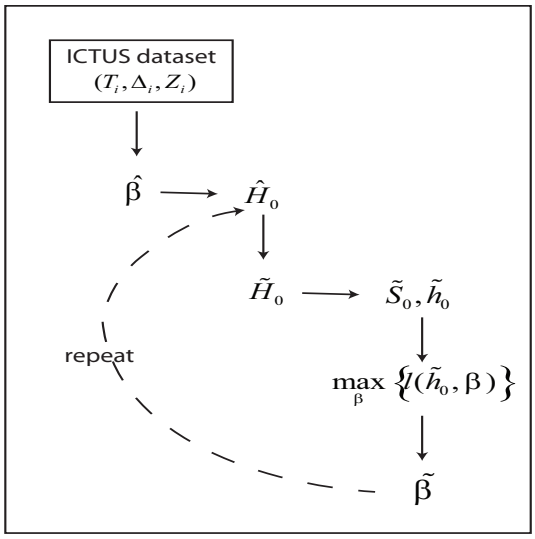

Figure 4.6: The Illustration of the schemes

The above procedure was implemented with the stopping criterion on subsequent likelihood values being not larger than a certain small number. In this experiment, it is set at 10^{-15}. Note that even with a stopping criterion smaller than 10^{-15}, the estimate of the covariate coefficient still converges. Using R software, the following table provides estimated values of six significant covariates after each iteration.

lth	Treat	Age	BMIhi	Hypl	Dbet	Mi.H
1	5.8815E−02	4.2531E−02	−3.4452E−01	3.6432E−01	7.1546E−01	5.2380E−01
2	5.8813E−02	4.2293E−02	−3.4452E−01	3.6432E−01	7.1546E−01	5.2380E−01
3	5.8811E−02	4.2060E−02	−3.4452E−01	3.6432E−01	7.1546E−01	5.2379E−01
4	5.8809E−02	4.1832E−02	−3.4452E−01	3.6432E−01	7.1546E−01	5.2379E−01
5	5.8807E−02	4.1608E−02	−3.4453E−01	3.6432E−01	7.1546E−01	5.2379E−01
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1357	2.8584E−02	3.0767E−02	−3.9433E−01	3.2094E−01	7.3859E−01	5.3816E−01

Table 4.6: The Change in Covariates Coefficients Value due to LCM

The last row in Table 4.6 indicates the estimate of the covariate values after the last adjustment. Figure 4.7(b) shows how the values of the log-likelihood behaves and (c) an enlargement of panel (b). In panel (c), the loglikelihood curves shows a zig-zag pattern. In the first iteration, we compute the log-likelihood which consists of the partial likelihood estimate of β with the estimate of LCM's baseline hazard. The value equals -1900.838. In the next iteration, we maximize the log-likelihood over all possible β by fixing the LCM's baseline hazard. In this process, surely we will have a log-likelihood value which is bigger than in previous step. It is recorded that in second iteration, the log-likelihood value equals -1900.131. However, in next iteration, the loglikelihood goes down. This pattern repeats until it reaches convergence. After the 511th iteration, the log-likelihood values remains constant at -1902.114. Figure 4.7(a) provides the change in estimated Treatment coefficient. We see that the estimated coefficient for Treatment in the end of the iteration decreases around 50% from the estimated coefficient using the traditional estimator. The second highest change can be found in covariate Age whilst the covariate Dbet and Mi.H differs a little from their initial values.

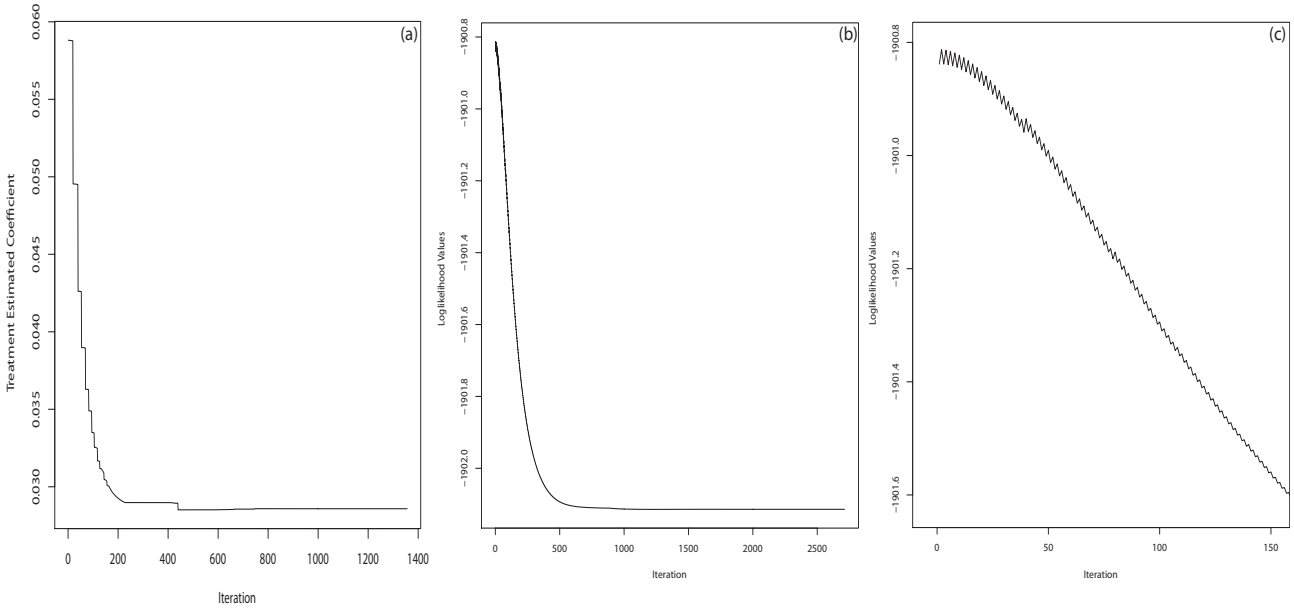

Figure 4.7: (a)The change in estimated covariates coefficients during 1357 iteration (b) The Log-Likelihood Values (c) The Log-Likelihood Values (zoom-in)

The Breslow estimator for the cumulative baseline hazard in the ICTUS data involves 6 significant covariates mentioned earlier in Chapter 2. Since we are mainly interested in Treatment covariate, we would like to see the change in the cumulative baseline hazard for Treatment due to the LCM adjustment. In this way, we should treat the remaining five covariates such that it represents the center of the data. One way to do such a thing is to use a quantity known as risk score. In a model containing p covariates, the risk score for j-th subject in the dataset is

$$\widehat{r}_j(\boldsymbol{z}_j, \widehat{\boldsymbol{\beta}}) = \sum_{k=1}^{p} \widehat{\beta}_k z_{jk}.$$

Typically, the quantiles or other measure such as the average of the risk score can be obtained by a common routine in most statistics software. Thus, the corresponding cumulative hazard will be

$$\widehat{H}(t_j, \widehat{r}_q, \widehat{\boldsymbol{\beta}}) = \left[\widehat{H}_0(t_j)\right] \mathrm{e}^{\widehat{r}_q} \quad , j = 1, \ldots, n$$

where q is an empirical quantile of the risk score. Further details on how to calculate the estimated value of the risk score can be found in [6].

Figure 4.8 shows the change in the cumulative baseline hazard for Treatment while setting the other covariates in median risk score. The lowest curves is the baseline cumulative hazard based on Cox model (black curve) and its LCM (red curve). The highest curves are the curves obtained after the last iteration. This picture provides an information that adjustment of the LCM shifts the Breslow estimate of the cumulative baseline hazard upward.

In Figure 4.9, an illustration regarding the estimate of the covariate coefficient is given. The curves in both of these figures are the cumulative hazard of patient with different treatment in median risk score. As can be seen for this picture, the distance between Treatment 0 and 1 is getting smaller in last iteration than before the iteration. While the baseline hazard estimate is shifted upward the regression coefficient estimate is shifted downward. This facts lead us to a conclusion that this procedure goes in the wrong direction in maximizing the log-likelihood.

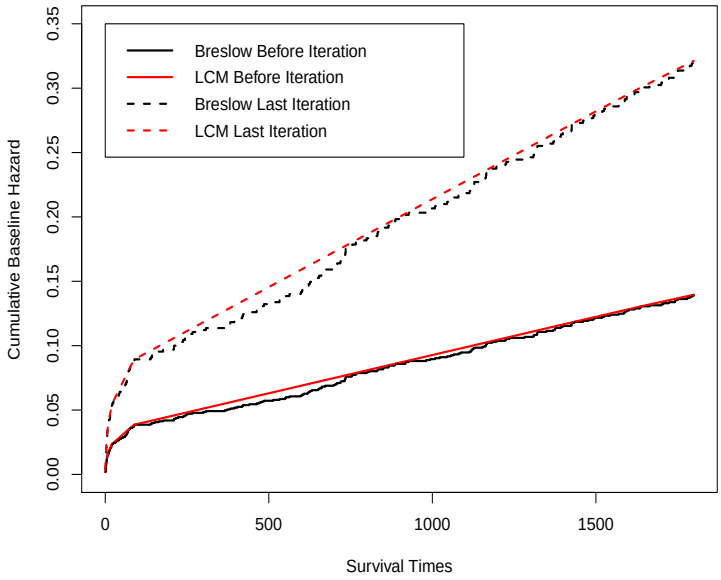

Figure 4.8: Simulation of the Baseline Hazard Change

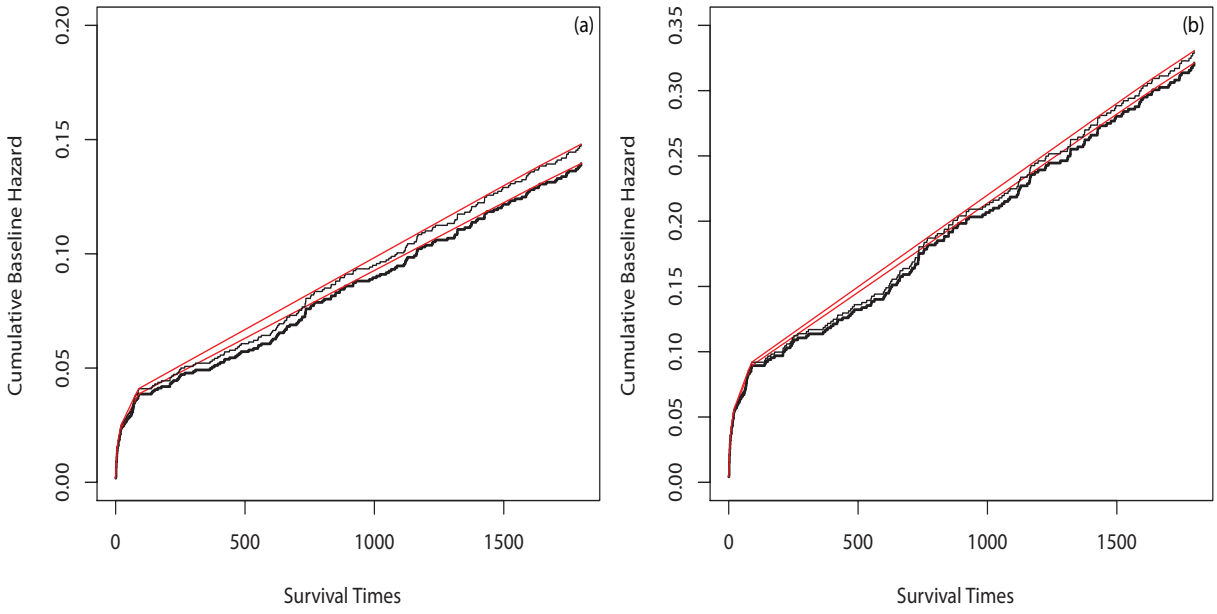

Figure 4.9: The Cumulative Hazard with the Effect of Treatment at the Median Risk Score (a) Before Iteration (b) In the last Iteration

4.5 Analysis on the Regression Coefficients Change in ICTUS Dataset

Details regarding the iterating schemes and the resulting change of the estimate of the covariate coefficient due to substitution of the LCM's baseline hazard are interesting to be analyzed. The change in the estimated covariates coefficients for some covariates are high while for some others are quite small. In addition, the cumulative baseline hazard is shifted upward in each iteration. Even the log-likelihood values decreases. By doing this, we are not even approximating the maximum of the log-likelihood.

Estimating the covariate coefficient corresponding to the LCM ($\widetilde{\beta}$) is done by maximizing the full log-likelihood function (2.14). For simplicity, we use one covariate (z) which can take any of two values, either 0 or 1. Suppose that we already have the estimate for the baseline distribution $\widehat{h}_0$. We could also plug the information about the baseline distribution into the following log-likelihood for n individuals.

$$L(\widehat{h}_0, \beta) = \sum_{i=1}^{n} \left\{ \delta_i \ln \widehat{h}_0(t_i) + \delta_i z_i' \boldsymbol{\beta} - \mathrm{e}^{z_i'\beta} \widehat{H}_0(t_i) \right\}. \tag{4.4}$$

If this function is maximized over all possible β then one way to do that is by differentiating the log-likelihood function with respect to β and set it equal to 0.

$$\frac{\partial L}{\partial \beta}(\widehat{h}_0, \beta) = \sum_{i=1}^{n} z_i \left(\delta_i - \widehat{H}_0(t_i) \mathrm{e}^{z_i'\beta} \right) = 0$$

Whenever $z_i = 0$ the summand in the above expression equals 0. The above expression can then be simplified as follows:

$$\sum_{i \in \{z_i = 1\}} \left(\delta_i - \widehat{H}_0(t_i) \mathrm{e}^{\beta} \right) = 0.$$

Thus, the estimate of the covariate coefficient ($\widehat{\beta}$) that maximizes the log-likelihood can be expressed in the following formula,

$$\mathrm{e}^{\widehat{\beta}} = \frac{\sum_{i \in \{z_i = 1\}} \delta_i}{\sum_{i \in \{z_i = 1\}} \widehat{H}_0(t_i)}. \tag{4.5}$$

The numerator in the equation (4.5) is fixed whenever the LCM is applied to the estimate of the cumulative baseline hazard $\widehat{H}_0$. Thus, we always have the following:

$$\widetilde{H}_0 \geq \widehat{H}_0 \Rightarrow \widetilde{\beta} \leq \widehat{\beta} \tag{4.6}$$

The abovementioned facts regarding the iterating schemes contradicts our hope at the beginning. We know that the Breslow estimates the true distribution. Since LCM gives a continuous estimate of the cumulative baseline hazard, we expect to see a better approximation of the true quantities by using this estimator than the traditional estimator does. However, iterating the LCM shifts the cumulative baseline hazard upward, which as a consequence makes the distance to the true distribution larger. In addition, the estimate of the covariate coefficient is getting smaller in each iteration. On the basis of these facts, we suggest to avoid the iterating procedure to improve the estimation of the Cox's quantities. This suggestion might raise the question whether the LCM is a better estimator if it is implemented without any iteration. In order to address this question, we performed an experiment in Section 4.6.

4.6 Simulation on the Regression Coefficient Distribution

In this section, we will investigate the performance of the LCM-based estimator in estimating the regression coefficient without the iterations. In the experiment, we need to build a dataset which follows a certain Cox model. The Weibull distribution will be used to model the hazard function since we can vary the shape of the hazard function. We deal with three types of the Weibull cumulative hazard's shapes as mentioned in the beginning, namely the strong concave, less concave and linear. These types were chosen since the corresponding cumulative hazard shape would be properly fitted with the LCM. After having the hazard, we take one covariate which can be either two values, 0 or 1. Having this so-called Weibull-Cox dataset, we will compare the estimates obtained by the traditional estimator and by the LCM. In more detail, we are interested in obtaining an information about the distribution of those estimates.

The resulting distribution of $\widehat{\beta}$ and $\widetilde{\beta}$ for different shape of cumulative hazards are illustrated by the measures listed in Table 4.7. In this table, we used the dataset with 10% censoring.

Type		$\widehat{\beta}$	$\widetilde{\beta}$
Strong Concave	mean	0.9959	0.9892
	sd	0.0425	0.0423
Less Concave	mean	0.9984	0.9818
	sd	0.0558	0.0559
Linear	mean	0.9951	0.8746
	sd	0.0479	0.0790

Table 4.7: The Resulting Variation of $\widehat{\beta}$ and $\widetilde{\beta}$ in the Dataset with 10% Censoring

Type		$\widehat{\beta}$	$\widetilde{\beta}$
Strong Concave	mean	0.9988	0.9856
	sd	0.0994	0.0997
Less Concave	mean	1.0193	0.9934
	sd	0.1168	0.1171
Linear	mean	1.0111	0.9623
	sd	0.0862	0.0865
ICTUS-Based	mean	1.0123	1.0055
	sd	0.0879	0.0880

Table 4.8: The Resulting Variation of $\widehat{\beta}$ and $\widetilde{\beta}$ in The Dataset Consists of 80% Censored Survival Times

As can be seen from Table 4.7 and 4.8, the resulting estimate of the covariate coefficient obtained by the Cox model has only a slight difference than the estimator using LCM, especially in the case of a concave cumulative hazard (either the strong concave or the less concave). The censoring percentage in the dataset does not notably influence the estimate resulting from LCM. This conclusion is supported by the distribution comparison provided in Table 4.8 which is generated from the Weibull-Cox dataset with around 60% censoring. Either with zero or 60% censoring, the estimates of the regression coefficient using LCM or the partial likelihood have similar behaviour. An extraordinary result occurs in the linear case. It appears that in the linear case, the LCM is not a good estimator for the regression coefficient. If we take a closer look at one iteration of the linear case, we see the trend illustrated in Figure 4.10.

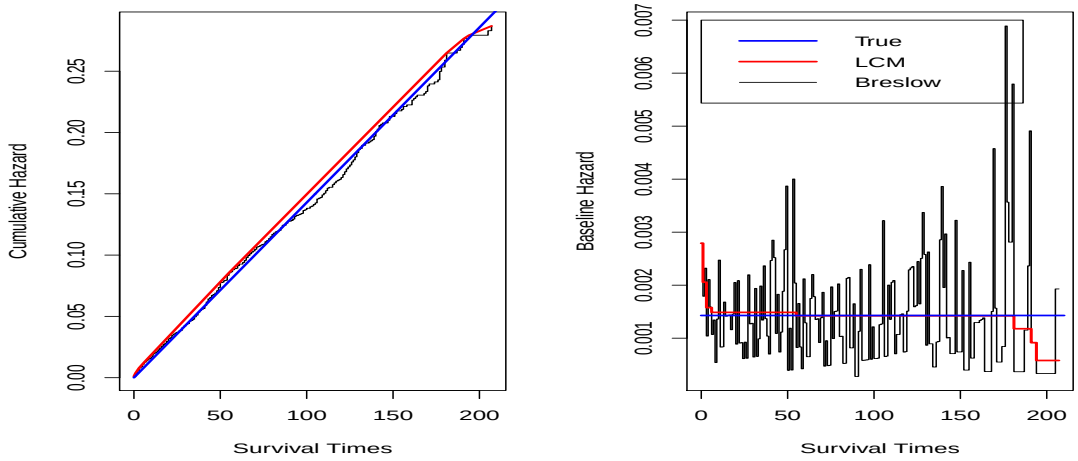

Figure 4.10: Example of Cumulative Hazard and Hazard in One Step Simulation from Linear Cumulative Hazard

In Figure 4.10, we see that the LCM estimate for the baseline hazard does not approximate the true hazard at the boundaries of follow-up times. Its estimate values are higher than the true hazard near 0 and lower than the true hazard near time 200. Whilst it has a constant hazard in the interior of the follow-up time domain. This is the consequence for estimating the constant hazard with some decreasing estimate. In the case of concave cumulative hazard (either in strong or less concave case), the LCM has a shape that is already close to the true hazard. Therefore, the distribution of the $\widetilde{\beta}$ will not change much comparing with the distribution of $\widehat{\beta}$.

4.6.1 Inserting the True Hazard Information

On the basis of the comparison between true hazard shape and the LCM's estimate, especially in linear case, we intend to investigate another thing. Suppose that we know completely the true underlying hazard. Instead of plugging the estimate of the baseline hazard into the full likelihood, we insert the true hazard. If we do the likelihood maximizing procedure over all possible β we want to investigate whether it will results in a distribution which has a smaller standard deviation than the traditional estimator. For this purpose, we use the three shapes parameters of Weibull-Cox dataset. The results can be seen in Table 4.9 and visualized in Figure 4.11.

Type		$\widehat{\beta}$	$\widetilde{\beta}$
Strong Concave	mean	1.0047	0.9977
	sd	0.1000	0.0512
Less concave	mean	1.0029	0.9985
	sd	0.0985	0.0542
Linear	mean	1.0038	0.9973
	sd	0.1008	0.0530
ICTUS-Based	mean	1.0113	0.9978
	sd	0.1813	0.0971

Table 4.9: The Resulting Variation of $\widehat{\beta}$ and $\widetilde{\beta}$ in The Dataset Consists of 80% Censored Survival Times, with The True Hazard

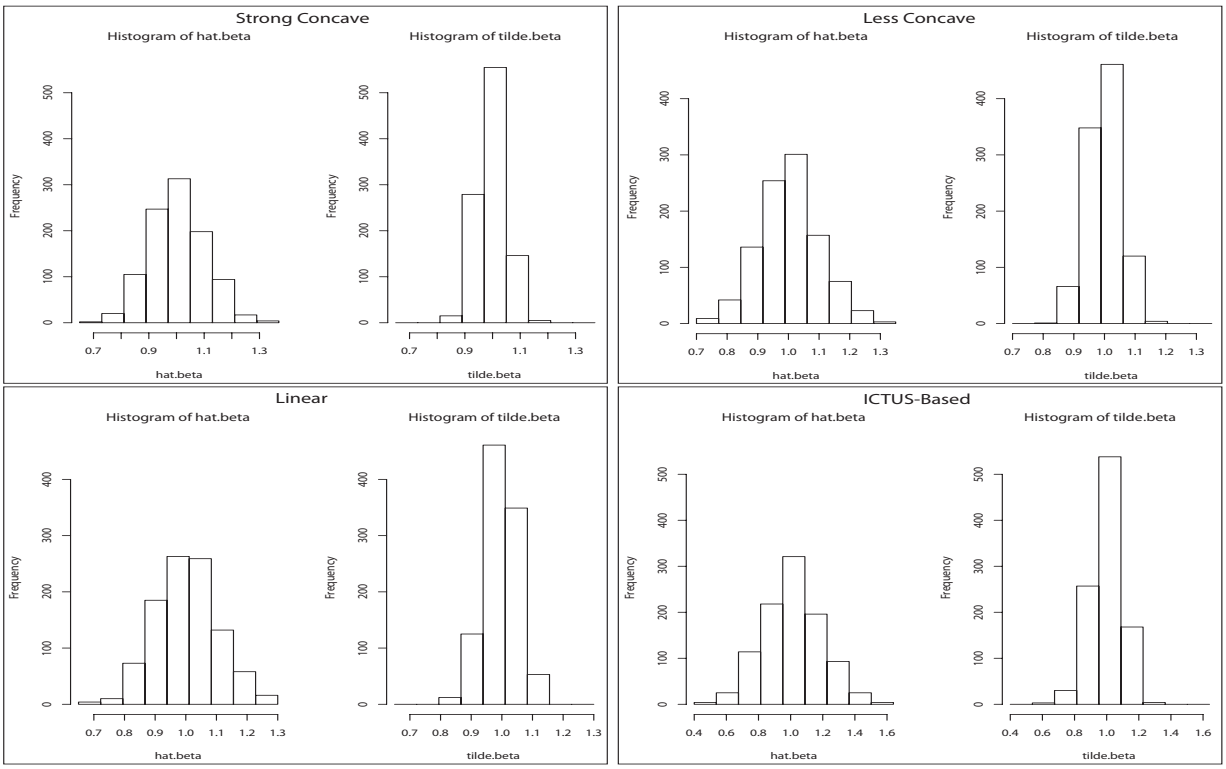

Figure 4.11: Distribution Comparison for the Coefficient Regression between Cox's Partial Likelihood and LCM with True Hazard

These results confirm our hope that whenever we know the true underlying hazard, estimating the regression coefficient using LCM will be a better approach compared to the Breslow estimator, even in the linear case. Based on this fact, we believe that the bad results in the distribution of the regression coefficient in linear case is caused by the boundary problems. We hope that if we can find some correction method for this boundary problem, we will have a better estimate for the regression coefficient. We are going to use this information on treating the estimate of the regression coefficient without any information regarding the true baseline hazard.

4.6.2 Correction on the Baseline Hazard's Estimate

From the previous explanation regarding the shape of the LCM baseline hazard in the linear case and the knowledge of inserting true hazard, we intuitively think that LCM performance in estimating the coefficient regression can be improved by correcting the values on the boundaries. A study in [15] asserts the rate of convergence of the monotone failure rate is $n^{1/3}\left[\widetilde{h}_0(t_0) - h_0(t_0)\right]$ for a fixed t_0. Thus, we might use this information as a way to correct the LCM. We set the LCM's baseline hazard on the interval $[0, n^{-1/3}]$ equals to the LCM's baseline hazard at point $n^{-1/3}$. As can be seen in Table 4.10, the correction on the interval $[0, n^{-1/3}]$ does not improve the LCM in the less concave and linear case. In the strong concave case, the performance of the LCM is getting worse. This is reasonable as a rigorous check shows that in the strong concave case, the LCM slope's segment decreases rapidly on that interval. The correction forces the nature of high baseline hazard into a far lower baseline hazard. This explains why we have a high estimated coefficient regression in this way.

Another intuitive way to correct the LCM baseline hazard is by correcting it on the interval $[0, t^*]$, where t^* is an average position where LCM has a jump for the first time. However, since LCM jumps is

Type		$\widehat{\beta}$	$\widetilde{\beta}_1$	$\widetilde{\beta}_2$
Strong Concave	mean	1.0105	0.9939	1.0858
	sd	0.1040	0.1033	0.1090
Less Concave	mean	1.0123	0.9710	0.9816
	sd	0.0973	0.1116	0.1124
Linear	mean	0.9940	0.9224	0.9225
	sd	0.0976	0.1016	0.1016
ICTUS-Based	mean	0.9945	0.9712	0.9763
	sd	0.0901	0.1092	0.1093

Table 4.10: The Distribution Comparison of the Estimated Coeffient Regression with Hazard Correction on the Interval $[0, n^{-1/3}]$

also influenced by the shape of the baseline hazard, we cannot fixed this t^*. In the strong concave case, the first jumps in LCM are close to 0 and it is getting further from 0 in the case of less concave and linear cumulative hazard. Thus we cannot fix a t^* that best represents the first jumps for all shapes of the concave cumulative hazard.

Now we focus only on the linear case. Varying the correction point t^* does not improve the estimation. By trial and error, we could not find a t^* which can improve the standard deviation. By this, we conclude that the traditional estimator for the covariate coefficient performs better than the LCM.

4.7 The Covariate Effect

In the above experiments, the true regression coefficient is fixed (equals 1). In the proportional hazard model, this value can be interpreted as the hazard ratio for subjects with different type of covariate (see equation (2.12)). For example, we use treatment type as the covariate which can take either of two values i.e., 0 and 1. Thus, the regression coefficient equals 1 means that the risk of subject with treatment 1 is more than twice the risk of subjects with treatment 0. In this situation, the treatment effect is considered to be important. Further in this subsection, the term important and unimportant refer to the covariate effect. Suppose that we fix the unimportant regression coefficient (for example 0.05) and do the similar experiment as before to investigate the resulting distribution of the $\widetilde{\beta}$.

Type		$\widehat{\beta}$	$\widetilde{\beta}$
Strong Concave	mean	0.0163	-0.0059
	sd	0.1246	0.1244
Less Concave	mean	0.0014	-0.0540
	sd	0.1194	0.1385
Linear	mean	0.0022	-0.0813
	sd	0.1113	0.1275
ICTUS-Based	mean	0.0079	-0.0097
	sd	0.1017	0.0974

Table 4.11: The Resulting Variation of $\widehat{\beta}$ and $\widetilde{\beta}$ in the Case of Insignificant Covariate

As can be seen in Table 4.11, we obtain a similar change in standard deviation for all three types of cumulative hazards as the case where the true regression coefficient is important. Notice that as in Table

4.8, the standard deviation in the strong concave and the less concave case are slightly different. In the linear case, we still see this large difference. However, the mean of $\widetilde{\beta}$ shifts considerably even in the concave cumulative hazard, see the means for $\widetilde{\beta}$ in the strong concave and the less concave case. Nevertheless, this situation is reasonable. Consider again the equation (4.5). The estimate of the covariate coefficient depends on the number of subjects in the dataset who show an event at time i and the cumulative baseline hazard at time t_i. Both of these quantities are conditioned on the covariate equal to 1. In the case of an unimportant true regression coefficient, the risk of having an event for covariate value 1 equals the risk to have an event for subjects with covariate value 0. In such a circumstance, we have the number of subjects who have an event approximately equal with the subjects who do not show an event. In the case of important coefficient, the risk for subject with covariate value 1 is bigger than those with covariate value 0. Consequently, the number of subjects who experience an event will be larger than the those who do not show an event. This lead to a conclusion that the numerator of equation (4.5) in the unimportant case is smaller than the numerator in the important case. Thus, a smaller increase in the cumulative baseline hazard (due to applying the LCM) will result in a bigger change of the covariate coefficient. This is explaining the biggest change in covariate Treatment after applying LCM to the cumulative baseline hazard in ICTUS dataset.

4.8 Summary

The purpose of the works done in this section is to investigate the advantages and disadvantages of using the LCM to estimate the parameters in the Cox model. In the beginning, we hope that by incorporating the knowledge about the shape of the baseline hazard, we will have a better estimator for these parameters. However, our simulation and analysis proves that the LCM-based estimator only performs better in estimating the baseline hazard. In estimating the regression coefficient in the strictly decreasing hazard, this method works similarly as the partial likelihood estimator. It turns out that this method performs bad in the case of linear hazard. It also gives a bias estimate of the regression coefficient in the case of unimportant covariate. In later chapter, another shape-constrained estimator is introduced and we will investigate the performance of this estimator in estimating the parameters in the Cox model.

Chapter 5

The Maximum Likelihood Estimator for the Monotone Baseline Hazard

In Chapter 4, we investigated the effect of monotonicing the cumulative baseline hazard of the estimate of the baseline hazard and the regression coefficients in Cox model. The purpose was to investigate whether incorporating the information about the shape of the baseline hazard would result in a better estimate for the parameters in the Cox model. In estimating the baseline hazard, this method is proven to give a better estimate than Breslow's. However, inserting the information about the shape constrained baseline hazard into the full likelihood gives a worse estimate of the regression coefficient.

A worse estimate in regression coefficient is affected by the way the LCM constructs the baseline hazard. The estimate for the regression coefficient is obtained by maximizing the full likelihood by first inserting the baseline hazard which was estimated by LCM. The maximization problem depends on the value of the cumulative hazard. Since LCM gives a cumulative baseline hazard which is always larger than the Breslow's cumulative hazard, it turns out that the regression coefficient is getting smaller (see (4.5)). The corresponding likelihood is even smaller than the likelihood for the Breslow's estimator. Thus, in this way, we did not reach the maximum value of the likelihood.

In this chapter, we will use a shape constrained estimator for the baseline hazard which is introduced in [16]. This estimator is the solution for the maximum likelihood problem and has been proven to be a consistent estimator. Before we compute this estimator, Section 5.1 provides an introduction of the underlying principle of this MLE.

5.1 Regression Under Order Restriction

Given a finite observations $y_1, y_2, \ldots, y_n$ and a function p which depends on the value of the observation. For a given points $(y_1, p(y_1)), (y_2, p(y_2)), \ldots, (y_n, p(y_n))$, the most common way to capture a curve that best fits those points is via regression which is in sense of least squares. In specific set, suppose Y is a finite set $\{y_1, y_2, \ldots, y_n\}$ with the simple order $y_1 \leq y_2 \leq \ldots \leq y_n$. A function f on Y is defined to be isotonic with respect to the ordering if it satisfies $f(y_1) \leq f(y_2) \leq \ldots \leq f(y_k)$. It is called antitonic with respect to the ordering if it satisfies $f(y_1) \geq f(y_2) \geq \ldots \geq f(y_n)$. If g is a given function on Y, then a function g^* on Y is the isotonic regression of g with weights w if and only if g^* is isotonic and g^* minimizes

$$\sum_{y \in Y} [g(y) - f(y)]^2 w(y) \tag{5.1}$$

in the class of all isotonic function f on Y. Thus, it is quite obvious to expect the isotonic regression in form of the weighted average

$$\frac{\sum_{y\in Y} w(y)g(y)}{\sum_{y\in Y} w(y)}$$

in order to satisfy the minimization problem (5.1). The similar result also hold in the case of antitonic regression. The exact formula for the iso(anti)tonic regression cannot be derived. However, there is a graphical interpretation for the solution which will be explained later on. Since in this manuscript we will deal mostly with antitonic regression, we will only concentrate on the term antitonic regression.

To visualize the solution of antitonic regression, consider a simple ordered set Y mentioned above. The plot of the points

$$P_j = (W_j, G_j), \quad P_0 = (0,0), \quad j = 1, 2, \ldots, n$$

where

$$G_j = \sum_{i=1}^{j} w(y_j)g(y_j) \quad \text{and} \quad W_j = \sum_{i=1}^{j} w(y_j), j = 1, 2, \ldots, n \tag{5.2}$$

in the Cartesian plane constitutes a cumulative sum diagram (CSD) of the given function g with weights w. The slope of the points P_{j-1} and $P_j, j = 1, 2, \ldots, n$ is simply $g(x)$. The antitonic regression of g is given by the slope of the LCM of the CSD.

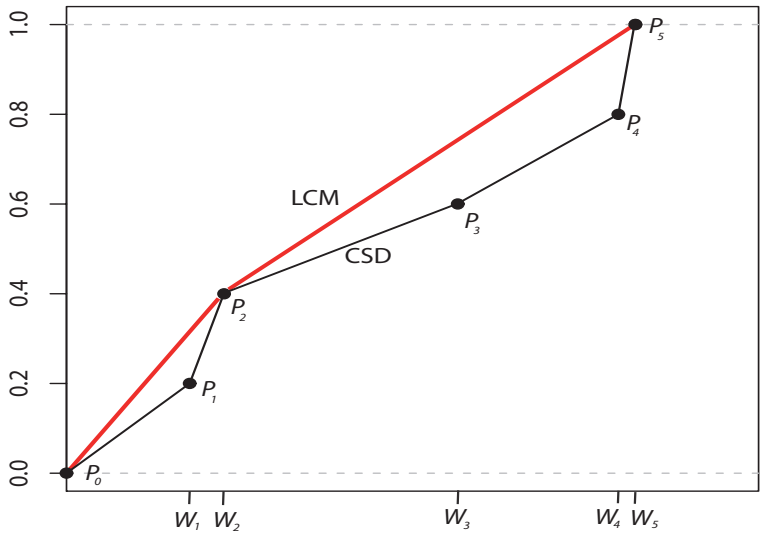

Figure 5.1: The CSD and LCM

5.2 The MLE for the Baseline Hazard Under Order Restriction

Now, we want to estimate the baseline hazard function through the MLE. In the end, it turns out that the likelihood maximization problem is an antitonic regression problem as explained in Section 5.1. The idea is to mimic the derivation of the isotonic estimator which is introduced in [16] with the antitonic estimator. Recall that the total likelihood function of the Cox model is given by equation (2.13). By inserting the relation (2.5), we obtain the following expression for the Cox's likelihood.

$$\ell(h_0, \boldsymbol{\beta}) = \prod_{i=1}^{n} [h(t_i, \boldsymbol{z}_i, \boldsymbol{\beta})]^{\delta_i} [\exp(-H(t_i, \boldsymbol{z}_i, \boldsymbol{\beta})] \tag{5.3}$$

We want to rewrite the likelihood (5.3) in the ordered survival times. For that purpose we use the convention that all censored observations are censored at the preceding uncensored failure times. This convention has been explained in Chapter 2. By the convention, if we have n ordered follow-up time, whenever we have censored follow-up times between subsequent survival times, they will be shifted back to the preceding survival time. As a consequence, we have a constant hazard between the subsequent survival times.

Suppose the ordered distinct survival times are denoted as $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$ and there are d_i subjects fail at $t_{(i)}$ and r_i subjects which are censored. Firstly, we focus on the first term in (5.3). The censored subjects (with $\delta_i = 0$) turns the first term into 1. The baseline hazard at time $t_{(i)}$ is $h_0(t_{(i)})$, thus for d_i subjects who fail at this particular time, $h_0(t_{(i)})$ is their baseline hazard. The uncensored subjects ($\delta_i = 1$)then contributes to the following expression

$$h_0^{d_i}(t_{(i)}) \exp(\mathbf{s}_i' \boldsymbol{\beta})$$

with $\mathbf{s}_i$ is the sum of covariates for the subjects which fail at time $t_{(i)}$.

Now,we focus on the second term in (5.3). The cumulative hazard for the Cox model is formulated as $H_0(t_i) \exp(\boldsymbol{z}_i' \boldsymbol{\beta})$. Since the cumulative baseline hazard at time $t_{(i)}$ is just an integration of the baseline hazard from 0 to time $t_{(i)}$, all subjects which have the follow-up time at time $t_{(i)}$ will have the same cumulative baseline hazard, which is $H_0(t_{(i)})$. Thus, the second form can be reformulated as follows.

$$\exp\left\{-H_0(t_{(i)}) \sum_{l \in H(t_{(i)})} \exp(\boldsymbol{z}_l' \boldsymbol{\beta})\right\} = \exp\left\{-\int_0^{t_{(i)}} h_0(u)du \sum_{l \in H(t_{(i)})} \exp(\boldsymbol{z}_l' \boldsymbol{\beta})\right\},$$

with $H(t_{(i)})$ is the label for all subjects who have follow-up time at $t_{(i)}$, either censored or uncensored. Then, the formula for the Cox's likelihood in terms of the ordered distinct survival times is as follows.

$$\ell(h_0, \boldsymbol{\beta}) = \prod_{i=1}^{k} h_0^{d_i}(t_{(i)}) \exp(\mathbf{s}_i' \boldsymbol{\beta}) \exp\left\{-\int_0^{t_{(i)}} h_0(u)du \sum_{l \in H(t_{(i)})} \exp(\boldsymbol{z}_l' \boldsymbol{\beta})\right\} \tag{5.4}$$

By fixing $\boldsymbol{\beta}$ as known parameters, the MLE for $h_0(t_{(i)})$ cannot be obtain directly by maximizing this likelihood. This is due to the baseline hazard values which can be chosen infinitely large over a small interval. One way to avoid this problem is by assuming that the baseline hazard values are constant for each interval. We choose a subclass of decreasing hazard which is bounded above (say, by M) and show that there is a hazard function that could maximize (5.4).

Since we already assume that the baseline hazard which maximize the likelihood (5.4) are in the subclass of decreasing hazard, the hazard should be in step function.

$$h_1 \geq h_2 \geq \ldots \geq h_k > 0, \qquad h_i = h_0(t_{(i)}).$$

Thus, the likelihood in (5.4) can be reformulated as

$$\ell(h_0, \boldsymbol{\beta}) = \prod_{i=1}^{k} h_i^{d_i} \exp(\mathbf{s}_i' \boldsymbol{\beta}) \exp\left\{-\sum_{j=1}^{i} h_j (t_{(j+1)} - t_{(j)}) \sum_{l \in H(t_{(i)})} \exp(\boldsymbol{z}_l' \boldsymbol{\beta})\right\}$$

By rearranging the the following term

$$-\sum_{j=1}^{i} h_j (t_{(j+1)} - t_{(j)}) \sum_{l \in H(t_{(i)})} \exp(\boldsymbol{z}_l' \boldsymbol{\beta}) = \sum_{j=1}^{k-1} h_j \left(t_{(j+1)} - t_{(j)}\right) \sum_{l \in R(t_{(i+1)})} \exp(\boldsymbol{z}_l' \boldsymbol{\beta})$$

we obtain the following likelihood.

$$\ell(h_0, \boldsymbol{\beta}) = \prod_{i=1}^{k} h_0^{d_i}(t_{(i)}) \exp(\mathbf{s}_i'\boldsymbol{\beta}) \exp\left\{ -\sum_{j=1}^{k-1} h_j \left(t_{(j)} - t_{(j-1)}\right) \sum_{l \in R(t_{(i+1)})} \exp(\boldsymbol{z}_l'\boldsymbol{\beta}) \right\}$$

The corresponding log-likelihood becomes

$$L(h_0, \boldsymbol{\beta}) = \sum_{i=1}^{k-1} d_i \log h_i + \mathbf{s}_i'\boldsymbol{\beta} - h_i(t_{(i+1)} - t_{(i)}) \sum_{l \in R(t_{i+1})} \exp(\boldsymbol{z}_l'\boldsymbol{\beta})$$

Since the maximization does not depend on the second term, the problem is reduced into maximization the following form

$$L = \sum_{i=1}^{k-1} (g(i) \log h_i - h_i) w(i), \tag{5.5}$$

subject to

$$h_1 \geq h_2 \geq \ldots \geq h_k > 0$$

where the function $g(i)$ and $w(i)$ in equation (5.5) are as follows

$$g(i) = \frac{d_i}{\left(t_{(i+1)} - t_{(i)}\right) \sum_{l \in R(t_{(i+1)})} \exp(\boldsymbol{z}_l'\boldsymbol{\beta})} \text{ and } w(i) = \left(t_{(i+1)} - t_{(i)}\right) \sum_{l \in R(t_{(i+1)})} \exp(\boldsymbol{z}_l'\boldsymbol{\beta}) \tag{5.6}$$

The maximization problem (5.5) has similar form as maximization of the likelihood over monotone hazard introduced in [9] Section 7.4. Theorem 1.5.7 in [9] guarantees that the solution of this maximum likelihood exist. One way to obtain the solution is by forming the cumulative sum diagram of $P_j = (W_j, G_j), j = 1, 2, \ldots, n$ as defined in Section 5.1 with the corresponding $w(j)$ and $g(j)$ are defined in (5.6). The left hand slope of the LCM of this CSD is the solution of (5.5). This solution is the MLE of the decreasing baseline hazard.

5.3 The MLE of Baseline Hazard Effect on the Parameters of the Cox Model

As we did in Chapter 4, we want to find the regression coefficients and the baseline hazard which maximize the total likelihood function in the Cox model. By incorporating this MLE of the baseline hazard, we want to investigate the distribution of the resulting baseline hazard and the regression coefficients in Cox model. In separate sections we will discuss the differences between this estimator and the traditional one.

By this MLE of the baseline hazard, we can assert that for every particular regression coefficient we can always find a baseline hazard which maximize the total likelihood of Cox model with the restriction that the hazard is decreasing. In order to obtain these estimates, we use the scheme which is illustrated in Figure 5.2. By choosing a certain parametric distribution for the survival and censoring variable, we then generate a Cox dataset with a fixed regression coefficient (as before, we set this regression coefficient $\beta = 1$). For simplicity, we use only one covariate which can take either of two values (0 and 1). Applying partial likelihood estimation for the regression coefficient, we obtain a value of this estimator, say $\widehat{h_0}$. We fix this value and maximize the likelihood over all possible β. From this step, we obtain a baseline hazard

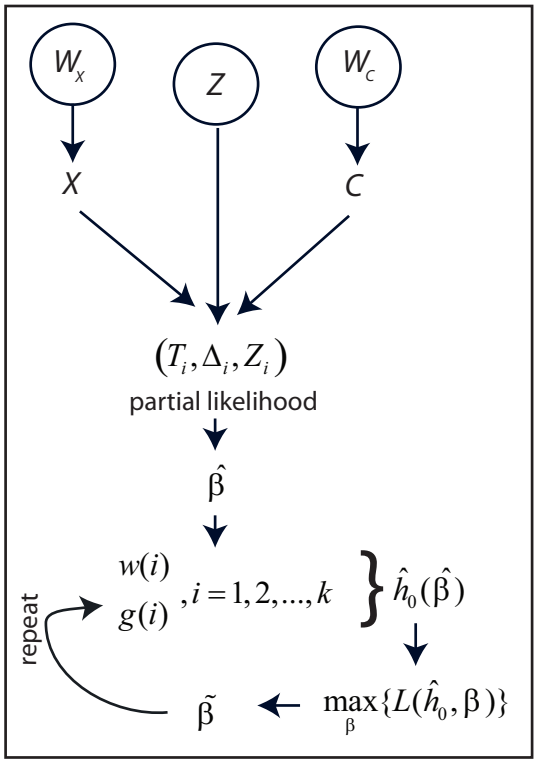

Figure 5.2: The Schemes for MLE of the Baseline Hazard

estimate. By fixing this baseline hazard estimate in the log-likelihood and maximize it over all possible β, we obtain the maximizing β. We repeat this procedure until the likelihood converges. We did this procedure for 1000 times using the three shapes of Weibull distribution as in Chapter 4. The procedure is also applied to the Weibull distribution which is based on the ICTUS dataset. In each iteration, the sample size equals 2000 we use the stopping criteria for the iteration to be the subsequent likelihood is smaller than 10^{-20}.

5.4 Effects on the Baseline Hazard

In this section, we do simulations to investigate the MLE consistency in estimating the baseline hazard. We use again several types of the Weibull distribution with a decreasing hazard as explained in Chapter 4. As we did in Chapter 4, the quality of this estimator was assessed using the bias, variance and mean squared error at several points. As a global measure, the mean integrated square error is used.

The distribution comparison between the baseline hazards in strong concave which are obtained by using the MLE and the Breslow estimator are illustrated in Figure 5.3. We can see that the MLE has a closer distance to the true baseline hazard than the Breslow estimator. We could also compare their performance numerically with the measures as listed in Table 5.1. Near zero, the Breslow estimator and the MLE show relatively similar performance. As it has been explained in Chapter 4, this is also the consequence of rounding up the follow-up time to the nearest integer. As time increases, we see that the MLE performance is much better than the Breslow estimator. The MISE also confirms that overall, the distance between the MLE estimator is closer to the true hazard compared to the Breslow estimator.

The illustration of the comparisons for other types of Weibull distribution can be seen in Figure B.1 and Figure B.2 in Appendix B. The numerical comparisons are listed from Table 5.2 to Table 5.4. In all

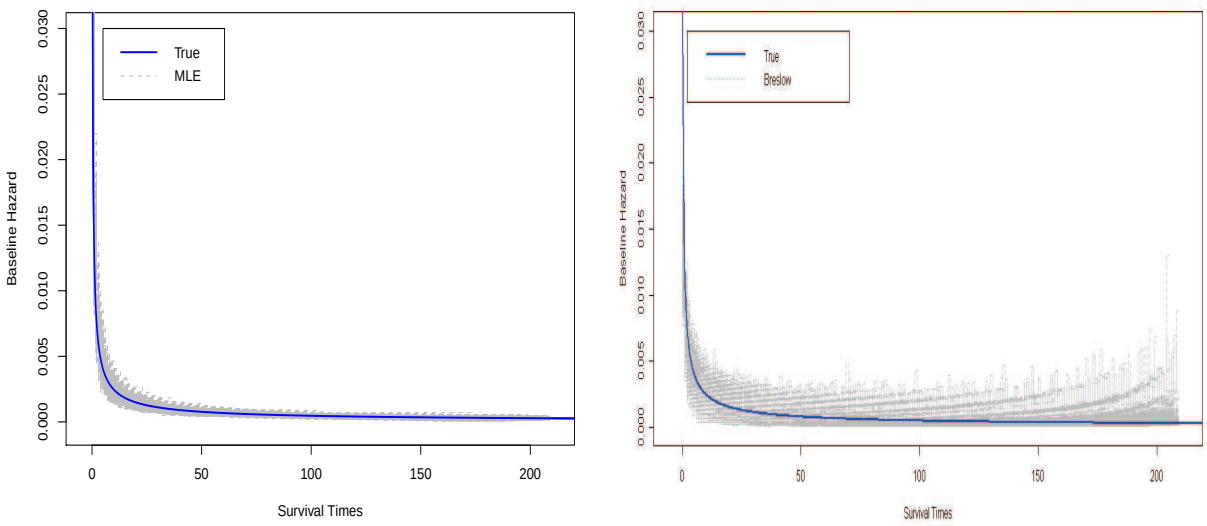

Figure 5.3: The MLE (Top Left) and the Larger Version (Top Right) and Breslow (Bottom) Estimator for the Baseline Hazard in Strong Concave Case

Points	Bias.hat	Var.hat	MSE.hat	Bias.tilde	Var.tilde	MSE.tilde
0.01	-3.41E-01	4.81E-06	1.16E-01	-3.15E-01	2.08E-05	9.91E-02
0.1	-5.43E-02	4.81E-06	2.95E-03	-2.80E-02	2.08E-05	8.07E-04
1	-5.01E-03	2.67E-06	2.77E-05	8.40E-04	5.37E-06	6.07E-06
5	-6.33E-04	1.04E-06	1.44E-06	1.09E-04	6.81E-07	6.92E-07
10	-2.03E-04	7.33E-07	7.73E-07	4.56E-05	2.14E-07	2.15E-07
20	-5.69E-05	4.76E-07	4.79E-07	1.41E-05	7.24E-08	7.25E-08
50	2.15E-05	2.75E-07	2.76E-07	3.72E-06	2.22E-08	2.22E-08
75	-3.88E-06	1.87E-07	1.87E-07	1.34E-06	1.11E-08	1.11E-08
100	3.02E-06	1.54E-07	1.54E-07	2.25E-06	8.16E-09	8.16E-09
125	1.72E-05	1.42E-07	1.42E-07	-8.16E-07	7.46E-09	7.46E-09
150	2.23E-05	1.62E-07	1.62E-07	-3.92E-06	7.47E-09	7.48E-09
175	3.68E-05	1.24E-07	1.25E-07	-2.02E-05	8.33E-09	8.73E-09
190	1.83E-04	2.30E-07	2.63E-07	-1.50E-05	7.75E-09	7.96E-09
200	4.62E-04	6.60E-07	8.72E-07	-4.46E-06	7.71E-09	7.72E-09
210	6.28E-04	1.15E-06	1.55E-06	5.32E-06	7.71E-09	7.73E-09

MISE.hat	MISE.tilde
4.11E-04	3.50E-04

Table 5.1: The Numerical Comparisons Between the Breslow Estimator (hat) and the MLE (tilde) in Strong Concave Case

cases of Weibull distributions used in this investigation, the MLE is preferable. It has smaller variance than the Breslow. In global measures, we also see that the MLE gives a smaller distance to the true hazard. Note that the MISE for the MLE of the strong concave case and the ICTUS-based Weibull are relatively in similar order as the Breslow estimator. This suggests that in both cases, the Breslow estimate has already a similar shape to the MLE. Even in this case, MLE is still preferable due to its smaller variance.

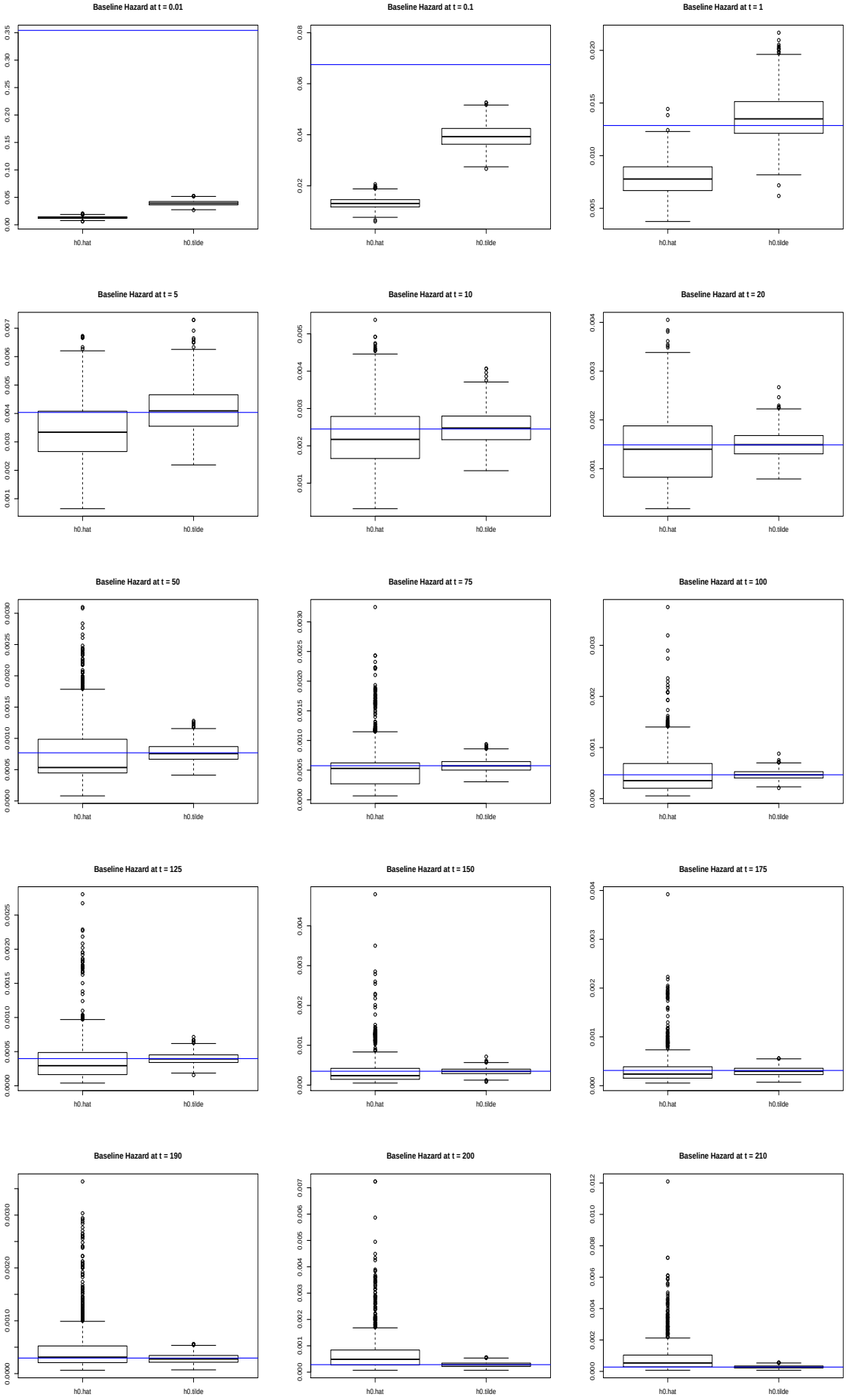

Figure 5.4: Pointwise Distribution of the Baseline Hazard in Strong Concave Case

Points	Bias.hat	Var.hat	MSE.hat	Bias.tilde	Var.tilde	MSE.tilde
0.01	-6.97E-03	9.90E-07	4.96E-05	-7.07E-03	4.21E-07	5.04E-05
0.1	-2.52E-03	9.90E-07	7.35E-06	-2.62E-03	4.21E-07	7.28E-06
1	-5.37E-04	8.18E-07	1.11E-06	-1.98E-04	3.58E-07	3.97E-07
5	-1.05E-04	6.85E-07	6.96E-07	5.74E-05	1.53E-07	1.56E-07
10	-6.30E-05	6.22E-07	6.25E-07	3.73E-05	9.81E-08	9.94E-08
20	-4.12E-06	5.65E-07	5.64E-07	1.88E-05	5.26E-08	5.29E-08
50	-6.67E-06	5.41E-07	5.41E-07	7.76E-06	2.45E-08	2.45E-08
75	1.11E-05	6.20E-07	6.19E-07	7.55E-06	1.93E-08	1.94E-08
100	-4.84E-05	5.46E-07	5.48E-07	-2.07E-06	1.68E-08	1.68E-08
125	5.40E-06	5.68E-07	5.67E-07	-7.23E-06	1.56E-08	1.57E-08
150	2.65E-05	7.24E-07	7.24E-07	-2.37E-05	1.78E-08	1.83E-08
175	1.47E-05	7.98E-07	7.98E-07	-6.34E-05	2.60E-08	3.00E-08
190	1.12E-05	7.96E-07	7.95E-07	-1.35E-04	4.39E-08	6.21E-08
200	1.29E-04	1.07E-06	1.08E-06	-2.05E-04	5.81E-08	1.00E-07
210	9.00E-04	2.17E-06	2.97E-06	-2.09E-04	5.56E-08	9.91E-08

MISE.hat	MISE.tilde
1.33E-04	9.72E-06

Table 5.2: The Numerical Comparisons Between the Breslow Estimator (hat) and the MLE (tilde) in Less Concave Case

Points	Bias.hat	Var.hat	MSE.hat	Bias.tilde	Var.tilde	MSE.tilde
0.01	6.33E-05	4.32E-07	4.35E-07	2.01E-04	5.86E-08	9.89E-08
0.1	6.33E-05	4.32E-07	4.35E-07	2.01E-04	5.86E-08	9.89E-08
1	1.82E-05	3.67E-07	3.67E-07	1.95E-04	5.68E-08	9.48E-08
5	1.79E-05	4.17E-07	4.17E-07	1.21E-04	2.95E-08	4.41E-08
10	-2.09E-05	4.18E-07	4.18E-07	8.47E-05	2.23E-08	2.95E-08
20	-1.09E-06	4.93E-07	4.93E-07	5.57E-05	1.92E-08	2.22E-08
50	-2.00E-05	5.62E-07	5.62E-07	1.55E-05	1.53E-08	1.55E-08
75	1.42E-05	6.95E-07	6.95E-07	-5.52E-06	1.44E-08	1.44E-08
100	3.14E-05	8.09E-07	8.10E-07	-2.31E-05	1.47E-08	1.52E-08
125	2.32E-05	8.52E-07	8.52E-07	-4.59E-05	1.53E-08	1.74E-08
150	2.72E-05	1.09E-06	1.09E-06	-7.97E-05	1.83E-08	2.47E-08
175	9.17E-06	1.47E-06	1.47E-06	-1.44E-04	3.35E-08	5.43E-08
190	6.28E-07	1.70E-06	1.69E-06	-2.14E-04	5.14E-08	9.70E-08
200	9.74E-05	1.62E-06	1.63E-06	-3.19E-04	8.56E-08	1.87E-07
210	1.16E-03	4.29E-06	5.62E-06	-3.63E-04	8.72E-08	2.19E-07

MISE.hat	MISE.tilde
1.85E-04	7.56E-06

Table 5.3: The Numerical Comparisons Between the Breslow Estimator (hat) and the MLE (tilde) in Linear Case

Points	Bias.hat	Var.hat	MSE.hat	Bias.tilde	Var.tilde	MSE.tilde
0.01	-1.60E-02	5.81E-07	2.56E-04	-1.53E-02	6.15E-07	2.34E-04
0.1	-3.88E-03	5.81E-07	1.56E-05	-3.18E-03	6.15E-07	1.07E-05
10	-3.32E-05	1.43E-07	1.44E-07	9.07E-06	2.53E-08	2.54E-08
50	-8.63E-07	4.62E-08	4.62E-08	1.53E-06	3.02E-09	3.02E-09
100	-1.30E-06	2.91E-08	2.91E-08	-9.64E-07	1.32E-09	1.31E-09
200	2.67E-06	1.76E-08	1.76E-08	-8.45E-08	5.58E-10	5.57E-10
300	-2.82E-06	1.22E-08	1.22E-08	3.30E-07	3.66E-10	3.66E-10
400	-5.93E-07	1.18E-08	1.18E-08	-1.21E-07	2.32E-10	2.31E-10
500	-5.23E-06	9.00E-09	9.02E-09	1.71E-07	1.77E-10	1.77E-10
1000	-1.26E-06	6.23E-09	6.23E-09	-6.60E-07	8.18E-11	8.22E-11
1200	2.96E-06	7.17E-09	7.17E-09	-3.68E-07	6.76E-11	6.77E-11
1500	-2.29E-06	3.73E-09	3.73E-09	-1.15E-06	6.95E-11	7.08E-11
1800	5.66E-05	1.64E-08	1.96E-08	-1.85E-05	1.56E-10	4.99E-10
1825	5.69E-05	1.64E-08	1.96E-08	-1.82E-05	1.56E-10	4.87E-10

MISE.hat	MISE.tilde
2.36E-05	3.67E-06

Table 5.4: The Numerical Comparisons Between the Breslow Estimator (hat) and the MLE (tilde) in ICTUS-Based Weibull

5.5 Effect on the Regression Coefficient

In Section 4.6.1, we investigated the estimate the regression coefficient by inserting the true hazard. We have seen that in this case, the estimate for the regression coefficient has a smaller variance than the partial likelihood estimator. Now, that we already derived a consistent estimator for the baseline hazard, we would like to assess its performance on estimating the regression coefficient via log-likelihood maximization. For this purpose we use the scheme illustrated in Figure 5.2.

As before, we want to investigate the performance of the MLE in estimating the regression coefficient in several Weibull distribution with decreasing hazard. We first generate the so-called Weibull-Cox dataset. Afterwards, we maximize the partial likelihood to obtain the estimate for the regression coefficient $\widehat{\beta}$. We use this value to maximize the log-likelihood over all possible decreasing hazard. We then do the iteration of maximum likelihood which is illustrated in Figure 5.2. The resulting distribution for the regression coefficient based on 1000 iterations can be seen in Table 5.5 and Figure 5.5.

For all types of Weibull distribution with decreasing hazard, we see that the regression coefficients

Type		$\widehat{\beta}$	$\widetilde{\beta}$
ICTUS-based	mean	1.0005	0.9542
	sd	0.0945	0.0960
Strong Concave	mean	1.0015	0.9922
	sd	0.0603	0.0600
Less Concave	mean	1.0051	0.9896
	sd	0.0620	0.0611
Linear	mean	0.9987	0.9696
	sd	0.0599	0.0576

Table 5.5: The Distribution for The Estimated Coefficient Regression using the MLE for the Baseline Hazard

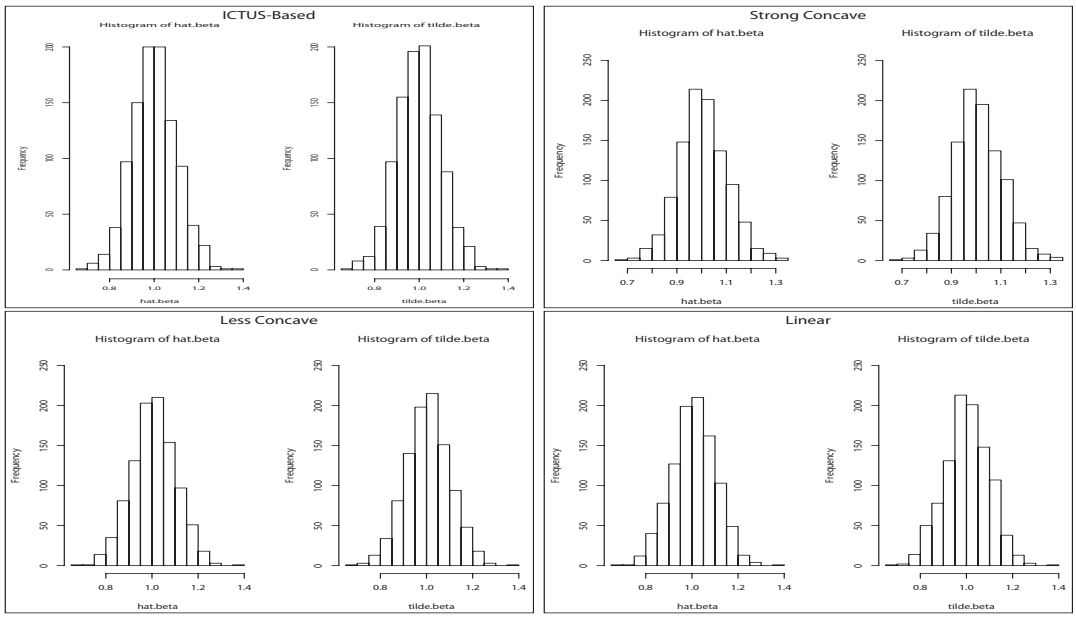

Figure 5.5: The Estimated Coefficient Regression Distribution Using the MLE for the Baseline Hazard

which are obtained by maximizing the log-likelihood function with the baseline hazard estimated by the MLE, have a distribution which is similar to partial likelihood estimator. From this point, we see that incorporating the shape of the baseline hazard in the sense that this baseline hazard is the MLE yields a better estimate than using the LCM. But still, in terms of their standard deviation, the MLE for the baseline hazard gives an estimate with a standard deviation which differs slightly from the partial likelihood estimator.

5.6 Application on the ICTUS Dataset

In the previous sections, we have investigated the performance of the MLE for the baseline hazard in estimating the baseline hazard and the regression coefficient in the Cox model. In conclusion, we have a better estimator for the parameters in Cox model in the sense that those estimators simultaneously maximizing the full likelihood. In this section, we want to estimate the parameters in the Cox model for the ICTUS data. As in Chapter 4, we use the Cox model which involves 5 significant covariates and Treatment. Applying the scheme in Figure 5.2 above gives the baseline hazard and all regression coefficient which maximizes the log-likelihood. We set the stopping criterion as before. It turns out that in 242 iterations, this procedure converges and the regression coefficient values for several iterations are listed in Table 5.6.

The first row in Table 5.6 shows all regression coefficient which were obtained from the partial likelihood estimator. We see that in the end, the change in all regression coefficient are not as big as we found in the procedure in Chapter 4. Figure 5.6 illustrates the log-likelihood values for every iteration and the change in estimated coefficient for Treatment covariate. The change in other covariates can be seen in Appendix C. We see that in this way, we have an increasing log-likelihood values for each iterations which suggests that our procedure is going on the right way.

The comparison of the Breslow estimator (before iteration) and the MLE in the last iteration are given in the level of baseline hazard and the cumulative baseline hazard (see Figure 5.7). An interesting point here is that actually in the first iteration, the MLE for the baseline hazard looks similar with the LCM in

Iteration	Treat	Age	BMIhi	Hypl	Dbet	Mi.H
1	0.0588	0.0425	-0.3445	0.3643	0.7155	0.5238
2	0.0362	0.0430	-0.3035	0.3273	0.6806	0.5175
3	0.0368	0.0432	-0.3029	0.3292	0.6798	0.5159
4	0.0380	0.0434	-0.3023	0.3293	0.6802	0.5175
5	0.0398	0.0436	-0.2968	0.3312	0.6807	0.5174
6	0.0390	0.0439	-0.3002	0.3311	0.6791	0.5148
⋮	⋮	⋮	⋮	⋮	⋮	⋮
240	0.0678	0.0558	-0.2497	0.3705	0.6569	0.4987
241	0.0678	0.0558	-0.2497	0.3705	0.6569	0.4987
242	0.0678	0.0558	-0.2497	0.3705	0.6569	0.4987

Table 5.6: The Change in Regression Coefficients in ICTUS Dataset due to Implementation of MLE Iterations

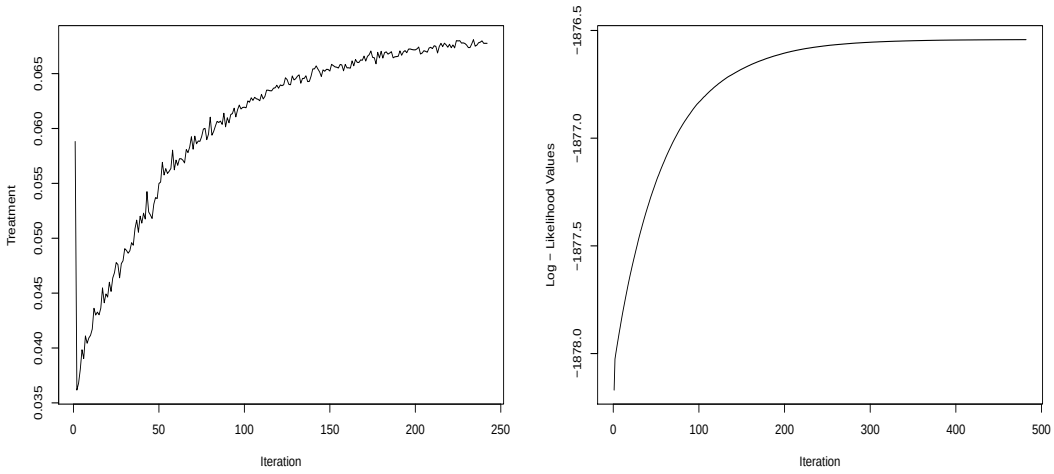

Figure 5.6: The Change in Estimated Treatment Coefficient (Left) and the Log-Likelihood Values in Each Iteration (Right)

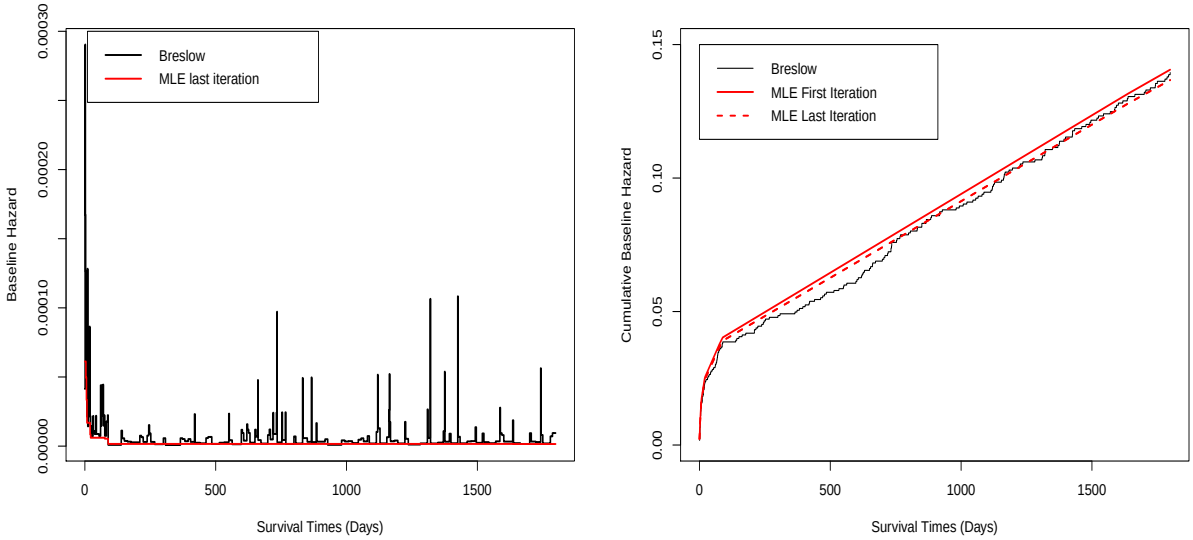

Figure 5.7: Comparison Between the Breslow Estimator and MLE (Left)in the Baseline Hazard Level; (Right) in the Cumulative Baseline Hazard Level

Chapter 4. In the end of the iteration, the MLE procedure gives an estimate for the cumulative baseline hazard which is close to the Breslow estimate. In Figure 5.8, we also illustrates the cumulative hazard curve for both treatment groups in median risk score at the first iteration and in the last iteration. This concludes that MLE for the baseline hazard is preferable than the LCM in estimating the parameters in Cox model especially in the case where we restrict the baseline hazard to be monotone.

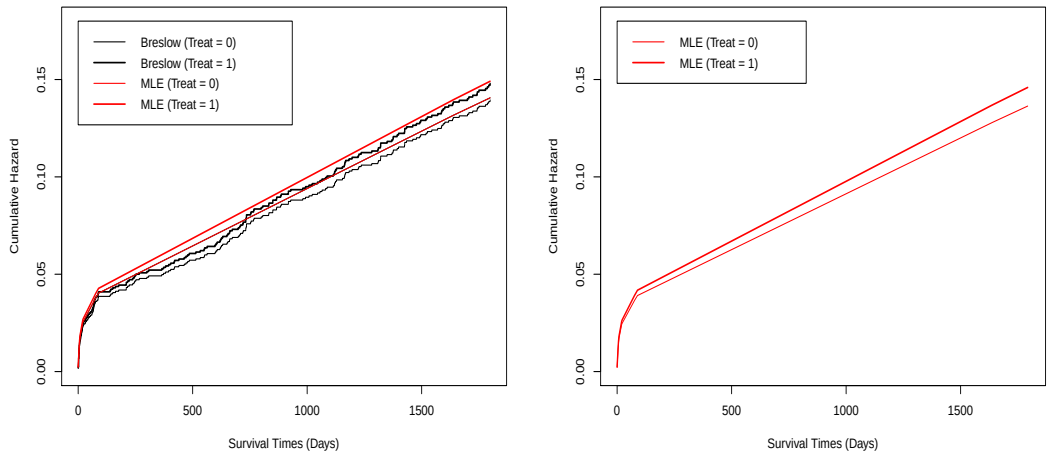

Figure 5.8: (Left:) The Comparison Between the Breslow Estimator and the MLE in the First Iteration at Median Risk Score. (Right:) The MLE in the Last Iteration at Median Risk Score

5.7 Summary

In this section, we investigated the effect of another shape constrained estimator (MLE for the baseline hazard) in estimating the Cox parameters. Our main purpose is to find a better estimate for the regression coefficient. In Chapter 4, we have seen that LCM-based estimator performs better in estimating the baseline hazard, but not in estimating the regression coefficient. In the end, we see that by applying this MLE for the baseline hazard, we obtain an estimate for the regression coefficient which has similar variance as the partial likelihood estimator. This concludes that even we have a maximum likelihood estimator that incorporates the information about the shape of the baseline hazard, we still could not improve the estimate for the regression coefficient.

Chapter 6

Conclusion

The purpose of this thesis can be divided into two main subjects. First, we want to test whether estimating the hazard of patients with myocardial infarction is nonincreasing. The second purpose is to investigate the effect of applying a shape constrained estimator to a semiparametric model. These issues are closely related with the medical perspective. For the traditional estimate, either the Kaplan-Meier estimator or the Breslow estimator, there is no assumption regarding the shape of the hazard curve. While from a biological perspective, the hazard is decreasing. As introduced in [10], we have an estimator for a decreasing hazard.

This estimator is reasonable to estimate the decreasing hazard in ICTUS dataset as confirmed by the statistical test in Chapter 3. Assuming that this shape constraint on the hazard is valid, we investigated the advantages and disadvantages of this estimator, especially in the Cox model. We investigated the effect in estimating the parameters in the Cox model, the regression coefficient and the baseline hazard. The LCM gives a better estimate for the Cox's baseline hazard. By means of simulation, we investigate pointwise distribution as well as a global measure. Both of these measures confirm that the LCM estimate for the baseline hazard gives a smaller distance to the true hazard than the Breslow estimator. The regression coefficient in the Cox model represents the importance of the covariates. We aim to investigate whether incorporating the knowledge of a decreasing hazard would yield a better estimate for the regression coefficient. In Chapter 4, we see that this shape contrained estimator which is obtained by the derivative of the LCM does not improve the estimation. It is even getting worse in the extreme case, such as a linear cumulative hazard.

Since applying the LCM does not give a better estimate for the regression coefficient, we use another method introduced in [16] to estimate the baseline hazard. In principle this estimator is the MLE for the shape-restricted baseline hazard. We insert this estimator to find the maximum likelihood estimate for the regression coefficient. In comparison with the LCM, this method gives a better estimate for the regression coefficient. But still the standard deviation of this estimator are similar with the partial likelihood estimator. This concludes that partial likelihood is still the best way to estimate the regression coefficient of Cox model.

We conclude that based on the investigations in this manuscript, the MLE is still preferable in estimating the Cox parameters than the LCM. If we compare the performance of the MLE and LCM in estimating the baseline hazard for all distribution with decreasing hazard, we see that the MLE performs better than the LCM (see the boxplots in Appendix A for the LCM and Appendix B for the MLE). Near zero, the MLE gives an estimate for the baseline hazard with smaller variance than the LCM. Also, we notice that the right-skewness of LCM at these points are relatively similar to the Breslow estimate. Another important

comparison can be found in the ICTUS dataset. The MLE iteration procedure converges to the maximum values and the corresponding parameters are close to the traditional estimator while this is not the case in the LCM-based estimator.

Appendix A

The LCM for the Baseline Hazard in Several Weibull Distributions

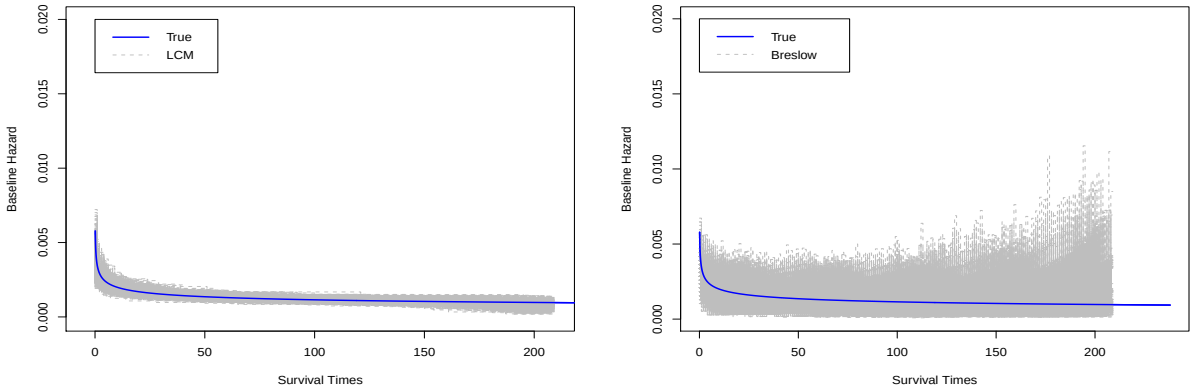

Figure A.1: The LCM (Left) and the Breslow (Right) Estimate of the Baseline Hazard in the Less Concave Cumulative Hazard

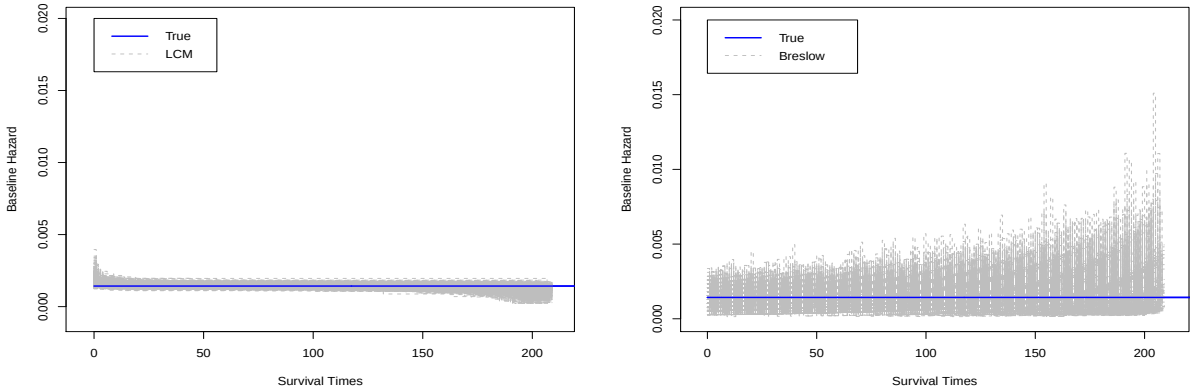

Figure A.2: The LCM (Left) the Breslow (Right) Estimate of the Baseline Hazard in the Linear Cumulative Hazard

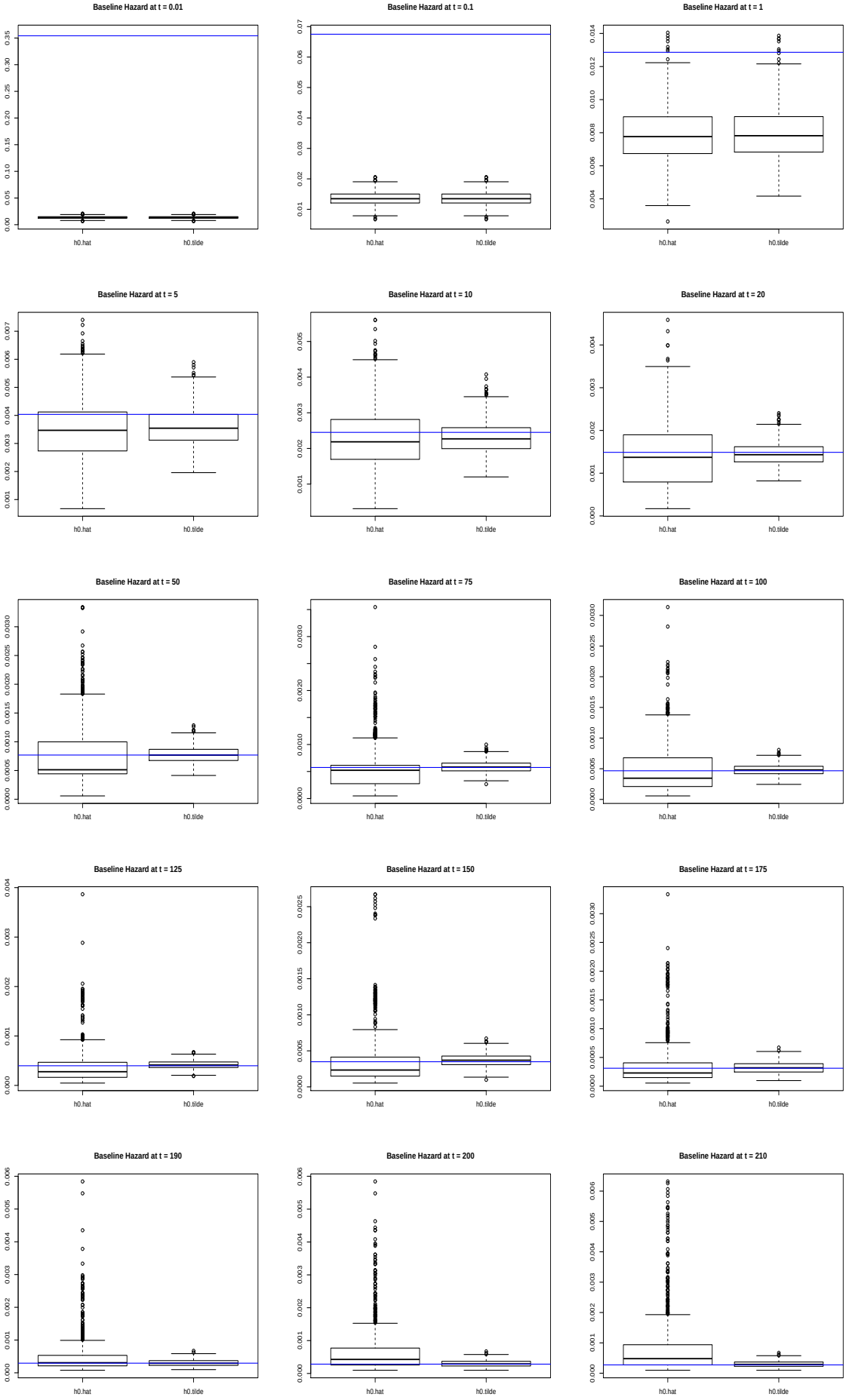

Figure A.3: Pointwise Distribution Comparison of the Baseline Hazard in Strong Concave Case

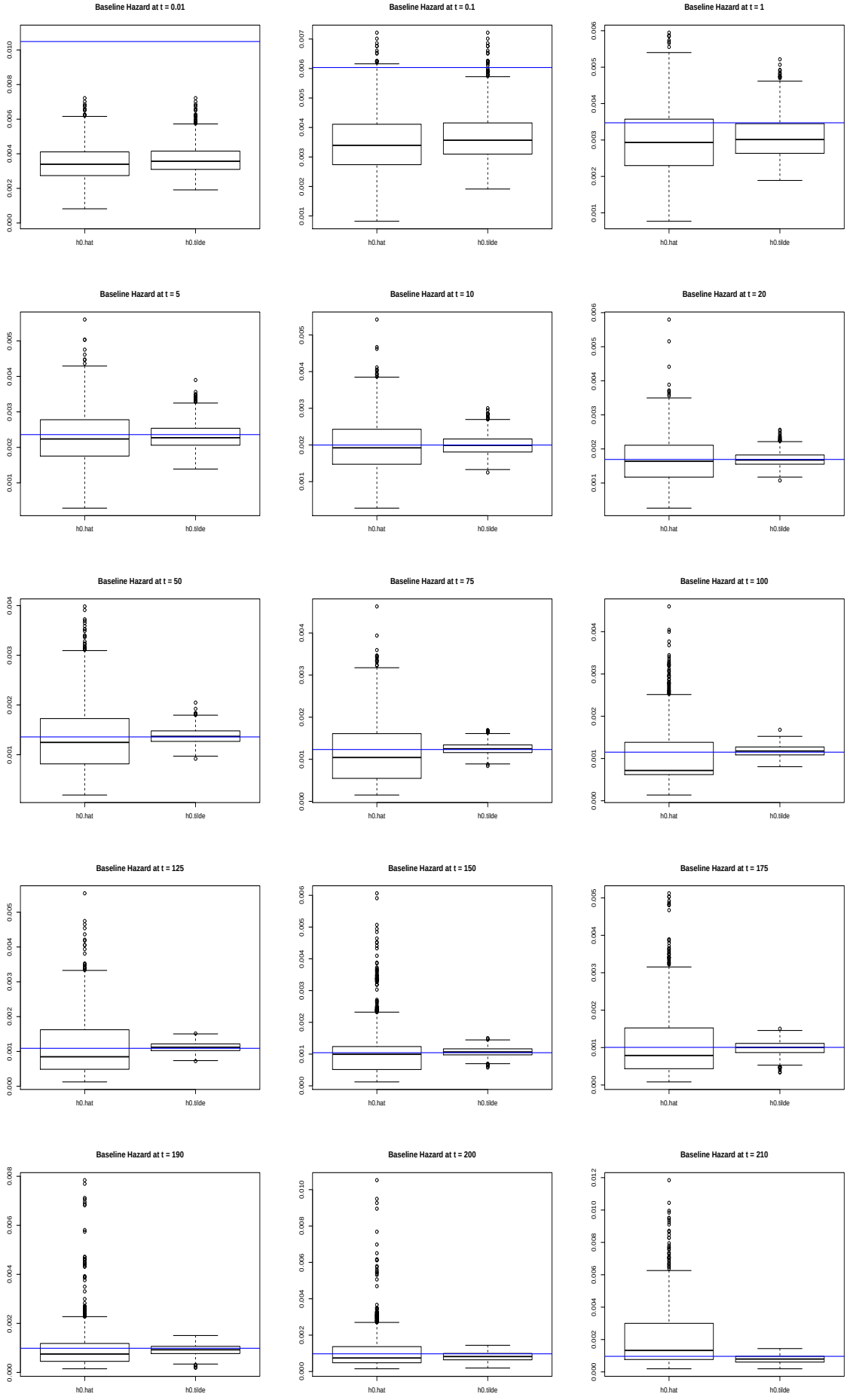

Figure A.4: Pointwise Distribution Comparison of the Baseline Hazard in Less Concave Case

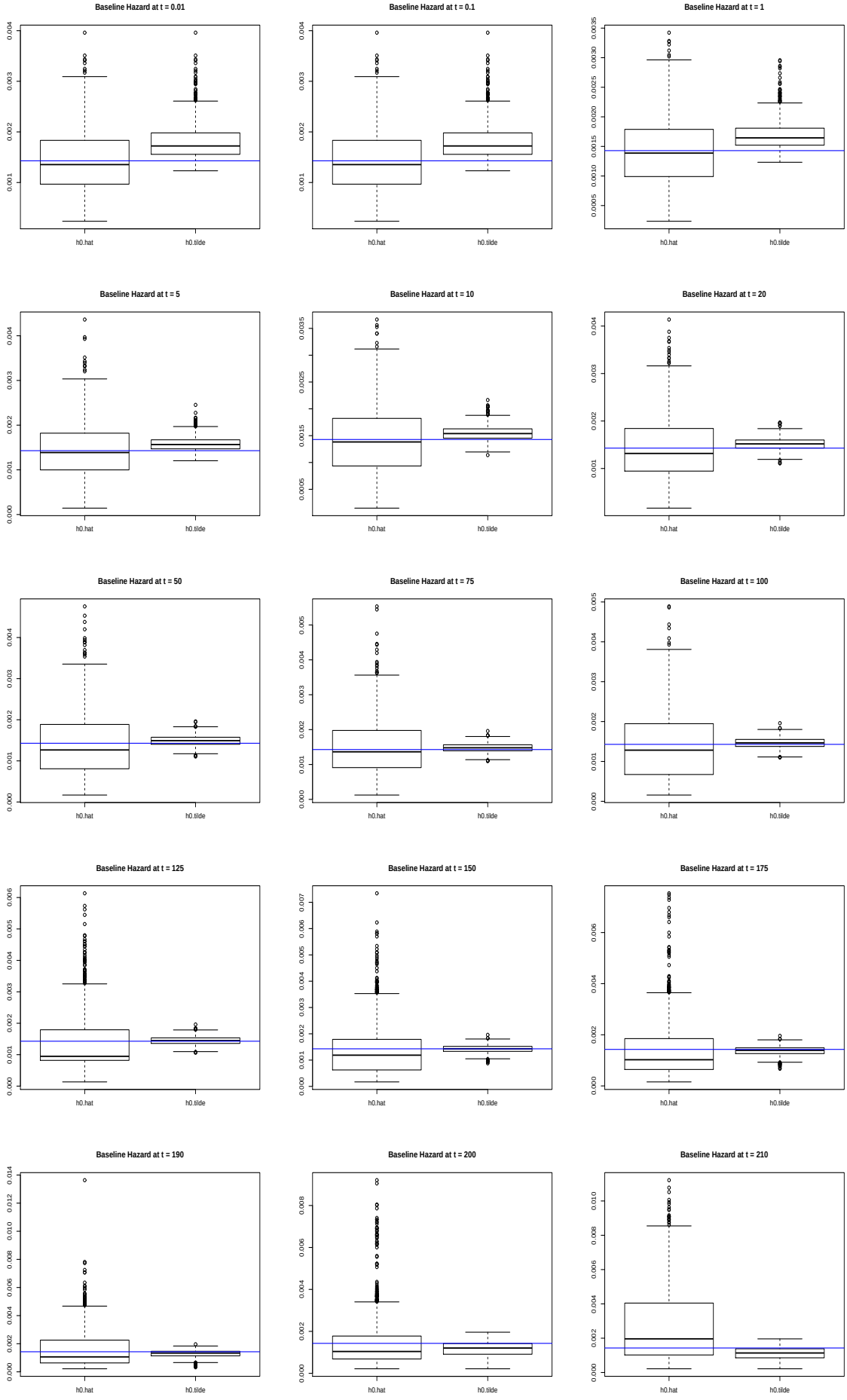

Figure A.5: Pointwise Distribution Comparison of the Baseline Hazard in Linear Case

Appendix B

The MLE for the Baseline Hazard in Several Weibull Distributions

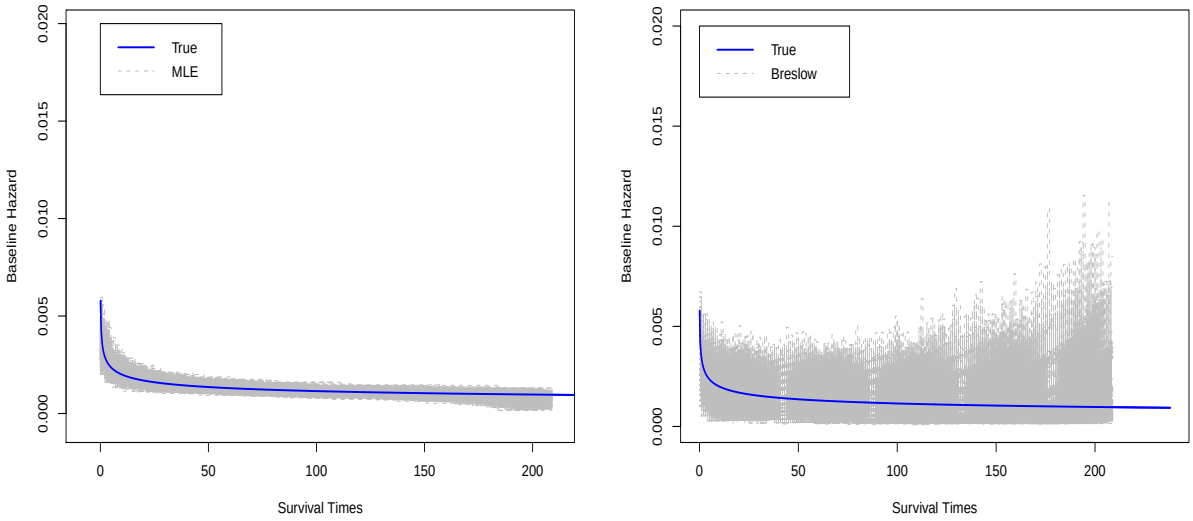

Figure B.1: The MLE (Left) and Breslow (Right) Estimator for the Baseline Hazard in Less Concave Case

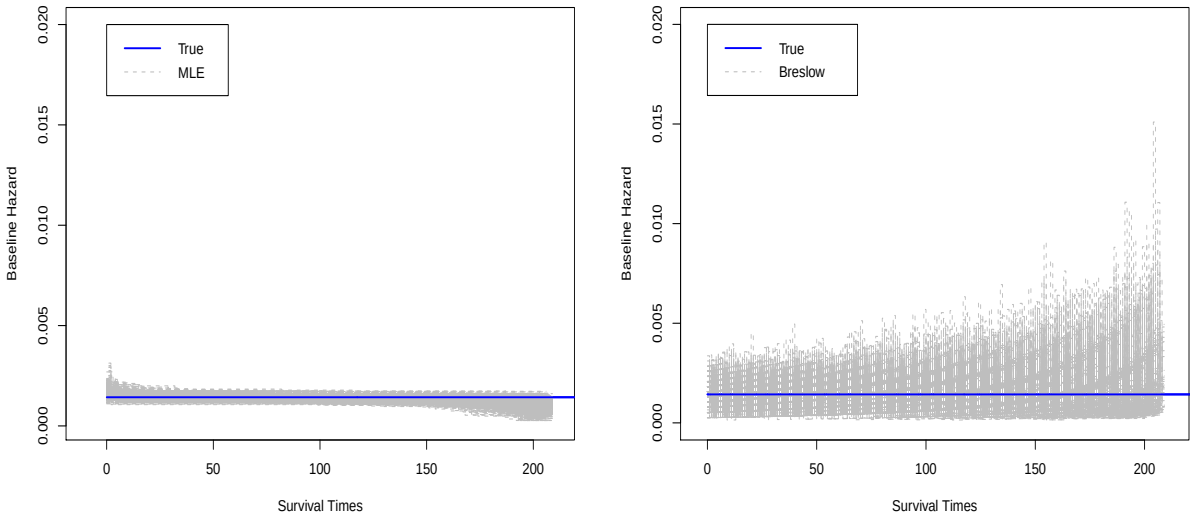

Figure B.2: The MLE (Left) and Breslow (Right) Estimator for the Baseline Hazard in Linear Case

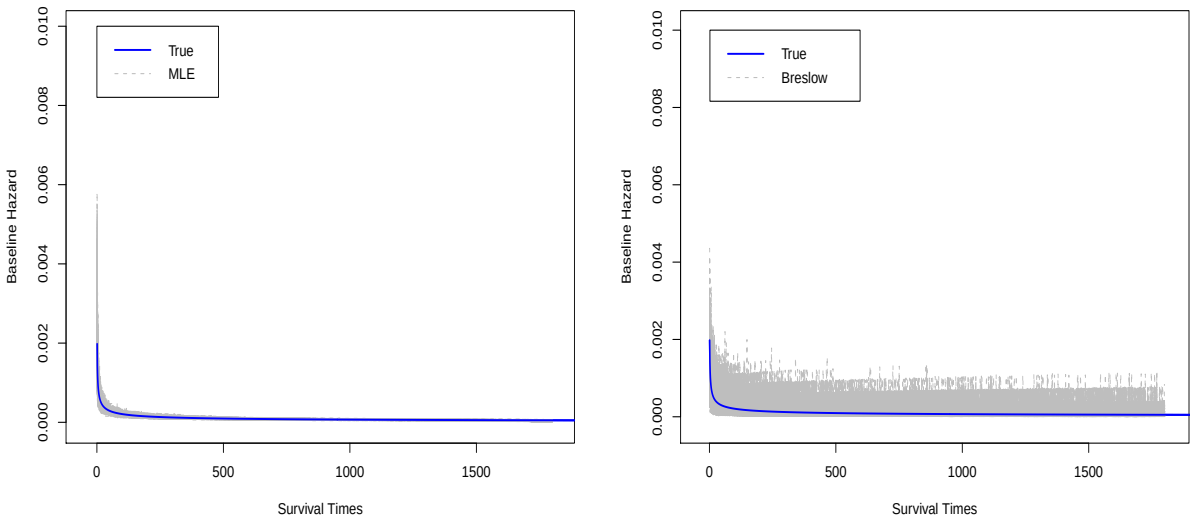

Figure B.3: The MLE (Left) and Breslow (Right) Estimator for the Baseline Hazard in ICTUS-Based Weibull Case

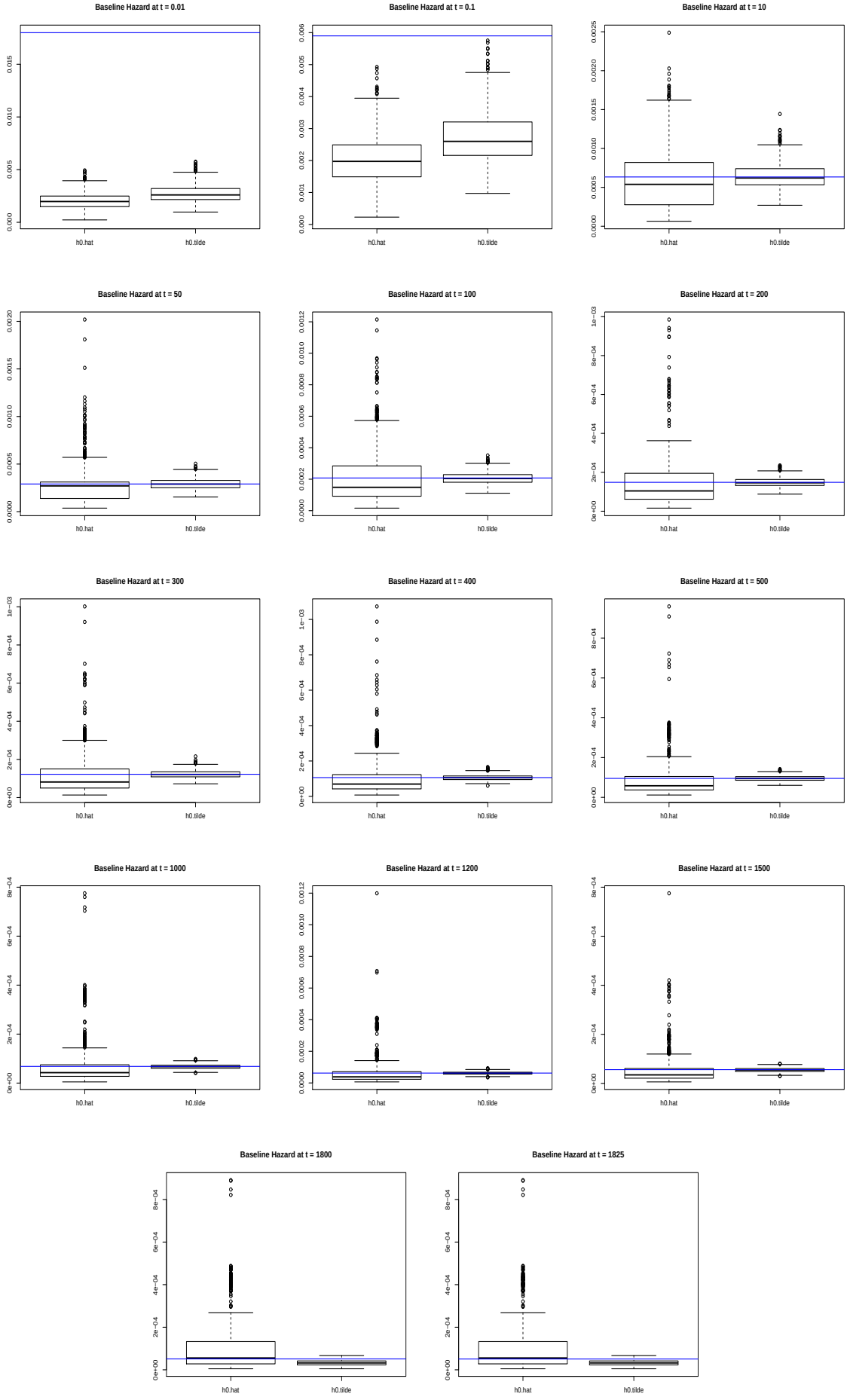

Figure B.4: Pointwise Distribution of the Baseline Hazard in ICTUS-Based Weibull Case

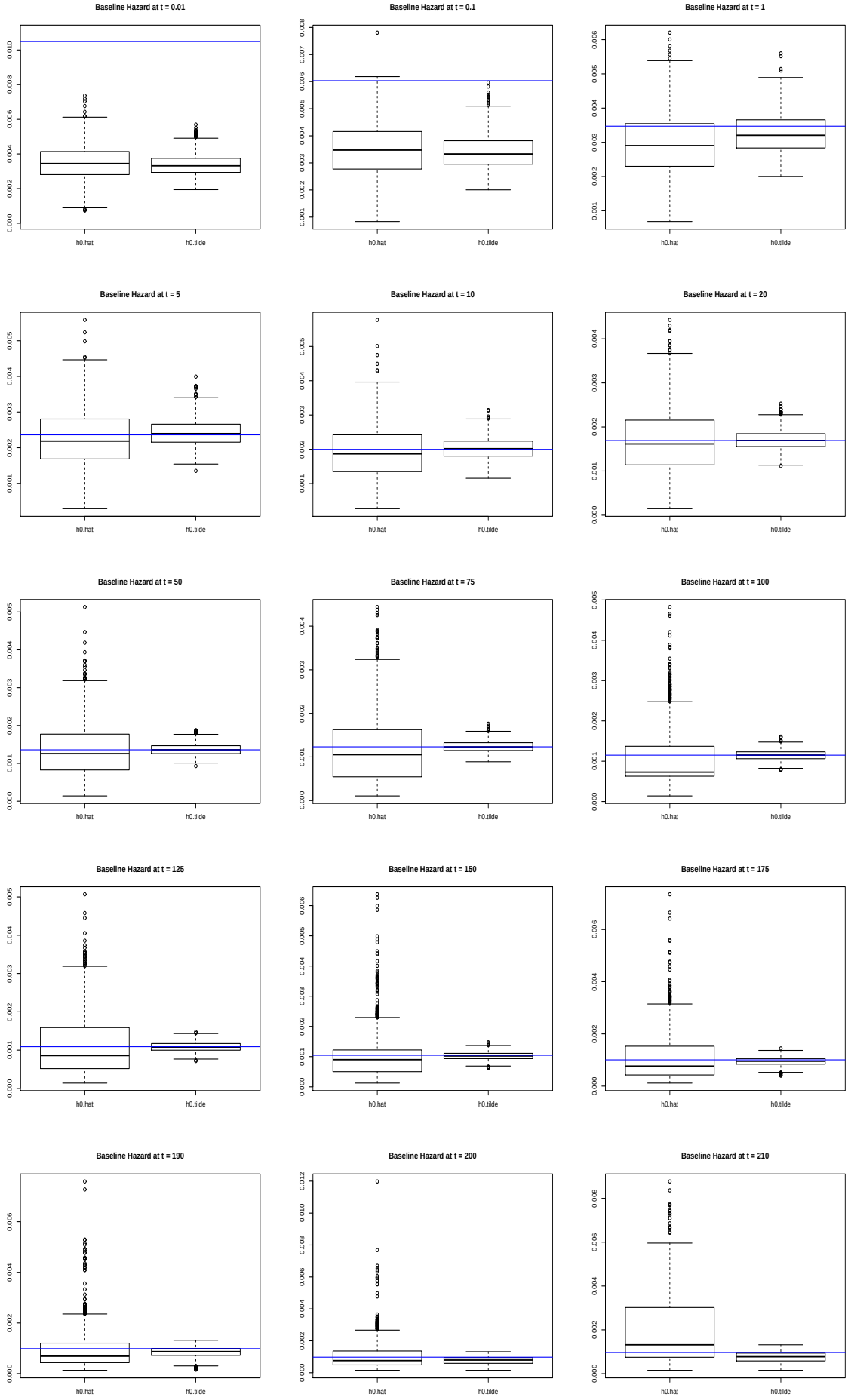

Figure B.5: Pointwise Distribution of the Baseline Hazard in Less Concave Case

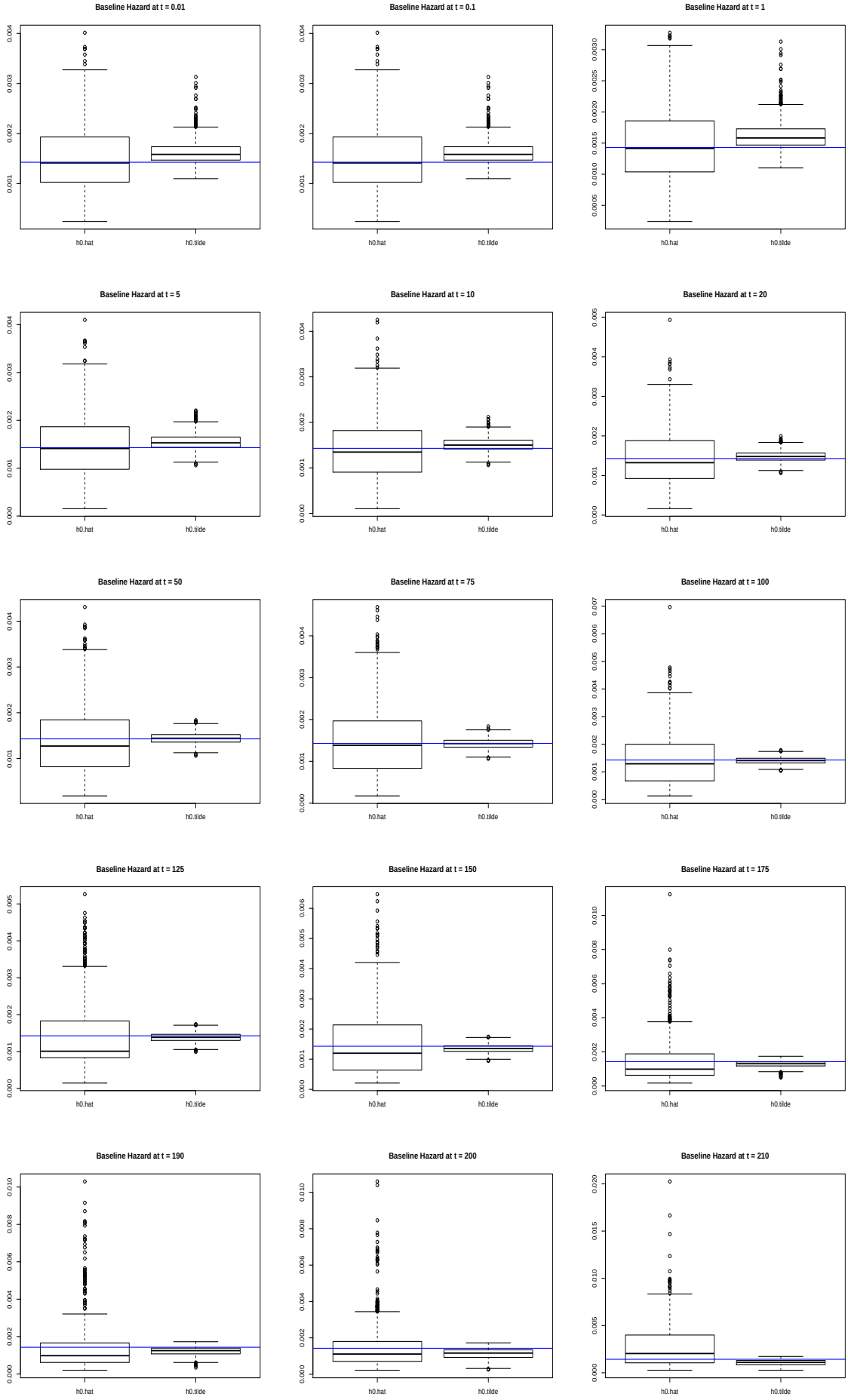

Figure B.6: Pointwise Distribution of the Baseline Hazard in Linear Case

Appendix C

Comparison Between LCM and MLE Performance in the ICTUS Dataset

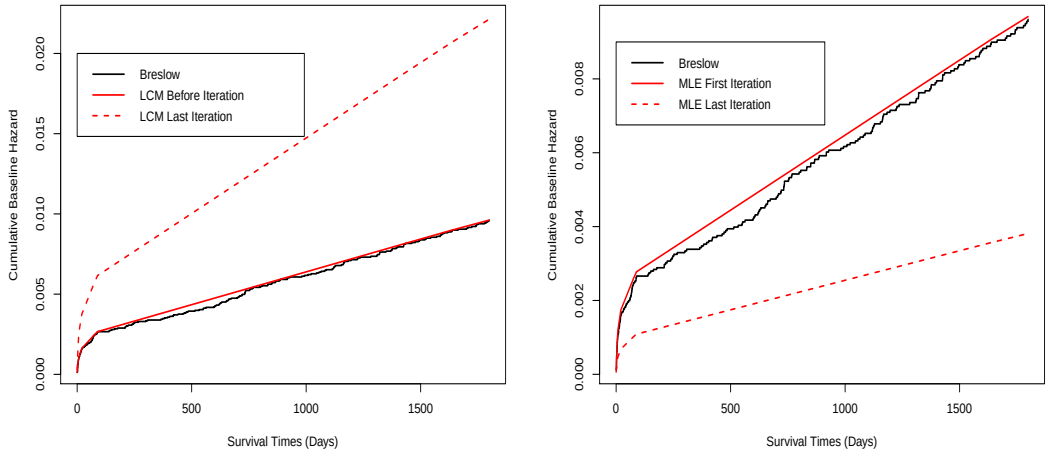

Figure C.1: The Cumulative Baseline Hazard Estimate Using the LCM (Left) and Using the MLE (Right) at the First and Last Iteration

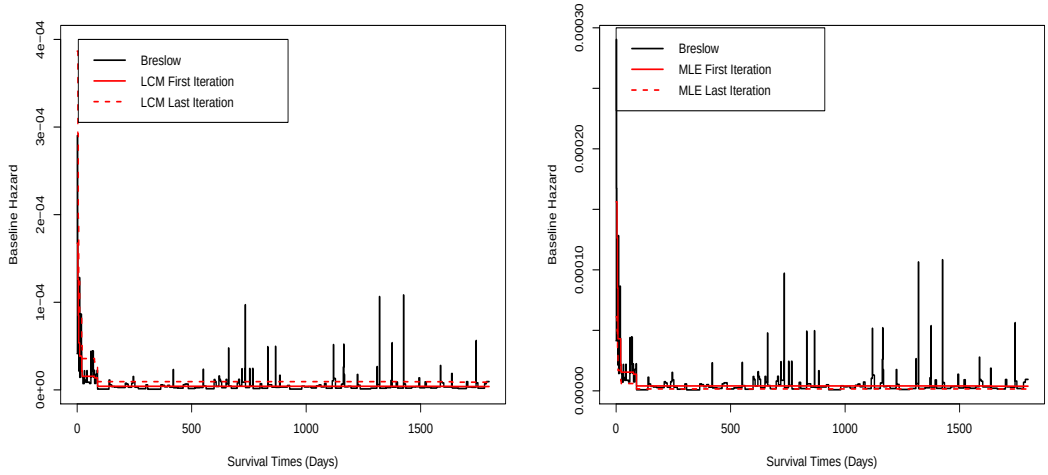

Figure C.2: The Baseline Hazard Estimate Using the LCM (Left) and Using the MLE (Right) at the First and Last Iteration

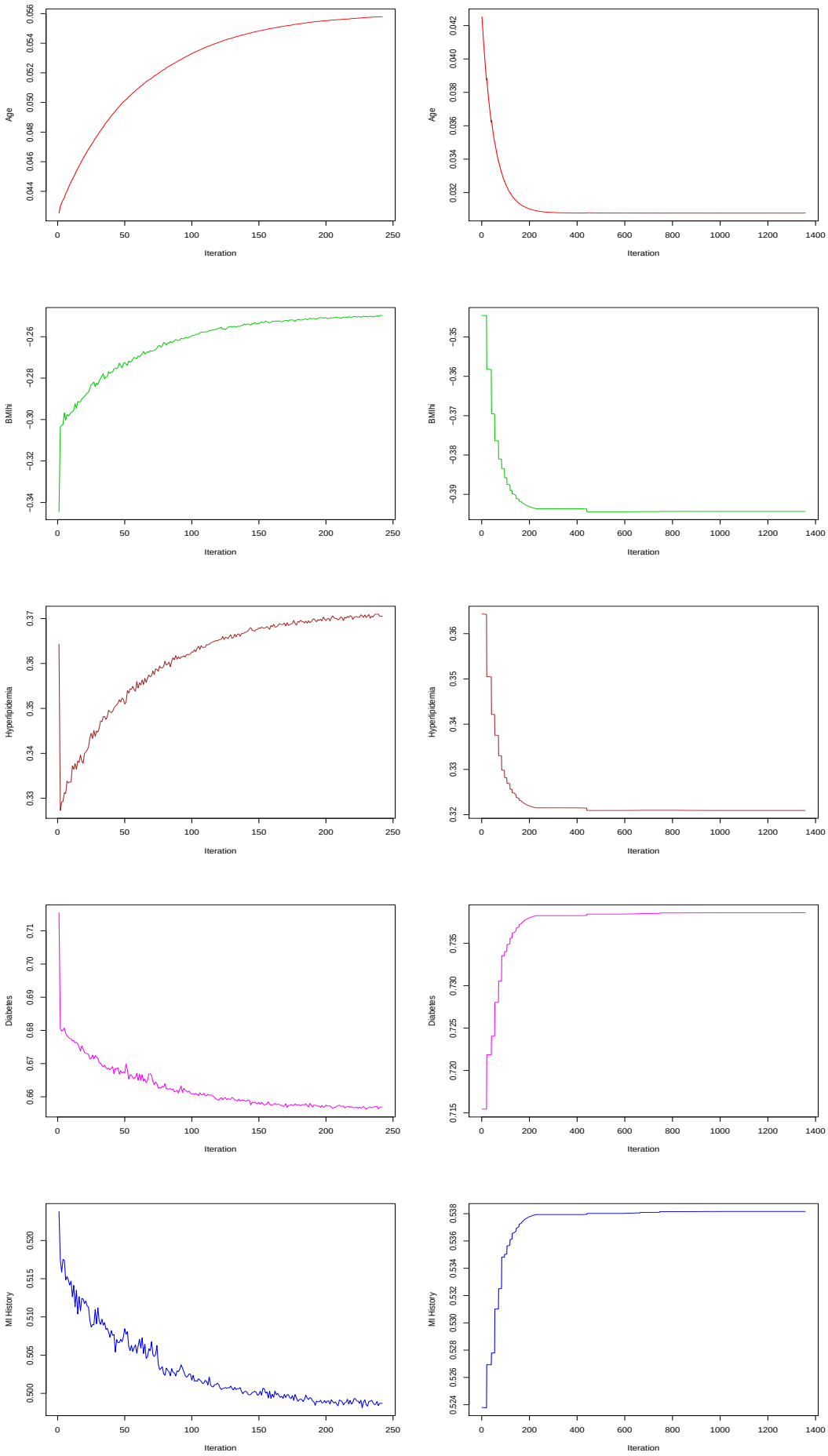

Figure C.3: The Changes in the Significant Regression Coefficient Estimates of the ICTUS Dataset

References

[1] Damman, P. *et. al.* 5-year clinical outcomes in the ictus (invasive versus conservative treatment in unstable coronary syndromes) trial: A randomized comparison of an early invasive versus selective invasive management in patients with non-st-elevation acute coronary syndrome. *Journal of American College of Cardiology*, 2010, http://content.onlinejacc.org/cgi/reprint/j.jacc.2009.11.026v1.pdf.

[2] Cox, D.R. Regression models and life tables (with discussion). *Journal of Royal Statistical Society, Series B*, 34:187–220, 1982.

[3] Cox, D.R. and Oakes, D. *Analysis of Survival Data*. Chapman and Hall, New York, 1984.

[4] Breslow, N.E. Covariance analysis of censored survival data. *Biometrics*, 30:89–100, 1974.

[5] Efron, B. The efficiency of cox's likelihood function for censored data. *Journal of American Statistical Association*, 72:557–565, 1977.

[6] Hosmer,D.W and Lemeshow,S. *Applied Survival Analysis: Regression Modelling of Time to Event Data*. Willey Interscience, New York, 1999.

[7] http://en.wikipedia.org/wiki/Body_mass_index.

[8] Grenander, U. On the theory of mortality measurement.ii. *Skand. Aktuarietidskr*, 39:125–153, 1956.

[9] Robertson, T., Wright, F.T and Dykstra, R.L. *Order Restricted Statistical Inference*. John Wiley & Sons, New York, 1988.

[10] Huang, Y. and Zhang, C.H. Estimating a monotone density from censored observations. *The Annals of Statistics*, 22 No. 3:1256–1274, Sep, 1994.

[11] Huang, J. and Wellner, J.A. Estimation of a monotone density or monotone hazard under random censoring. *Scandinavian Journal of Statistics*, 22 No. 1:3 – 33, Mar, 1995.

[12] Manly, B.F.J. *Randomization,Bootstrap and Monte-Carlo Methods in Biology, 2nd ed.* Chapman & Hall, London, 1997.

[13] Heller, G. and Simonoff, J.S. Prediction in censored survival data: A comparison of the proportional hazards and linear regression models. *Biometrics*, 48:101–115, 1992.

[14] Halabi, S. and Singh, B. Sample size determination for comparing several survivalcurves with unequal allocations. *Statistis in Medicine*, 23:1793–1815, 2004.

[15] Rao, P. Estimations for distribution with monotone failure rate. *The Annals of Mathematical Statistics*, 41, No. 2:507–519, 1970.

[16] Chung, D. and Chang, M.N. An isotonic estimator of the baseline hazard function in cox's regression model under order restriction. *Statistics and Probability Letters*, 21:223 – 228, 1994.