

# LOG HAZARD REGRESSION

by

Huiying Sun

Ph.D, Harbin Institute of Technology, Harbin, CHINA, 1991.

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Statistics

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September, 1999

©Huiying Sun, 1999

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Statistics

The University of British Columbia  
Vancouver, Canada

Date Sept. 29, 1988

## Abstract

We propose using regression splines to estimate the two log marginal hazard functions of bivariate survival times, where each time could be censored. The method is a modified version of Kooperberg, Stone and Truong's (JASA, 1995) hazard regression for estimating a univariate survival time. We derive an approach to find standard errors for estimates of the difference of the log hazard functions. The approach is inspired by Wei, Lin, and Weissfeld (JASA, 1989).

We also propose procedures for testing the four hypotheses that the marginals follow an exponential or Weibull distribution and that the two failure times have the same distribution or have proportional hazards.

A simulation study is conducted to assess the performance of our estimates and test procedures. We study the effects of the censoring rates, correlation levels, and number of knots. The regression is applied to the data set of the Diabetic Retinopathy Study (Diabetic Retinopathy Study Research Group, 1981). Our analysis for the data set matches study results of Huster, Brookmeyer, and Self (1989).

# Contents

Abstract	ii
Table of Contents	iii
List of Tables	vi
List of Figures	vii
Acknowledgments	xv
<b>1 Introduction</b>	<b>1</b>
<b>2 Hazard Regression</b>	<b>5</b>
2.1 Definitions and Notation . . . . .	5
2.1.1 Failure Time Distribution . . . . .	5
2.1.2 Censored Survival Data . . . . .	7
2.2 Log Hazard Regression . . . . .	8
2.2.1 Model and Estimates . . . . .	8
2.2.2 Conditions on $\mathbf{b}$ and $\boldsymbol{\beta}$ . . . . .	12
2.3 Consistency and Normality of $\hat{\boldsymbol{\beta}}$ . . . . .	15
2.4 Proofs of the Lemmas . . . . .	18

2.5	Log Hazard Regression of Paired Failure Times . . . . .	26
2.5.1	Model and Notation . . . . .	26
2.5.2	Consistency and Normality of $\hat{\beta}$ . . . . .	27
<b>3</b>	<b>Regression Space</b>	<b>32</b>
3.1	Cubic Splines . . . . .	33
3.2	The Regression Space $\mathcal{B}$ . . . . .	33
3.3	Numerical Implementation . . . . .	38
3.4	Knot Selection . . . . .	40
3.5	Hypothesis Testing . . . . .	42
3.5.1	Univariate Case . . . . .	42
3.5.2	Bivariate Case . . . . .	43
<b>4</b>	<b>Application to the Diabetic Retinopathy Study</b>	<b>46</b>
4.1	Data Description . . . . .	46
4.2	Data Analysis . . . . .	47
4.2.1	Model 1 . . . . .	50
4.2.2	Model 2 . . . . .	54
4.3	Other Models Used to fit the Eye Data . . . . .	60
4.3.1	Exponential Model . . . . .	60
4.3.2	Weibull Model . . . . .	63
4.3.3	Cox Proportional Hazard Model . . . . .	64
<b>5</b>	<b>Simulation</b>	<b>68</b>
5.1	Description of the Simulation Study . . . . .	69
5.2	The Univariate Case . . . . .	74
5.2.1	Exponential Model . . . . .	74
5.2.2	Weibull Model . . . . .	77

5.2.3	The B-spline Model . . . . .	79
5.3	The Bivariate Case . . . . .	81
5.3.1	Generation of Paired Dependent Data . . . . .	81
5.3.2	Proportional Hazards Model . . . . .	83
5.3.3	Non-proportional Hazards Model . . . . .	86
5.3.4	Effect of the Number of Knots . . . . .	87
5.4	Summary of Simulations . . . . .	88
<b>6</b>	<b>Conclusion</b>	<b>135</b>
	<b>Bibliography</b>	<b>139</b>
	<b>Appendix</b>	<b>142</b>
<b>A</b>	<b>Simulation Results for Bivariate Data with 50% Censoring</b>	<b>142</b>

# List of Tables

4.1	Test statistic and p-value for testing that the failure times of the treated eyes and the untreated eyes have the same distribution. . . . .	59
4.2	$\chi^2$ test statistics and p-values for testing that failure times are exponentially distributed. . . . .	63
4.3	$z$ statistics and p-values for testing that failure times have Weibull distributions. . . . .	63
4.4	Test statistic and p-value for testing that the two hazards of the treated eyes and the untreated eyes are proportional. . . . .	64
4.5	Test results from the Cox proportional hazards model for the eye data. .	64

# List of Figures

4.1	Histograms of observed times of the eye data. . . . .	48
4.2	Scatter plots of the observations of the eye data. . . . .	49
4.3	The estimated density functions of the eye data for Model 1. . . . .	51
4.4	The estimated survival functions of the eye data for Model 1. . . . .	52
4.5	The estimated hazard functions of the eye data for Model 1 with pointwise 95% confidence intervals. . . . .	53
4.6	The estimated log ratio of the hazard functions of the eye data with pointwise 95% confidence intervals for Model 1. . . . .	54
4.7	The estimated density functions of the eye data for Model 2. . . . .	55
4.8	The estimated survival functions of the eye data for Model 2. . . . .	56
4.9	The estimated hazard function of the untreated eye for Model 2. . . . .	56
4.10	The estimated log ratio of the hazard functions of the eye data with pointwise 95% confidence intervals for Model 2. . . . .	57
4.11	The estimated density functions of the failure times of the untreated eyes. . . . .	58
4.12	The estimated survival functions of the untreated eyes. . . . .	58
4.13	The estimated hazard functions of the failure times of the untreated eyes. . . . .	59
4.14	The estimated survival curves of the treated eye. . . . .	61
4.15	The estimated survival curves of the untreated eye. . . . .	62



4.16	The estimated log hazards ratio from Model 1 and the Cox proportional hazards model with the pointwise 95% confidence intervals from the Cox proportional hazards model. . . . .	66
4.17	The estimated log hazards ratio from Model 1 and the Cox proportional hazards model with the pointwise 95% confidence intervals from Model 1. . . . .	66
4.18	The estimated log hazards ratio from Model 2 and the Cox proportional hazards model with the pointwise 95% confidence intervals from the Cox proportional hazards model. . . . .	67
4.19	The estimated log hazards ratio from Model 2 and the Cox proportional hazards model with the pointwise 95% confidence intervals from Model 2. . . . .	67
5.1	The histograms of the simulated failure times, censoring times, non-censored failure times, and observed censoring times for exponential data as in model (5.2). . . . .	91
5.2	Log hazard of the exponential distribution (5.2) and the pointwise quartiles and empirical mean of the 500 estimated log hazards. . . . .	92
5.3	The pointwise standard deviations of the estimated log hazard for the exponential model (5.2) and the pointwise quartiles and empirical mean of the 500 estimated standard errors. . . . .	93
5.4	The quantiles of the pointwise z values of the estimated log hazards for the exponential distribution (5.2). . . . .	94
5.5	Histograms and qq-plots of the normalized estimate of $\beta$ from the Quantile Knots methods with three knots for the exponential model (5.2). . . . .	95
5.6	Empirical distribution functions of the p-values for testing that the failure times are exponentially distributed. . . . .	96

5.7	The histograms of the simulated failure times, censoring times, non-censored failure times, and observed censoring times for Weibull data as in model (5.3). . . . .	97
5.8	Log hazard of the Weibull distribution (5.3) and the pointwise quartiles and empirical mean of the estimated log hazards. . . . .	98
5.9	The pointwise standard deviations of the estimated log hazards for the Weibull distribution (5.3) and the pointwise quartiles and empirical mean of the 500 estimated standard errors. . . . .	99
5.10	The quantiles of the pointwise z values of the estimated log hazards for the Weibull model (5.3). . . . .	100
5.11	The histograms and the qq-plots of the standardized estimates of the parameter $\beta$ from the method 5.1 with six knots for the Weibull model (5.3). . . . .	101
5.12	Empirical distribution functions of the p-values for testing the hypothesis that the failure time distribution is exponential. . . . .	102
5.13	Empirical distribution functions of the p-values for testing that the failure times have a Weibull distribution. . . . .	103
5.14	The histograms of the simulated failure times, censoring times, non-censored failure times, and observed censoring times for the data as in the B-spline model (5.4). . . . .	104
5.15	Log hazard of the the B-spline data as in model (5.4) and the pointwise quartiles and empirical mean of the estimated log hazards. . . . .	105
5.16	The pointwise standard deviations of the estimated log hazards for the B-spline model (5.4) and the pointwise quartiles and empirical mean of the estimated standard errors. . . . .	106

5.17	The quantiles of the pointwise $z$ values of the estimated log hazards for the B-spline model (5.4). . . . .	107
5.18	The histograms and the qq-plots of the standardized estimates of the parameter $\beta$ from the True Knots method for the B-spline model (5.4). .	108
5.19	Empirical distribution functions of the $p$ -values for testing the hypothesis that the failure times have a Weibull distribution. . . . .	109
5.20	The plots of the correlation coefficients versus $\theta$ for the data as in the proportional hazards and non-proportional hazards models . . . . .	110
5.21	The histograms of the marginal failure times, and non-censored marginal failure times for the data generated according to the proportional hazards model in Section 5.1. The censoring rate for the “treatment” is 25%. . .	111
5.22	The true log hazards ratio of the data from the proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the 500 estimated log hazards ratios. . . . .	112
5.23	The standard deviations of the estimated log hazards ratio for data generated according to the proportional hazards model. The censoring rate is 25% for the “treatment”. . . . .	113
5.24	The quantiles of the pointwise $z$ values of the estimated log hazards ratio for the data generated according to the proportional hazards model. The censoring rate is 25% for the “treatment”. . . . .	114
5.25	Empirical distribution functions of the $p$ -values for testing the hypothesis that the two failure times of the data generated as in the proportional hazards model in Section 5.1 have proportional hazards. . . . .	115
5.26	The histograms of the marginal failure times, and the non-censored marginal failure times for the data from the non-proportional hazards model in Section 5.1. The censoring rate for the “treatment” is 25%. . . . .	116

5.27	The true log hazards ratio of the data from the non-proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the estimated log hazards ratios. . . . .	117
5.28	The standard deviations of the estimated log hazards ratio for data generated according to the non-proportional hazards model. The censoring rate is 25% for the “treatment”. . . . .	118
5.29	The quantiles of the pointwise z values of the estimated log hazards ratio for the data generated according to the non-proportional hazards model. The censoring rate is 25% for the “treatment”. . . . .	119
5.30	Empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the non-proportional hazards model in Section 5.1 have proportional hazards. . . . .	120
5.31	The true log hazards ratio of the data from the proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the estimated log hazards ratios by the Same Knots method. . .	121
5.32	The true log hazards ratio of the data from the proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the estimated log hazards ratios by the Different Knots method.	122
5.33	The true log hazards ratio of the data from the non-proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the estimated log hazards ratios by the Same Knots method. . .	123
5.34	The true log hazards ratio of the data from the non-proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the estimated log hazards ratios by the Different Knots method.	124

5.35	The standard deviations of the estimated log hazards ratio by the Same Knots method for data generated according to the proportional hazards model. The censoring rate is 25% for the "treatment". . . . .	125
5.36	The standard deviations of the estimated log hazards ratio by the Different Knots method for data generated according to the proportional hazards model. The censoring rate is 25% for the "treatment". . . . .	126
5.37	The standard deviations of the estimated log hazards ratio by the Same Knots method for data generated according to the non-proportional hazards model. The censoring rate is 25% for the "treatment". . . . .	127
5.38	The standard deviations of the estimated log hazards ratio by the Different Knots method for data generated according to the non-proportional hazards model. The censoring rate is 25% for the "treatment". . . . .	128
5.39	The quantiles of the pointwise z values of the estimated log hazards ratio by the Same Knots method for the data generated according to the proportional hazards model. The censoring rate is 25% for the "treatment".	129
5.40	The quantiles of the pointwise z values of the estimated log hazards ratio by the Different Knots method for the data generated according to the proportional hazards model. The censoring rate is 25% for the "treatment".	130
5.41	The quantiles of the pointwise z values of the estimated log hazards ratio by the Same Knots method for the data generated according to the non-proportional hazards model. The censoring rate is 25% for the "treatment".	131
5.42	The quantiles of the pointwise z values of the estimated log hazards ratio by the Different Knots method for the data generated according to the non-proportional hazards model. The censoring rate is 25% for the "treatment". . . . .	132

5.43	By the Same Knots method empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the proportional hazards model in Section 5.1 have proportional hazards. . . . .	133
5.44	By the Same Knots method empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the non-proportional hazards model in Section 5.1 have proportional hazards. . . . .	134
A.1	The histograms of the marginal failure times, and non-censored marginal failure times for the data generated according to the proportional hazards model in Section 5.1. The censoring rate for the “treatment” is 50%. . .	143
A.2	The true log hazards ratio of the data as in the proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the 500 estimated log hazards ratios. . . . .	144
A.3	The standard deviations of the estimated log hazards ratio for data generated according to the proportional hazards model. The censoring rate is 50% for the “treatment”. . . . .	145
A.4	The quantiles of the pointwise z values of the estimated log hazards ratio for the data generated according to the proportional hazards model. The censoring rate is 50% for the “treatment”. . . . .	146
A.5	Empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the proportional hazards model in Section 5.1 have proportional hazards. . . . .	147
A.6	The histograms of the marginal failure times, and the non-censored marginal failure times for the data from the non-proportional hazards model in Section 5.1. The censoring rate for the “treatment” is 50%. . . . .	148

A.7	The true log hazards ratio of the data as in the non-proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the 500 estimated log hazards ratios. . . . .	149
A.8	The standard deviations of the estimated log hazards ratio for data generated according to the non-proportional hazards model. The censoring rate is 50% for the “treatment”. . . . .	150
A.9	The quantiles of the pointwise z values of the estimated log hazards ratio for the data generated according to the non-proportional hazards model. The censoring rate is 50% for the “treatment”. . . . .	151
A.10	Empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the non-proportional hazards model in Section 5.1 have proportional hazards. . . . .	152

# Acknowledgments

Foremost, I would like to express my sincere thanks to my supervisor, Nancy Heckman, for her guidance, support, and patience throughout the development of this thesis. Her numerous advices will be eternally appreciated.

I would also like to thank James Zidek for his thoughtful comments on this manuscript. Additionally, I am grateful to all faculty members and office staff for being kind and helpful throughout my Master's program.

Finally, I would like to thank all graduate students for letting me share their knowledge and expertise and making my two years at UBC such a pleasant experience.



# Chapter 1

## Introduction

Bivariate failure time data arise when study subject units are paired. Examples of paired units include eyes or ears from the same person, twins, and father and son. Since there exist natural relationships between the two subjects in one pair, the two failure times within the same pair might be correlated. Moreover, in such studies, either or both failure times might not be observed because of censoring. A well-know study involving paired failure times is the Diabetic Retinopathy Study, Diabetic Retinopathy Study Research Group (1981). The dependence, along with the censoring, greatly complicates the analysis of the data. This data set will be discussed in Chapter 4.

In univariate survival analysis, classical methods such as the Kaplan-Meier estimator and the Cox proportional regression model are based on the hazard function  $h(t)$ ,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)},$$

where  $S$  is the survivor function and  $f$  is the density function, see Section 2.1. If  $T$  is a discrete random variable and  $T$  can take on values  $0 \leq t_1 < t_2 < \dots$ , then the hazard function is

$$h(t_j) = P(T = t_j | T \geq t_j) = \frac{P(T = t_j)}{S(t_j)} = 1 - \frac{S(t_{j+1})}{S(t_j)}, j = 1, 2, \dots.$$

And thus the survival function can be written as

$$S(t) = \prod_{j:t_j \leq t} (1 - h(t_j)),$$

see Lawless (1982). An analogous formula relating  $S$  and  $h$  exists for  $T$  continuous, see Gill (1992). But in the bivariate case, as Gill discussed, we do not have a nice formula relating the hazard and survivor functions, since there is no canonical way to define past, present, and future at “time”  $t$ .

Traditionally, there are two main approaches to analyze censored paired data. One approach is to try non-parametric estimation of the bivariate survivor function. The typical works on this approach are Dabrowska (1988) and Prentice and Cai (1992). Dabrowska extended the univariate Kaplan-Meier approach by defining a bivariate hazard and using it to estimate the joint survival function. The marginals of Dabrowska’s estimate are given by the univariate Kaplan-Meier estimates. She proved the consistency of the estimates but did not give estimates for the covariance of the estimates. Prentice and Cai (1992) did not define a joint hazard. Instead, they gave a representation of the bivariate survivor function in terms of the marginal survivor functions and the covariance

$$\text{cov}\left(\int_0^{T_1} h_1(t)dt, \int_0^{T_2} h_2(t)dt\right),$$

where  $T_1$  and  $T_2$  are paired failure times and  $h_k$  is the hazard of  $T_k$ ,  $k = 1, 2$ . They proposed an estimate of this bivariate survivor function and proved the consistency of their estimate but they did not give estimates of standard errors. Lin and Ying (1993) presented a simple estimator of the bivariate survival function and also estimated the covariance function of their estimates. They assumed univariate censoring, that is, that there is one censoring time and it affects both failure times.

The other approach to bivariate survival analysis is to extend the Cox proportional hazards model, Cox (1972), to paired data. Recall the Cox proportional model has

hazard function

$$h(t|z) = h_0(t) \exp(\beta z)$$

where  $h_0$  is the baseline hazard,  $z$  is a covariate vector, and  $\beta$  is unknown. The Cox proportional model assumes that all observations are independent. Holt and Prentice (1974) assume that within a pair, hazards are proportional according to a covariate, but the baseline hazard can depend on the pair. The effect of the covariate, that is, the value of  $\beta$ , is the same across pairs.

Clayton (1978) and Oakes (1982) presented a fully parametric model for paired survival data. This model assumes that each marginal follows the Cox proportional model with respect to some covariates. They used an additional parameter to describe the association within a pair. Huster, Brookmeyer, and Self (1989) extended the Clayton-Oakes model to allow censoring. They also discussed another approach for inducing the correlation within a pair. They obtained the parameter estimates from an independence working model, that is, univariate estimates are used in the model to get the parameter estimates, and then they estimated variance robustly.

Wei, Lin, and Weissfeld (1989) proposed a model for multivariate failure times. They assumed each marginal distribution of the failure times follows a Cox proportional hazards model with respect to some covariates. They showed that the resulting estimators of the parameters for covariate effects are asymptotically jointly normal and gave a consistent estimate of the covariance matrix.

Lee, Wei, and Amato (1991) used a different way to extend the Cox proportional model to paired data. For paired data  $(T_{1i}, T_{2i})$ , with treatment indicator  $z_i$ , they showed the consistency and asymptotic normality of the estimated regression coefficient in the Cox proportional model. In this case the  $z_i$ 's might be dependent. But the corresponding variance-covariance estimate may no longer be valid due to the dependence between the members in the pairs. They proposed a correct variance-covariance estimate taking

account of the correlation within a pair.

All of the proposed models do not estimate the baseline hazard functions, treating it as a nuisance parameter.

With the development of smoothing theory and methods, many people have used splines in survival analysis, that is, to approximate density functions, survival functions, hazard functions, or baseline hazard functions, in the presence of censored data. For example, Abrahamowicz, Ciampi, and Ramsay (1992) used B-splines to estimate the density and O'Sullivan (1988) used smoothing splines to estimate log hazard functions.

Kooperberg, Stone, and Troung (1995) used linear B-splines and their tensor products to estimate the conditional log-hazard function as a function of  $t$  and a covariate vector  $z$ . Their model contains the Cox proportional hazards model as a submodel. Since cubic splines can provide a better approximation than linear splines, Kooperberg et. al. also proposed a model in which the log hazard function is estimated with cubic B-splines. Unfortunately, because of the complication in estimation and model selection, they did not consider covariates in this model. Nor did they provide standard errors for the estimated log hazard function.

In this thesis, we use cubic B-splines to estimate log hazards in univariate survival data and log hazard ratios in bivariate data. We use Kooperberg's approach to get estimates, but we provide estimates of standard errors.

In Chapter 2, we define the log hazard regression model, the estimates of the log hazard functions and ratios, and prove the asymptotic properties of the estimates. In Chapter 3, we discuss the properties of cubic B-splines functions and how to choose knots which define these functions. In Chapter 4, we use our model to analyze the data set from the Diabetic Retinopathy Study. We present our simulation study in Chapter 5.

# Chapter 2

## Hazard Regression

In this chapter we define a parametric regression model for log hazard functions of censored failure times. We start with introducing definitions and notation in the first section. The regression model and the estimates are defined in Section 2.2. We show the consistency and normality of the estimates in Section 2.3 and leave the proofs of two required lemmas to Section 2.4. Finally, in Section 2.5, we define the regression model for bivariate failure times and discuss inference based on the estimates.

### 2.1 Definitions and Notation

#### 2.1.1 Failure Time Distribution

All functions, unless stated otherwise, are defined over the interval  $R^+ \equiv [0, \infty)$ . Let  $T$  be a nonnegative continuous random variable representing the failure time of individuals in some population. Let  $f$  denote the *density function* of  $T$  and let the *distribution*

function be

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

The probability of an individual surviving till time  $t$  is given by the *survivor function*

$$S(t) = P(T \geq t) = \int_t^\infty f(u) du = 1 - F(t).$$

The *hazard function* is defined as

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)},$$

which specifies the instantaneous rate of failure at time  $t$ , given that the individual survives until at least time  $t$ .

The functions  $f, F, S$ , and  $h$  give mathematically equivalent specifications of the distribution of  $T$ . Since  $f(t) = -S'(t)$ , by the definition of  $h$ ,

$$h(t) = -\frac{d}{dt} \log S(t).$$

Thus

$$\log S(t) - \log S(0) = -\int_0^t h(u) du$$

and since  $S(0) = 1$ ,

$$S(t) = \exp \left( -\int_0^t h(u) du \right) \tag{2.1}$$

and

$$f(t) = h(t) \exp \left( -\int_0^t h(u) du \right). \tag{2.2}$$

For our log hazard regression, we denote the *log hazard function* by  $\alpha$ , that is,

$$\alpha(t) = \log h(t). \tag{2.3}$$

It is easy to derive expressions for  $S(t)$  and  $f(t)$  in terms of  $\alpha(t)$ .

### 2.1.2 Censored Survival Data

Consider  $n$  individuals in some population. Let  $T_i$  be the failure time of individual  $i$ , and  $C_i$  a *censoring time*, that is, a time beyond which individual  $i$  can not be observed. The variable  $T_i$  will be observed whenever  $T_i \leq C_i$ . Let  $X_i = \min(T_i, C_i)$  and

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

That is,  $\delta_i$  indicates whether the failure time  $T_i$  is censored or not and  $X_i$  is equal to  $T_i$  if  $T_i$  is observed, and to  $C_i$  if it is not. The data from observations on  $n$  individuals consist of the pairs  $(X_i, \delta_i)$ ,  $i = 1, \dots, n$ . For the data we first make the following assumptions.

#### Assumptions

- (1)  $(X_i, \delta_i)$  are independent random vectors with an identical distribution;
- (2) the failure times  $T_i$  and the censoring times  $C_i$  are independent.

We denote the survivor functions and the density functions of failure time  $T_i$  and the censoring times  $C_i$  by  $S, S_c, f$  and  $f_c$ , respectively. By Lawless (1982), under the Assumptions (1) and (2),

$$P(X < t, \Delta = \delta) = \begin{cases} \int_0^t S(u) f_c(u) du & \delta = 0, \\ \int_0^t f(u) S_c(u) du & \delta = 1. \end{cases}$$

The probability is a measure defined on  $[0, \infty) \times \{0, 1\}$ . We denote it by  $\nu$ . Then the density function of  $(X_i, \delta_i)$  is defined by

$$f(t, \delta) = f(t)^\delta S(t)^{1-\delta} f_c(t)^{1-\delta} S_c(t)^\delta. \quad (2.4)$$

It is easy to check that  $f$  is the Radon-Nikodym derivative of  $\nu$  with respect to the  $\sigma$ -finite cross-product measure  $\mu$  defined on  $[0, \infty) \times \{0, 1\}$  such that the marginal of  $\mu$  on  $[0, \infty)$  is the Lebesgue measure and the marginal on  $\{0, 1\}$  is counting measure.

## 2.2 Log Hazard Regression

In this section we fit a parametric model for the log hazard function  $\alpha$  of failure time  $T$ . First we define the model in a general form and determine the estimates that we will use. Then, in Section 2.2.2, we discuss the conditions on the model.

### 2.2.1 Model and Estimates

Let  $G$  be a  $p$ -dimensional linear space of real continuous functions defined on  $[0, \infty)$ , and let  $\{b_j, j = 1, \dots, p\}$  be a basis of  $G$ . We call  $G$  the *regression space*. The *log hazard regression* is a model for the log hazard function, that is,

$$\alpha(t|\boldsymbol{\beta}) = \sum_{j=1}^p \beta_j b_j(t) = \boldsymbol{\beta}^T \mathbf{b}(t), \quad (2.5)$$

where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  is unknown and  $\mathbf{b}(t) = (b_1(t), \dots, b_p(t))^T$ .

Given  $\boldsymbol{\beta}$ , by Equations (2.3), (2.1), and (2.2), we have the hazard function, survivor function, and density function

$$h(t|\boldsymbol{\beta}) = \exp \left( \sum_{j=1}^p \beta_j b_j(t) \right), \quad (2.6)$$

$$S(t|\boldsymbol{\beta}) = \exp \left( - \int_0^t \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du \right), \quad (2.7)$$

$$f(t|\boldsymbol{\beta}) = \exp \left( \sum_{j=1}^p \beta_j b_j(t) \right) \exp \left\{ - \int_0^t \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du \right\}, \quad (2.8)$$

respectively.

Note that to ensure that (2.7) defines a non-degenerate survivor function, it is necessary and sufficient that

$$\int_0^t \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du < \infty \quad (2.9)$$

for some  $t > 0$  and

$$\int_0^\infty \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du = +\infty. \quad (2.10)$$



So we need some conditions on  $\mathbf{b}$  and  $\boldsymbol{\beta}$ . For now we assume that (2.9) and (2.10) hold, and we will postpone discussing the conditions until Section 2.2.2.

For  $n$  observations  $(X_i, \delta_i), i = 1, \dots, n$ , we make the following assumptions, in addition to Assumptions (1) and (2).

### Assumptions

(3) the log hazard function  $\alpha$  of  $T_i$  satisfies model (2.5);

(4) the distribution of the censoring times  $C_i$  does not involve  $\boldsymbol{\beta}$  and the density  $f_C$  of  $C$  is bounded.

Then by (2.4), the density function of  $(X_i, \delta_i)$  is

$$f(t, \delta | \boldsymbol{\beta}) = f(t | \boldsymbol{\beta})^\delta S(t | \boldsymbol{\beta})^{1-\delta} f_c(t)^{1-\delta} S_c(t)^\delta. \quad (2.11)$$

We denote the joint distribution function of  $(X_i, \delta_i)$  by  $L(\mathbf{X} | \boldsymbol{\beta})$ , that is,

$$L(\mathbf{X} | \boldsymbol{\beta}) = \prod_{i=1}^n f(X_i, \delta_i | \boldsymbol{\beta}),$$

and we choose  $\hat{\boldsymbol{\beta}}$  to maximize  $L$ .

Let

$$f_q(t, \delta | \boldsymbol{\beta}) = f(t | \boldsymbol{\beta})^\delta S(t | \boldsymbol{\beta})^{1-\delta} \quad (2.12)$$

and

$$L_q(\boldsymbol{\beta} | \mathbf{X}) = \prod_{i=1}^n f_q(X_i, \delta_i | \boldsymbol{\beta}).$$

Then

$$\begin{aligned} L(\mathbf{X} | \boldsymbol{\beta}) &= \prod_{i=1}^n f_q(X_i, \delta_i | \boldsymbol{\beta}) f_c(X_i)^{1-\delta_i} S_c(X_i)^{\delta_i} \\ &= L_q(\boldsymbol{\beta} | \mathbf{X}) \prod_{i=1}^n f_c(X_i)^{1-\delta_i} S_c(X_i)^{\delta_i}. \end{aligned}$$

Since  $S_c$  and  $f_c$  do not involve  $\boldsymbol{\beta}$ , we estimate  $\boldsymbol{\beta}$  by  $\hat{\boldsymbol{\beta}}$ , the maximizer of the function  $L_q$ .

We call the partial derivative of  $\log L_q$  with respect to  $\boldsymbol{\beta}$  the *score function*, denoted by  $\mathbf{U}^{(n)}(\boldsymbol{\beta}|\mathbf{X})$ , which is a vector with  $p$  components

$$\mathbf{U}^{(n)}(\boldsymbol{\beta}|\mathbf{X})_m = \frac{\partial}{\partial \beta_m} \log L_q(\boldsymbol{\beta}|\mathbf{X}) = \sum_{i=1}^n \frac{\partial}{\partial \beta_m} \log f_q(X_i, \delta_i|\boldsymbol{\beta}), \quad (2.13)$$

$m = 1, 2, \dots, p$ . The *observed information matrix*, denoted by  $I^{(n)}(\boldsymbol{\beta}|\mathbf{X})$ , is a  $p \times p$  matrix with  $(ml)$ -th entry

$$I^{(n)}(\boldsymbol{\beta}|\mathbf{X})_{ml} = -\frac{\partial^2}{\partial \beta_m \partial \beta_l} \log L_q(\boldsymbol{\beta}|\mathbf{X}) = -\sum_{i=1}^n \frac{\partial^2}{\partial \beta_m \partial \beta_l} \log f_q(X_i, \delta_i|\boldsymbol{\beta}). \quad (2.14)$$

We denote the *information matrix*, in the usual way, by  $I(\boldsymbol{\beta})$ , which is a  $p \times p$  matrix with the  $(ml)$ -th entry

$$I(\boldsymbol{\beta})_{ml} = -E \left\{ \frac{\partial^2}{\partial \beta_m \partial \beta_l} \log f_q(X_1, \delta_1|\boldsymbol{\beta}) \right\}. \quad (2.15)$$

By definition of  $f_q$  and Equations (2.8) and (2.7), we can write  $f_q, \log f_q, \mathbf{U}^{(n)}, I^{(n)}$ , and  $I(\boldsymbol{\beta})$  in terms of the basis functions  $b_1, \dots, b_p$ . We have

$$f_q(t, \delta|\boldsymbol{\beta}) = \left[ \exp \left( \sum_{j=1}^p \beta_j b_j(t) \right) \right]^\delta \exp \left\{ - \int_0^t \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du \right\}; \quad (2.16)$$

and

$$\log f_q(t, \delta|\boldsymbol{\beta}) = \delta \sum_{j=1}^p \beta_j b_j(t) - \int_0^t \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du. \quad (2.17)$$

Thus, under regularity conditions,

$$\begin{aligned} & \frac{\partial}{\partial \beta_m} \log f_q(X_i, \delta_i|\boldsymbol{\beta}) \\ &= \frac{\partial}{\partial \beta_m} \left\{ \delta_i \sum_{j=1}^p \beta_j b_j(X_i) - \int_0^{X_i} \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du \right\} \\ &= \delta_i b_m(X_i) - \int_0^{X_i} b_m(u) \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du; \end{aligned}$$

and

$$\frac{\partial^2}{\partial \beta_m \partial \beta_l} \log f_q(X_i, \delta_i|\boldsymbol{\beta}) = - \int_0^{X_i} b_m(u) b_l(u) \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du.$$

Hence

$$U^{(n)}(\beta|\mathbf{X})_m = \sum_{i=1}^n \left\{ \delta_i b_m(X_i) - \int_0^{X_i} b_m(u) \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du \right\} \quad (2.18)$$

and

$$I^{(n)}(\beta|\mathbf{X})_{ml} = \sum_{i=1}^n \int_0^{X_i} b_m(u) b_l(u) \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du. \quad (2.19)$$

Since for any vector  $\mathbf{v} = (v_1, \dots, v_p)^T \neq \mathbf{0}$ ,

$$\mathbf{v}^T I^{(n)}(\beta|\mathbf{X}) \mathbf{v} = \sum_{i=1}^n \int_0^{X_i} (\mathbf{v}^T \mathbf{b}(u))^2 \exp \left( \sum_{j=1}^p \beta_j b_j(u) \right) du > 0,$$

$I^{(n)}(\beta|\mathbf{X})$  is positive definite (i.e., the second partial derivative of  $\log L_q$  is negative definite). Therefore, if  $\hat{\beta}$  exists, it is the unique solution of the  $p$  equations

$$U^{(n)}(\beta|\mathbf{X}) = \mathbf{0}. \quad (2.20)$$

In the next chapter, we will show how to numerically solve (2.20) for  $\hat{\beta}$ .

**Note:** By the relationship between  $L_q$  and  $L$ ,

$$\frac{\partial}{\partial \beta_m} \log L(\beta|\mathbf{X}) = \frac{\partial}{\partial \beta_m} \log L_q(\beta|\mathbf{X}).$$

Therefore,

$$U^{(n)}(\beta|\mathbf{X})_m = \sum_{i=1}^n \frac{\partial}{\partial \beta_m} \log f(X_i, \delta_i|\beta)$$

and

$$I^{(n)}(\beta|\mathbf{X})_{ml} = - \sum_{i=1}^n \frac{\partial^2}{\partial \beta_m \partial \beta_l} \log f(X_i, \delta_i|\beta)$$

which are the usual definitions of the score function and the observed information matrix.

### 2.2.2 Conditions on $\mathbf{b}$ and $\beta$

As we mentioned in the last section, we need some conditions on the basis  $\mathbf{b}$  or on  $\beta$  such that (2.9) and (2.10) are true. A sufficient condition that (2.9) and (2.10) hold is that  $\mathbf{b}$  is *bounded*. We say a function  $\mathbf{u} : R^+ \rightarrow R^p$  is bounded if there exists  $0 < M < \infty$  such that

$$\|\mathbf{u}(t)\|^2 \equiv \sum_{j=1}^p |u_j(t)|^2 < M,$$

for all  $t \in R^+$ .

**Proposition 2.2.1** *If  $\mathbf{b}$  is bounded, then*

- (i) (2.9) and (2.10) hold for all  $t$  and  $\beta$  and
- (ii) there exists a number  $M(\beta, p)$  such that

$$f(t|\beta) \leq M(\beta, p) \exp\left(-\frac{t}{M(\beta, p)}\right). \quad (2.21)$$

**Proof.** Since  $\mathbf{b}$  is bounded, there exists a number  $0 < M < \infty$ , dependent on  $\beta$  and  $p$ , such that  $|\sum_{j=1}^p \beta_j b_j(u)| < M$ . Hence

$$\frac{1}{\exp(M)} \leq \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) \leq \exp(M). \quad (2.22)$$

It follows that for any  $t$

$$\int_0^t \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) du \leq \int_0^t \exp(M) du < \infty$$

and

$$\int_0^\infty \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) du \geq \int_0^\infty \frac{1}{\exp(M)} du = \infty.$$

So we proved that (2.9) and (2.10) are true.

Next, let  $M(\beta, p) = \exp(M)$ , then, by the definition of  $f$  from (2.8) and from (2.22),

$$\begin{aligned} f(t|\beta) &= \exp\left(\sum_{j=1}^p \beta_j b_j(t)\right) \exp\left\{-\int_0^t \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) du\right\} \\ &\leq M(\beta, p) \exp\left(-\frac{t}{M(\beta, p)}\right), \end{aligned}$$

which completes the proof of (ii).  $\square$

We would like the regression space  $G$  to contain the log hazard for the commonly used Weibull distribution. Recall that the density function of a Weibull distribution is

$$f_w(t) = \gamma \lambda^\gamma t^{\gamma-1} \exp\left(-(\lambda t)^\gamma\right), \quad t > 0, \lambda > 0, \gamma > 1.$$

The corresponding log hazard function is

$$\alpha_w(t) = \log(\gamma \lambda^\gamma) + (\gamma - 1) \log(t).$$

Thus, if  $b_p(t) = \log(t)$  and  $b_{p-1} \equiv 1$ , then model (2.5) includes the Weibull distributions.

More generally, we consider the basis  $b_1, \dots, b_{p-1}$  bounded and  $b_p(t) = \log(t + c)$ , where  $c \geq 0$  is a constant. We call  $b_p$  *log tail* and  $\{b_j, j = 1, \dots, p\}$  *bounded plus log tail*. This basis doesn't satisfy the assumption of Proposition 2.2.1. But we have the following result.

**Proposition 2.2.2** *If  $\mathbf{b}$  is bounded plus log tail and  $\beta_p > -1$ , then*

(i) *(2.9) and (2.10) hold for all  $t$  and*

(ii) *there exists a number  $M(\boldsymbol{\beta}, p)$  such that*

$$f(t|\boldsymbol{\beta}) \leq M(\boldsymbol{\beta}, p)(t + c)^{\beta_p} \exp\left(-\frac{(t + c)^{\beta_p+1}}{M(\boldsymbol{\beta}, p)}\right) \quad (2.23)$$

**Proof.** Since  $\mathbf{b}$  is bounded plus log tail, there exists  $M$  such that  $\left|\sum_{j=1}^{p-1} \beta_j b_j(u)\right| < \log(M)$ . Then

$$\frac{(u + c)^{\beta_p}}{M} \leq \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) \leq M(u + c)^{\beta_p}. \quad (2.24)$$

Therefore,

$$\int_0^t \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) du \leq \int_0^t M(u + c)^{\beta_p} du < \infty$$

for any  $t$  and

$$\int_0^\infty \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) du \geq \int_0^\infty \frac{(u+c)^{\beta_p}}{\exp(M)} du = \infty$$

as  $\beta_p > -1$ . Then (i) is proved. To prove (ii), using the definition of  $f$  from (2.8) and using (2.24), we have

$$\begin{aligned} f(t|\boldsymbol{\beta}) &= \exp\left(\sum_{j=1}^p \beta_j b_j(t)\right) \exp\left\{-\int_0^t \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) du\right\} \\ &\leq M(t+c)^{\beta_p} \exp\left(\frac{c^{\beta_p+1}}{(\beta_p+1)M}\right) \exp\left(-\frac{(t+c)^{\beta_p+1}}{(\beta_p+1)M}\right) \\ &\equiv C_1(t+c)^{\beta_p} \exp\left(-\frac{(t+c)^{\beta_p+1}}{C_2}\right) \\ &\leq M(\boldsymbol{\beta}, p)(t+c)^{\beta_p} \exp\left(-\frac{(t+c)^{\beta_p+1}}{M(\boldsymbol{\beta}, p)}\right), \end{aligned}$$

where  $M(\boldsymbol{\beta}, p) = \max\{C_1, C_2\}$ . So (ii) is true.  $\square$

From now on, unless stated otherwise, we assume the  $n$  observations  $(X_i, \delta_i)$ ,  $i = 1, \dots, n$ , satisfy Assumptions (1) - (4) defined in the previous sections and either

### Assumption

(5a)  $\mathbf{b}$  is bounded and  $\boldsymbol{\beta} \in R^p$

or

(5b)  $\mathbf{b}$  is bounded plus log tail and  $\boldsymbol{\beta} \in R^{p-1} \times (-1, \infty)$ .

We call  $R^p$  the *parameter space*, if (5a) holds and  $R^{p-1} \times (-1, \infty)$  the *parameter space* if (5b) holds.

**Note 1** Since (2.9) holds for all  $t > 0$ ,  $f(t) > 0$  for all  $t > 0$ . So the support of  $f$  is  $R^+$ , which will be used for the proof of Theorem 2.3.1 in the next section.

**Note 2** It is easy to see that if  $\mathbf{b}$  is bounded then for any compact set  $B$  in the parameter space, there exists  $M$  dependent on  $B$  and  $p$  such that

$$f(t|\boldsymbol{\beta}) \leq M \exp\left(-\frac{t}{M}\right)$$

for all  $\beta \in B$ . If  $\mathbf{b}$  is bounded plus log tail, then for any compact set  $B$  in the parameter space, there exist  $M$  and  $\beta'_p > -1$ , dependent on  $B$  and  $p$ , such that

$$f(t|\beta) \leq M(t+c)^{\beta'_p} \exp\left(-\frac{(t+c)^{\beta'_p+1}}{M}\right)$$

for any  $\beta \in B$ . That is, on a compact set  $B$  in the parameter space, the density function  $f(t|\beta)$  of  $T$  can be dominated by a function of  $t$ .

## 2.3 Consistency and Normality of $\hat{\beta}$

In this section we use Theorem 5.1 in Lehmann and Casella (1998) to show the following theorem.

**Theorem 2.3.1** *Let  $(X_1, \delta_1), \dots, (X_n, \delta_n)$  satisfy Assumptions (1) - (5). Then, for  $\beta_0$ , the true parameter,*

- (i)  $\hat{\beta} \rightarrow \beta_0$  in probability;
- (ii)  $\sqrt{n}(\hat{\beta} - \beta_0) \implies N(\mathbf{0}, [I(\beta_0)]^{-1})$ ;
- (iii)  $\frac{1}{n}I^{(n)}(\beta|\mathbf{X}) \rightarrow I(\beta_0)$  in probability provided  $\beta \rightarrow \beta_0$  in probability.

We first state the conditions of their theorem.

### Conditions

- A. For different values of  $\beta$ , the distributions  $P_\beta$  of the observations are distinct.
- B. The distributions  $P_\beta$  have common support.
- C. The observations  $(X_1, \delta_1), \dots, (X_n, \delta_n)$  are i.i.d. with probability density  $f(t, \delta|\beta)$  with respect to a  $\sigma$ -finite measure  $\mu$ .

D. There exists an open set  $\omega$  of the parameter space  $\Omega$  containing the true parameter point  $\beta_0$  such that for almost all  $(X, \delta)$ , the density  $f(X, \delta|\beta)$  admits all third derivatives  $(\partial^3/\partial\beta_j\partial\beta_k\partial\beta_l)f(X, \delta|\beta)$  for all  $\beta \in \omega$ .

E. The first and second logarithmic derivatives of  $f$  satisfy the equations

$$E\left\{\frac{\partial}{\partial\beta} \log f(X, \delta|\beta)\right\} = 0,$$

and

$$I(\beta)_{ml} = E\left\{-\frac{\partial^2}{\partial\beta_m\partial\beta_l} \log f(X, \delta|\beta)\right\} = E\left\{\frac{\partial}{\partial\beta_m} \log f(X, \delta|\beta) \cdot \frac{\partial}{\partial\beta_l} \log f(X, \delta|\beta)\right\},$$

$$l, m \in \{1, \dots, p\}.$$

F.  $I(\beta)_{jk}$  is finite and the matrix  $I(\beta)$ , a  $p \times p$  matrix with the  $(ml)$ -th entry  $I(\beta)_{ml}$ , is positive definite for all  $\beta \in \omega$ .

G. There exist functions  $M_{jkl}$  such that

$$\left|(\partial^3/\partial\beta_j\partial\beta_k\partial\beta_l) \log f(X, \delta|\beta)\right| \leq M_{jkl}(X)$$

for all  $\beta \in \omega$ , where  $E_{\beta_0}(M_{jkl}(X)) < \infty$  for all  $j, k, l \in \{1, \dots, p\}$ .

**Theorem 2.3.2** (Theorem 5.1 in Lehmann, E.L. and Casella, G, 1998). Let  $(X_1, \delta_1), \dots, (X_n, \delta_n)$  be i.i.d., each with a density  $f(t, \delta|\beta)$  which satisfies the above conditions. Then with probability tending to 1 and as  $n \rightarrow \infty$ , there exists a solution  $\hat{\beta}$  of the likelihood equation

$$\sum_{i=1}^n \frac{\partial \log f(X_i, \delta_i|\beta)}{\partial\beta} = 0$$

such that

(a)  $\hat{\beta}_j$  is consistent for estimating  $\beta_j$ ,

(b)  $\sqrt{n}(\hat{\beta} - \beta)$  is asymptotically normal with mean zero and covariance matrix  $[I(\beta)]^{-1}$ ,

and



(c)  $\hat{\beta}_j$  is asymptotically efficient in the sense that  $\sqrt{n}(\hat{\beta}_j - \beta_j)$  converges in distribution to a random normal variable with mean zero and variance  $[I(\boldsymbol{\beta})_{jj}]^{-1}$ .

**Proof of Theorem 2.3.1.**

Suppose  $\hat{\boldsymbol{\beta}}$  is the solution of the equation

$$0 = \sum_{i=1}^n \frac{\partial \log(f_q(t, \delta | \boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{\partial \log(f(t, \delta | \boldsymbol{\beta}))}{\partial \boldsymbol{\beta}}.$$

To show (i) and (ii), we need only check the conditions of Theorem 2.3.2.

First, by Assumptions 1 and 2 and the definition of  $f(\cdot, \cdot | \boldsymbol{\beta})$  from (2.11),  $(X_1, \delta_1), \dots, (X_n, \delta_n)$  and  $f(\cdot, \cdot | \boldsymbol{\beta})$  satisfy condition C.

Second, since  $\{b_1, \dots, b_p\}$  is a basis, from (2.16), the definition of  $f_q$  in terms of the basis functions, for different value of  $\boldsymbol{\beta}$ ,  $f_q$  is distinct. From Note 1, after Propositions 2.2.1 and 2.2.2, the support of  $f_q$  is  $[0, \infty)$ . By the relationship between  $f(\cdot, \cdot | \boldsymbol{\beta})$  and  $f_q$  from (2.11) and (2.12), conditions A and B hold.

Then, from (2.11) and (2.12),

$$\frac{\partial \log f(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \beta_l} = \frac{\partial \log f_q(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \beta_l}, \quad (2.25)$$

$l = 1, \dots, p$ , since  $S_c$  and  $f_c$  do not involve  $\boldsymbol{\beta}$ . Hence condition D follows from a direct calculation of the derivatives of  $f_q$  from the definition (2.16).

Next for condition F,  $I(\boldsymbol{\beta})_{jk}$  is finite follows from Lemma 2.3.4 below. As shown in the calculations after (2.19),  $I(\boldsymbol{\beta})$  is positive definite.

Finally, conditions E and G follow from Lemmas 2.3.4 and 2.3.5 below, respectively.

Therefore, (a) and (b) in Theorem 2.3.2 yield (i) and (ii), respectively.

To prove (iii), we need to use (i) and the bounds in Lemma 2.3.4 below. The proof is as in Theorem 3.10 of Lehmann and Cassella (1998). Thus, we have finished proving Theorem 2.3.1.

Using (ii) and (iii) in the above theorem and Slutsky's theorem, we have the following Corollary.

**Corollary 2.3.3** Let  $(X_1, \delta_1), \dots, (X_n, \delta_n)$  satisfy Assumptions (1) - (5). Then

$$[nI^{(n)}(\hat{\beta}|\mathbf{X})]^{\frac{1}{2}}(\hat{\beta} - \beta) \Rightarrow N(\mathbf{0}, I),$$

where  $I$  is the  $p \times p$  identity matrix.

The following two lemmas are used in the proof of Theorem 2.3.1.

**Lemma 2.3.4** Suppose that  $\mathbf{b}$  satisfies Assumption (5a) (or Assumption (5b)) and  $B$  is a compact set in  $R^p$  (or in  $R^{p-1} \times (-1, \infty)$ ). Then there exists  $0 < M < \infty$ , dependent on  $B$  and  $p$  (or on  $B, p$ , and  $c$ ), such that for any  $\beta \in B$ ,  $s_k \in \{1, \dots, p\}; k = 1, \dots, m; m = 0, 1, 2, 3$ ,

$$\left| \frac{\partial^m \log f_q(X_1, \delta_1 | \beta)}{\partial \beta_{s_1} \cdots \partial \beta_{s_m}} \right| < MK(X_1), \quad (2.26)$$

where

$$K(X_1) = \begin{cases} 1 + X_1 & \text{if } \mathbf{b} \text{ is bounded} \\ 1 + |\log(X_1 + c)| + (X_1 + c)^{\beta_p + 4} & \text{if } \mathbf{b} \text{ is bounded plus log tail.} \end{cases} \quad (2.27)$$

Furthermore,  $E\{K(X_1)\} < \infty$ .

**Lemma 2.3.5** Let  $\beta_0$  be the true values of the parameters in model (2.5). Then

$$\begin{aligned} \text{i) } & E \left\{ \frac{\partial \log f_q(X_1, \delta_1 | \beta)}{\partial \beta} \Big|_{\beta_0} \right\} = \mathbf{0}; \\ \text{ii) } & E \left\{ \frac{\partial \log f_q(X_1, \delta_1 | \beta)}{\partial \beta_m} \Big|_{\beta_0} \frac{\partial \log f_q(X_1, \delta_1 | \beta)}{\partial \beta_l} \Big|_{\beta_0} \right\} = -E \left\{ \frac{\partial^2 \log f_q(X_1, \delta_1 | \beta)}{\partial \beta_m \partial \beta_l} \Big|_{\beta_0} \right\}. \end{aligned}$$

We will give the proofs of the above lemmas in the next section.

## 2.4 Proofs of the Lemmas

Now we prove Lemmas 2.3.4 and 2.3.5.

For convenience, we first calculate the partial derivatives of  $\log f_q$  with respect to  $\beta$  in terms of the basis functions. By the definition of  $\log f_q$  from (2.17), for  $s_k \in \{1, \dots, p\}; k = 1, \dots, m$ ,

$$\frac{\partial^m \log f_q(X_1, \delta_1 | \beta)}{\partial \beta_{s_1} \cdots \partial \beta_{s_m}} = \begin{cases} \delta_1 \sum_{j=1}^p \beta_j b_j(X_1) - \int_0^{X_1} \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) du, & m = 0; \\ \delta_1 b_{s_1}(X_1) - \int_0^{X_1} b_{s_1}(u) \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) du, & m = 1; \\ -\int_0^{X_1} b_{s_1}(u) \cdots b_{s_m}(u) \exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) du, & m = 2, 3. \end{cases} \quad (2.28)$$

#### Proof of Lemma 2.3.4

If  $\mathbf{b}$  is bounded, since  $B$  is a compact set,  $\sum_{j=1}^p |\beta_j b_j(u)|$  is bounded on  $B$ , and so is  $\exp\left(\sum_{j=1}^p \beta_j b_j(u)\right)$ . Hence, by (2.28), we can find  $0 < M < \infty$ , dependent on  $B$  and  $p$ , such that (2.26) holds.

If  $\mathbf{b}$  is bounded plus log tail, then  $b_1, \dots, b_{p-1}$  are bounded. Hence there exists  $1 < M_1 < \infty$  such that for  $\beta \in B$

$$\max \left\{ \left| \sum_{j=1}^{p-1} \beta_j b_j(u) \right|, \sum_{j=1}^{p-1} |b_j(u)| \right\} < M_1, \quad (2.29)$$

which implies

$$\exp\left(\sum_{j=1}^p \beta_j b_j(u)\right) < e^{M_1}(u+c)^{\beta_p}. \quad (2.30)$$

Then for  $m = 0$ , by (2.28) and (2.30), (2.26) is true. To prove (2.26) for  $m = 1, 2, 3$ , it suffices to show that there exists  $M < \infty$  such that

$$\int_0^{X_1} |b_{s_1}(u) \cdots b_{s_m}(u)| (u+c)^{\beta_p} du < M(1 + (X_1 + c)^{\beta_p + 4}), \quad (2.31)$$

$s_k \in \{1, \dots, p\}; k = 1, \dots, m$ . This clearly holds for  $s_k \in \{1, \dots, p-1\}$ , since  $b_j$  is bounded,  $j = 1, \dots, p-1$ . But the log tail  $\log(t+c)$  is unbounded, which makes the proof of (2.31) more complicated. So we need to investigate the properties of the integral

$$\int_0^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du,$$

$m = 1, 2, 3$ . Noting  $\beta_p > -1$  and  $c \geq 0$ , we have the following easily proven facts.

**Fact 1** for any  $\ell > 0$  and any compact set  $B \subset (-1, \infty)$ , there exists  $M_1 < \infty$  such that

$$\sup_{u \in (0,1], \beta_p \in B} |\log(u+c)|^\ell (u+c)^{\frac{\beta_p+1}{2}} < M_1;$$

**Fact 2** for any compact set  $B \subset (-1, \infty)$ ,  $\sup_{\beta_p \in B} \int_0^1 (u+c)^{\frac{\beta_p-1}{2}} du < \infty$ ;

**Fact 3**  $\log(u+c) < u+c$  for  $u \in [1, \infty)$ ,

**Fact 4** for any  $\beta' > 0, \beta'' > -1$ , and  $M' > 0$ ,

$$\int_0^\infty (u+c)^{\beta''} \exp \left\{ -(u+c)^{\beta'}/M' \right\} du < \infty.$$

We will use **Facts 1 - 3** to prove that

i) there exists  $M_2 < \infty$  such that for  $\beta \in B$ ,

$$\int_0^1 |\log(u+c)|^m (u+c)^{\beta_p} du < M_2,$$

$$m = 1, 2, 3;$$

ii) there exists  $M_3 < \infty$  such that for  $X_1 > 1$  and  $\beta \in B$ ,

$$\int_1^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du < M_3 (X_1 + c)^{\beta_p+4},$$

$$m = 1, 2, 3.$$

If i) and ii) hold, then, when  $X_1 \leq 1$ ,

$$\begin{aligned} & \int_0^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du \\ & \leq \int_0^1 |\log(u+c)|^m (u+c)^{\beta_p} du < M_2, \end{aligned}$$

and when  $X_1 > 1$ ,

$$\begin{aligned} & \int_0^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du \\ & = \int_0^1 |\log(u+c)|^m (u+c)^{\beta_p} du + \int_1^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du \\ & \leq M_2 + M_3 (X_1 + c)^{\beta_p+4} \leq M(1 + (X_1 + c)^{\beta_p+4}) \end{aligned}$$

for  $M = \max\{M_2, M_3\}$ . So (2.31) would hold, which implies that (2.26) would be true.

Proof of i): Since  $B$  is compact, using **Facts 1** and **2**, we can find  $M'_2, M''_2 < \infty$  such that for  $\beta \in B$ ,

$$|\log(u+c)|^m (u+c)^{\frac{\beta_p+1}{2}} < M'_2,$$

$m = 1, 2, 3$ , for  $u \in (0, 1)$  and

$$\int_0^1 (u+c)^{\frac{\beta_p-1}{2}} du \leq M''_2.$$

Then for  $\beta \in B$ ,

$$\begin{aligned} & \int_0^1 |\log(u+c)|^m (u+c)^{\beta_p} du \\ & \leq \int_0^1 |\log(u+c)|^m (u+c)^{\frac{\beta_p+1}{2}} (u+c)^{\frac{\beta_p-1}{2}} du \\ & \leq M'_2 \int_0^1 (u+c)^{\frac{\beta_p-1}{2}} du \leq M'_2 M''_2 = M_2 < \infty, \end{aligned}$$

$m = 1, 2, 3$ .

Proof of ii): By **Fact 3** and (2.29),

$$\begin{aligned} & \int_1^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du \\ & \leq \int_1^{X_1} (u+c)^3 (u+c)^{\beta_p} du \\ & = \frac{1}{\beta_p+4} \left( (X_1+c)^{\beta_p+4} - (1+c)^{\beta_p+4} \right) \\ & \leq M_3 (X_1+c)^{\beta_p+4} \end{aligned}$$

$m = 1, 2, 3$ , for some  $M_3 > 0$ .

Now that we have proven that (2.26) is true, we show that  $E\{K(X_1)\} < \infty$ .

First suppose that  $\mathbf{b}$  is bounded. Then

$$E\{K(X_1)\} = E\{1 + X_1\} \leq E\{(1 + T_1)\},$$

since, by definition,  $X_1 \leq T_1$ . By Proposition 2.4.1 below,  $E\{T_1\} < \infty$ .

Now consider  $\mathbf{b}$  bounded plus log tail. Write

$$K(X_1) = K^*(X_1) + K^{**}(X_1),$$

where

$$K^*(X_1) = 1 + (X_1 + c)^{\beta_p + 4} + |\log(X_1 + c)|I\{X_1 + c \geq 1\}$$

and

$$K^{**}(X_1) = |\log(X_1 + c)|I\{X_1 + c < 1\}.$$

Since  $K^*$  is a non-decreasing function,

$$E\{K^*(X_1)\} \leq E\{K^*(T_1)\} < E\{K(T_1)\},$$

which is finite by Proposition 2.4.2 below. To show  $E(K^{**})$  is bounded, note that, by Assumption 4,  $f_c$  is bounded, and by (ii) in Proposition 2.2.2,  $f(t|\boldsymbol{\beta})$  is bounded on  $[0, 1]$ . Thus, the marginal density of  $X_1$ ,

$$f_{X_1}(x) = f(x|\boldsymbol{\beta})S_C(x) + S(x|\boldsymbol{\beta})f_C(x),$$

is bounded. Then we have

$$E\{K^{**}(X_1)\} \leq \int_0^1 |\log(x + c)|f_{X_1}(x)dx < \infty$$

since  $\int_0^1 |\log(x + c)|dx < \infty$ . Therefore,

$$E\{K(X_1)\} = E\{K^*(X_1)\} + E\{K^{**}(X_1)\} < \infty,$$

which completes the proof of Lemma 2.3.4.

The following propositions were used in the proof of Lemma 2.3.4.

**Proposition 2.4.1** *If  $\mathbf{b}$  is bounded, then for any  $a > 0$ ,  $E(T^a) < \infty$ .*

**Proof.** Since  $\mathbf{b}$  is bounded, by (ii) in Proposition 2.2.1, there exists  $0 < M(\boldsymbol{\beta}, p) < \infty$  such that

$$f(t|\boldsymbol{\beta}) \leq M(\boldsymbol{\beta}, p) \exp\left(-\frac{t}{M(\boldsymbol{\beta}, p)}\right).$$

From this and **Fact 4** in the proof of Lemma 2.3.4, we have

$$E(T^a) = \int_0^\infty t^a f(t|\boldsymbol{\beta}) dt \leq \int_0^\infty M(\boldsymbol{\beta}, p) t^a \exp\left(\frac{-t}{M(\boldsymbol{\beta}, p)}\right) dt < \infty.$$

□

**Proposition 2.4.2** *If  $\mathbf{b}$  is bounded plus log tail and  $\beta_p > -1$ , then*

(i) *for any  $a > 0$ ,  $E(T^a) < \infty$ , and*

(ii)  *$E(|\log(T + c)|) < \infty$ .*

**Proof.** Since  $b_1, \dots, b_{p-1}$  are bounded, by (ii) in Proposition 2.2.2, there exists  $0 < M(\boldsymbol{\beta}, p) < \infty$  such that (2.23) hold. Let

$$g(t, \boldsymbol{\beta}) = \exp\left(-\frac{(t + c)^{\beta_p + 1}}{M(\boldsymbol{\beta}, p)}\right).$$

Then from (2.23) and **Fact 4** in the proof of Lemma 2.3.4,

$$E(T^a) = \int_0^\infty t^a f(t|\boldsymbol{\beta}) dt \leq M(\boldsymbol{\beta}, p) \int_0^\infty t^a (t + c)^{\beta_p} g(t, \boldsymbol{\beta}) dt < \infty,$$

as  $\beta_p > -1$ . Thus (i) is proved.

To prove (ii), using (2.23) and **Facts 1 - 4** in the proof of Lemma 2.3.4, we have

$$\begin{aligned} E(|\log(T + c)|) &= \int_0^\infty |\log(t + c)| f(t|\boldsymbol{\beta}) dt \\ &\leq M(\boldsymbol{\beta}, p) \int_0^\infty |\log(t + c)| (t + c)^{\beta_p} g(t, \boldsymbol{\beta}) dt \\ &\leq M(\boldsymbol{\beta}, p) \left\{ \int_0^1 |\log(t + c)| (t + c)^{\frac{\beta_p + 1}{2}} (t + c)^{\frac{\beta_p - 1}{2}} g(t, \boldsymbol{\beta}) dt + \int_1^\infty (t + c)^{\beta_p + 1} g(t, \boldsymbol{\beta}) dt \right\} \\ &< M(\boldsymbol{\beta}, p) \left\{ M_1 \int_0^1 (t + c)^{\frac{\beta_p - 1}{2}} g(t, \boldsymbol{\beta}) dt + \int_1^\infty (t + c)^{\beta_p + 1} g(t, \boldsymbol{\beta}) dt \right\} < \infty, \end{aligned}$$

which finishes the proof of (ii). □

**Remark:** The following proposition will be used in the proof of Lemma 2.3.5 and Section 2.5. The proof is omitted since it uses the same procedures and arguments as in the proofs of Propositions 2.4.1 and 2.4.2.

**Proposition 2.4.3** *Suppose that  $\mathbf{b}$  satisfies the assumptions of Lemma 2.3.4. Then  $E\{[K(X_1)]^2\} < \infty$ .*

**The proof of Lemma 2.3.5**

To prove i), write

$$\begin{aligned} E \left\{ \frac{\partial \log f_q(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}_0} \right\} &= E \left\{ \frac{\partial \log f(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}_0} \right\} \\ &= \int \frac{\partial f(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}_0} d\mu = \left( \frac{\partial}{\partial \boldsymbol{\beta}} \int f(X_1, \delta_1 | \boldsymbol{\beta}) d\mu \right) \Big|_{\boldsymbol{\beta}_0} = \frac{\partial E(\mathbf{1})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}_0} = \mathbf{0}. \end{aligned}$$

To justify the interchange of differentiation and integration, we must prove that there exists a neighborhood  $B$  of  $\boldsymbol{\beta}_0$  such that uniformly on  $B$ , the partial derivatives of  $f(X_1, \delta_1 | \boldsymbol{\beta})$  with respect to  $\boldsymbol{\beta}$  can be dominated by an integrable function of  $(X_1, \delta)$ . Then by the Dominated Convergence Theorem, we can exchange the order of derivative and integral.

To do this, we choose  $a > 0$  such that  $B = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq a\}$  is compact in the parameter space. By Lemma 2.3.4, there exists  $M$  such that for  $\boldsymbol{\beta} \in B$ ,

$$\left| \frac{\partial \log f_q(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \beta_l} \right| \leq MK(X_1), \quad (2.32)$$

$l = 1, \dots, p$ , and  $E(K(X_1)) < \infty$ . Hence, from (2.25) and (2.32), for  $\boldsymbol{\beta} \in B$ ,

$$\left| \frac{\partial f(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \beta_l} \right| = f(X_1, \delta_1 | \boldsymbol{\beta}) \left| \frac{\partial \log f(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \beta_l} \right| \leq MK(X_1)f(X_1, \delta_1 | \boldsymbol{\beta}).$$

Thus it suffices to find an integrable function of  $(X_1, \delta)$  to dominate  $K(X_1)f(X_1, \delta_1 | \boldsymbol{\beta})$ .

By the definition of  $f(\cdot, \cdot | \boldsymbol{\beta})$  from (2.11),

$$K(t)f(t, 1 | \boldsymbol{\beta}) = K(t)f(t | \boldsymbol{\beta})S_c(t) \leq K(t)f(t | \boldsymbol{\beta}) \leq K(t)f^*(t),$$



where  $f^*$  is the bound on  $f$  given in Note 2 after Proposition 2.2.2. It is easy to show that  $Kf^*$  is integrable.

To bound  $K(X_1)f(X_1, 0|\beta) = K(X_1)S(X_1|\beta)f_c(X_1)$ , it suffices to bound  $K(X_1)S(X_1|\beta)$ , since, by assumption 4,  $f_c$  is bounded. By the definition of  $S(t|\beta)$  from (2.7) and by the proofs of Propositions 2.2.1 and 2.2.2, there exists  $M > 0$  such that, for  $\beta \in B$ ,

$$S(t|\beta) \leq \exp\left(-\frac{t}{M}\right) \equiv S^*(t)$$

if  $\mathbf{b}$  is bounded, and

$$S(t|\beta) \leq M \exp\left(-\frac{(t+c)^{\beta'_p+1}}{M}\right) \equiv S^*(t)$$

if  $\mathbf{b}$  is bounded plus log tail, where  $\beta'_p$  is the lower bound of  $\beta_p$  for  $\beta \in B$ . One easily shows that  $KS^*$  is integrable.

To prove ii), provided integration and differentiation can be interchanged, we can write

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta_m} E \left\{ \frac{\partial \log f(X_1, \delta_1|\beta)}{\partial \beta_l} \right\} = \frac{\partial}{\partial \beta_m} \int \frac{\partial \log f(X_1, \delta_1|\beta)}{\partial \beta_l} f(X_1, \delta_1|\beta) d\mu \\ &= \int \frac{\partial}{\partial \beta_m} \left( \frac{\partial \log f(X_1, \delta_1|\beta)}{\partial \beta_l} f(X_1, \delta_1|\beta) \right) d\mu \\ &= \int \left\{ \frac{\partial \log f(X_1, \delta_1|\beta)}{\partial \beta_m} \cdot \frac{\partial \log f(X_1, \delta_1|\beta)}{\partial \beta_l} + \frac{\partial^2 \log f(X_1, \delta_1|\beta)}{\partial \beta_m \partial \beta_l} \right\} f(X_1, \delta_1|\beta) d\mu \\ &= E \left\{ \frac{\partial \log f_q(X_1, \delta_1|\beta)}{\partial \beta_m} \cdot \frac{\partial \log f_q(X_1, \delta_1|\beta)}{\partial \beta_l} \right\} + E \left\{ \frac{\partial^2 \log f_q(X_1, \delta_1|\beta)}{\partial \beta_m \partial \beta_l} \right\}. \end{aligned}$$

Therefore,

$$E \left\{ \frac{\partial}{\partial \beta_m} \log f_q(X_1, \delta_1|\beta) \frac{\partial}{\partial \beta_l} \log f_q(X_1, \delta_1|\beta) \right\} = -E \left\{ \frac{\partial^2}{\partial \beta_m \partial \beta_l} \log f_q(X_1, \delta_1|\beta) \right\}.$$

To exchange integration and differentiation, we need to bound  $\partial^2 \log f(X_1, \delta_1|\beta)/\partial \beta_m \partial \beta_l$ . By i) and Proposition 2.4.3, this can be done with an argument similar to that in the proof of i). We omit the calculations here.

## 2.5 Log Hazard Regression of Paired Failure Times

Now we consider the individuals with paired failure times. We first introduce the notation and define the log hazard regression model and estimates for paired failure data. Then, in Section 2.5.2, we show the consistency and normality of our estimates and give a consistent estimate of the covariance matrix of the estimates.

### 2.5.1 Model and Notation

Consider  $n$  subjects with paired failure times for each. For  $k = 1, 2$ , and  $i = 1, \dots, n$ , let  $T_{ki}$  and  $C_{ki}$  be the  $k$ th failure time and censoring time of the  $i$ th individual, respectively. Let  $X_{ki} = \min(T_{ki}, C_{ki})$  and

$$\delta_{ki} = \begin{cases} 1 & \text{if } T_{ki} \leq C_{ki}, \\ 0 & \text{if } T_{ki} > C_{ki}. \end{cases}$$

Then  $(X_{ki}, \delta_{ki})$  is the observation of the  $k$ th failure of the  $i$ th individual.

**Note:** the censoring times may be the same, i.e.,  $C_{1i} = C_{2i}$ .

We refer to  $f_k$ ,  $F_k$ ,  $S_k$ ,  $h_k$ , and  $\alpha_k$  as the corresponding density, distribution, survivor, hazard, and log hazard functions, respectively. We fit a log hazard regression model for the two log hazard functions separately. That is, we suppose that

$$\alpha_1(t|\boldsymbol{\beta}_1) = \sum_{j=1}^{p_1} \beta_{1j} b_{1j}(t)$$

and

$$\alpha_2(t|\boldsymbol{\beta}_2) = \sum_{j=1}^{p_2} \beta_{2j} b_{2j}(t).$$

For  $k = 1, 2$ , we estimate  $\boldsymbol{\beta}_k$  by  $\hat{\boldsymbol{\beta}}_k$  as defined in Section 2.2.1. Referring to the univariate case, we denote the regression basis by  $\mathbf{b}_k$ . As in (2.16), we let

$$f_{kq}(t, \delta|\boldsymbol{\beta}_k) = f_k(t|\boldsymbol{\beta}_k)^\delta S_k(t|\boldsymbol{\beta}_k)^{1-\delta}.$$

We also let  $U_k^{(n)}(\beta_k | \mathbf{X}_k)$ ,  $I_k^{(n)}(\beta_k | \mathbf{X}_k)$ , and  $I_k(\beta_k)$  be the corresponding score functions, observed information matrix, and information matrix, respectively, see (2.13), (2.14) and (2.15). Similarly, we denote by  $K_k$  the corresponding functions defined in (2.27). Let

$$\underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \quad \hat{\underline{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}. \quad (2.33)$$

We denote the  $i$ th summand in  $U_k^{(n)}$  by  $u_{ki}(\beta_k)$ , that is,

$$u_{ki}(\beta_k) = \frac{\partial}{\partial \beta_k} \log f_{kq}(X_{ki}, \delta_{ki} | \beta_k),$$

$k = 1, 2$ . Let

$$U^{(n)}(\beta_1, \beta_2 | \mathbf{X}_1, \mathbf{X}_2) = \begin{pmatrix} U_1^{(n)}(\beta_1 | \mathbf{X}_1) \\ U_2^{(n)}(\beta_2 | \mathbf{X}_2) \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} u_{1i}(\beta_1) \\ u_{2i}(\beta_2) \end{pmatrix},$$

and define

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \text{cov}(u_{11}(\beta_1), u_{11}(\beta_1)) & \text{cov}(u_{11}(\beta_1), u_{21}(\beta_2)) \\ \text{cov}(u_{21}(\beta_2), u_{11}(\beta_1)) & \text{cov}(u_{21}(\beta_2), u_{21}(\beta_2)) \end{pmatrix}. \quad (2.34)$$

## 2.5.2 Consistency and Normality of $\hat{\underline{\beta}}$

In this section, we show the consistency and the normality of the estimates  $\hat{\underline{\beta}}$ . The main result, given in Theorem 2.5.2, relies on the asymptotic normality of  $U^{(n)}(\hat{\beta}_1, \hat{\beta}_2 | \mathbf{X}_1, \mathbf{X}_2)$  given in the following.

**Proposition 2.5.1** *Suppose, marginally, each of the failure times satisfies Assumptions (1) through (5). Then*

$$\frac{1}{\sqrt{n}} U^{(n)}(\beta_1, \beta_2 | \mathbf{X}_1, \mathbf{X}_2) \Rightarrow N(0, \Sigma).$$

**Proof.** Since  $\mathbf{u}_{ki}(\boldsymbol{\beta}_k), i = 1, \dots, n$ , are i.i.d. and, by i) in Lemma 2.3.5,  $E(\mathbf{u}_{ki}(\boldsymbol{\beta}_k)) = 0$ , using the multivariate central limit theorem, we have

$$\frac{1}{\sqrt{n}} \mathbf{U}^{(n)}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 | \mathbf{X}_1, \mathbf{X}_2) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} \mathbf{u}_{1i}(\boldsymbol{\beta}_1) \\ \mathbf{u}_{2i}(\boldsymbol{\beta}_2) \end{pmatrix} \Rightarrow N(0, \Sigma).$$

□

**Theorem 2.5.2** *Suppose, marginally, each of the failure times satisfies Assumptions (1) through (5). Then*

(i)  $\hat{\boldsymbol{\beta}} \rightarrow \boldsymbol{\beta}$  in probability;

(ii)  $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow N(0, Q)$ , where

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = \begin{pmatrix} I_1(\boldsymbol{\beta}_1)^{-1} \Sigma_{11} I_1(\boldsymbol{\beta}_1)^{-1} & I_1(\boldsymbol{\beta}_1)^{-1} \Sigma_{12} I_2(\boldsymbol{\beta}_2)^{-1} \\ I_2(\boldsymbol{\beta}_2)^{-1} \Sigma_{21} I_1(\boldsymbol{\beta}_1)^{-1} & I_2(\boldsymbol{\beta}_2)^{-1} \Sigma_{22} I_2(\boldsymbol{\beta}_2)^{-1} \end{pmatrix}. \quad (2.35)$$

**Proof.** (i) For any  $\epsilon > 0$ , by Theorem 2.3.1,

$$P(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| > \epsilon) \leq P(\|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\| > \epsilon) + P(\|\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\| > \epsilon) \rightarrow 0,$$

as  $n \rightarrow \infty$ .

(ii) By expanding  $\mathbf{U}^{(n)}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2 | \mathbf{X}_1, \mathbf{X}_2)$  in a Taylor series about  $\boldsymbol{\beta}$ , we have

$$\mathbf{U}^{(n)}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2 | \mathbf{X}_1, \mathbf{X}_2) = \begin{pmatrix} \mathbf{U}_1^{(n)}(\boldsymbol{\beta}_1 | \mathbf{X}_1) \\ \mathbf{U}_2^{(n)}(\boldsymbol{\beta}_2 | \mathbf{X}_2) \end{pmatrix} - \begin{pmatrix} I_1^{(n)}(\boldsymbol{\beta}_1^* | \mathbf{X}_1)(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \\ I_2^{(n)}(\boldsymbol{\beta}_2^* | \mathbf{X}_2)(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \end{pmatrix},$$

where  $\boldsymbol{\beta}^* \equiv (\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*)^T$  is on a line segment between  $\hat{\boldsymbol{\beta}}$  and  $\boldsymbol{\beta}$ . Since

$$\mathbf{U}^{(n)}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2 | \mathbf{X}_1, \mathbf{X}_2) = \mathbf{0},$$

$$\begin{pmatrix} \mathbf{U}_1^{(n)}(\boldsymbol{\beta}_1 | \mathbf{X}_1) \\ \mathbf{U}_2^{(n)}(\boldsymbol{\beta}_2 | \mathbf{X}_2) \end{pmatrix} = \begin{pmatrix} I_1^{(n)}(\boldsymbol{\beta}_1^* | \mathbf{X}_1) & 0 \\ 0 & I_2^{(n)}(\boldsymbol{\beta}_2^* | \mathbf{X}_2) \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix},$$

i.e.,

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n} I_1^{(n)}(\beta_1^* | \mathbf{X}_1) & 0 \\ 0 & \frac{1}{n} I_2^{(n)}(\beta_2^* | \mathbf{X}_2) \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}} U_1^{(n)}(\beta_1 | \mathbf{X}_1) \\ \frac{1}{\sqrt{n}} U_2^{(n)}(\beta_2 | \mathbf{X}_2) \end{pmatrix}.$$

By (iii) in Theorem 2.3.1,

$$\left[ \frac{1}{n} I_k^{(n)}(\beta_k^* | \mathbf{X}_k) \right]^{-1} \rightarrow I_k(\beta_k)^{-1}$$

in probability, since  $\beta_k^* \rightarrow \beta_k$  in probability,  $k = 1, 2$ . Using this and Proposition 2.5.1, by Slutsky's theorem, we have

$$\sqrt{n}(\hat{\beta} - \beta) \Rightarrow N(0, Q).$$

□

To estimate  $Q$ , let

$$\hat{Q}_{kl} = \left[ \frac{1}{n} I_k^{(n)}(\hat{\beta}_k | \mathbf{X}_k) \right]^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{u}_{ki}(\hat{\beta}_k) \mathbf{u}_{li}(\hat{\beta}_l)^T \right) \left[ \frac{1}{n} I_l^{(n)}(\hat{\beta}_l | \mathbf{X}_l) \right]^{-1}, \quad (2.36)$$

$k, l = 1, 2$ , and

$$\hat{Q} = \begin{pmatrix} \hat{Q}_{11} & \hat{Q}_{12} \\ \hat{Q}_{21} & \hat{Q}_{22} \end{pmatrix}.$$

In the next theorem we show that  $\hat{Q}$  is a consistent estimate of  $Q$ .

**Theorem 2.5.3** *Suppose, marginally, each of the failure times satisfies Assumptions (1) - (5) for  $k = 1, 2$ . Then  $\hat{Q}$  is a consistent estimate of  $Q$ .*

**Proof.** Let  $\hat{\beta}_0$  be the true values of the parameter. From Theorem 2.3.1,

$$\frac{1}{n} I_k^{(n)}(\hat{\beta}_k | \mathbf{X}_k) \rightarrow I_k(\beta_{k0})$$

in probability,  $k = 1, 2$ . Therefore, we need only prove that

$$\frac{1}{n} \sum_{i=1}^n \mathbf{u}_{ki}(\hat{\beta}_k) \mathbf{u}_{li}(\hat{\beta}_l)^T \rightarrow \Sigma_{kl} \quad (2.37)$$

in probability,  $k, l = 1, 2$ . By the law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{u}_{ki}(\boldsymbol{\beta}_{k0}) \mathbf{u}_{li}(\boldsymbol{\beta}_{l0})^T \rightarrow E(\mathbf{u}_{k1}(\boldsymbol{\beta}_{k0}) \mathbf{u}_{l1}(\boldsymbol{\beta}_{l0})^T) = \Sigma_{kl},$$

in probability. Hence, to prove (2.37), it is sufficient to prove that

$$\frac{1}{n} \sum_{i=1}^n \mathbf{u}_{ki}(\hat{\boldsymbol{\beta}}_k) \mathbf{u}_{li}(\hat{\boldsymbol{\beta}}_l)^T - \frac{1}{n} \sum_{i=1}^n \mathbf{u}_{ki}(\boldsymbol{\beta}_{k0}) \mathbf{u}_{li}(\boldsymbol{\beta}_{l0})^T \rightarrow 0$$

in probability.

Let  $u_{kis}(\boldsymbol{\beta}_k)$  be the  $s$ th component of  $\mathbf{u}_{ki}(\boldsymbol{\beta}_k)$ , i.e.,

$$u_{kis}(\boldsymbol{\beta}_k) = \frac{\partial}{\partial \beta_{ks}} \log f_{kq}(X_{ki}, \delta_{ki} | \boldsymbol{\beta}_k),$$

$s = 1, \dots, p_k; k = 1, 2; i = 1, \dots, n$ . We need to prove

$$\frac{1}{n} \sum_{i=1}^n u_{kis}(\hat{\boldsymbol{\beta}}_k) u_{lij}(\hat{\boldsymbol{\beta}}_l) - \frac{1}{n} \sum_{i=1}^n u_{kis}(\boldsymbol{\beta}_{k0}) u_{lij}(\boldsymbol{\beta}_{l0}) \rightarrow 0, \quad (2.38)$$

in probability, for  $s = 1, \dots, p_k; j = 1, \dots, p_l; k, l = 1, 2$ . Using the mean value theorem

applied to  $\sum_{i=1}^n u_{kis}(\hat{\boldsymbol{\beta}}_k) u_{lij}(\hat{\boldsymbol{\beta}}_l)$  and  $\sum_{i=1}^n u_{kis}(\boldsymbol{\beta}_{k0}) u_{lij}(\boldsymbol{\beta}_{l0})$ , we have

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n u_{kis}(\hat{\boldsymbol{\beta}}_k) u_{lij}(\hat{\boldsymbol{\beta}}_l) - \frac{1}{n} \sum_{i=1}^n u_{kis}(\boldsymbol{\beta}_{k0}) u_{lij}(\boldsymbol{\beta}_{l0}) \right| \\ & \leq \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^{p_k} \left| \frac{\partial}{\partial \beta_{km}} (u_{kis}(\boldsymbol{\beta}_k) u_{lij}(\boldsymbol{\beta}_l)) \right|_{\underline{\boldsymbol{\beta}}^*} |\hat{\beta}_{km} - \beta_{k0m}| \\ & \quad + \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^{p_l} \left| \frac{\partial}{\partial \beta_{lm}} (u_{kis}(\boldsymbol{\beta}_k) u_{lij}(\boldsymbol{\beta}_l)) \right|_{\underline{\boldsymbol{\beta}}^*} |\hat{\beta}_{lm} - \beta_{l0m}| \\ & \leq \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^{p_k} \left| \frac{\partial}{\partial \beta_{km}} (u_{kis}(\boldsymbol{\beta}_k) u_{lij}(\boldsymbol{\beta}_l)) \right|_{\underline{\boldsymbol{\beta}}^*} \sum_{m=1}^{p_k} |\hat{\beta}_{km} - \beta_{k0m}| \\ & \quad + \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^{p_l} \left| \frac{\partial}{\partial \beta_{lm}} (u_{kis}(\boldsymbol{\beta}_k) u_{lij}(\boldsymbol{\beta}_l)) \right|_{\underline{\boldsymbol{\beta}}^*} \sum_{m=1}^{p_l} |\hat{\beta}_{lm} - \beta_{l0m}| \end{aligned}$$

where  $\underline{\boldsymbol{\beta}}^*$  is on a line segment between  $\hat{\boldsymbol{\beta}}$  and  $\boldsymbol{\beta}_0$ . Thus, since  $\hat{\boldsymbol{\beta}} \rightarrow \boldsymbol{\beta}_0$  in probability, to prove (2.38), it suffices to show that  $\frac{1}{n} \sum_{i=1}^n \sum_{m=1}^{p_k} \left| \frac{\partial}{\partial \beta_{km}} (u_{kis}(\boldsymbol{\beta}_k) u_{lij}(\boldsymbol{\beta}_l)) \right|$  is bounded in probability around  $\boldsymbol{\beta}_0$ . That means we need to find  $M < \infty$  and a neighborhood of  $\boldsymbol{\beta}_0$  such that

$$P \left\{ \sup_{\boldsymbol{\beta} \in B} \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^{p_k} \left| \frac{\partial}{\partial \beta_{km}} (u_{kis}(\boldsymbol{\beta}_k) u_{lij}(\boldsymbol{\beta}_l)) \right| < M \right\} \rightarrow 1$$

when  $n \rightarrow \infty$ ,  $k, l = 1, 2$ .

We choose a number  $a > 0$  such that  $B = \{\underline{\beta} : \|\underline{\beta} - \underline{\beta}_0\| \leq a\}$  is a compact set in the parameter space. By Lemma 2.3.4, there exist  $M_k$  and  $M_l$  such that for  $\underline{\beta} \in B, k, l = 1, 2, k \neq l, m = 1, \dots, p_k$ ,

$$\begin{aligned} & \left| \frac{\partial}{\partial \beta_{km}} (u_{kis}(\beta_k) u_{lij}(\beta_l)) \right| = \left| u_{lij}(\beta_l) \frac{\partial}{\partial \beta_{km}} u_{kis}(\beta_k) \right| \\ &= \left| \left( \frac{\partial}{\partial \beta_{lj}} \log f_{lq}(X_{li}, \delta_{li} | \beta_l) \right) \frac{\partial^2}{\partial \beta_{ks} \partial \beta_{km}} \log f_{kq}(X_{ki}, \delta_{ki} | \beta_k) \right| \\ &\leq M_k M_l K_k(X_{ki}) K_l(X_{li}). \end{aligned}$$

For  $k = l$ ,

$$\left| \frac{\partial}{\partial \beta_{km}} (u_{kis}(\beta_k) u_{lij}(\beta_l)) \right| \leq 2M_k^2 [K_k(X_{ki})]^2.$$

Thus,

$$\begin{aligned} & P \left\{ \sup_{\beta \in B} \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^{p_k} \left| \frac{\partial}{\partial \beta_{km}} (u_{kis}(\beta_k) u_{lij}(\beta_l)) \right| < M \right\} \\ &\geq P \left\{ 2p_k M_k M_l \frac{1}{n} \sum_{i=1}^n K_k(X_{ki}) K_l(X_{li}) \leq M \right\} \end{aligned}$$

which will converge to 1 for  $M$  sufficiently large, since

$$\frac{1}{n} \sum_{i=1}^n K_k(X_{ki}) K_l(X_{li}) \rightarrow E(K_k(X_{ki}) K_l(X_{li})) < \infty$$

by Proposition 2.4.3. □

As in the univariate case, we have the following Corollary.

**Corollary 2.5.4** *If  $\mathbf{b}_1$  and  $\mathbf{b}_2$  satisfy assumptions 1 through 5, then*

$$\sqrt{n} \hat{Q}^{-\frac{1}{2}}(\hat{\beta} - \beta) \Rightarrow N(\mathbf{0}, I),$$

where  $I$  is  $(p_1 + p_2) \times (p_1 + p_2)$  identity matrix.

## Chapter 3

# Regression Space

In the previous chapter we discussed the regression model for log hazard functions for a general regression space  $G$ . Here, we will use a family of extremely flexible functions, cubic splines, as the regression space. This family was used by Kooperberg, Stone, and Truong in 1995 for univariate log hazard regression. In this chapter we first give a brief introduction to cubic splines. Then we give the definition of a restricted cubic spline regression space  $\mathcal{B}$  and a method to construct a cubic B-spline basis of  $\mathcal{B}$ . In Section 3.3 we introduce the numerical computation method that we use to calculate the estimates of the log hazard regression model for paired failure data. In Section 3.4 we explain our methods for choosing knots for the log regression model.

Finally, in Section 3.5, we show that the log hazard regression model with the regression space  $\mathcal{B}$  can be used for hypothesis testing. In the univariate case, we can test the hypotheses that the failure times have an exponential or a Weibull distribution. In the bivariate case, we can test the hypotheses that the two failure times have the same distribution or that the two failures have proportional hazards.



### 3.1 Cubic Splines

**Definition 3.1.1** *The function  $\phi$  is a cubic spline on  $[a, b]$  with knots  $t_1, \dots, t_K$  ( $a < t_1 < \dots < t_K < b$ ) if  $\phi$  is a cubic polynomial on the subintervals  $[a, t_1], [t_1, t_2], \dots, [t_K, b]$  and  $\phi$  has 2 continuous derivatives on  $[a, b]$ . Denote the collection of these splines by  $Sp(t_1, \dots, t_K)$ .*

By De Boor (1978),  $Sp(t_1, \dots, t_K)$  is a linear space with dimension  $K+4$ . The *power basis*  $\{p_k, k = -3, \dots, K\}$  of  $Sp(t_1, \dots, t_K)$  is defined as:

$$p_0(t) = 1, \quad p_{-1}(t) = t, \quad p_{-2} = t^2, \quad p_{-3} = t^3,$$

$$p_1(t) = (t - t_1)_+^3, \quad \dots, \quad p_K(t) = (t - t_K)_+^3,$$

where

$$(t - t')_+^3 = \begin{cases} (t - t')^3 & \text{for } t \geq t' \\ 0 & \text{for } t < t'. \end{cases}$$

Therefore, a power basis representation of a cubic spline  $\phi \in Sp(t_1, \dots, t_K)$  is

$$\phi(t) = \sum_{k=-3}^K \theta_k p_k(t) = \sum_{k=0}^3 \theta_{-k} t^k + \sum_{k=1}^K \theta_k (t - t_k)_+^3. \quad (3.1)$$

The power basis is very easy to understand, but isn't used in computation since it has bad numerical properties (see De Boor, 1978). A numerically much better basis is the *B-spline basis*, which consists of functions that are zero except on some sequential subintervals. For more details about the B-spline basis, see De Boor (1978) and Shikin (1995). We will use a restricted B-spline basis, defined in Section 3.2, for our regression.

### 3.2 The Regression Space $\mathcal{B}$

In this section we will introduce the cubic spline regression spaces defined by Kooperberg et al. (1995) and we will give an algorithm to construct restricted B-spline bases in those spaces.

First, we assume the log hazard function  $\alpha(t|\beta)$  of a failure time to be a cubic spline from  $Sp(t_1, \dots, t_K)$  satisfying

(\*\*)  $\alpha$  is linear on  $[0, t_1]$  and is constant on  $[t_K, \infty)$ .

If we use the power basis representation  $\alpha = \sum_{k=-3}^K \theta_k p_k$ , then (\*\*) implies  $\theta_{-3}$  and  $\theta_{-2}$  equal 0, and it places three constraints on  $\theta_{-1}, \theta_1, \dots, \theta_K$ . This leaves a  $K + 4 - 2 - 3 = K - 1$  dimensional space. More precisely, define

$$\mathcal{P} = \text{span}\{p_{-1}, p_0, p_1, \dots, p_K\}$$

and

$$\mathcal{B} = \{\phi \in \mathcal{P} : \phi \text{ is linear on } [0, t_1] \text{ and constant on } [t_K, \infty)\}.$$

Then  $\dim(\mathcal{P}) = K + 2$  and  $\dim(\mathcal{B}) = K - 1$ .

Next, we give the definition of a restricted B-spline basis of  $\mathcal{B}$ . We assume that  $K \geq 3$ , and place restrictions on  $b_1, b_{K-1}$  and, if  $K > 3$ , on  $b_2, \dots, b_{K-2}$ .

**Definition 3.2.1** *The set of functions  $\{b_1, \dots, b_{K-1}\}$  in  $\mathcal{B}$  is called a restricted B-spline basis of  $\mathcal{B}$  if it has the following properties:*

1.  $b_1$  is linear but not constant on  $[0, t_1]$  and is zero on  $[t_3, \infty)$ ;
2. if  $K > 3$ , for  $1 < j < K - 2$ ,  $b_j$  is zero on  $[0, t_{j-1})$  and  $[t_{j+3}, \infty)$  but not zero on  $(t_{j-1}, t_j)$ ;
3. if  $K > 3$ ,  $b_{K-2}$  is zero on  $[0, t_{K-3})$ , not zero on  $(t_{K-3}, t_{K-2})$ , and a non-zero constant on  $[t_K, \infty)$ ;
4.  $b_{K-1}$  is a non-zero constant on  $[0, \infty)$ .

Clearly, any  $b_j$ 's satisfying 1 - 4 are bounded, a condition required in Section 2.2.1. Now we need to show that the restricted B-spline basis is well defined. That means that

we need to prove that  $b_j$ 's with the above properties exist and they are a basis of  $\mathcal{B}$ . We will give an algorithm for constructing  $b_j$ 's (this implies the existence), but first we will show that any  $b_j$ 's satisfying 1 - 4 are a basis.

**Theorem 3.2.2** *If  $b_1, \dots, b_{K-1}, K \geq 3$ , in  $\mathcal{B}$  have properties 1 - 4 of Definition 3.2.1, then they are a basis of  $\mathcal{B}$ .*

**Proof.** Since  $\mathcal{B}$  is a  $K - 1$  dimensional space, to show that  $\{b_1, \dots, b_{K-1}\}$  is a basis of  $\mathcal{B}$ , we need only prove that  $b_1, \dots, b_{K-1}$  are linearly independent. Let

$$\phi(t) = \sum_{j=1}^{K-1} \alpha_j b_j(t).$$

Suppose that  $\phi(t) \equiv 0$ . We are going to show that this implies that  $\alpha_j = 0, j = 1, \dots, K - 1$ . Consider  $t \in [0, t_1)$ . If  $K > 3$ , by properties 2 and 3,  $b_j(t) = 0$ , for  $j = 2, \dots, K - 2$ . If  $K = 3$ , then  $\phi(t) = \alpha_1 b_1(t) + \alpha_2 b_2(t)$ . Therefore, for  $K \geq 3$  and  $t \in [0, t_1)$

$$\phi(t) = \alpha_1 b_1(t) + \alpha_{K-1} b_{K-1}(t) = 0.$$

By property 4,  $b_{K-1}$  is non-zero constant, and by property 1,  $b_1$  is linear but not a constant. Hence we have

$$\alpha_1 = \alpha_{K-1} = 0. \tag{3.2}$$

Hence, if  $K = 3$ , then  $b_1$  and  $b_2$  are a basis of  $\mathcal{B}$ . Now we assume that  $K > 3$ . Consider  $t \in (t_1, t_2)$ . If  $K > 4$ , by properties 2 and 3,  $b_j(t) = 0, j = 3, \dots, K - 2$ . If  $K = 4$ , then

$$\phi(t) = \alpha_1 b_1(t) + \alpha_2 b_2(t) + \alpha_3 b_3(t).$$

Hence, for  $K \geq 4$  and  $t \in (t_1, t_2)$ , using (3.2), we have

$$\phi(t) = \alpha_1 b_1(t) + \alpha_2 b_2(t) + \alpha_{K-1} b_{K-1}(t) = \alpha_2 b_2(t) = 0.$$

So  $\alpha_2 = 0$  since  $b_2(t) \neq 0$  by property 2. By induction on  $j$ , using a similar explanation, for  $t \in (t_{j-1}, t_j)$ ,

$$\phi(t) = \alpha_j b_j(t) = 0,$$

which implies that  $\alpha_j = 0$  for  $j = 3, \dots, K-2$ . Hence we have that  $\alpha_j = 0, j = 1, \dots, K-1$ , which shows that  $\{b_1, \dots, b_{K-1}\}$  is a basis of  $\mathcal{B}$ .  $\square$

For the convenience of our constructing a restricted B-spline basis, we give the following result as a lemma.

**Lemma 3.2.3** *Fix  $J = 2, \dots, K-2$ , and let*

$$b(t) = \sum_{k=-3}^0 \theta_k p_k(t) + \sum_{k=J-1}^{J+2} \theta_k p_k(t).$$

*Then  $b \equiv c$  on  $[t_{J+2}, \infty)$  if and only if  $\theta_{-3}, \theta_{-2}, \theta_{-1}, \theta_0, \theta_{J-1}, \theta_J, \theta_{J+1}, \theta_{J+2}$  solve the linear system*

$$\begin{aligned} \theta_{-3} + \theta_{J-1} + \theta_J + \theta_{J+1} + \theta_{J+2} &= 0 \\ \theta_{-2} + 3t_{J-1}\theta_{J-1} + 3t_J\theta_J + 3t_{J+1}\theta_{J+1} + 3t_{J+2}\theta_{J+2} &= 0 \\ \theta_{-1} + 3t_{J-1}^2\theta_{J-1} + 3t_J^2\theta_J + 3t_{J+1}^2\theta_{J+1} + 3t_{J+2}^2\theta_{J+2} &= 0 \\ \theta_0 + t_{J-1}^3\theta_{J-1} + t_J^3\theta_J + t_{J+1}^3\theta_{J+1} + t_{J+2}^3\theta_{J+2} &= c. \end{aligned}$$

**Proof.** First regroup terms of  $b(t)$  by powers of  $t$ , so that

$$b(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3,$$

with  $\alpha$ 's depending on indicator functions involving the  $\theta_k$ 's and the  $t_j$ 's. We see that  $b \equiv c$  on  $[t_{J+2}, \infty)$  is equivalent to  $\alpha_0 = c, \alpha_1 = \alpha_2 = \alpha_3 = 0$ , which is equivalent to the above equations.  $\square$

Now we start our construction of a restricted B-spline basis of  $\mathcal{B}$ . Let  $b_j = \sum_{k=-1}^K \theta_{kj} p_k$ ,  $j = 1, \dots, K-1$ . We will find  $\theta_{kj}$ 's so that the  $b_j$ 's satisfy properties 1 - 4 of Definition 3.2.1.

**Step 1:** Let  $b_{K-1} = p_0$ , i.e.,  $\theta_{0(K-1)} = 1$  and  $\theta_{k(K-1)} = 0$ , for  $k = -1, 1, \dots, K$ .

**Step 2:** To define  $b_1 = \sum_{k=-1}^K \theta_{k1} p_k$ , let  $\theta_{k1} = 0$  if  $k > 3$ . As in the proof of Lemma 3.2.3, we regroup terms of  $b_1$  by powers of  $t$ . Then  $b_1 \equiv 0$  on  $[t_3, \infty)$  implies that  $\theta_{01}, \theta_{(-1)1}, \theta_{11}, \theta_{21}, \theta_{31}$  solve the following linear system

$$\begin{aligned} \theta_{11} + \theta_{21} + \theta_{31} &= 0 \\ t_1 \theta_{11} + t_2 \theta_{21} + t_3 \theta_{31} &= 0 \\ \theta_{(-1)1} + 3t_1^2 \theta_{11} + 3t_2^2 \theta_{21} + 3t_3^2 \theta_{31} &= 0 \\ \theta_{01} + t_1^3 \theta_{11} + t_2^3 \theta_{21} + t_3^3 \theta_{31} &= 0. \end{aligned}$$

Since the  $t_k$ 's are not equal to each other, there is a unique solution for the linear system once one of the  $\theta_{k1}, k = -1, \dots, 3$ , is fixed. We let  $\theta_{11} = 1$  and solve the linear system. Note that  $\theta_{(-1)1} \neq 0$  otherwise the linear system has only zero solution which would contradict  $\theta_{11} = 1$ . Thus,  $b_1$  is linear but not constant on  $[0, t_1)$  and  $b_1 \equiv 0$  on  $(t_3, \infty)$ .

To construct  $b_j, 1 < j < K - 1$ , we first construct  $b_j^*$  which is zero on  $[0, t_{j-1})$  and a non-zero constant on  $[t_{j+2}, \infty)$ . We will define the  $b_j^*$ 's as linear combinations of the  $b_j$ 's such that  $b_j$  is zero on  $[t_{j+3}, \infty), 1 < j < K - 2$ .

**Step 3** For  $1 < j < K - 1$ , let

$$b_j^* = \theta'_{(j-1)j} p_{j-1} + \theta'_{jj} p_j + \theta'_{(j+1)j} p_{j+1} + \theta'_{(j+2)j} p_{j+2}.$$

Then for any constant  $c$ , by Lemma 3.2.3,  $b_j^* \equiv c$  on  $[t_{j+2}, \infty)$  and  $\theta_{sj} = 0, s = -3, -2, -1, 0$  imply that  $\theta'_{(j-1)j}, \theta'_{jj}, \theta'_{(j+1)j}, \theta'_{(j+2)j}$  satisfy the linear system

$$\begin{aligned} \theta'_{(j-1)j} + \theta'_{jj} + \theta'_{(j+1)j} + \theta'_{(j+2)j} &= 0 \\ t_{j-1} \theta'_{(j-1)j} + t_j \theta'_{jj} + t_{j+1} \theta'_{(j+1)j} + t_{j+2} \theta'_{(j+2)j} &= 0 \\ t_{j-1}^2 \theta'_{(j-1)j} + t_j^2 \theta'_{jj} + t_{j+1}^2 \theta'_{(j+1)j} + t_{j+2}^2 \theta'_{(j+2)j} &= 0 \\ t_{j-1}^3 \theta'_{(j-1)j} + t_j^3 \theta'_{jj} + t_{j+1}^3 \theta'_{(j+1)j} + t_{j+2}^3 \theta'_{(j+2)j} &= c. \end{aligned}$$

Let  $\theta'_{(j-1)j} = 1$ . Then we can solve the first three equations for  $\theta'_{jj}, \theta'_{(j+1)j}, \theta'_{(j+2)j}$  and then set  $c$  to satisfy the fourth equation. We must show that  $c \neq 0$ . It is easy to see that, if  $c = 0$ , then the linear system in Lemma 3.2.3 has only the zero solution. Hence here  $\theta_{(j-1)j} = 1$  implies that  $c \neq 0$ . Then  $b_j^*$  is zero on  $[0, t_{j-1}]$ , non-zero on  $(t_{j-1}, t_j)$ , and a non-zero constant on  $(t_{j+2}, \infty)$ .

**Step 4:** Let  $b_{K-2} = b_{K-2}^*$ . Then  $b_{K-2}$  satisfies property 3.

**Step 5:** Recall that  $b_j^*$  is a non-zero constant on  $[t_{j+2}, \infty)$ . For  $1 < j < K - 1$ , let  $a_j$  be that constant on  $[t_{j+2}, \infty)$ . For  $1 < j < K - 2$ , if we let

$$b_j = b_j^* - \frac{a_j}{a_{j+1}} b_{j+1}^*$$

then  $b_j$  is zero on  $[t_{j+3}, \infty)$  and also on  $[0, t_{j-1})$ , since  $b_j^*$  and  $b_{j+1}^*$  are zero on  $[0, t_{j-1})$ . Also  $b_j$  is not zero on  $(t_{j-1}, t_j)$  since  $b_{j+1}^*$  is zero but  $b_j^*$  is not zero on  $(t_{j-1}, t_j)$ .

Thus  $b_1, \dots, b_{K-1}$  satisfy properties 1 - 4 and are a B-spline basis of  $\mathcal{B}$ .

From now on, for given knots, we use  $\mathcal{B}$  as our regression space with the restricted B-spline basis  $b_1, \dots, b_{K-1}$  defined as above. By a bounded basis plus log tail, we mean the basis is  $\{b_1, \dots, b_{K-1}, b_K\}$  and the regression space is  $\text{span}\{b_1, \dots, b_K\}$ , where  $b_K(t) \equiv \log(t + c)$ ,  $c \geq 0$  is a constant.

Since the basis  $\mathbf{b}$  is bounded or bounded plus log tail, all relevant results discussed in Chapter 2 can be used. To simplify notation, we will denote both regression spaces by  $\mathcal{B}$ .

### 3.3 Numerical Implementation

In this section we introduce the algorithm used to calculate the estimates  $\hat{\underline{\beta}}$  and  $\hat{Q}$  in the log hazard regression model of paired failure times.

By the definition of  $\hat{\underline{\beta}}$  from (2.33), the estimates  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are calculated separately. So we need only discuss the calculation of  $\hat{\beta}_1$ . We can use Kooperberg's *left*

algorithm (see Kooperberg, Stone, Truong, 1995) to find  $\hat{\beta}_1$ . They use the Newton-Raphson method for computing  $\hat{\beta}_1$  since  $\beta_1$  is the unique solution of the equations (2.20). Specifically, they start with an initial guess  $\hat{\beta}_1^{(0)}$ , then iteratively determine  $\hat{\beta}_1^{(k+1)}$  from  $\hat{\beta}_1^{(k)}$  according to the formula

$$\hat{\beta}_1^{(k+1)} = \hat{\beta}_1^{(k)} + [I^{(n)}(\hat{\beta}_1^{(k)} | \mathbf{X}_1)]^{-1} U^{(n)}(\hat{\beta}_1^{(k)} | \mathbf{X}_1),$$

and stop the iterations when  $\log L_q(\hat{\beta}_1^{(k+1)} | \mathbf{X}_1) - \log L_q(\hat{\beta}_1^{(k)} | \mathbf{X}_1) \leq \epsilon$ , where  $\epsilon$  is a given positive number chosen so that estimates with the desired accuracy can be obtained.

Thus the main numerical task in calculating  $\hat{\beta}$  is the computation of the log likelihood  $\log L_q(\beta_1 | \mathbf{X}_1)$ , score function  $U_1^{(n)}(\beta_1)$ , and observed information matrix  $I_1^{(n)}(\beta_1 | \mathbf{X}_1)$  for various values of  $\beta$ . By the definitions of  $\log L_q(\beta_1 | \mathbf{X}_1)$ ,  $U_1^{(n)}(\beta_1)$  and  $I_1^{(n)}(\beta_1 | \mathbf{X}_1)$ , see (2.17), (2.18), and (2.19), this computation involves the numerical approximation of

$$\sum_{i=1}^n \int_0^{X_{1i}} \psi(t) dt, \quad (3.3)$$

for  $\psi$  of the form

$$\psi(t) = b_{1m}(t) b_{1l}(t) \exp\left(\sum_{j=1}^{p_1} \beta_{1j} b_{1j}(t)\right),$$

$l, m \in \{1, \dots, p_1\}$ .

Kooperberg et. al. do not calculate  $\int_0^{X_{1i}} \psi(t) dt$  for each  $i$ . Rather, they rewrite (3.3) as

$$\int_0^\infty N(t) \psi(t) dt, \quad (3.4)$$

where  $N(t)$  is the number of  $i$  satisfying  $X_{ki} \geq t$ . The function  $N(\cdot)$  is piecewise constant, has jumps at the observations  $X_{k1}, \dots, X_{kn}$ , and equals zero to the right of the maximum observation. Then they rewrite (3.4) as

$$\sum_{\nu} \int_{a_{\nu-1}}^{a_{\nu}} N(t) \psi(t) dt,$$

where  $\{a_\nu\}$  is a finite grid of points containing all knots. Then they calculate  $\int_{a_{\nu-1}}^{a_\nu} N(t)\psi(t)dt$  numerically.

Now consider calculation of  $\hat{Q}$ . By the definition of  $\hat{Q}$  from (2.36), the calculation of  $\hat{Q}$  involves the calculation of  $\hat{\beta}_k, I_k^{(n)}(\hat{\beta}_k|\mathbf{X}_k)$ , and the summands  $\mathbf{u}_{ki}(\hat{\beta}_k)$ ,  $k = 1, 2$ ,  $i = 1, \dots, n$ . As we mentioned above, we can get  $\hat{\beta}_k$  from the *heft* algorithm directly. By slightly modifying the *heft* code, we can obtain  $I_k^{(n)}(\hat{\beta}_k|\mathbf{X}_k)$  from *heft* as well. But to evaluate  $\mathbf{u}_{ki}(\hat{\beta}_k)$ , we need to calculate  $\int_0^{X_{ki}} \psi(t)dt$  separately. We can not get these integrals from *heft* without rewriting its entire integration program. Instead, we use Gaussian Quadrature (see Abramowitz and Stegun 1964, p. 916) for the integration  $\int_0^{X_{ki}} \psi(t)dt$  to calculate  $\mathbf{u}_{ki}(\hat{\beta}_k)$ ,  $k = 1, 2$ .

### 3.4 Knot Selection

In Section 3.2, we give the method to define the regression space  $\mathcal{B}$  for given knots. In this section we introduce the methods that we use to choose the knots.

For the univariate case, we use the following two methods to select the knots.

- Choose knots by Kooperberg's *heft* algorithm.

For a data set, Kooperberg's *heft* algorithm can choose the knots for the model fitting. The algorithm chooses knots by a stepwise addition and stepwise deletion procedure. See Kooperberg et. al. 1995 for details. We hoped that *heft* would select knots well. But from our simulation study we find that there are some numerical calculation problems with the knots chosen by *heft*. If the ratio of the maximum of the knots to the minimum of the knots is too big, then the resulting log hazard estimate and the calculation of the estimated standard errors may be impossible, see Section 5.2.2. So we should not use the knots from *heft*, if we receive a warning message from the *heft* code, or the calculation of the estimated standard error is not possible.



- Choose the knots by the quantiles.

We choose the quantiles of the non-censored observations as the knots to define the regression space  $\mathcal{B}$ . This procedure is based on ideas in Kooperberg et. al's knot selection for their stepwise addition method and Abrahamowicz et. al's knot selection method, Abrahamowicz, Ciampi, and Ramsay (1992). Choosing knots equal to quantiles may also result in a large ratio of maximum of the knots to the minimum of the knots. As noted above, this causes numerical problems. We solve this problem by truncating the data, that is, if there are any warnings, we use the quantiles of a truncated data set. We truncate the observations which are greater than 80 in our simulations.

For paired data we use the following methods:

- Use different knots for modeling the two log hazard functions.

Choose knots for each failure time by the above two methods. We then use the two sets of knots to define the two regression spaces and marginally fit the log hazard regression models for the paired data.

- Use the the same knots for modeling the two log hazard functions.

In this method we need to choose one set of knots which defines one regression space for both log hazard functions. If we denote the ranges of non-censored observations  $\{X_{ki} : \delta_{ki} = 1\}$  by  $R_k$ ,  $k = 1, 2$ , then our knots must lie in  $R_1 \cap R_2$ . There are two way to choose this set of knots.

- a) Use the quantiles of the non-censored observations which lie in  $R_1 \cap R_2$ .
- b) Use the union of the knots selected for each marginal log hazard estimate. Denote the sets of knots selected for the two log hazards by  $\mathcal{K}_1$  and  $\mathcal{K}_2$ . Use  $(\mathcal{K}_1 \cup \mathcal{K}_2) \cap (R_1 \cap R_2)$  as the set of knots for the regression space.

## 3.5 Hypothesis Testing

In this section we show how to use the restricted B-spline basis and the regression space  $\mathcal{B}$  and given data to test hypotheses.

### 3.5.1 Univariate Case

In the univariate case, we can test:

- $H_0^e$ : the failure time has an exponential distribution;
- $H_0^w$ : the failure time has a Weibull distribution.

By the definition of the log hazard regression model in (2.5), the log hazard function is

$$\alpha(t|\boldsymbol{\beta}) = \sum_{j=1}^p \beta_j b_j(t).$$

If the failure time is exponentially distributed, then  $\alpha(t|\boldsymbol{\beta}) \equiv \text{constant}$ . Therefore, for the restricted B-spline basis defined in Section 3.2,  $(\{b_j, j = 1, \dots, K-1\}$  if  $\mathbf{b}$  is bounded or  $\{b_j, j = 1, \dots, K\}$  if  $\mathbf{b}$  is bounded plus log tail, where  $K$  is the number of knots), we can rewrite  $H_0^e$  as

$$H_0^e: \beta_j = 0 \text{ for } j \neq K-1.$$

Then we can write  $H_0^e$  as

$$H_0^e: X_e^T \boldsymbol{\beta} = 0$$

for an appropriate matrix  $X_e$ . By Corollary 2.3.3,

$$[nI^{(n)}(\hat{\boldsymbol{\beta}}|\mathbf{X})]^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow N(0, I).$$

Hence, under  $H_0^e$ ,

$$n(X_e^T \hat{\beta})^T (X_e^T [I^{(n)}(\hat{\beta}|\mathbf{X})]^{-1} X_e)^{-1} (X_e^T \hat{\beta}) \Rightarrow \begin{cases} \chi_{K-2}^2 & \text{if } \mathbf{b} \text{ is bounded;} \\ \chi_{K-1}^2 & \text{if } \mathbf{b} \text{ is bounded plus log tail.} \end{cases}$$

Therefore,  $n(X_e^T \hat{\beta})^T (X_e^T [I^{(n)}(\hat{\beta}|\mathbf{X})]^{-1} X_e)^{-1} (X_e^T \hat{\beta})$  can be used as a test statistic.

To test  $H_0^w$ , we need to use the restricted B-spline basis plus log tail  $b_K(t) \equiv \log(t)$ .

Thus, if the failure time has a Weibull distribution, then

$$\alpha(t|\beta) = \beta_{K-1} b_{K-1}(t) + \beta_K b_K(t).$$

Hence, testing  $H_0^w$  is equivalent to testing  $\beta_j = 0$  for  $j < K - 1$ . By a similar procedure as for testing  $H_0^e$ , we can find a matrix  $X_w$  such that

$$n(X_w^T \hat{\beta})^T (X_w^T [I^{(n)}(\hat{\beta}|\mathbf{X})]^{-1} X_w)^{-1} (X_w^T \hat{\beta}) \Rightarrow \chi_{K-2}^2$$

and we can use  $n(X_w^T \hat{\beta})^T (X_w^T [I^{(n)}(\hat{\beta}|\mathbf{X})]^{-1} X_w)^{-1} (X_w^T \hat{\beta})$  as our test statistic.

### 3.5.2 Bivariate Case

In bivariate case, we can test:

- $H_0$ : the two failure times have the same distribution;
- $H_0^p$ : the two failure times have proportional hazards.

For given paired data  $(X_{1i}, \delta_{1i}, X_{2i}, \delta_{2i}), i = 1, \dots, n$ , to test  $H_0$  and  $H_0^p$ , we model the log hazard functions of the two failure times with the same regression space  $\mathcal{B}$ . Then the log hazard regressions are

$$\begin{aligned} \alpha_1(t|\beta_1) &= \sum_{j=1}^p \beta_{1j} b_j(t) \\ \alpha_2(t|\beta_2) &= \sum_{j=1}^p \beta_{2j} b_j(t). \end{aligned}$$

If the two failure times have the same distribution, then

$$\alpha_1(t|\beta_1) - \alpha_2(t|\beta_2) = \sum_{j=1}^p (\beta_{1j} - \beta_{2j}) b_j(t) \equiv 0.$$

Therefore, testing  $H_0$  is equivalent to testing

$$H_0^*: \beta_{1j} - \beta_{2j} = 0 \text{ for } j = 1, \dots, p.$$

Then we can rewrite  $H_0^*$  as

$$H_0^* : X^T \underline{\beta} = \mathbf{0}$$

for an appropriate matrix  $X$ . For bases described in Section 3.2, by Corollary 2.5.4,

$$\sqrt{n} \hat{Q}^{-\frac{1}{2}}(\hat{\underline{\beta}} - \underline{\beta}) \Rightarrow N(0, I).$$

Then under  $H_0$ ,

$$n(X^T \hat{\underline{\beta}})^T (X^T \hat{Q} X)^{-1} (X^T \hat{\underline{\beta}}) \Rightarrow \begin{cases} \chi_{K-1}^2 & \text{if } \mathbf{b} \text{ is bounded;} \\ \chi_K^2 & \text{if } \mathbf{b} \text{ is bounded plus log tail.} \end{cases}$$

So, we use  $n(X^T \hat{\underline{\beta}})^T (X^T \hat{Q} X)^{-1} (X^T \hat{\underline{\beta}})$  as test statistic.

If the two hazards are proportional, then

$$\alpha_1(t|\beta_1) - \alpha_2(t|\beta_2) = \sum_{j=1}^p (\beta_{1j} - \beta_{2j}) b_j(t) \equiv \text{constant}.$$

Hence, for the regression spaces and  $b_j$ 's defined in Section 3.2, the test for  $H_0^p$  is equivalent to the test for  $\beta_{1j} - \beta_{2j} = 0$  for  $j \neq K - 1$ . Then we can rewrite  $H_0^p$  as

$$H_0^p : X^T \underline{\beta} = \mathbf{0}$$

for an appropriate matrix  $X$ . By Corollary 2.5.4,

$$\sqrt{n} \hat{Q}^{-\frac{1}{2}}(\hat{\underline{\beta}} - \underline{\beta}) \Rightarrow N(0, I).$$

Then under  $H_0^p$ ,

$$n(X^T \hat{\underline{\beta}})^T (X^T \hat{Q} X)^{-1} (X^T \hat{\underline{\beta}}) \Rightarrow \begin{cases} \chi_{K-2}^2 & \text{if } \mathbf{b} \text{ is bounded;} \\ \chi_{K-1}^2 & \text{if } \mathbf{b} \text{ is bounded plus log tail.} \end{cases}$$

Therefore,  $n(X^T \hat{\underline{\beta}})^T (X^T \hat{Q} X)^{-1} (X^T \hat{\underline{\beta}})$  can be used as test statistic.

**Note:** When  $T_{1i}$ 's and  $T_{2i}$ 's are independent, the Cox proportional hazards model

$$h_C(t) = h_T(t) \exp(\gamma) \tag{3.5}$$

provides an estimate of the relative risk  $\gamma$ . Thus under independence assumption, we can also test  $H_0$  by testing  $H_0^c : \gamma = 0$ . To our knowledge, there is no non-parametric test for proportional hazards for dependent paired data as we defined in Chapter 2. We might use the usual Cox proportional model, assuming the two failure times are independent. However, while the usual estimate of  $\gamma$  may be good, the standard errors are probably biased.

## Chapter 4

# Application to the Diabetic Retinopathy Study

In this chapter we apply the log hazard regression model to data collected to study the effect of laser treatment on diabetic retinopathy (see Diabetic Retinopathy Study Research Group, 1981). We give a description of this study in Section 4.1, and analyze the data in Section 4.2. Finally, in Section 4.3, we discuss analysis of this data set using other models.

### 4.1 Data Description

Diabetic retinopathy is a complication associated with diabetes mellitus, which consists of abnormalities in the microvasculature within the retina of the eye. It is the major cause of visual loss in many industrialized countries (Murphy and Patz, 1978).

The Diabetic Retinopathy Study (DRS) was funded by the National Eye Institute in 1971 to investigate the effectiveness of laser photocoagulation in delaying the onset

of blindness for diabetic retinopathy patients. One eye of each patient was randomly chosen to receive photocoagulation and the other eye was observed without treatment.

A total of 1,742 patients was followed over several years. The endpoint used to assess the treatment effect was the occurrence of visual acuity less than 5/200 at two consecutively completed 4-month follow-ups. The only data available are from the 197 patients defined as high-risk by DRS criteria.

Of the 197 pairs of observations, approximately one-half (101/197) of the untreated eyes and one-quarter (54/197) of the treated eyes achieved the outcome after 5 years of follow-up. The histograms of the censored and uncensored data for treatment and control (Figure 4.1) show that more untreated eyes than treated eyes failed during the study and many patients left the study after about 3 years (36 months) or more from the start time.

The correlation between the uncensored observations of the treatment group and the control group is 0.28. This indicates possible dependence between the two failure times. We show the scatter plots in Figure 4.2. For this data set, the two censoring times are identical, that is,  $C_{Ti} = C_{Ci}$ .

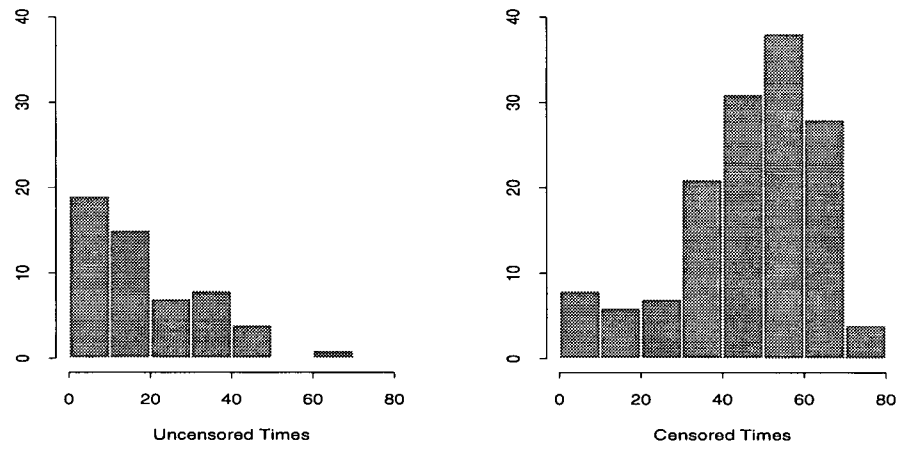
The primary goal of the DRS study was to assess the effectiveness of the laser photocoagulation treatment. A secondary goal was to assess the relative risk of blindness of the untreated and the treated eyes as a function of time.

## 4.2 Data Analysis

In this section we will address the following questions:

1. Do the failure times of the treated and the untreated eyes have the same distribution? That is, is there a treatment effect?

### Treatment



### Control

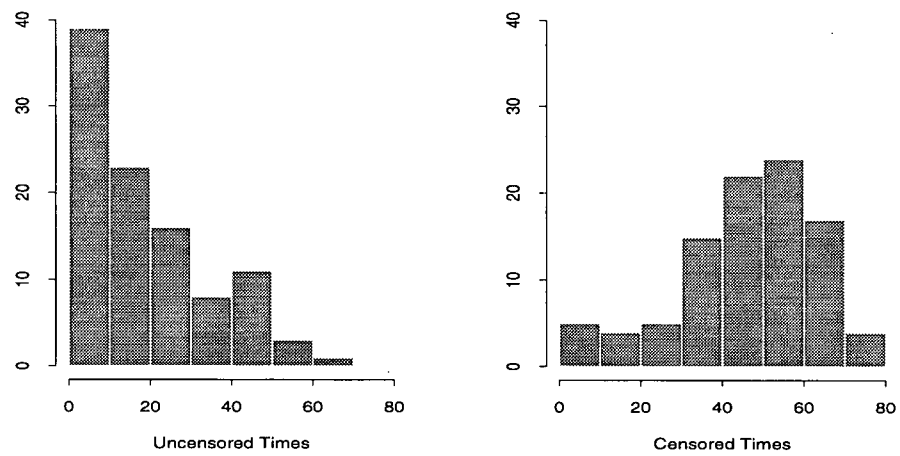


Figure 4.1: Histograms of observed times of the eye data.



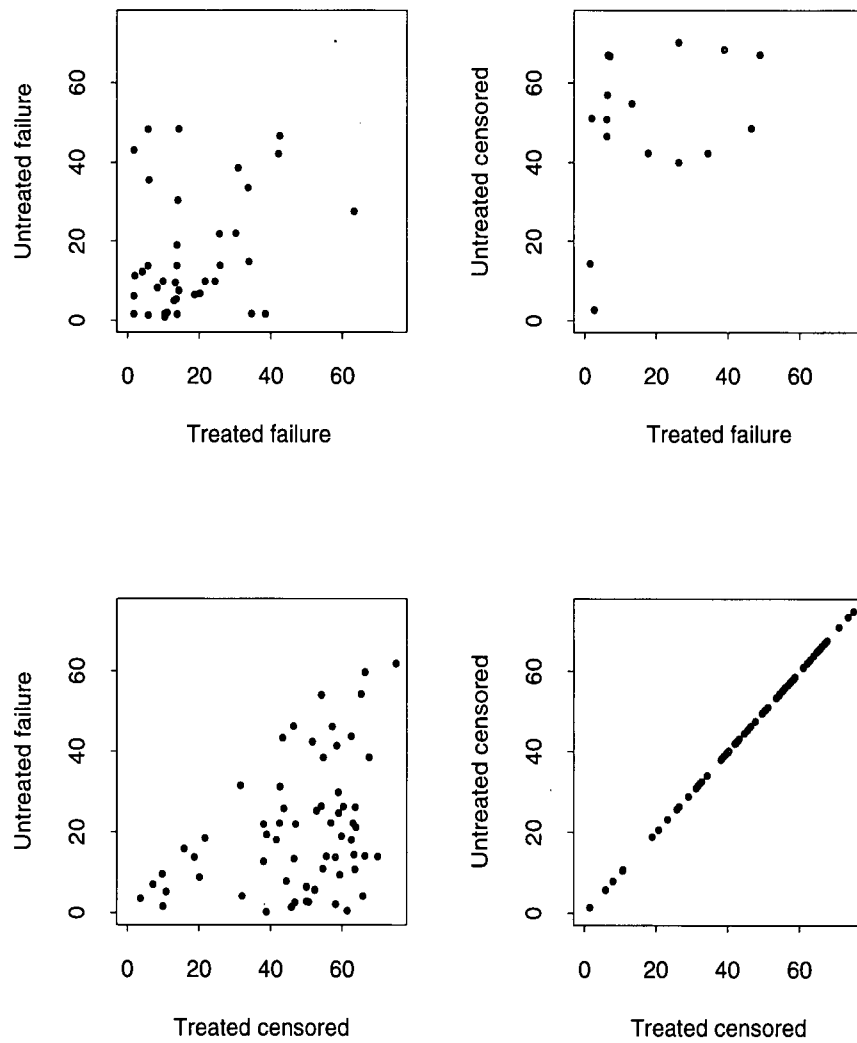


Figure 4.2: Scatter plots of the observations of the eye data.

2. What is the log hazards ratio? That is, what is the relative risk of blindness of the untreated eyes and treated eyes?

We carry out two analyses of the eye data. First, as described in Section 3.4 on knot selection, we let Kooperberg's *heft* algorithm choose the knots for the log hazard regression model (Model 1). This allows different knots to be used for the log hazards estimates of the failure times of the treated and the untreated eyes. In our second analysis (Model 2), we use the same regression space for the log hazards of the two failure times. The set of knots used is the union of the knots chosen by the *heft* algorithm for each failure time. With both models we can get an estimate of the log hazards ratio with 95% pointwise confidence intervals. We then test if there is a treatment effect, and if the two hazards are proportional with Model 2, as described in Section 3.5.

#### 4.2.1 Model 1

To estimate each failure time's log hazard, we use Kooperberg's *heft* algorithm with log tail  $\beta_p(t) \equiv \log(t)$  ( $c = 0$ ), and without specifying knots. The *heft* algorithm selects 1.5, 6.17, and 63.33 as knots for the regression model of the treated eyes and chooses no knots for the untreated eyes, which means that *heft* chooses a Weibull model for the failure time of the untreated eye. Then the log hazard regression model, Model 1, is:

$$\alpha_T(t|\boldsymbol{\beta}_T) = \beta_{T1}b_{T1}(t) + \beta_{T2}b_{T2}(t) + \beta_{T3}b_{T3}(t), \quad (4.1)$$

$$\alpha_C(t|\boldsymbol{\beta}_C) = \beta_{C1}b_{C1}(t) + \beta_{C2}b_{C2}(t). \quad (4.2)$$

As defined in Section 3.2,  $b_{T1}$  is linear on  $[0, 1.5)$ , cubic on  $[1.5, 6.17)$  and  $[6.17, 63.33)$ , constant on  $[63.33, \infty)$ ,  $b_{T2}$  and  $b_{C1}$  are constant functions, and  $b_{T3}(t) = b_{C2}(t) \equiv \log(t)$ . The estimates of  $\boldsymbol{\beta}_T$  and  $\boldsymbol{\beta}_C$  from *heft* are

$$\hat{\boldsymbol{\beta}}_T = (0.00017, -8.13, 0.69)^T$$

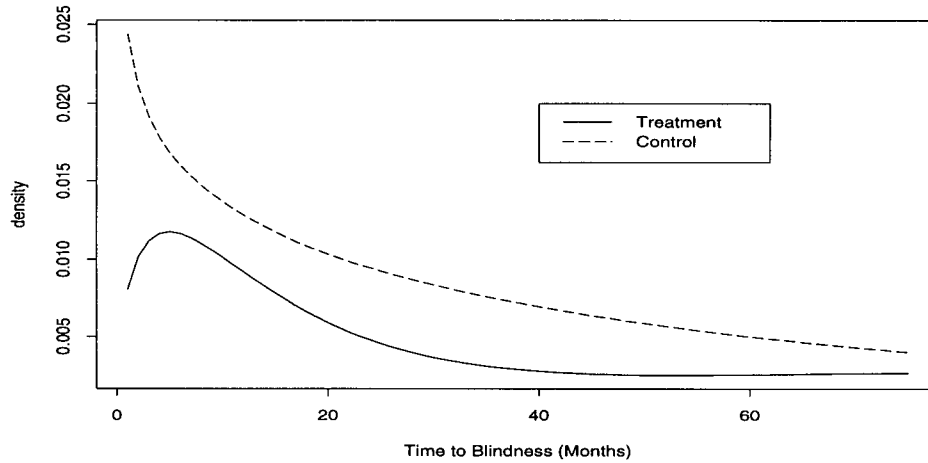


Figure 4.3: The estimated density functions of the eye data for Model 1.

and

$$\hat{\beta}_C = (-3.68, -0.18)^T,$$

respectively. The estimates of the marginal densities, survival functions, hazard functions, and the log hazards ratio are in Figures 4.3 to 4.6.

From the plot in Figure 4.3, we find that both estimated densities have a high value in the first twenty months. But the density corresponding to the untreated eye is higher than that of the treated eye during the observed time period. The estimated density of the failure time of the untreated eyes has its maximum at time zero, while the other density achieves its maximum at about month six. Since the regression model is for the log hazard, we calculate the estimated densities based on the estimate of log hazards as in equation (2.8). A pointwise confidence interval for the estimated densities would require integration of the upper and lower bounds of confidence intervals for the hazard function. This is difficult, so these confidence intervals have not been calculated.

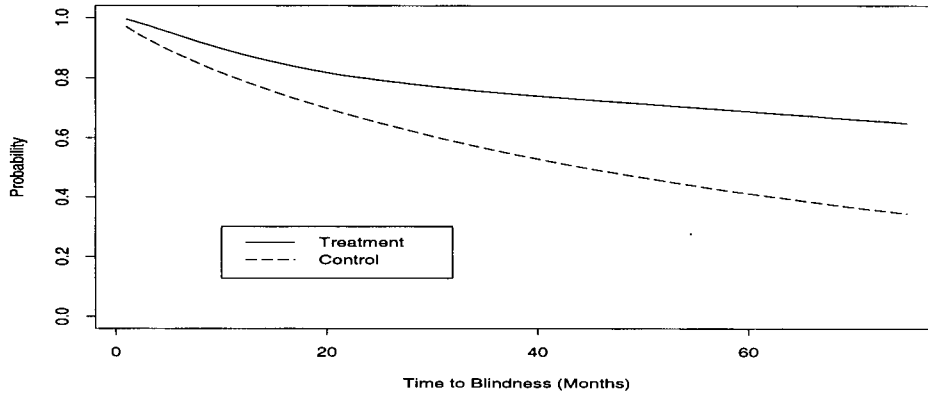


Figure 4.4: The estimated survival functions of the eye data for Model 1.

The estimated survival curves, in Figure 4.4, show that the estimated survival function of the treated eyes is always greater than that of the untreated eyes. Therefore, there appears to be a large treatment effect.

From the plots of the estimated hazard functions in Figure 4.5, we see that the risk of blindness in the untreated eye is much higher than that of the treated eye. The risk of blindness in the untreated eyes decreases with time while the treated eyes may have a maximal risk at  $t = 6$ . Based on the estimated hazard functions, we can see that the treatment may have delayed the onset of blindness for the patients for the first couple of months after the operation.

Figure 4.6 shows the estimated log hazards ratio with pointwise 95% confidence intervals. Values greater than zero indicate a higher risk of the control group and so it seems that the treatment has a beneficial effect. The estimate of log relative risk of blindness of the untreated eye to the treated eye is equal to 1.15 at  $t = 1$ , and decreases to 0.34 at about month six, then increases smoothly. After attaining its maximum value

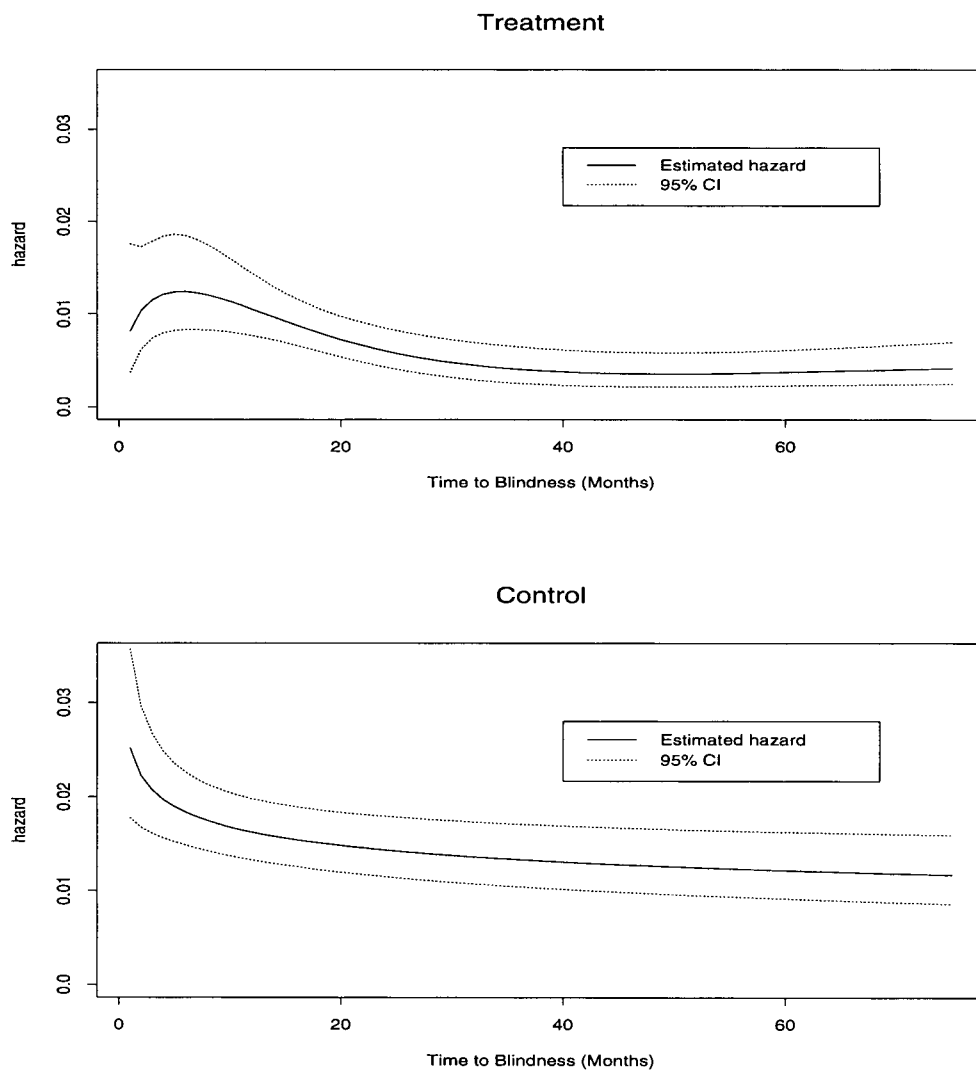


Figure 4.5: The estimated hazard functions of the eye data for Model 1 with pointwise 95% confidence intervals.

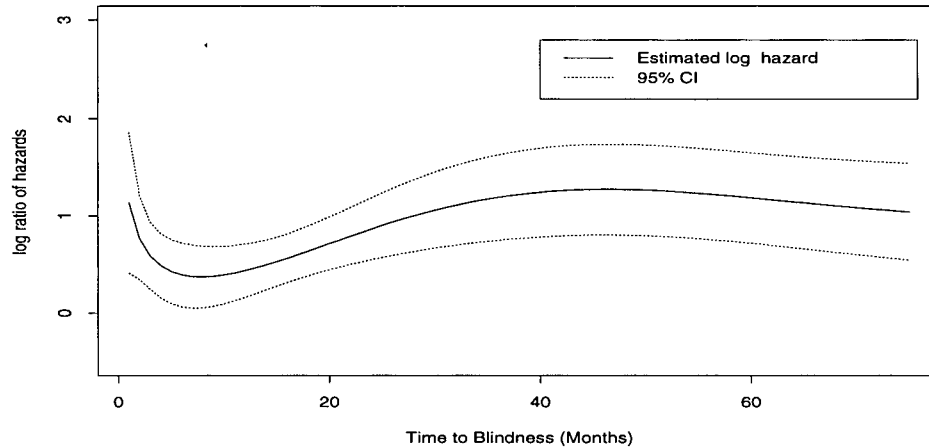


Figure 4.6: The estimated log ratio of the hazard functions of the eye data with pointwise 95% confidence intervals for Model 1.

of 1.186 at month 42, it decreases slowly.

Thus, based on Figure 4.6, the estimated log ratio of the two hazards gives an answer to questions 1 and 2 asked in the beginning of Section 4.2.

#### 4.2.2 Model 2

To test if there is a treatment effect, we assume the log hazard functions of the treated eye and the untreated eye are from the same regression space  $\mathcal{B}$ . So we will use the same restricted B-spline basis plus a log tail for the two regressions. To choose knots that define  $\mathcal{B}$ , we refer to the knots used for Model 1. As *heft* uses 1.5, 6.17, and 63.33 for the treatment group and no knots for the controls, we use 1.5, 6.17, and 63.33, the

union of the knots used for the two groups, as the knots to define our regression space  $\mathcal{B}$ . Then  $\mathcal{B}$  is a three dimensional space and the log hazard regression, Model 2, is:

$$\alpha_T(t|\beta_T) = \beta_{T1}b_1(t) + \beta_{T2}b_2(t) + \beta_{T3}b_3(t), \quad (4.3)$$

$$\alpha_C(t|\beta_C) = \beta_{C1}b_1(t) + \beta_{C2}b_2(t) + \beta_{C3}b_3(t), \quad (4.4)$$

where  $b_1 = b_{T1}$ ,  $b_2 = b_{T2}$ , and  $b_3 = b_{T3}$  as defined in Section 4.2.1. Using the *heft* algorithm with fixed knots, we get the estimates

$$\hat{\beta}_T = (0.00017, -8.13, 0.69)^T,$$

$$\hat{\beta}_C = (0.00007, -5.15, 0.16)^T.$$

As in Section 4.2.1, we calculate the estimates of the densities, survival, hazard functions, and the log hazards ratio. We give the plots in Figures 4.7 to 4.10. The results are similar to those from Model 1.

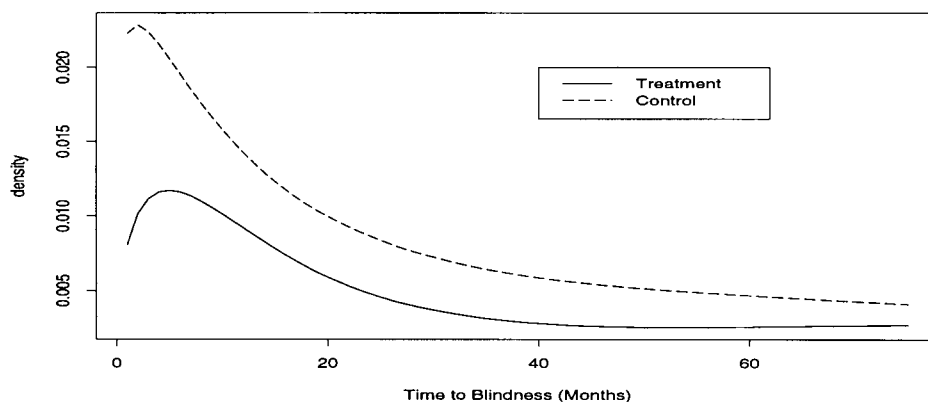


Figure 4.7: The estimated density functions of the eye data for Model 2.

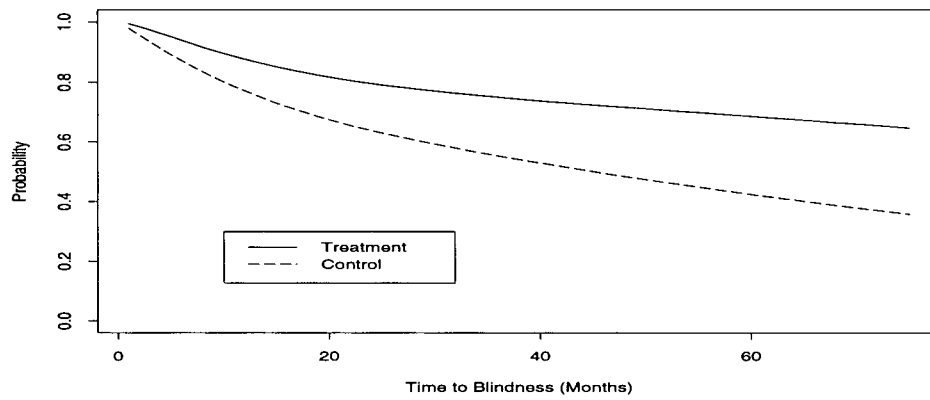


Figure 4.8: The estimated survival functions of the eye data for Model 2.

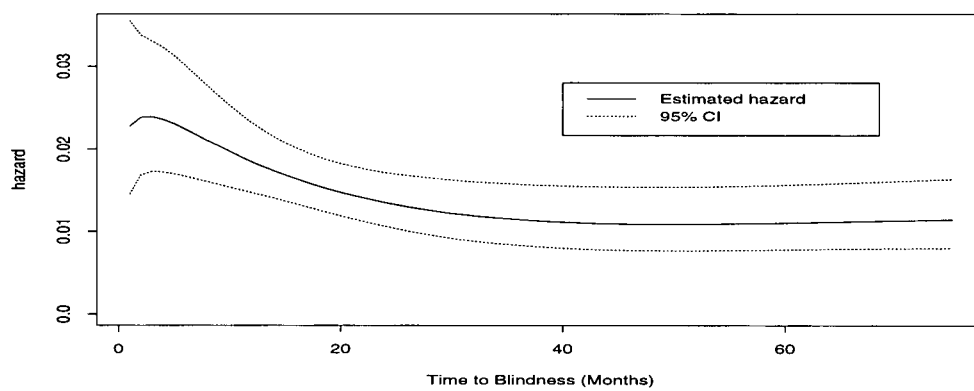


Figure 4.9: The estimated hazard function of the untreated eye for Model 2.



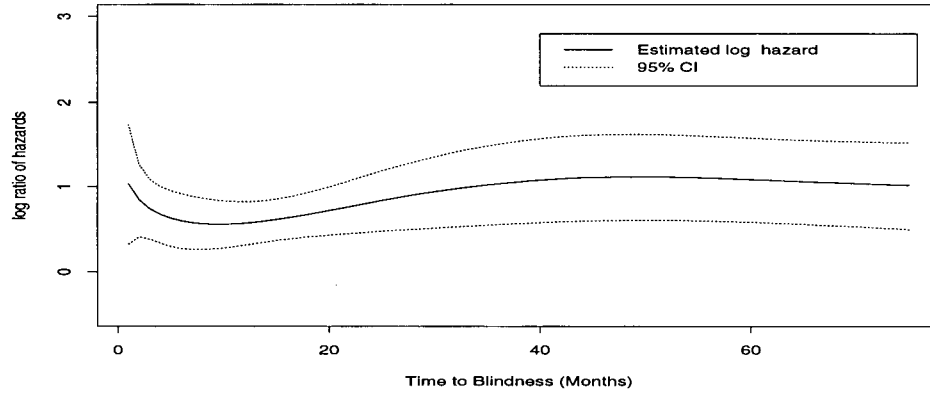


Figure 4.10: The estimated log ratio of the hazard functions of the eye data with pointwise 95% confidence intervals for Model 2.

Since Model 1 and Model 2 use the same regression space for the log hazard of the failure times of the treated eyes, the two models produce the same estimates for the failure times of the treated eyes. Thus we do not repeat the plots of these estimates. We compare the estimates from Model 1 and Model 2 for the untreated eyes in Figures 4.11 to 4.13.

From Figures 4.11 to 4.13 we see that Model 1 and Model 2 give almost the same estimates of the survival function of the untreated eyes. But the estimated densities and hazards look a little different.

Next we use Model 2 to test the hypothesis that the failure times of the treated eye and the untreated eye have the same distribution. By the definition of Model 2 from (4.3) and (4.4), this hypothesis is equivalent to

$$H_0 : \beta_{Tj} - \beta_{Cj} = 0, \quad j = 1, 2, 3.$$

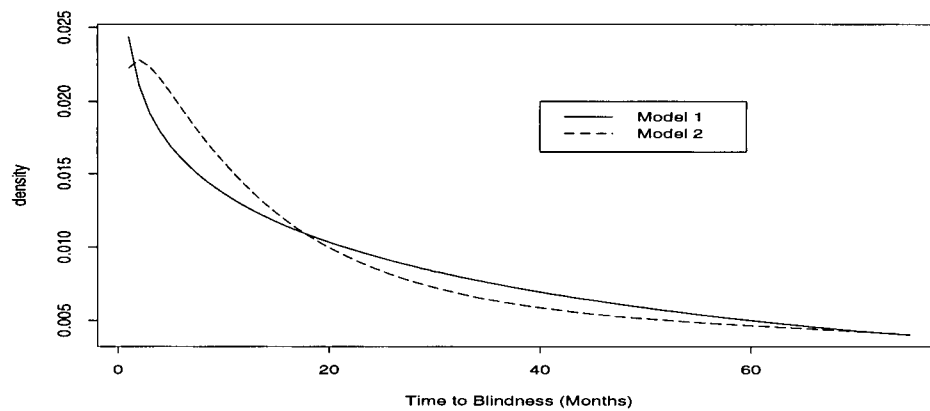


Figure 4.11: The estimated density functions of the failure times of the untreated eyes.

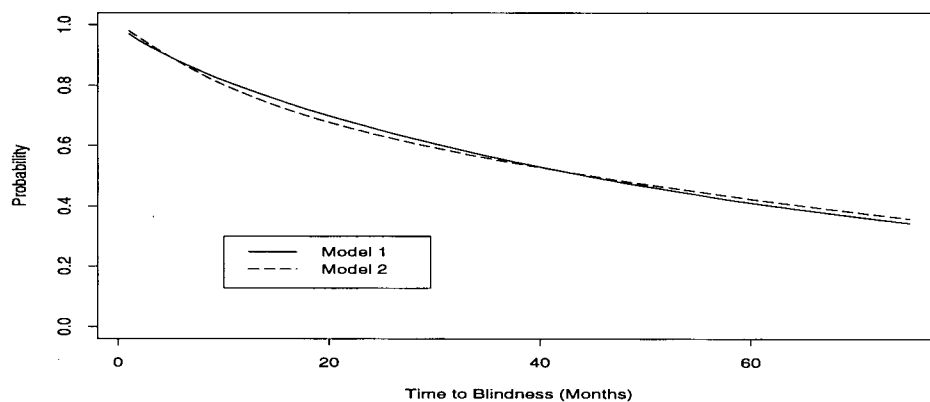


Figure 4.12: The estimated survival functions of the untreated eyes.

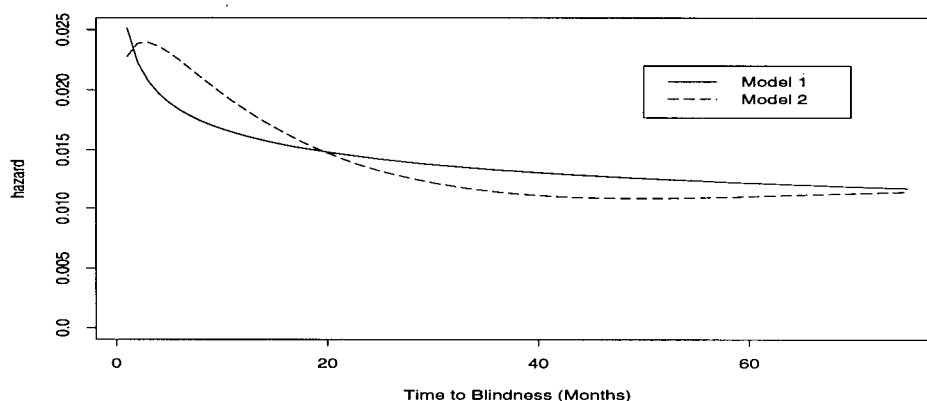


Figure 4.13: The estimated hazard functions of the failure times of the untreated eyes.

Using the test described in Section 3.5, we obtain the test statistics and p-values shown in Table 4.1.

Table 4.1: Test statistic and p-value for testing that the failure times of the treated eyes and the untreated eyes have the same distribution.

Test	$\chi^2_3$	p-value
$\beta_{Tj} - \beta_{Cj} = 0, j = 1, 2, 3$	48.56	0

Hence we may reject  $H_0$  and conclude that the two distributions are different. From our plots of the log hazards ratios in Figures 4.6 and 4.10, we conclude that the laser photocoagulation treatment had a significant effect in delaying the onset of blindness in patients with diabetic retinopathy.

## 4.3 Other Models Used to fit the Eye Data

Huster et. al. (1989), considered modeling the marginal distributions for the eye data as exponential and as Weibull or modeling the data via the Cox proportional hazards. We show the estimated marginal survival curves from fitting different models in Figures 4.14 and 4.15. The Kaplan-Meier estimate is also given. Note that for the treatment Model 1 and Model 2 are the same and for the control Model 1 is the Weibull. We see that all fit inside the confidence intervals of the Kaplan-Meier estimates except the estimates from the exponential model. But the estimated survival functions from our log hazard regression models are closer to the Kaplan-Meier estimates than those from the parametric models.

In this section we use Model 2, defined in Section 4.2.2, to decide which model(s) are appropriate to the eye data. Then, based on the test results, if we can choose a standard parametric model, say exponential, Weibull, or proportional hazards, we will use the selected model to test if there is a treatment effect.

### 4.3.1 Exponential Model

Since Model 2 from (4.3) and (4.4) includes the exponential distribution, we can use the model to test if the marginal distributions are exponential. As we discussed in Section 3.5, we test the two null hypotheses:

$$\beta_{Tj} = 0, j = 1, 3$$

and

$$\beta_{Cj} = 0, j = 1, 3.$$

Table 4.2 presents our test results.

Hence we reject the hypotheses that the failure times of the treated eye and the untreated eye are exponentially distributed.

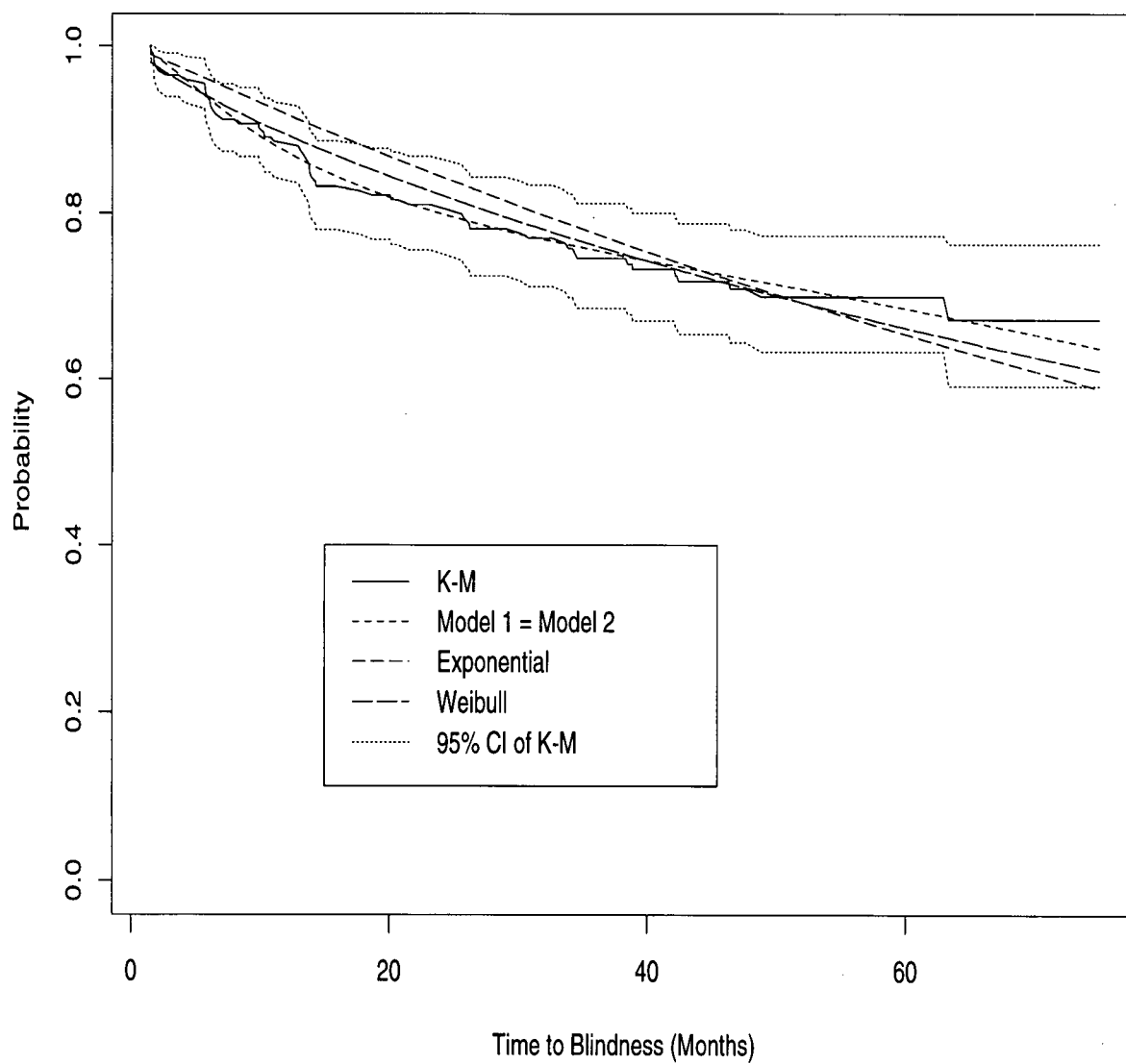


Figure 4.14: The estimated survival curves of the treated eye.

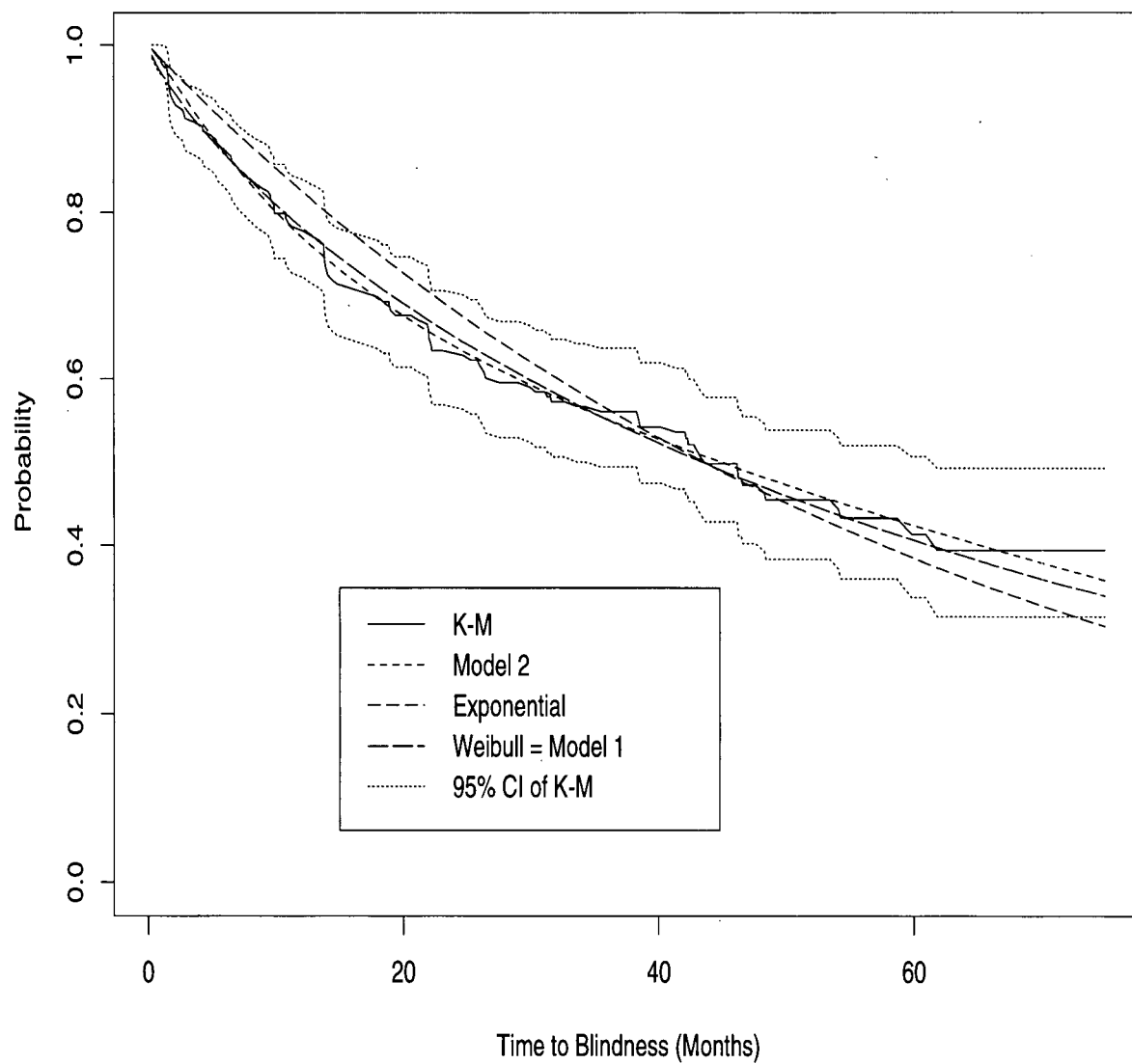


Figure 4.15: The estimated survival curves of the untreated eye.

Table 4.2:  $\chi^2$  test statistics and p-values for testing that failure times are exponentially distributed.

Test	$\chi^2_2$	p-value
$\beta_{Tj} = 0, j = 1, 3$	11.26	0.004
$\beta_{Cj} = 0, j = 1, 3$	8.14	0.017

### 4.3.2 Weibull Model

To test the hypotheses that the marginal distributions are Weibull, we test the two hypotheses:

$$\beta_{T1} = 0$$

and

$$\beta_{C1} = 0.$$

We show the test results in Table 4.3.

Table 4.3:  $z$  statistics and p-values for testing that failure times have Weibull distributions.

Test	$\hat{\beta}$	se	$z$	p-value
$\beta_{T1} = 0$	$1.73 \times 10^{-4}$	$6.28 \times 10^{-5}$	2.75	0.006
$\beta_{C1} = 0$	$6.94 \times 10^{-5}$	$3.59 \times 10^{-5}$	1.93	0.053

Thus we reject the hypothesis that the failure time of the treated eye has a Weibull distribution and we can not accept the hypothesis that the failure time of the untreated eye has a Weibull distribution.

### 4.3.3 Cox Proportional Hazard Model

Now we use Model 2 to test the hypothesis that the two hazards are proportional. As mentioned in Section 3.5, we need to test

$$H_0 : \beta_{Tj} - \beta_{Cj} = 0, \quad j = 1, 3,$$

since  $b_2$  is constant. The test statistics and p-values are shown in Table 4.4.

Table 4.4: Test statistic and p-value for testing that the two hazards of the treated eyes and the untreated eyes are proportional.

Test	$\chi^2_2$	p-value
$\beta_{Tj} - \beta_{Cj} = 0, j = 1, 3$	2.81	0.25

Thus we may assume that the two groups have proportional hazards, that is, we may assume that the relative risk of blindness of the untreated eye versus the treated eye is a constant.

Now we assume that the two failure times are independent and fit the Cox proportional hazards model (3.5) for the eye data. To test the treatment effect, we test  $\gamma = 0$ . The results are given in Table 4.5.

Table 4.5: Test results from the Cox proportional hazards model for the eye data.

Test	$\hat{\gamma}$	se	z-value	p-value
$\gamma = 0$	0.777	0.169	4.6	$4.2 \times 10^{-6}$

The Cox proportional hazards model also indicates that there is a treatment effect. Figures 4.16 to 4.19 compare the pointwise 95% confidence intervals of the log hazards



ratio from Models 1 and 2 and the Cox proportional hazards model, which assumes the two failure times are independent. We can see that the confidence interval from Model 2 is narrower than the confidence interval from the Cox proportional hazards model during the months 4 through 23, when the most failures in the both groups occurred.

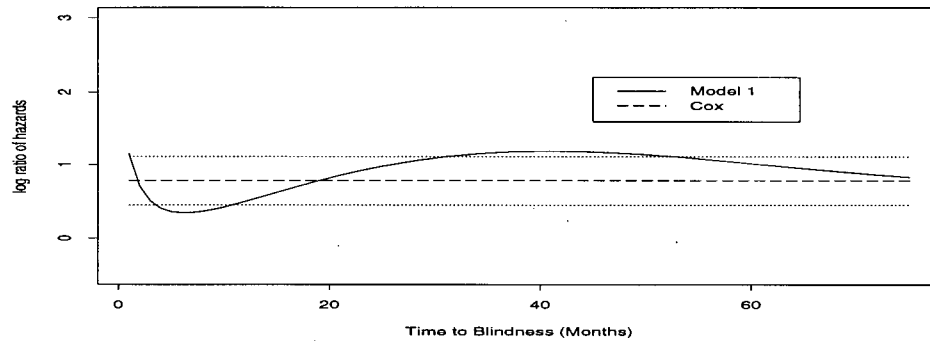


Figure 4.16: The estimated log hazards ratio from Model 1 and the Cox proportional hazards model with the pointwise 95% confidence intervals from the Cox proportional hazards model.

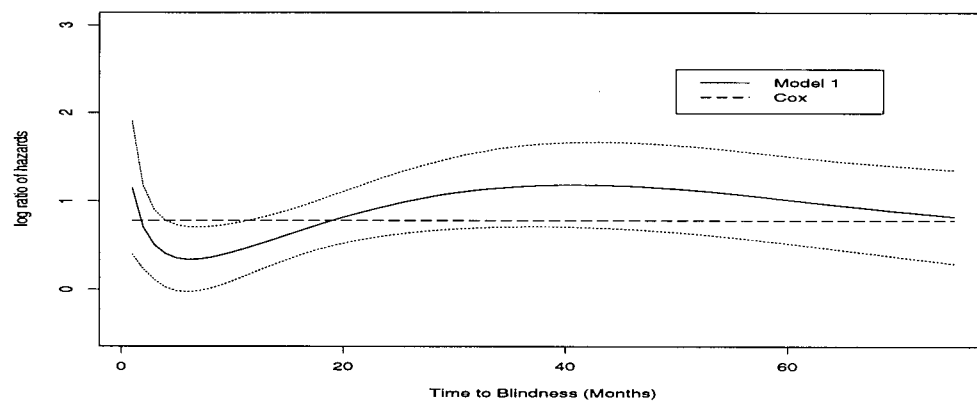


Figure 4.17: The estimated log hazards ratio from Model 1 and the Cox proportional hazards model with the pointwise 95% confidence intervals from Model 1.

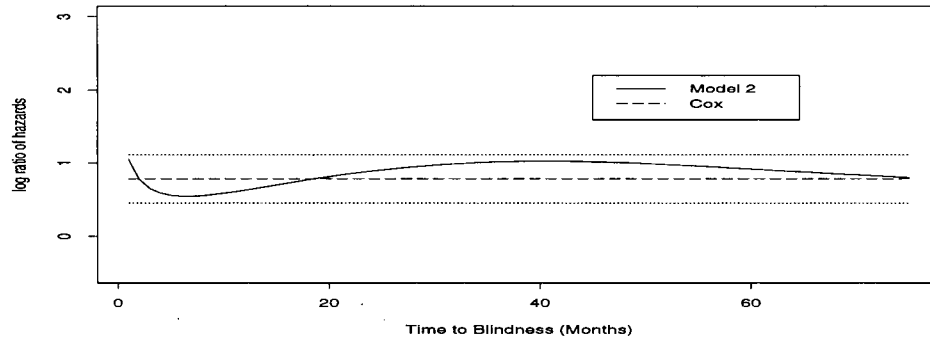


Figure 4.18: The estimated log hazards ratio from Model 2 and the Cox proportional hazards model with the pointwise 95% confidence intervals from the Cox proportional hazards model.

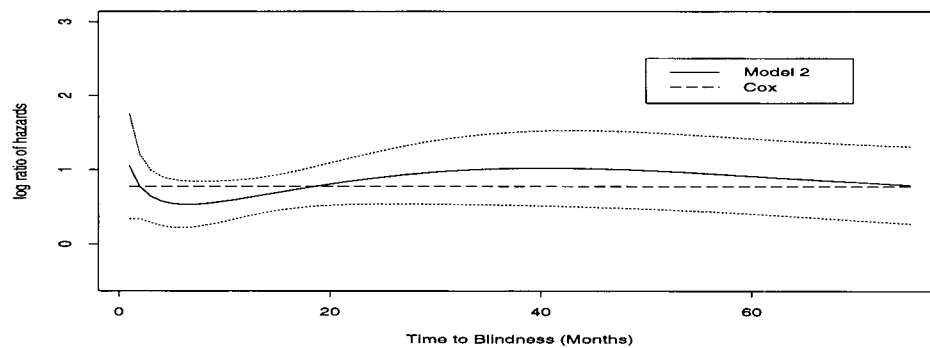


Figure 4.19: The estimated log hazards ratio from Model 2 and the Cox proportional hazards model with the pointwise 95% confidence intervals from Model 2.

# Chapter 5

## Simulation

This chapter contains discussion of a simulation study of estimates in the log hazard regression model. Our main aim is to check the bias and variability of our estimates for the log regression model. We consider different censoring rates and different correlation levels of the paired failure times. We also investigate the three test procedures: for exponential or Weibull marginal distributions and for proportional hazards, see Section 3.5.

Our study consists of two parts: the univariate case and the bivariate case. In the univariate case, presented in Section 5.2, we investigate the estimates of the marginal log hazard functions and the pointwise standard errors. We also test that the marginals follow exponential or Weibull distributions. In the bivariate case, presented in Section 5.3, we examine the estimated log hazards ratios and their estimated standard errors. We also test that the paired failure times follow the proportional hazards model.

In Section 5.1 we give a brief description of our data generation and model fitting. We present the simulation results in Sections 5.2 and 5.3 for the marginal and bivariate data, respectively.

In this simulation study we find that, in general, the estimates for the log hazards perform well except in the tails of the failure time distribution or when the censoring is high. The estimated standard errors slightly underestimate the true variability, which is fairly small. The estimated standard errors for the log hazards ratios do not depend on the correlation level too much. All of the estimates depend on the censoring rate. The lower the censoring, the better the results. We will give a summary of this study in Section 5.4.

## 5.1 Description of the Simulation Study

In this study for each distribution of  $T$  considered, we generate 200 pairs  $T$ s and  $C$ s in each simulation, where the  $T$ s are failure times, the  $C$ s are censoring times, and the  $T$ s and  $C$ s are independent. We choose two censoring rates  $c = 0.25$  or  $0.50$ , i.e.,

$$P(T > C) = 0.25 \quad \text{or} \quad P(T > C) = 0.50.$$

We generate the failure time  $T$  from distributions with log hazards of the form

$$\alpha(t|\boldsymbol{\beta}) = \beta_1 b_1(t) + \cdots + \beta_p b_p(t), \quad (5.1)$$

where  $\{b_1, \dots, b_{p-1}\}$  is a B-spline basis and  $b_p(t) \equiv \log(t)$ . We use three distributions for the marginal failure time  $T$ :

**Exponential :**

$$\alpha_T(t) \equiv -\log(20). \quad (5.2)$$

In this case, we choose the censoring time  $C$  with

$$\alpha_C(t) \equiv -\log(20)$$

for a 50% censoring rate or

$$\alpha_C(t) \equiv -\log(60)$$

for 25% censoring.

**Weibull :**

$$\alpha_T(t) = -3.68 - 0.18 \log(t). \quad (5.3)$$

The corresponding censoring time  $C$  has the log hazard

$$\alpha_C(t) = -3.68 - 0.18 \log(t)$$

for 50% censoring or

$$\alpha_C(t) = -4.69 - 0.18 \log(t)$$

for 25% censoring.

**B-spline :**

$$\alpha_T(t) = 0.00017 b_1(t) - 8.13 + 0.69 \log(t), \quad (5.4)$$

where  $b_1$  is linear on  $[0, 1.5)$ , cubic on  $[1.5, 6.17)$  and  $[6.17, 63.33)$ , and zero on  $[63.33, \infty)$ . See Sections 3.2 and 4.2 for the exact definition of  $b_1$ . The censoring time follows a Weibull with log hazard

$$\alpha_C(t) = -13.39 + 2 \log(t)$$

for 50% censoring or

$$\alpha_C(t) = -25.24 + 4 \log(t)$$

for a 25% censoring rate.

In all three cases, we choose the parameters for  $T$ 's distribution to fit the eye data in Chapter 4. We choose the exponential distribution with mean close to the empirical mean of the non-censored failure times of the treated eyes; the Weibull which is the distribution of the estimated distribution of the failure time of the untreated eye based on the given data. The B-spline model (5.4) is the estimated distribution of the failure times of the treated eye as calculated in Section 4.2. We use S-plus to generate data from an exponential or Weibull distribution. For the data as in the B-spline model (5.4), we first generate a random variable  $u \sim U(0, 1)$  then solve  $S(t) = u$  for  $t$  numerically. Since  $t$  is the  $(1 - u)$ th quantile of  $T$ 's distribution,  $t$  can be calculated via *heft*'s quantile function.

We use the following three methods to estimate the log hazard function.

**True Model** : Fit the data assuming it follows the true model. That is, estimate the parameters in (5.1) using the true  $b_j$ 's. In this case both the number and the locations of the knots do not depend on the data set.

**Quantile Knots** : Use the quantiles of the non-censored observations as the knots that define the  $b_j$ 's in (5.1). We consider two cases: three knots and six knots. In the three knots case, our knots are the quartiles. In the six knots case, our knots are the  $1/6, \dots, 5/6$  quantiles. Thus, the locations of the knots depend on the data set, but the number of the knots does not.

**Flexible Knots** : Choose the knots that define the  $b_j$ 's in (5.1) by Kooperberg's *heft* algorithm. In this case both the number and locations of the knots depend on the data set.

Thus we have six models for  $(T, C)$ , from the three distributions for the failure times and the two censoring times for each distribution. For each of these six models we run 500 simulations and calculate the above four estimates of the log hazard functions.

In the bivariate case, we will use the Clayton method to generate  $(T_1, T_2)$  from marginals for  $T_1$  and  $T_2$  and with a parameter  $\theta$ , (see Section 5.3.1). We consider two types of paired data:

**Proportional Hazards** : The failure time  $T_1$  and censoring time  $C_1$ , say for the treatment, are as in the univariate B-spline model. Then the censoring rates for  $T_1$  are 0.5 and 0.25. The failure time  $T_2$ , say for the control, is from the distribution with the log hazard

$$\alpha_2(t) = 0.00017 b_1(t) - 7.13 + 0.69 \log(t). \quad (5.5)$$

Then  $T_1$  and  $T_2$  have the proportional hazards. We assume that the treatment and control have the same censoring time  $C = C_1 = C_2$ .

**Non-proportional Hazards** : The failure time  $T_1$  and censoring time  $C_1$  of the treatment are as in the univariate B-spline model. The failure time  $T_2$  of the control is from the exponential model (5.2). Thus  $T_1$  and  $T_2$  do not have proportional hazards. Again we also assume that the treatment and the control have the same censoring time  $C$ . So the censoring rates of  $T_1$  are 0.5 and 0.25.

The correlation level of  $T_1$  and  $T_2$  is determined by a parameter  $\theta$ . We will give more details about  $\theta$  in Section 5.3.1. In this study we consider three correlation levels between  $T_1$  and  $T_2$ :  $\theta = 1$ , in which case  $T_1$  and  $T_2$  are independent;  $\theta = 1.5$  and  $\theta = 2.5$ . The bigger  $\theta$ , the more  $T_1$  and  $T_2$  are correlated.

We use the following three methods to estimate the log hazards ratios of the paired data.

**Cox** : Assume  $T_1$  and  $T_2$  are independent and have proportional hazards. Use the Cox proportional hazards model, as implemented in Splus's *coxph*.



**Same Knots** : Use the Quantile Knots method with the same six knots for  $T_1$  and  $T_2$  to estimate each marginal log hazard. The knots are the 1/7, 2/7, 3/7, 4/7, 5/7, 6/7 quantiles of the union of the non-censored observations of  $T_1$  and  $T_2$  which are less than 80. See our explanation in Section 3.4.

**Different Knots** : Use the Quantile Knots method with six knots for estimating  $T_1$ 's log hazard and six knots for  $T_2$ 's when  $T_1$  and  $T_2$  have proportional hazards, and three knots for  $T_2$ 's when  $T_1$  and  $T_2$  do not have proportional hazard. Then the knots are different for  $T_1$  and  $T_2$  in each simulation.

Hence in the bivariate case we have 12 models for  $(T_1, T_2, C)$ : two pairs of marginals for  $T_1, T_2$ , two censoring rates, and three correlation levels. For each of the 12 models we run 500 simulations and use the above three methods to estimate the log hazards ratios.

In all of these simulations, we will look at:

- plots of the pointwise averages and quantiles of the estimates to study the bias and variability of the estimates.
- plots of the pointwise standard deviations of the estimates to study the variability of the estimates.
- plots of the quantiles of the pointwise  $z$  values of the estimates to assess the reliability of the estimates. The pointwise  $z$  values are constructed as

$$z(t) = \frac{\text{estimate} - \text{true}}{\text{estimated SE}}.$$

- plots of the empirical distribution of the p-values for testing that the failure times have exponential or Weibull marginal distributions and for testing that the two failure times have proportional hazards.

Plots appear at the end of this chapter and in the appendix.

## 5.2 The Univariate Case

We present our simulation results for the models (5.2), (5.3), and (5.4) below in Sections 5.2.1, 5.2.2, and 5.2.3, respectively.

### 5.2.1 Exponential Model

In this section we apply the True Model, Quantile Knots, and Flexible Knots estimation methods to the data generated from the exponential distribution (5.2). Figure 5.1 shows the histograms of the failure times, the censoring times, the non-censored data and the censored data for those distributions. Since 60 is the 98th percentile of non-censored observations under 25% censoring and the 99.7th percentile of the non-censored observations under 50% censoring, we only plot estimates for values of  $t$  between 0 and 60. See Figures 5.2 to 5.4.

Note that in the Flexible Knots estimation method, if no knots are chosen and the coefficient of the log tail is zero, then an exponential model is fit. This occurs 468 out of 500 times for 25% censoring and 481 out of 500 times for 50% censoring. Hence the estimates from the Flexible Knots method are usually the same as from the True Model. So we do not include the results from the Flexible Knots estimation method in these figures.

Figure 5.2 shows the quartiles and empirical mean of the estimated log hazards from the True Model, Quantile Knots, and Flexible Knots methods and the true log hazard function for the exponential model (5.2). As expected, we find that the estimates from the True Model method are the closest to the true log hazard. For the Quantile Knots method, using three knots seems better than using six knots, which is somewhat surprising. In general, we expect that the more knots are used, the smaller the bias of the estimates but the larger the variability. The six knots estimates are more variable

but they are also more biased. This may be since the true distribution is exponential and so no knots are needed. Using more knots would not reduce the bias but it would increase the variability of the estimates. It is not surprising that the higher the censoring rate, the higher the bias of the estimates.

Figure 5.3 shows the pointwise standard deviations of the estimated log hazards and the quartiles and the empirical mean of the estimated standard errors of the estimated log hazards. It is normal that the True Model method has the smallest pointwise standard deviations. The standard deviations using six knots are bigger than when using three knots. The higher the censoring rate, the bigger the standard deviations. Comparing the the pointwise standard deviations of the estimated log hazards with the pointwise quantiles of the estimated standard errors, we find that the bias of the estimated standard errors is the smallest with True Model method and the largest when using six knots. The higher censoring rate causes a bigger bias of the estimated standard errors except with the True Model method.

We give the 97.5 and 2.5 percentiles and the quartiles of the pointwise z-values of the 500 estimated log hazards in Figure 5.4. We expect that they are between  $-2$  and  $2$ . We find that they are not out of this range too much for all estimates. Thus, pointwise confidence intervals based on the estimates would be reliable.

Figure 5.5 presents the histograms and the qq-plots of

$$z_i = \frac{\hat{\beta}_i - \beta_i}{\hat{se}_{\beta_i}},$$

$i = 1, 2, 3$ , the standardized estimates of  $\beta$  with the Quantile method using three knots. Those figures show that the estimated  $\beta$ s are approximately normal. For the method using six knots, the graphs look the same. We do not show them here.

We also check three test procedures that the failure times follow an exponential distribution. Each procedure involves fitting the log hazard by same regression model and then using a  $\chi^2$  test that some regression parameters are equal to zero. First we fit

the data from model 5.2 to a Weibull model, that is,

$$\alpha(t|\boldsymbol{\beta}) = \beta_1 \log(t) + \beta_2.$$

Then we use  $(\hat{\beta}_1/\hat{se}(\beta_1))^2$  as the test statistic to test  $\beta_1 = 0$ . This is the usual parametric test of exponential versus Weibull. When using the Quantile Knots method, equation (5.1) can be written as

$$\alpha(t|\boldsymbol{\beta}) = \beta_1 b_1(t) + \beta_2 b_2(t) + \beta_3 b_3(t)$$

when using three knots and

$$\alpha(t|\boldsymbol{\beta}) = \beta_1 b_1(t) + \cdots + \beta_5 b_5(t) + \beta_6 b_6(t)$$

when using six knots. As discussed in Section 3.5.1, the test is equivalent to test that

$$\beta_1 = \beta_3 = 0$$

when using three knots and

$$\beta_1 = \cdots = \beta_4 = \beta_6 = 0$$

when using six knots. Using the  $\chi^2$  statistics discussed in Section 3.5.1, we can calculate p-values of the test for each simulation.

In Figure 5.6, we give plots of the empirical distributions of the p-value to test the hypothesis that the failure times have an exponential distribution. The plots in the top row in Figure 5.6 show the p-values for the test of exponential versus Weibull, and those in the second and third rows are the p-values for the test of exponential versus B-spline model as in (5.4) when using three and six knots, respectively. Since the null hypothesis is true, the p-value should be uniformly distributed. Thus we expect that each plot would be a straight line. We find that the test works well and it does not depend on the knot selection or the censoring rates too much. However, the test rejects a little too often when the censoring rate is high.

### 5.2.2 Weibull Model

As in Section 5.2.1, we first present the histograms of the data in Figure 5.7. From the histograms we can see that, for both censoring rates, most non-censored failure times were less than 150. In fact 150 is about the 98th percentile for 50% censoring and the 95th percentile for 25% censoring of the observations. So in Figures 5.8 to 5.10, we end our plots at time 150.

As we mentioned in Section 3.4, the large range of the distribution of non-censored failure times may cause numerical problems for the Quantiles Knots and the Flexible Knots methods. This happens in our simulations for model (5.3) when we use the Flexible Knots method. Using the knots chosen by the *left* algorithm we get four warning messages under 25% censoring and ten warning messages under 50% censoring. Moreover, we can not carry out the calculation for the estimated standard error for any simulated data sets due to the big ranges of the knots. Hence with the Flexible Knots methods we can only get the results of the estimated log hazards for simulated data sets that did not produce warning messages. We can not get any results related to the estimated standard errors. However, the Quantile Knots Method works well for the data as in model (5.3).

Figure 5.8 shows summary plots of the estimated log hazards compared to the true log hazard of the model (5.3). We find that under 25% censoring the True Model and Quantile Knots estimation methods perform well. The estimates with the Flexible Knots method are highly biased when time is large, especially for high censoring.

When we look at the pointwise standard deviations in Figure 5.9, as expected, the True Model method gives the smallest standard deviations. The Quantile Knots method using six knots has the largest standard deviations.

When we compare the pointwise quartiles of our estimated standard error with the pointwise standard deviations of the estimated log hazards, we find that all the estimates

look fine, except that the estimates are a little bit too small after time 40. The censoring rate effect is the same as in the exponential model: the higher the censoring rate, the bigger the bias and variability of the estimates of the standard errors, particularly for small values of time.

Figure 5.10 shows the pointwise 2.5 and 97.5 percentiles and quartiles of the empirical  $z$  values of the estimated log hazards. They are as expected for standard normal random variables.

In Figure 5.11, we give the histograms and qq-plots of the estimates of  $\beta$  from the Quantile Knots method with six knots under 50% censoring. From the plots we can see that the estimates of  $\beta$  are approximately normal. For the True Model and three knots estimates, the graphs look the same. We do not present them here.

For test procedures, we check not only the tests that data follow an exponential distribution as in Section 5.2.1, but also test that data follow a Weibull distribution. For the Weibull test we fit the data using  $b_j$ 's from the Quantile Knots estimation method. We use the test statistics discussed in Section 3.5.1 to test the hypothesis that the data follow a Weibull distribution. The test is equivalent to testing that  $\beta_1 = 0$  when using three knots and that  $\beta_1 = \dots = \beta_4 = 0$  when using six knots. Figures 5.12 and 5.13 show the plots of the empirical distribution functions of the p-values for testing the null hypotheses that the distribution is exponential and that the distribution is Weibull, respectively. We expect a high proportion of small p-values in Figure 5.12 and a straight line in Figure 5.13.

From Figure 5.12, we see that under 25% censoring the True Model method and the Quantile Knots method with six knots has the lowest power. Increasing censoring to 50% results in lower power for all methods.

From Figure 5.13, we find that the test for Weibull works very well and it depends on neither the number of the knots nor the censoring rates.

### 5.2.3 The B-spline Model

We first look at the histograms of the failure and censoring distributions, the censored failure times, and the non-censored failure times presented in Figure 5.14. From the histograms for the non-censored failure times, we see a high proportion of times on the interval  $[0, 20]$ , with the remaining times almost uniformly distributed on  $[20, 150]$  for censoring rate 25% and on  $[20, 100]$  for censoring rate 50%.

As in the Weibull model (5.3) case, the large range of the distribution of non-censored failure times causes some numerical problems for the B-spline model when we use the Quantiles Knots and the Flexible Knots methods. Using the Flexible Knots method, that is, using the knots chosen by the *heft* algorithm, we can calculate only 113 estimated standard errors out of 500 simulations under 25% censoring and 306 under 50% censoring. The same problem occurs when we use six knots with the Quantile Knots method. To avoid those problems, when we use six knots we choose the knots which are the quantiles of the non-censored failure times less than 80. We present the simulation results in Figures 5.2.15 to 5.2.19. The results for the Flexible Knots method are based on the estimates that we can get.

By comparing the estimated log hazards with the true hazard function in Figure 5.15, we find that the biases of all estimates are fairly small when time is less than 70 under 25% censoring and when time is less than 40 under 50% censoring. We also find that those two numbers, 70 and 40, are close to the medians, 70.82 and 36.89, of the non-censored failure times in the two models. The estimates with the Flexible method perform well until time 100 under 25% censoring. We expect that using larger knots would get better estimates at tail part if there were not numerical calculation problems.

Figure 5.16 gives the pointwise standard deviations of the estimated log hazards and the quantiles and the empirical mean of estimated standard errors. From those plots we find that the True Model method and using three knots give the similar pointwise

standard deviations which are smaller than the pointwise standard deviations with using six knots. But the difference between using six knots and using three knots is no as big as with data generated according to the Exponential and Weibull models. All of the estimates have the biggest variability at time zero. The censoring rate effect is as the same as for data generated as in models (5.2) and (5.3). The biases of the estimated standard error are very small except with the Flexible Knots method.

Next we look at the plot of the 2.5 and 97.5 quantiles and the quartiles of the pointwise  $z$  values of the estimated log hazards from the True Model and Quantile Knots methods in Figure 5.17. We find that under 25% censoring the estimates look reliable for time is less than 70. Under 50% censoring the reliability of the estimates is not too bad as the time is less than 40, which match what we see from Figure 5.2.14. Different from in the Exponential model and Weibull model cases, for the B-spline model, the estimates with using six knots look more reliable than those with three knots. It does not surprise us since we expect using six knots would obtain better estimates than using three knots. To see if we can rely on the number of knots used by *heft* algorithm, we also check the numbers of knots used by *heft* algorithm. The average of the numbers of knots used by *heft* in the 500 simulations is 5.18, which is close to 6.

The histograms and the qq-plots of the estimated parameters  $\hat{\beta}_1, \hat{\beta}_2$  and  $\hat{\beta}_3$ , in Figure 5.18, show that these estimates look normally distributed but not with mean zero.

We also test the hypothesis that the data follow a Weibull distribution. For the Quantile Knots estimation method, the test procedures are the same to those we used for the data from the Weibull distribution (5.3) in Section 5.2.2. For the True Model method, we test  $\beta_1 = 0$ . For the Flexible Knots method we test  $\beta_1 = \cdots = \beta_{K-2} = 0$ , where  $K$  is the number of the knots chosen by the *heft* for each simulation. The test statistics are the same as we discussed in Section 3.5.1. Figure 5.19 shows that, under 25% censoring, by all the four estimation methods the test procedures for that the failure



times have a Weibull distribution have a high power except that by using six knots.

## 5.3 The Bivariate Case

In this section we generate the paired data  $(T_1, C_1, T_2, C_2)$ .  $T_1$  and  $T_2$  are from (5.2), (5.3) or (5.4). We assume  $C_1 = C_2$ , and the censoring times are defined according to the distribution of  $T_1$  from (5.2), (5.3) or (5.4). We are interested in investigating our estimated log hazards ratios and the estimates of their standard errors. We will also test for proportional hazards.

In Section 5.3.1 we introduce the method we use to generate dependent data for given marginal distributions. Then we present the simulation results in Sections 5.3.2 to 5.3.4.

### 5.3.1 Generation of Paired Dependent Data

Clayton (1978) proposed a family of bivariate distributions for survival times. Let  $S_1$  and  $S_2$  denote the marginal survivor functions for each member of a pair of failure times  $(T_1, T_2)$ . The joint survivor function for the Clayton model with parameter  $\theta$  is

$$S(t_1, t_2, \theta) = \begin{cases} [S_1(t_1)^{1-\theta} + S_2(t_2)^{1-\theta} - 1]^{-\frac{1}{\theta-1}}, & t_1, t_2 > 0, \theta > 1 \\ S_1(t_1)S_2(t_2), & t_1, t_2 > 0, \theta = 1 \end{cases} \quad (5.6)$$

When  $\theta = 1$ , the failure times are independent.

The parameter  $\theta$  can be written in terms of two conditional hazard functions. If we denote the hazard for the conditional distribution of  $T_1$  given  $T_2 = t_2$  and the hazard for  $T_1$  given  $T_2 > t_2$  by  $h_{T_1|T_2=t_2}(t_1)$  and  $h_{T_1|T_2>t_2}(t_1)$ , respectively, then in this model  $h_{T_1|T_2=t_2}(t_1)/h_{T_1|T_2>t_2}(t_1) = \theta$ . See Clayton (1978).

For given  $S_1, S_2$  and  $\theta$  we generate bivariate data  $(T_1, T_2)$  satisfying the Clayton model using the following:

(A) Generate  $T_1 = t_1$  according to  $S_1$ ;

(B) Generate  $T_2 = t_2$  according to the conditional distribution of  $T_2$  given  $T_1 = t_1$ .

(A) can be carried out as in the univariate case. For (B), when  $\theta > 1$ , the conditional probability is (abusing notation slightly),

$$\begin{aligned}
 P(T_1 > t_2 | T_1 = t_1) &= \frac{P(T_2 > t_2 \text{ and } T_1 = t_1)}{P(T_1 = t_1)} = \frac{\partial S(t_1, t_2, \theta) / \partial t_1}{S'_1(t_1)} \\
 &= \frac{-\frac{1}{\theta-1} [S_1(t_1)^{1-\theta} + S_2(t_2)^{1-\theta} - 1]^{-\frac{1}{\theta-1}-1} (1-\theta) S_1(t_1)^{-\theta} S'_1(t_1)}{S'_1(t_1)} \\
 &= [S_1(t_1)^{1-\theta} + S_2(t_2)^{1-\theta} - 1]^{\frac{\theta}{1-\theta}} S_1(t_1)^{-\theta} \\
 &= 1 - F_{T_2|T_1=t_1}(t_2).
 \end{aligned}$$

We will generate  $U = u$ , a uniform  $[0, 1]$ , and solve  $u = F_{T_2|T_1=t_1}(t_2)$  for  $t_2$ :

$$\begin{aligned}
 u &= 1 - P(T_1 > t_2 | T_1 = t_1) \\
 &= 1 - [S_1(t_1)^{1-\theta} + S_2(t_2)^{1-\theta} - 1]^{\frac{\theta}{1-\theta}} S_1(t_1)^{-\theta}.
 \end{aligned}$$

Solving for  $S_2(t_2)$ , we have

$$S_2(t_2) = \left\{ [(1-u)^{\frac{1-\theta}{\theta}} - 1] S_1(t_1)^{1-\theta} + 1 \right\}^{\frac{1}{1-\theta}}. \quad (5.7)$$

Now solve Equation 5.7 for  $t_2$ . Let  $p_2$  denote 1 minus the right part of Equation 5.7 equal  $p_2$ . Then  $t_2$  is the  $p_2$ th quantile of the distribution function  $1 - S_2$ . If  $S_2$  has a standard distribution say mean one exponential, then  $t_2 = \text{qexp}(p_2)$  in Splus. For nonstandard distributions, solve  $S_2(t_2) = 1 - p_2$  for  $t_2$  numerically.

Thus, for given marginal survivor functions  $S_1$  and  $S_2$  and the parameter  $\theta$ , we can generate a random bivariate variable  $(T_1, T_2) = (t_1, t_2)$  satisfying the Clayton model by the following procedure.

1. Generate a random variable  $T_1 = t_1$  according to  $S_1$ ;
2. Calculate  $S_1(t_1)$ ;
3. Generate a random variable  $U = u \sim U(0, 1)$  ;
4. Calculate

$$p_2 = 1 - \left\{ [(1 - u)^{\frac{1-\theta}{\theta}} - 1] S_1(t_1)^{1-\theta} + 1 \right\}^{\frac{1}{1-\theta}};$$

5. Solve  $S_2(t_2) = 1 - p_2$  for  $t_2$ .

Our choices of marginal distributions  $S_1$  and  $S_2$  follow the proportional hazards and non-proportional hazards model as defined in Section 5.1. Figure 5.20 presents the plots of the correlation coefficients of  $T_1$  and  $T_2$  vs  $\theta$ , where  $T_k$  has survivor function  $S_k$ ,  $k = 1, 2$ . In our simulation studies, we consider these models with  $\theta = 1, \theta = 1.5$ , and 2.5. When  $\theta = 1$ ,  $T_1$  and  $T_2$  are independent.

### 5.3.2 Proportional Hazards Model

We present our simulation results for the proportional hazards model in this section. We address the effect of the correlation level and the comparisons of the Cox, Same Knots, and Different Knots estimates of the log hazards ratio. Recall that the Cox estimate assumes that  $\theta = 1$ . Since we had studied the effect of the censoring rate on the estimates of log hazards in the univariate case, we only present the results for 25% censoring here and give the results for 50% censoring in the Appendix.

Figure 5.21 shows the histograms of the two marginals and the non-censored failure times. We can see that there are very few non-censored observations larger than 150. For the same reason as in the univariate case, we only plot results in the the Proportional Hazards model up to time 150.

First we look at the plots of the pointwise quartiles and average of the 500 estimated log hazards ratios in Figure 5.22. Comparing them with the true log ratio, we find that the estimates from the Cox method have the smallest bias, as we expected. The plots show that at time values within the range of knots, the biases of the estimates from the Same Knots method and the Different Knots method are almost the same except that the estimates from the Different Knots method have bigger bias at time zero. It looks like the correlation level does not affect the Same Knots and Different Knots estimates too much. But it is interesting to see that the estimates from the Cox method have the smallest bias when  $\theta = 1.5$ .

Then we look at the plots of the pointwise standard deviations of the 500 estimated log hazards ratios in Figure 5.23. We see that the estimates of the log hazards ratio from the Cox method have the smallest variability. There is no big difference between the variabilities of the estimated log hazards ratio from the other two methods. We compare Figure 5.23 with Figure 5.16, which shows the variability of the marginal log hazards. It seems that near time 0 the variabilities of the estimated log hazards ratio with the Same Knots and Different Knots methods are similar to the variabilities of the estimated marginal log hazards.

When we compare the estimated standard errors with the pointwise standard deviations in Figure 5.23, we find that the estimates of the standard error from the Cox methods have the smallest bias when  $\theta$  is 1 and 1.5, that is when the paired data are independent or only slightly correlated. But when  $\theta$  is 2.5, the estimated standard errors from the other two methods have smaller bias than the Cox method. The reason for the increasing bias of the estimated standard errors from the Cox method might be the violation of the assumption of independence of the two failure times. Over all three correlation levels, the biases of the estimated standard errors from the Same Knots method and the Different Knots method are very close to each other.

When we look at the plots of the quantiles of the pointwise  $z$  values of the estimated log hazards ratios, Figure 5.24, we note that the estimated log hazards ratios from the Cox method are more reliable than the estimates from the other two methods, and the best case for the Cox method is when  $\theta = 1.5$ . For the Same Knots and Different Knots methods, the reliability of the estimates of the log hazards ratio is fairly good when time  $t$  is between 5 and 70, which is about the range of the knots. Once more the reliability of the estimates from the Same Knots and Different Knots methods does not depend on the correlation levels, which does not surprise us because our estimated standard errors take the correlation into account.

Finally, we test the hypothesis that the failure times have proportional hazards by first estimating the two log hazards using the Same Knots method. We then test  $\beta_{1j} - \beta_{2j} = 0$ ,  $j = 1, 2, 3, 4, 6$ , as indicated in Section 3.5.2. Using the  $\chi^2$  test statistics discussed in Section 3.5.2, we can calculate p-values for the test.

We look at Figure 5.25, the plots of the empirical distributions of the p-value for testing that the failure times have proportional hazards. The null hypothesis is rejected far more than we expected. We expected each plot would be a straight line since the null hypothesis is true and the p-value should be uniformly distributed. An interesting result is that the distribution curve is closest to the line  $y = x$  when the data are highly correlated ( $\theta = 2.5$ ).

Through the simulation results, we would like to say that for the data as in the proportional hazards model, the Cox method works well. The performance of the Same Knots and Different Knots methods is reasonable. The estimated standard errors from all three methods are good but those from the Cox model deteriorate as  $\theta$  increases.

### 5.3.3 Non-proportional Hazards Model

Figure 5.26 shows the histograms of the two marginals and the non-censored failure times for the data from the non-proportional hazards model. As with the proportional hazard model, we only plot results in the the non-proportional hazards model up to time 150.

First we look at Figure 5.27, the plots of the true log ratio and the pointwise quantiles and average of the 500 estimated log hazards ratios. As we expected, the Cox method does not work. The estimates from the other two methods are very close. These estimates have smallest bias when the data are independent. But it is hard to see any difference between  $\theta = 1.5$  and  $\theta = 2.5$ .

Next we look at Figure 5.28, the plot of the pointwise standard deviations of the 500 estimated log hazards ratios and the pointwise quartiles and mean of the 500 estimated standard errors. Since the Cox method does not work for estimating log hazards ratios for the non-proportional hazards data, we only discuss the results from the other two methods here. We find that the pointwise standard deviations from the Same Knots method are a little bit bigger than those from the Different Knots method. But the standard deviation from the Different Knots method has a bigger jump at  $t = 0$ . When we look at the variability of the estimated standard errors, we find that the bias difference between the Same Knots method and the Different Knots method is very small within the range of knots. The standard errors from the Same Knots and Different Knots methods are too small for  $t \geq 70$ . The Different Knots method performs slightly better in this range. The bias and the variability of the estimated standard errors do not depend on the correlation level of the paired data.

We also check the plots of the quantiles of the pointwise  $z$  values of the estimated log hazards ratios from the three methods, Figure 5.29. Those plots show that the quantile curves for the Same Knots and Different Knots methods mainly lie in the range  $(-2, 2)$

with the best performance when the data are independent. It is not surprising that the Cox method performs poorly since the failure times do not have proportional hazards.

Now we consider testing the hypothesis that the failure times have proportional hazards. The test procedures are the same as those we used for the data sets from the proportional model in Section 5.3.2. When we look at the plot of the empirical distributions of the p-values, Figure 5.30, we find that the test power is not high but the lowest power is when  $\theta = 1$ .

### 5.3.4 Effect of the Number of Knots

In this section we study how the number of knots used in each simulation affects the estimates of the log hazards ratios and the standard errors. The data sets are generated as in the proportional hazards model and the non-proportional hazards model. For the data set from proportional hazards model, we use three, six, and nine knots to estimate the marginal hazards of  $T_1$  and  $T_2$ . For the data generated from the non-proportional hazards model, we use three, six, and nine knots for failure time  $T_1$  but only three knots for  $T_2$ . For the data from proportional hazards model, the true hazard of  $T_2$  has three knots, see (5.5). In contrast, for the data from non-proportional hazards model,  $T_2$  is exponentially distributed. As we found in the univariate simulation (Section 5.2), *heft* will rarely choose more than three knots to fit this exponential. Therefore, we do not need to study fits with more than three knots.

The simulation results are as we expected, that is, the more knots, the smaller the bias of the estimates but the bigger the variability of the estimates. We show the results in Figures 5.31 to 5.37. From those graphs we find that the estimates of the log hazard ratio by using three knots have the smallest variability but the biggest bias. There is no big difference in bias between using six knots and using nine knots but the estimates using nine knots have the biggest variability.

When we look at the estimated standard errors, we find that the estimates using six knots have the smallest bias and variability.

Figures 5.39 to 5.42 show the plots of the quantiles of the pointwise  $z$  values of the estimated log hazards ratios from the three methods. It can be seen that the performance of the estimates using six knots is similar to those with nine knots and superior to those using three knots.

Figures 5.43 and 5.44 present the  $p$ -values to test that the two failure times have proportional hazards for the data as in the proportional model and non-proportional model, respectively. We find that, for both data sets, the number of knots does not effect the test too much but the test performs better when  $T_1$  and  $T_2$  are dependent than when  $T_1$  and  $T_2$  are independent.

Overall, for the data generated as in the proportional hazards and non-proportional hazards models, using six knots seems better than using three knots or using nine knots. Of course, the value of six may also depend on the sample size and the true data distribution.

## 5.4 Summary of Simulations

Through this simulation study, we find the following.

In the univariate case:

1. With all the estimation methods, the estimates of the log hazard perform well within the range of knots used. However, if the range of knots is large, the Flexible Knots method might cause some numerical calculation problems in estimating the log hazards and in calculating standard errors.
2. With the Quantile Knots method, the estimates of standard errors perform well



except that they slightly underestimate the variability of the estimated log hazards when the censoring rate is high.

3. The censoring rate affects the bias and variability of the estimated log hazards. The smaller the censoring rate, the better the estimates.
4. For all marginal models, the True Model method gives the best estimates for the log hazards.
5. The estimates using three knots look a little bit better than those using six knots for the data generated as in the exponential distribution (5.2) and the Weibull distribution (5.3). But for the data generated as in the B-spline model (5.4), the estimates with six knots look more reliable.
6. The Flexible Knots method gives better estimated log hazards than other methods when the ratio of the largest knot to the smallest knot is not too big. Otherwise calculation of the estimated standard errors and the estimated log hazards causes numerical problems.
7. The test procedures for testing that the failure times follow an exponential and Weibull distribution perform well.

In the bivariate case:

1. With the Same Knots and Different Knots estimation methods, the estimates of the log hazards ratio perform well within the range of knots used. There is not a big difference between the estimates from the two estimation methods.
2. For the data with proportional hazards, the Cox method gives the least-biased estimates of the log hazards ratios when the failure times are independent. For the data from the non-proportional hazards model, the Cox proportional model does not work at all.

3. The test that the failure times have proportional hazards does not perform well.

We expect a bigger sample size might improve the test procedure.

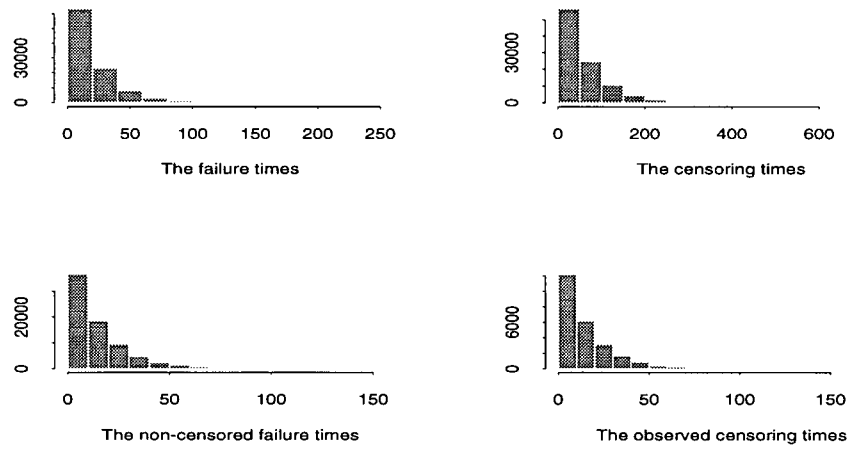
4. The correlation levels do not affect the estimated log hazards ratios too much.

But they affect the test that the failure times have proportional hazards. The test has the lowest power when the two failure times are independent. Further simulations are required to better understand the effects of correlation.

In this simulation we also check the test procedures for the hypothesis that the two failure times have the same distribution. But there are some numerical problems when we calculate p-values. We do not present the result here.

One more thing we would like to point out is that our estimated formulae for the standard errors assume for non-random knots. The procedures used in our simulations have random knots except for the True Model method in the univariate case. So this might be a reason that the estimated standard errors are a bit small.

$c = 0.25$



$c = 0.5$

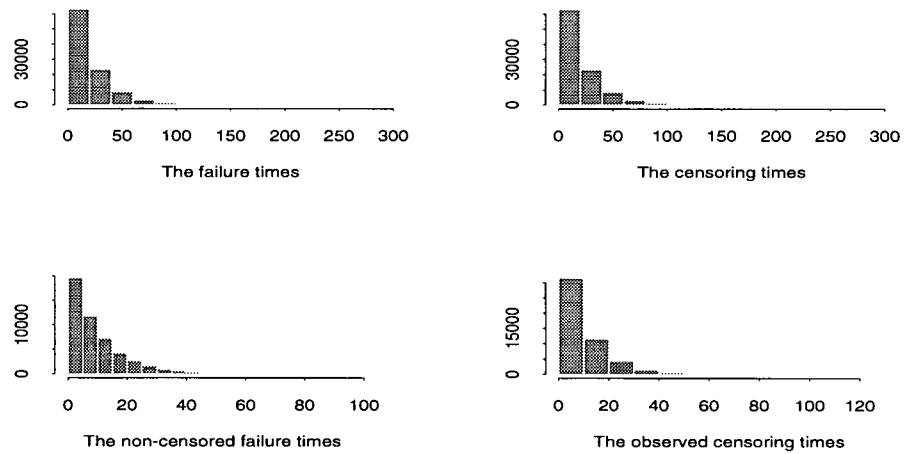


Figure 5.1: The histograms of the simulated failure times, censoring times, non-censored failure times, and observed censoring times for exponential data as in model (5.2).

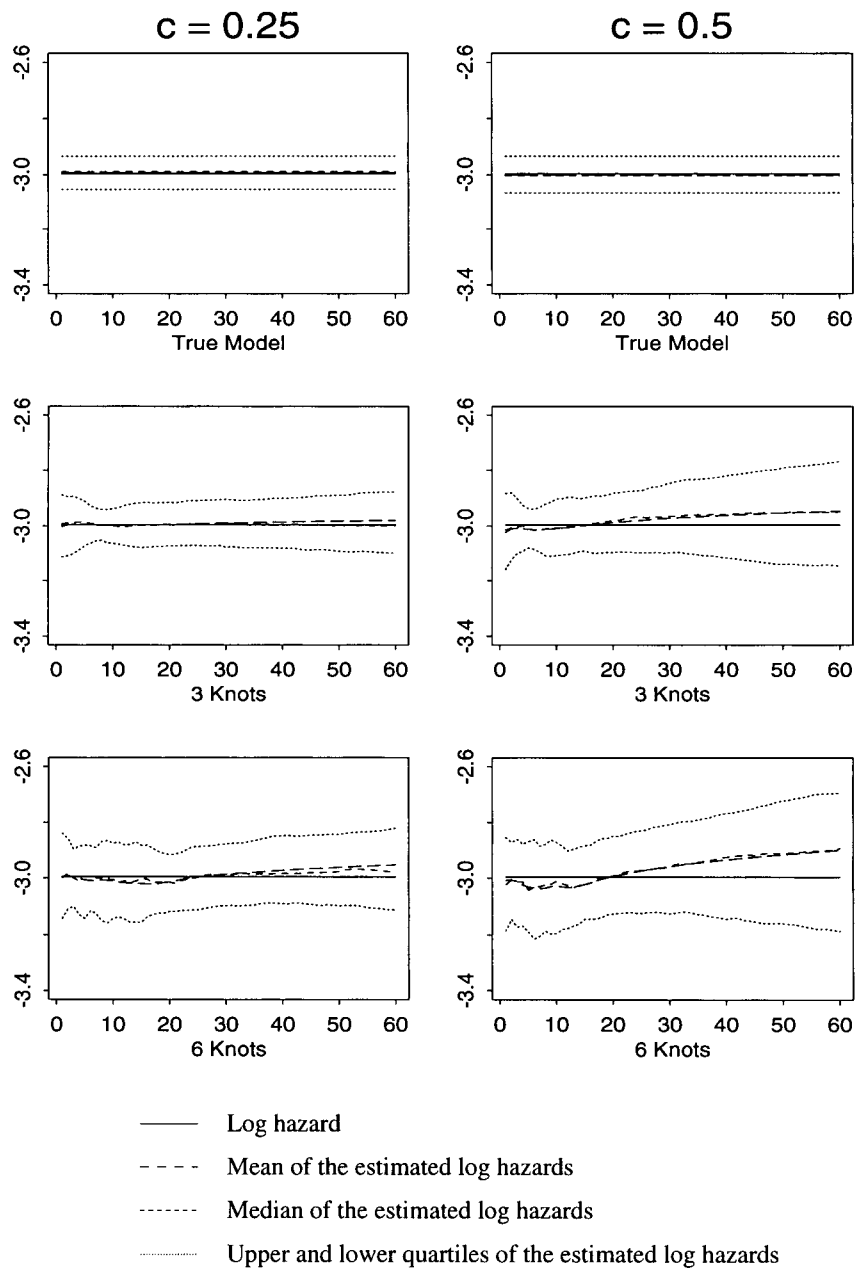


Figure 5.2: Log hazard of the exponential distribution (5.2) and the pointwise quartiles and empirical mean of the 500 estimated log hazards.

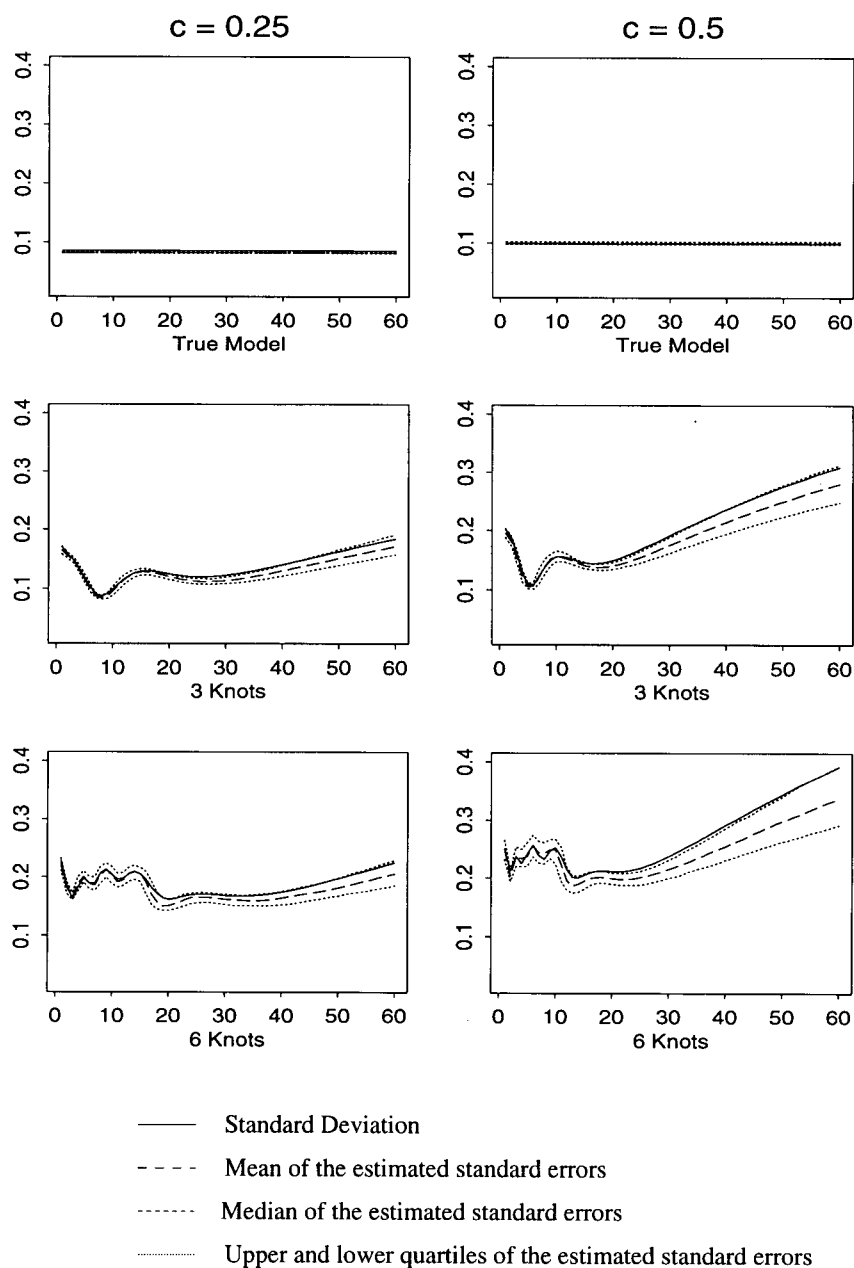


Figure 5.3: The pointwise standard deviations of the estimated log hazard for the exponential model (5.2) and the pointwise quartiles and empirical mean of the 500 estimated standard errors.

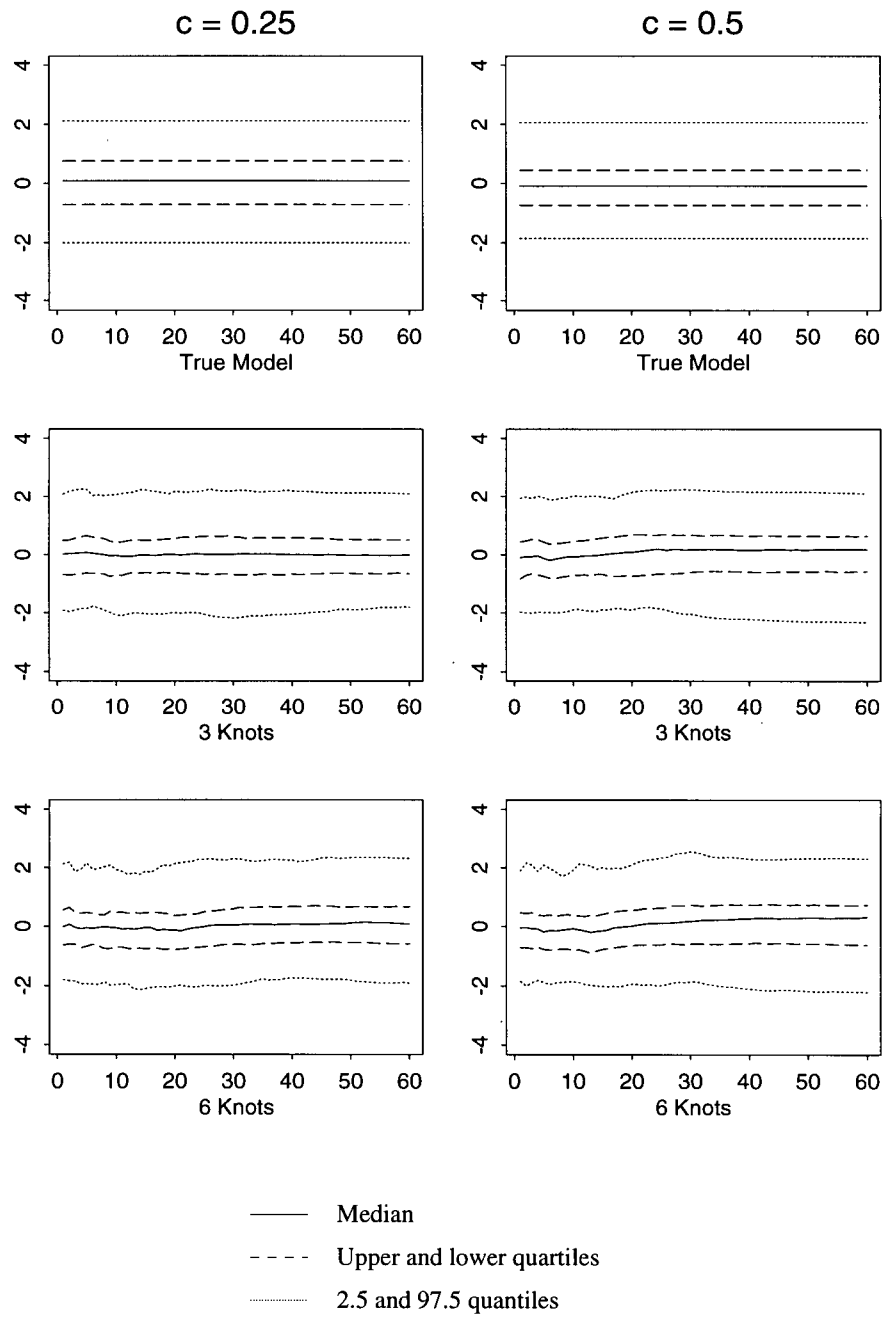


Figure 5.4: The quantiles of the pointwise z values of the estimated log hazards for the exponential distribution (5.2).

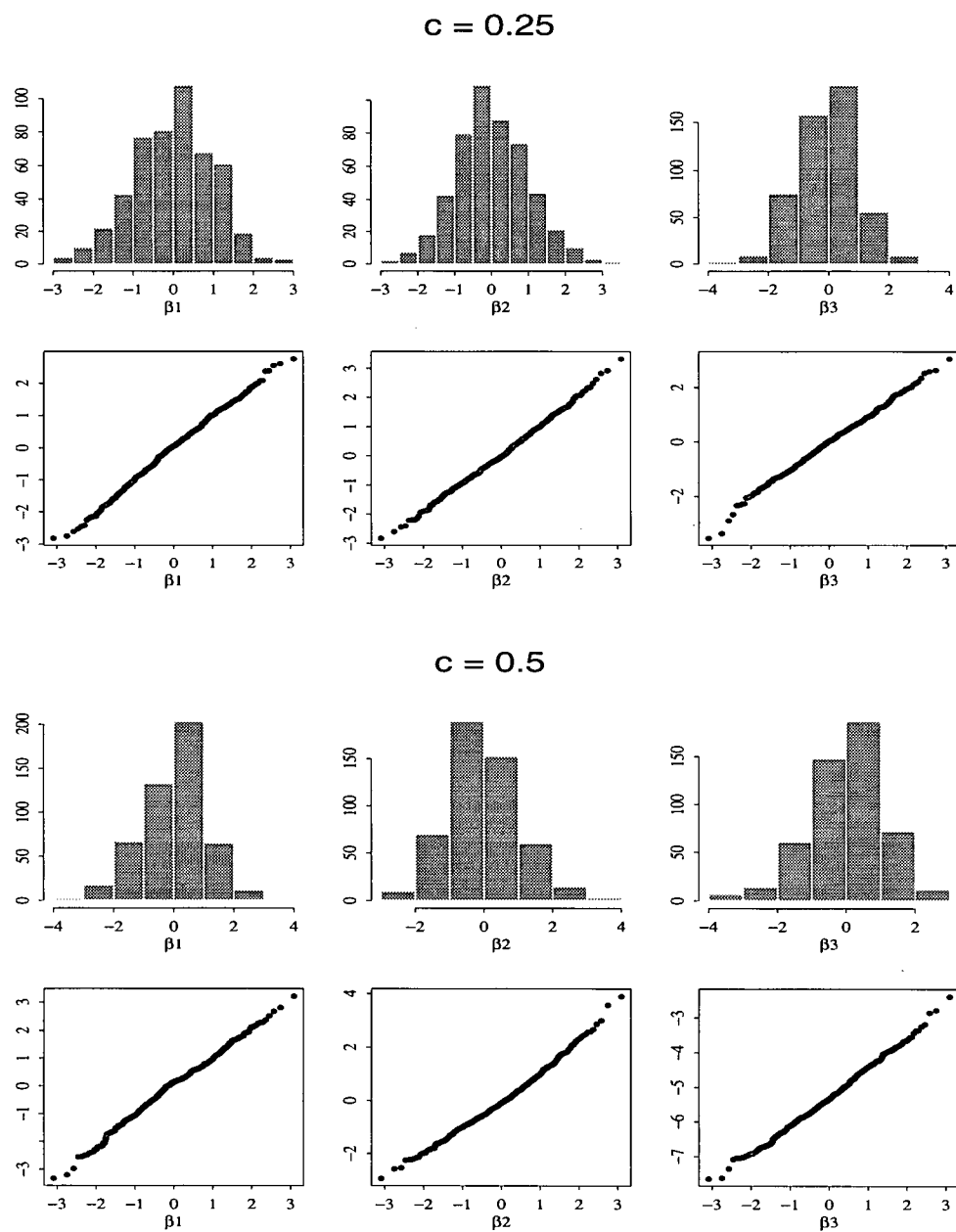


Figure 5.5: Histograms and qq-plots of the normalized estimate of  $\beta$  from the Quantile Knots methods with three knots for the exponential model (5.2).

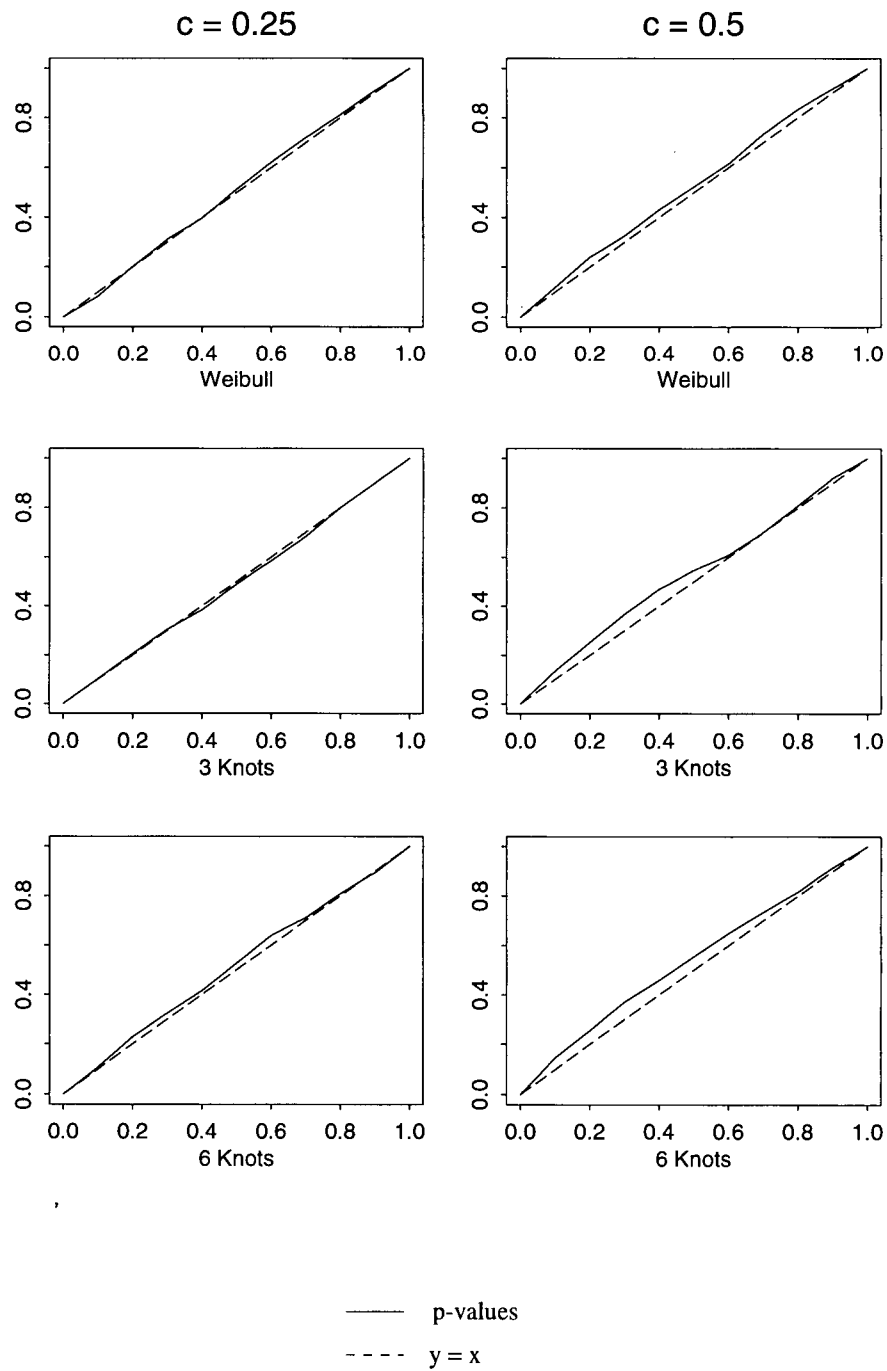
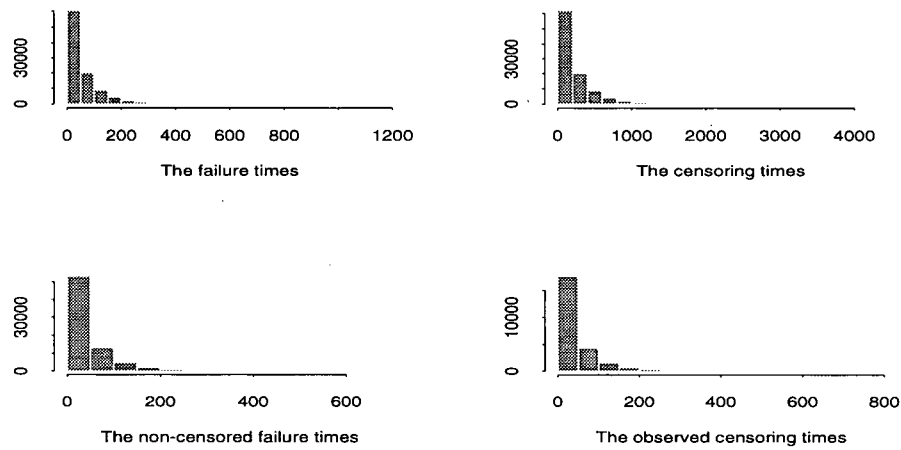


Figure 5.6: Empirical distribution functions of the p-values for testing that the failure times are exponentially distributed.



$c = 0.25$



$c = 0.5$

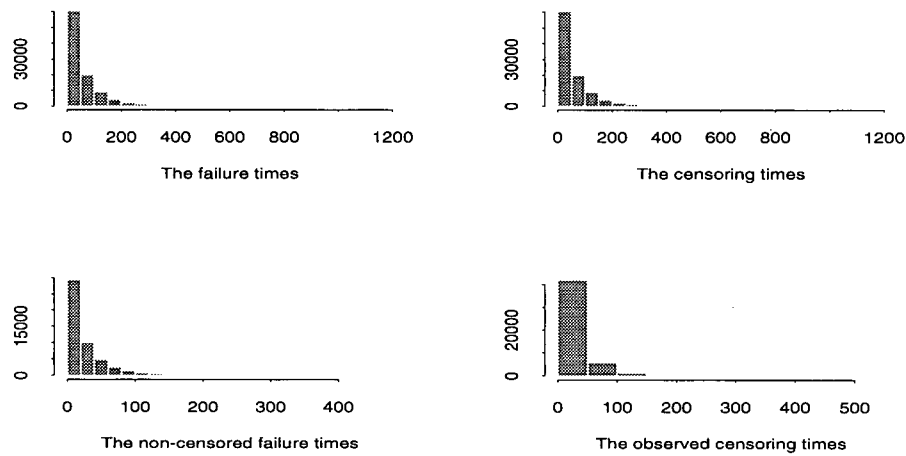


Figure 5.7: The histograms of the simulated failure times, censoring times, non-censored failure times, and observed censoring times for Weibull data as in model (5.3).

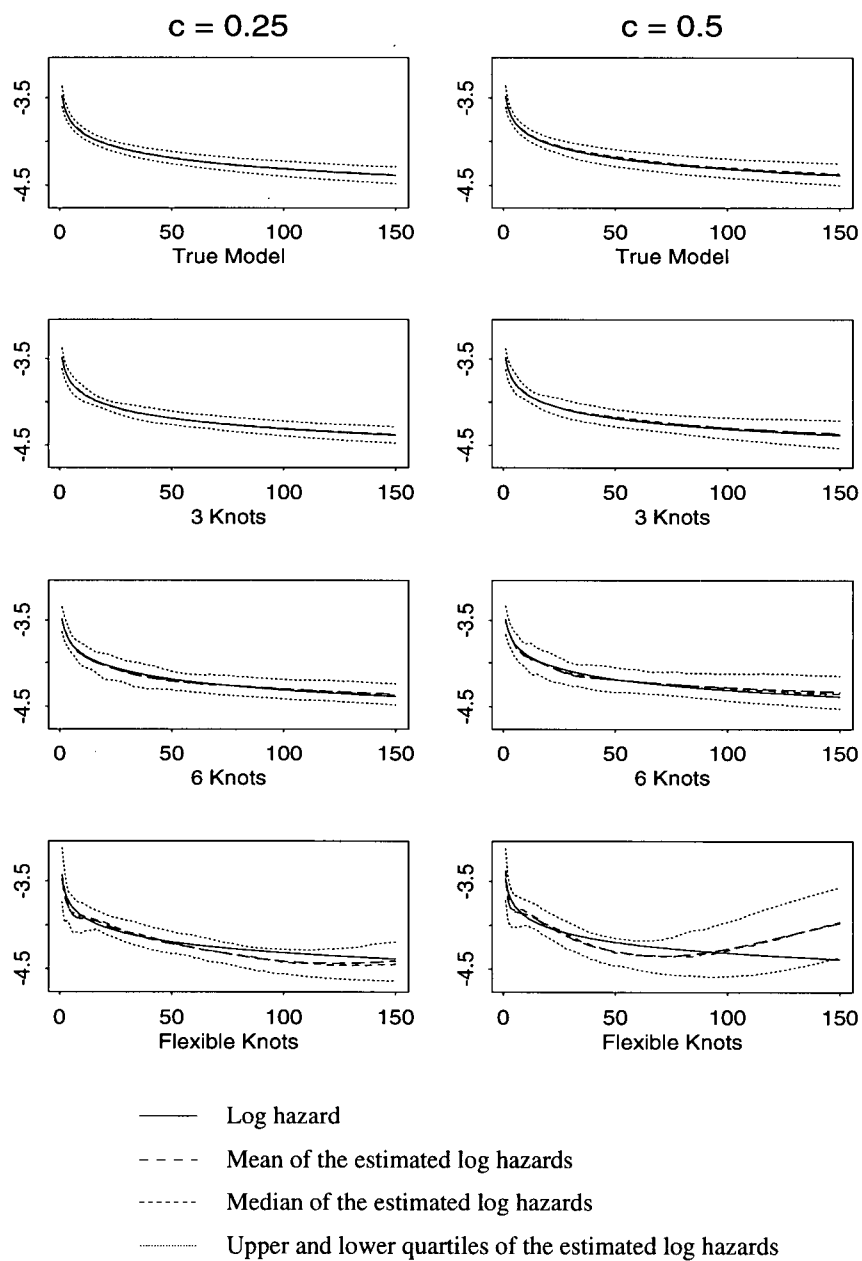


Figure 5.8: Log hazard of the Weibull distribution (5.3) and the pointwise quartiles and empirical mean of the estimated log hazards.

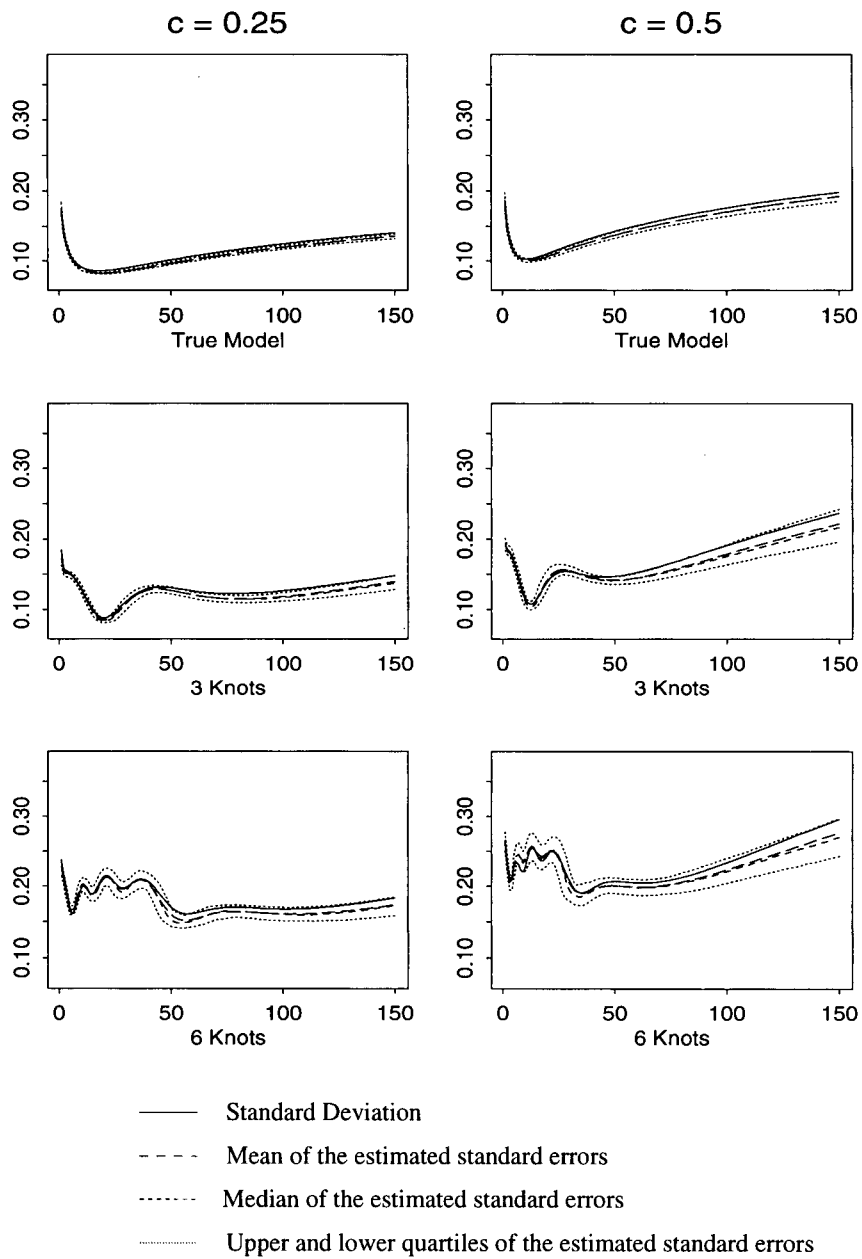


Figure 5.9: The pointwise standard deviations of the estimated log hazards for the Weibull distribution (5.3) and the pointwise quartiles and empirical mean of the 500 estimated standard errors.

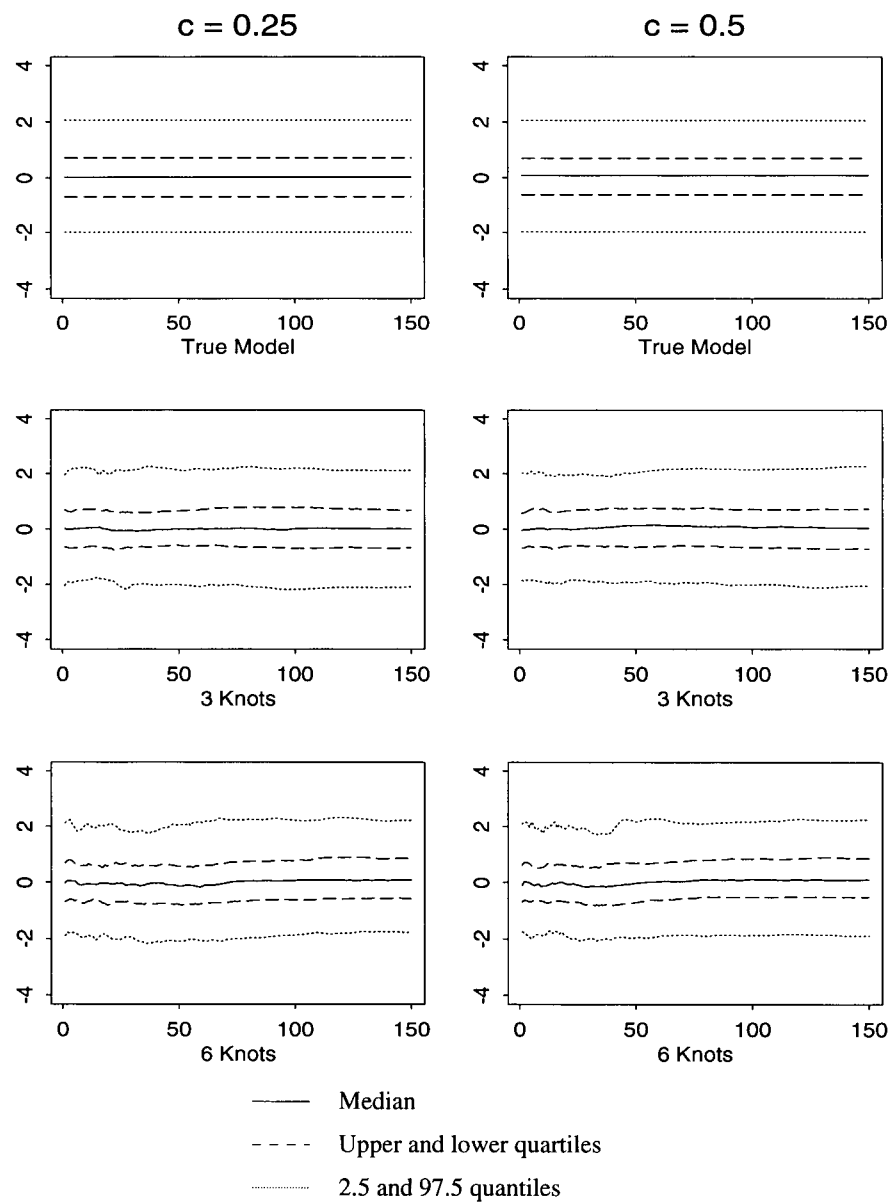


Figure 5.10: The quantiles of the pointwise z values of the estimated log hazards for the Weibull model (5.3).

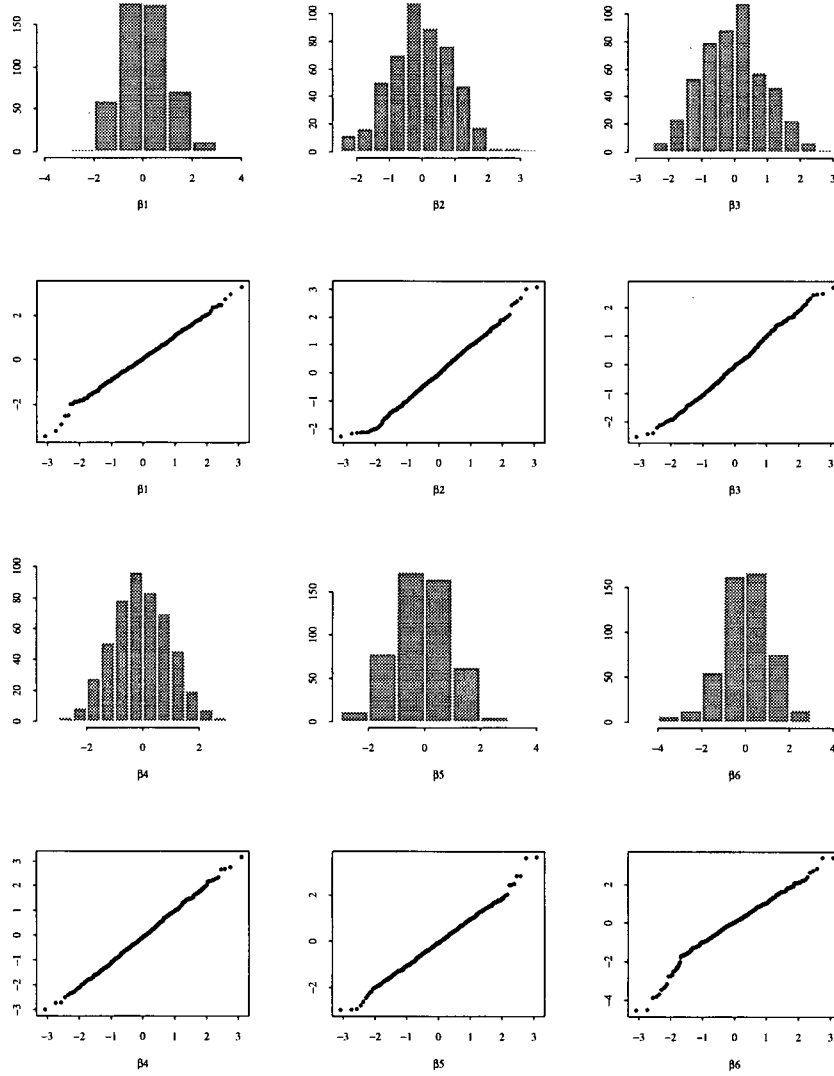


Figure 5.11: The histograms and the qq-plots of the standardized estimates of the parameter  $\beta$  from the method 5.1 with six knots for the Weibull model (5.3).

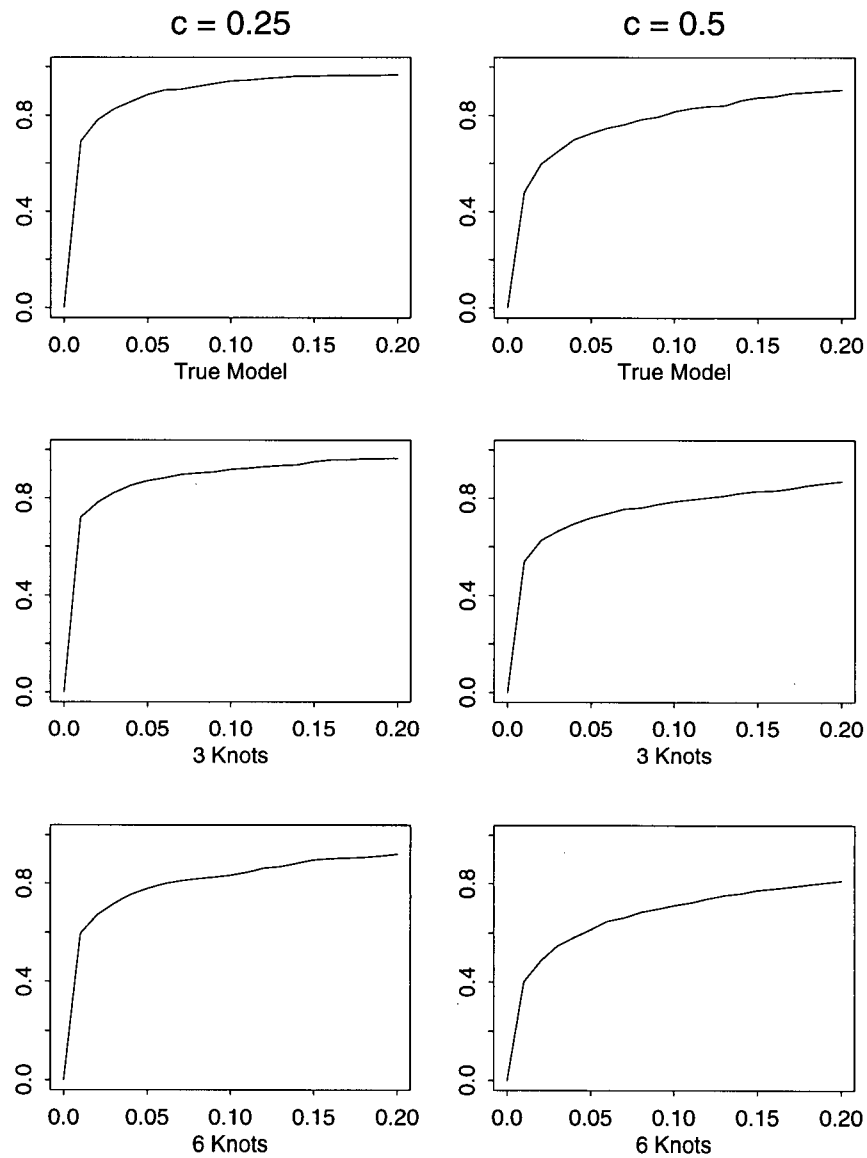


Figure 5.12: Empirical distribution functions of the p-values for testing the hypothesis that the failure time distribution is exponential.

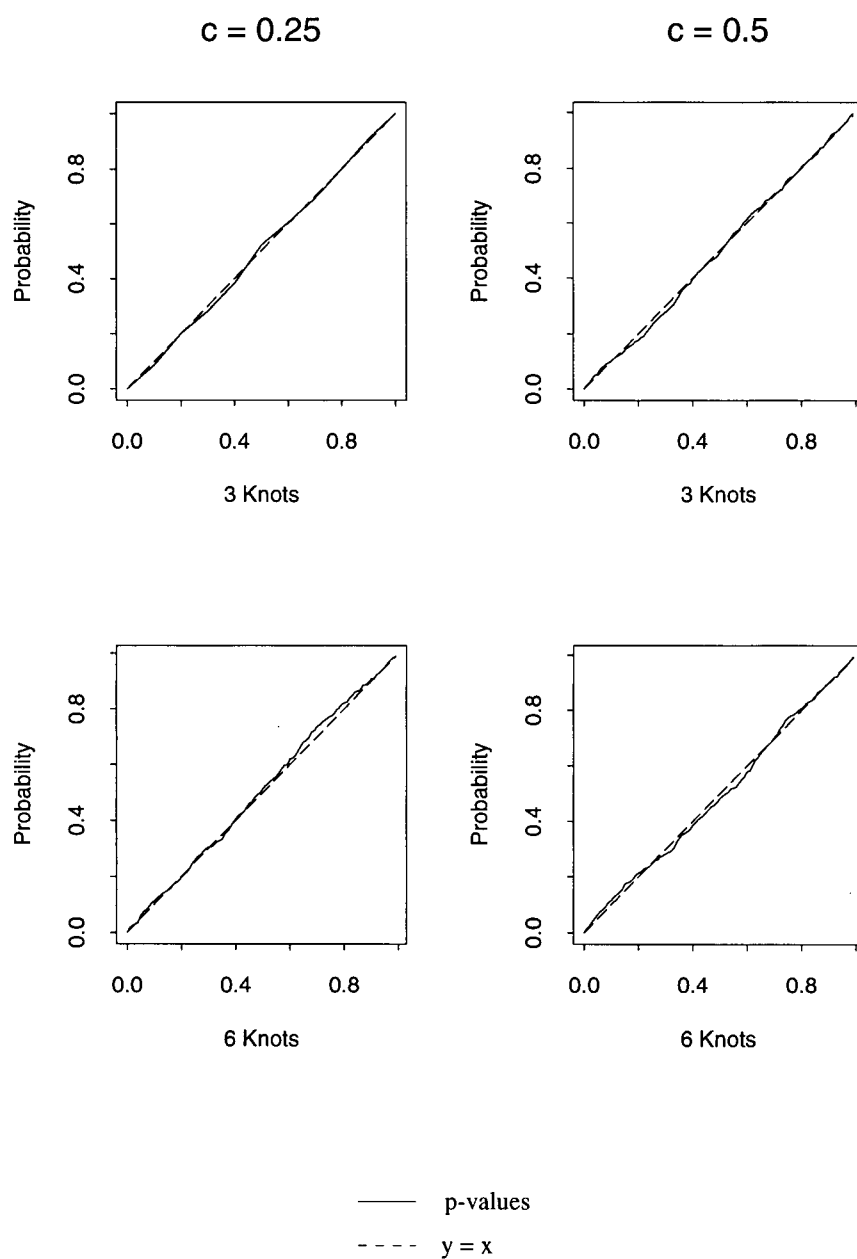
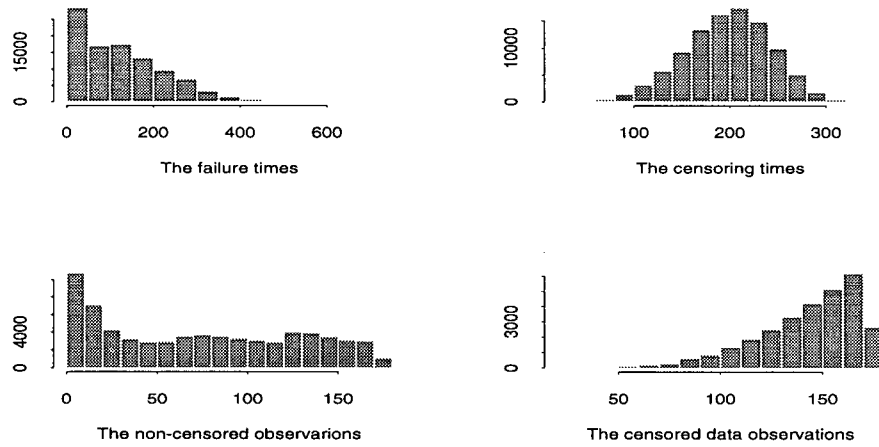


Figure 5.13: Empirical distribution functions of the p-values for testing that the failure times have a Weibull distribution.

$c = 0.25$



$c = 0.5$

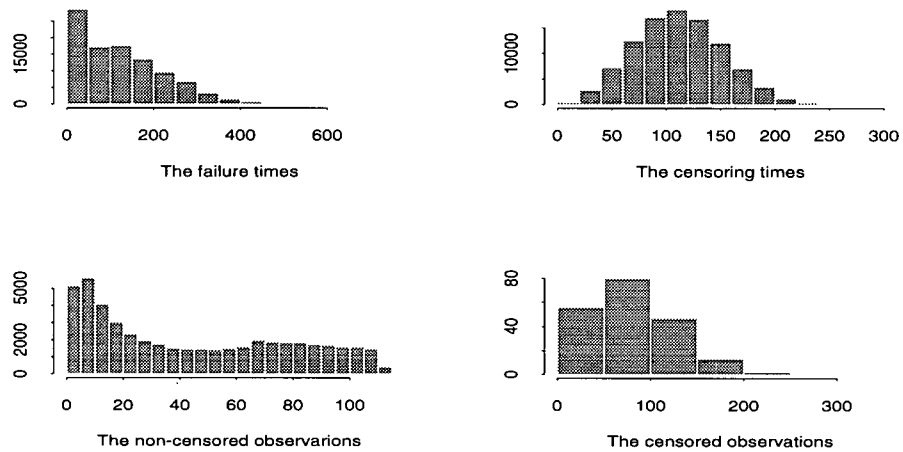


Figure 5.14: The histograms of the simulated failure times, censoring times, non-censored failure times, and observed censoring times for the data as in the B-spline model (5.4).



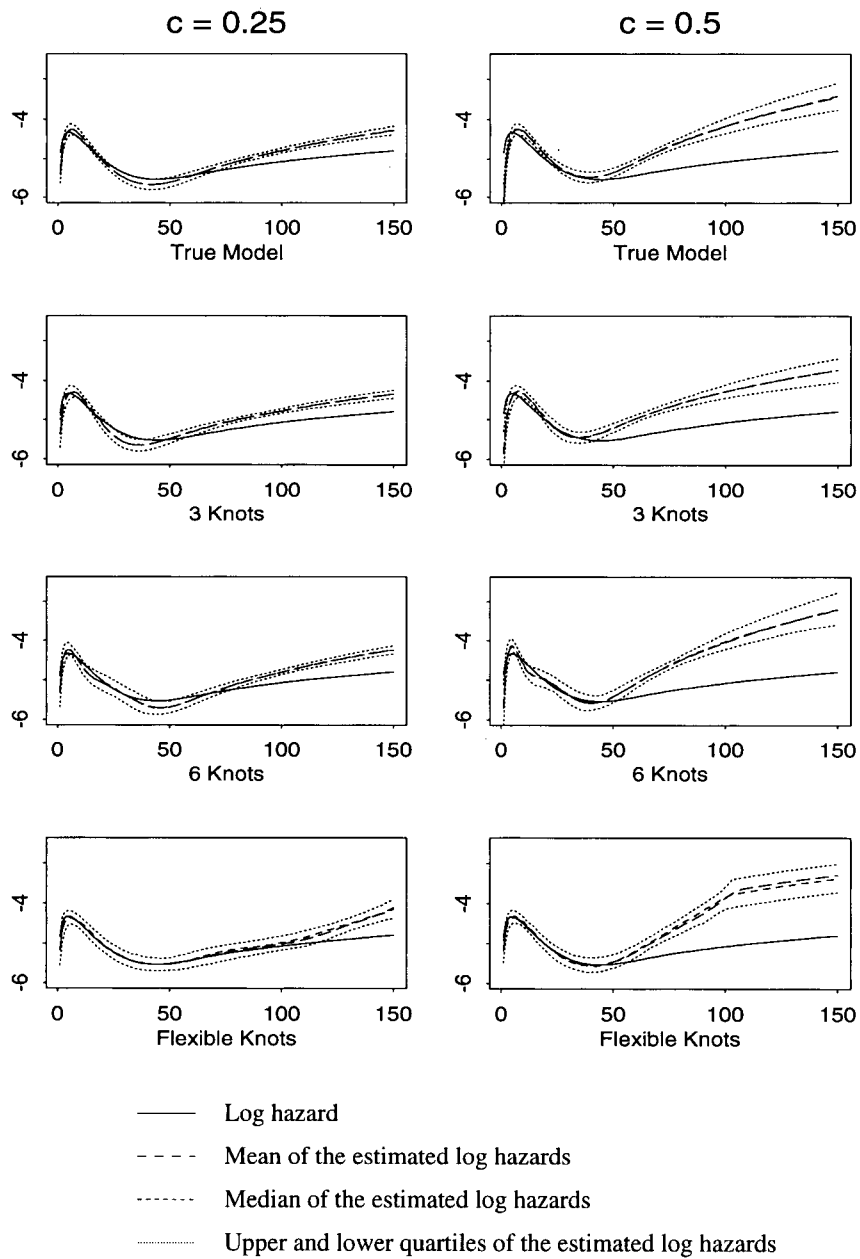


Figure 5.15: Log hazard of the the B-spline data as in model (5.4) and the pointwise quartiles and empirical mean of the estimated log hazards.

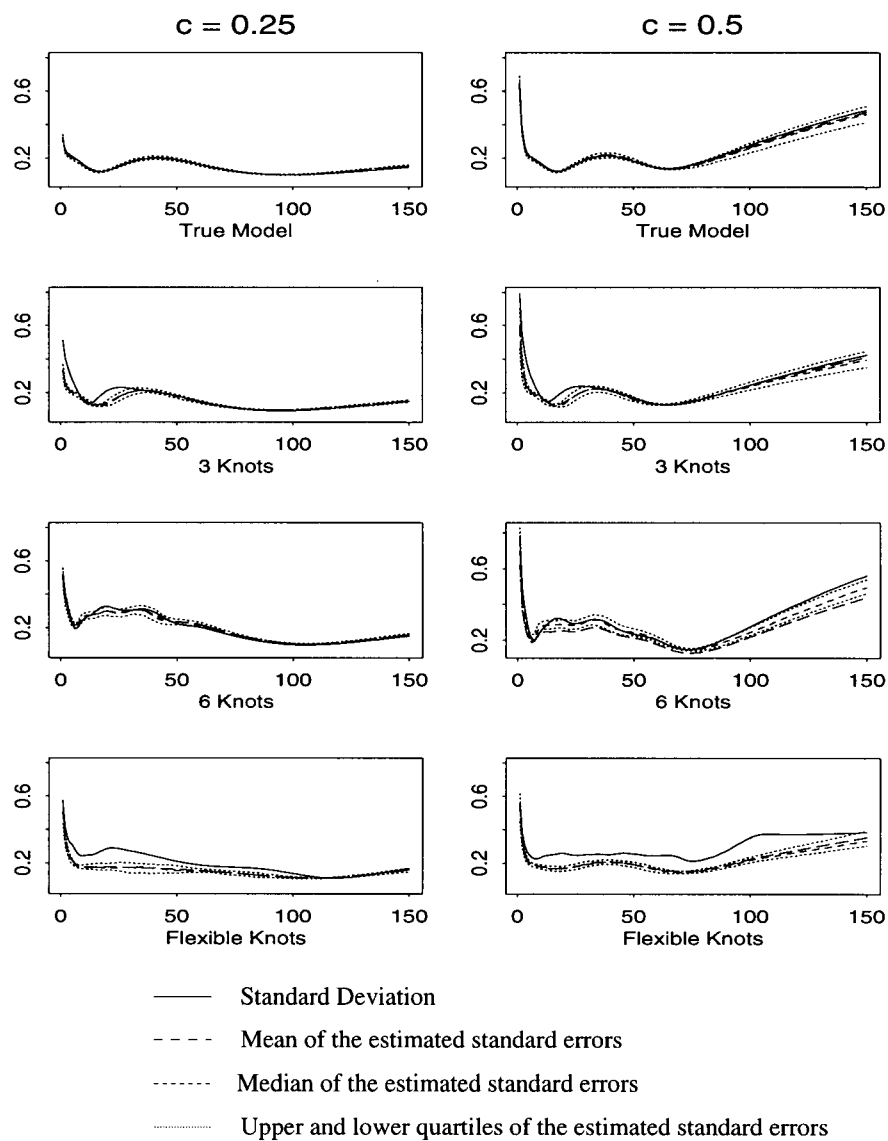


Figure 5.16: The pointwise standard deviations of the estimated log hazards for the B-spline model (5.4) and the pointwise quartiles and empirical mean of the estimated standard errors.

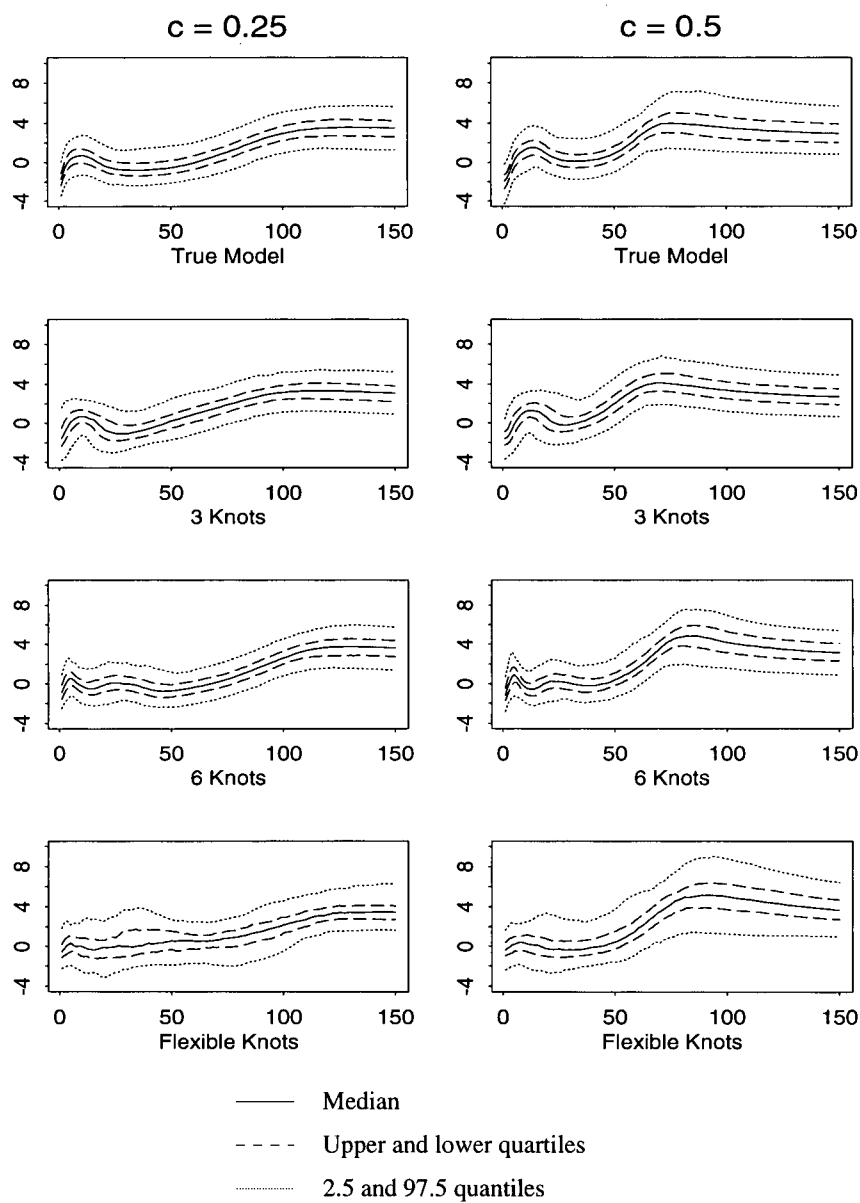


Figure 5.17: The quantiles of the pointwise z values of the estimated log hazards for the B-spline model (5.4).

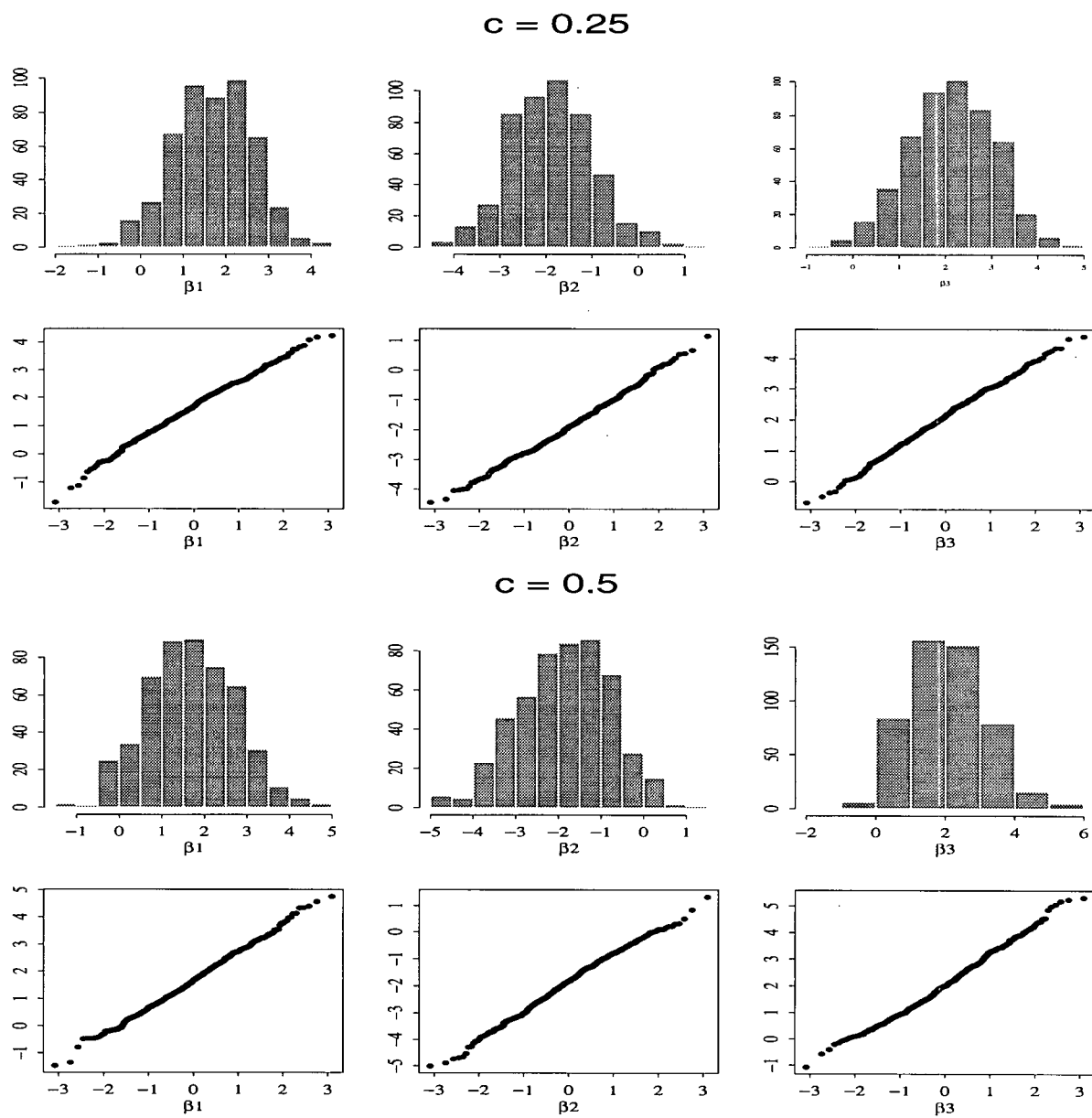


Figure 5.18: The histograms and the qq-plots of the standardized estimates of the parameter  $\beta$  from the True Knots method for the B-spline model (5.4).

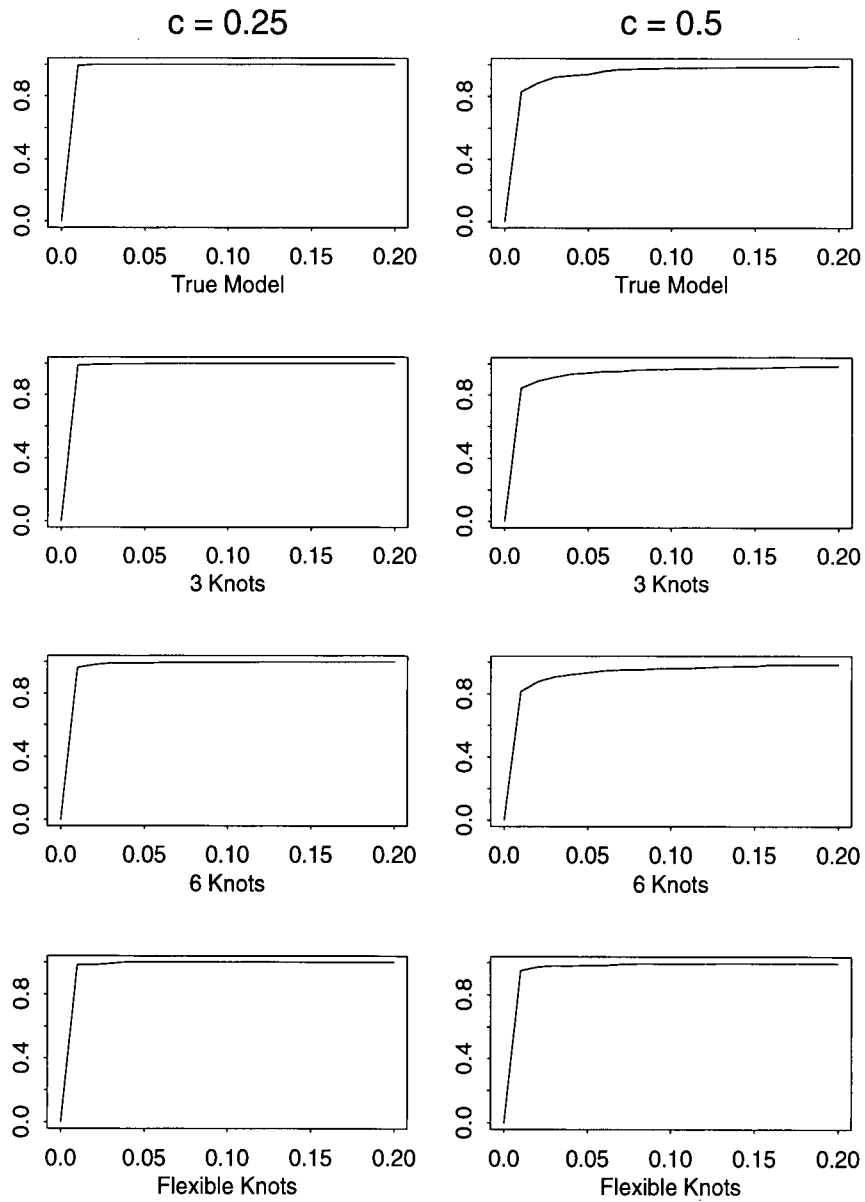


Figure 5.19: Empirical distribution functions of the p-values for testing the hypothesis that the failure times have a Weibull distribution.

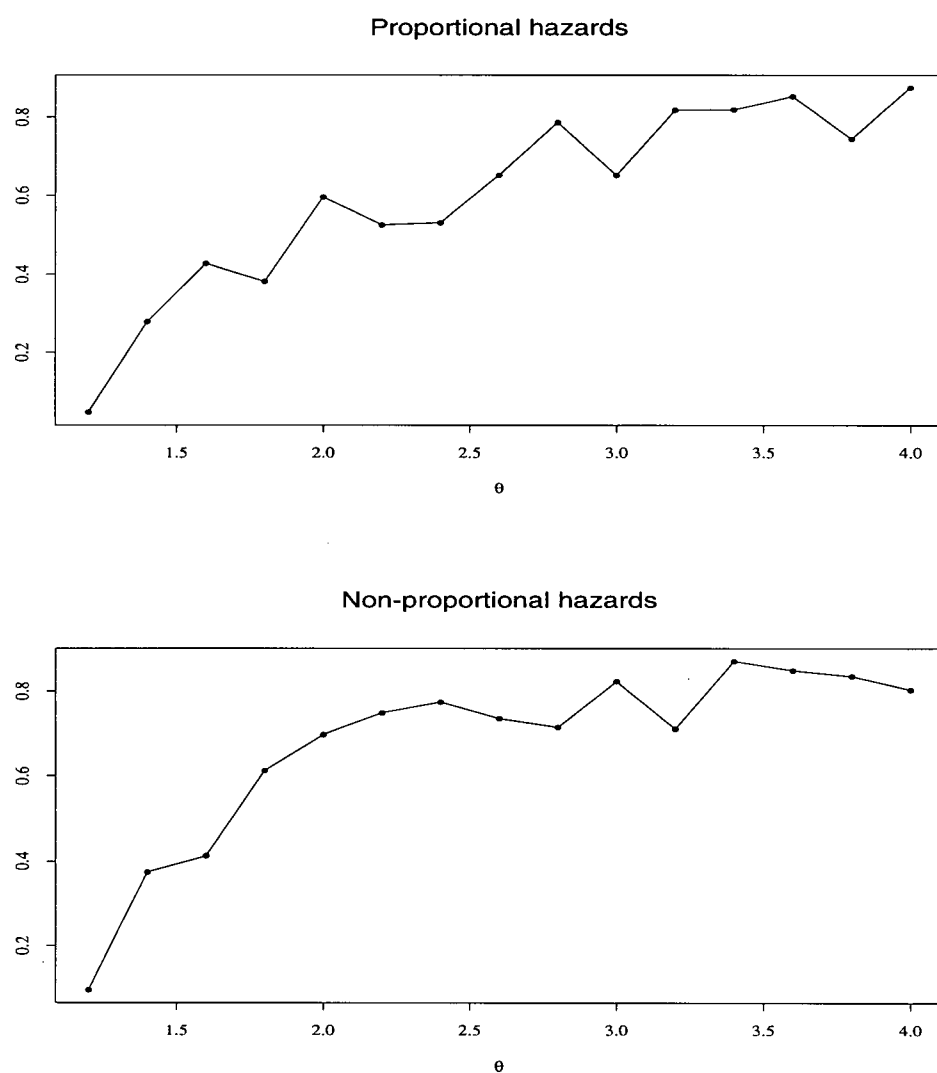


Figure 5.20: The plots of the correlation coefficients versus  $\theta$  for the data as in the proportional hazards and non-proportional hazards models

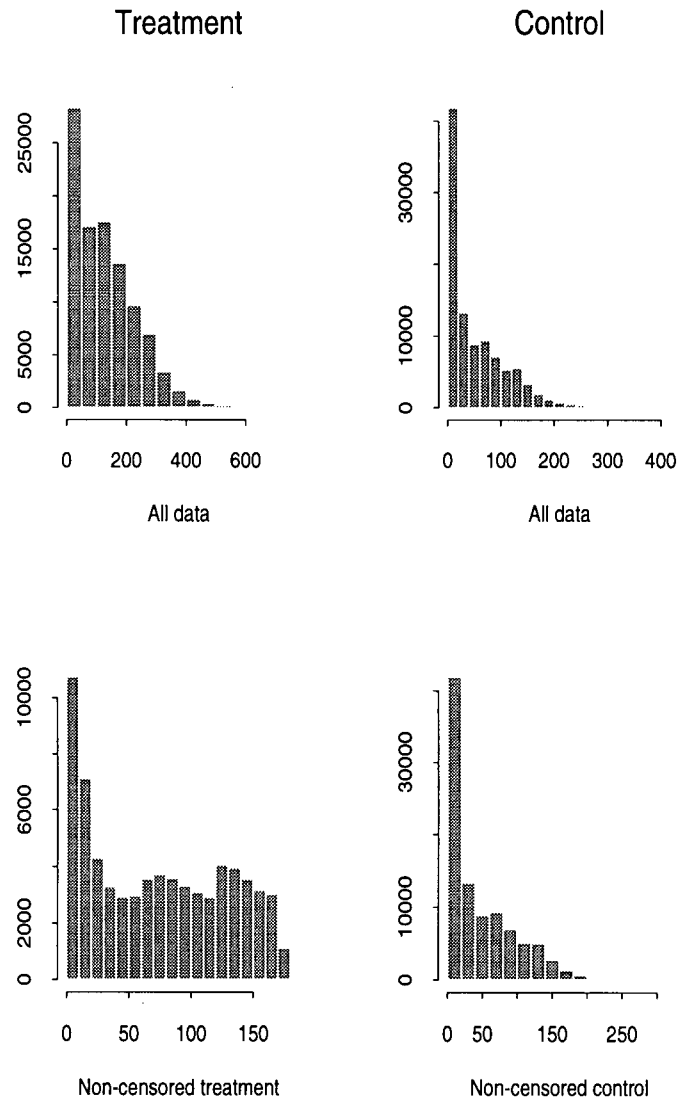


Figure 5.21: The histograms of the marginal failure times, and non-censored marginal failure times for the data generated according to the proportional hazards model in Section 5.1. The censoring rate for the “treatment” is 25%.

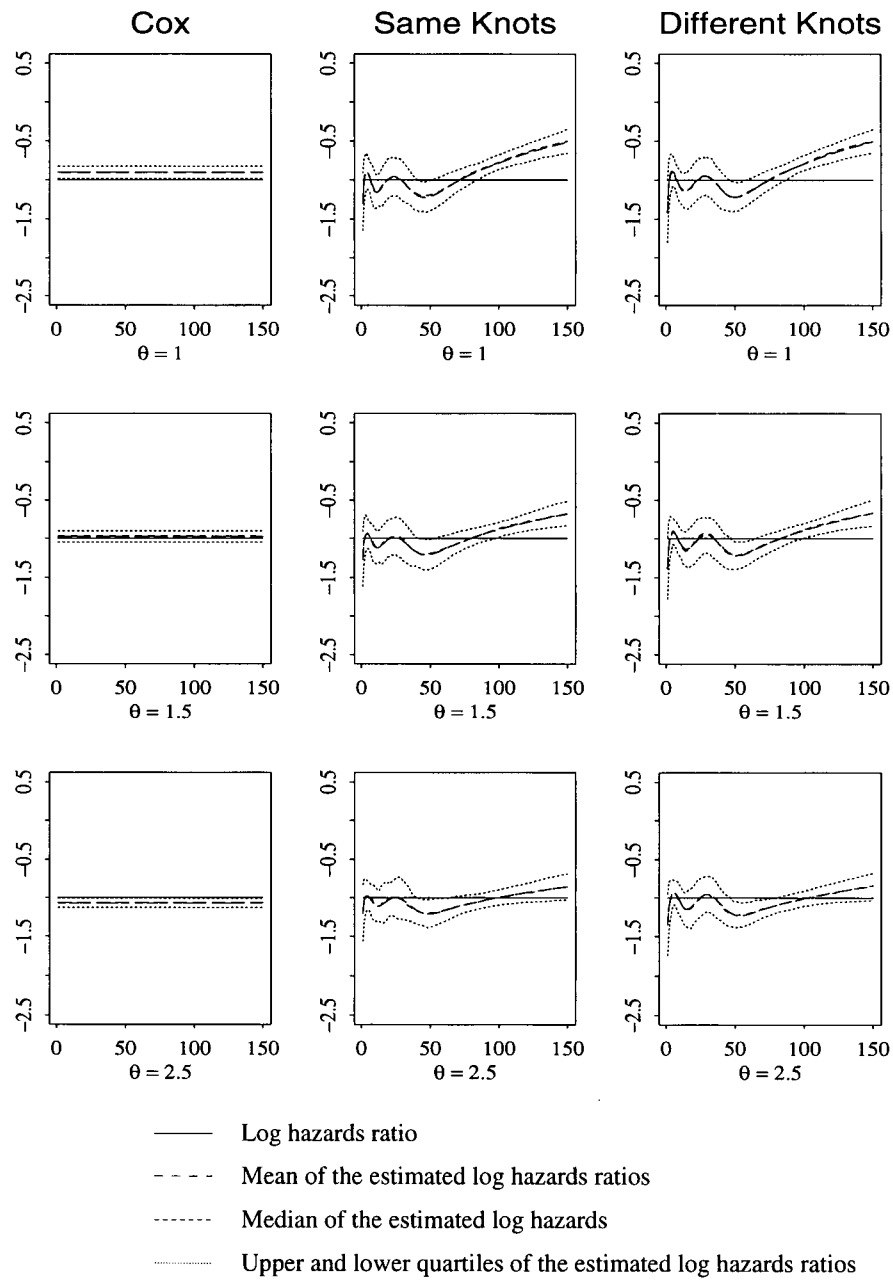


Figure 5.22: The true log hazards ratio of the data from the proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the 500 estimated log hazards ratios.



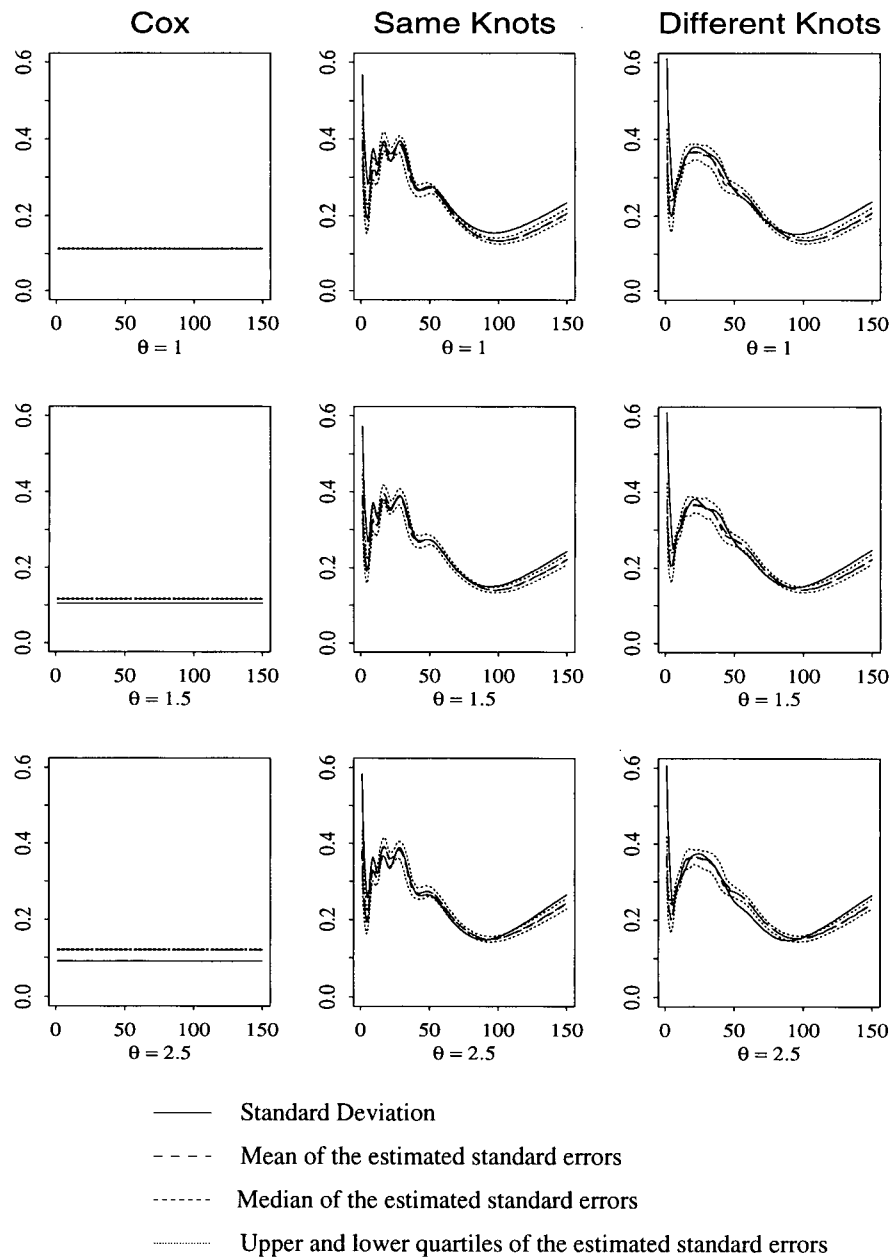


Figure 5.23: The standard deviations of the estimated log hazards ratio for data generated according to the proportional hazards model. The censoring rate is 25% for the “treatment”.

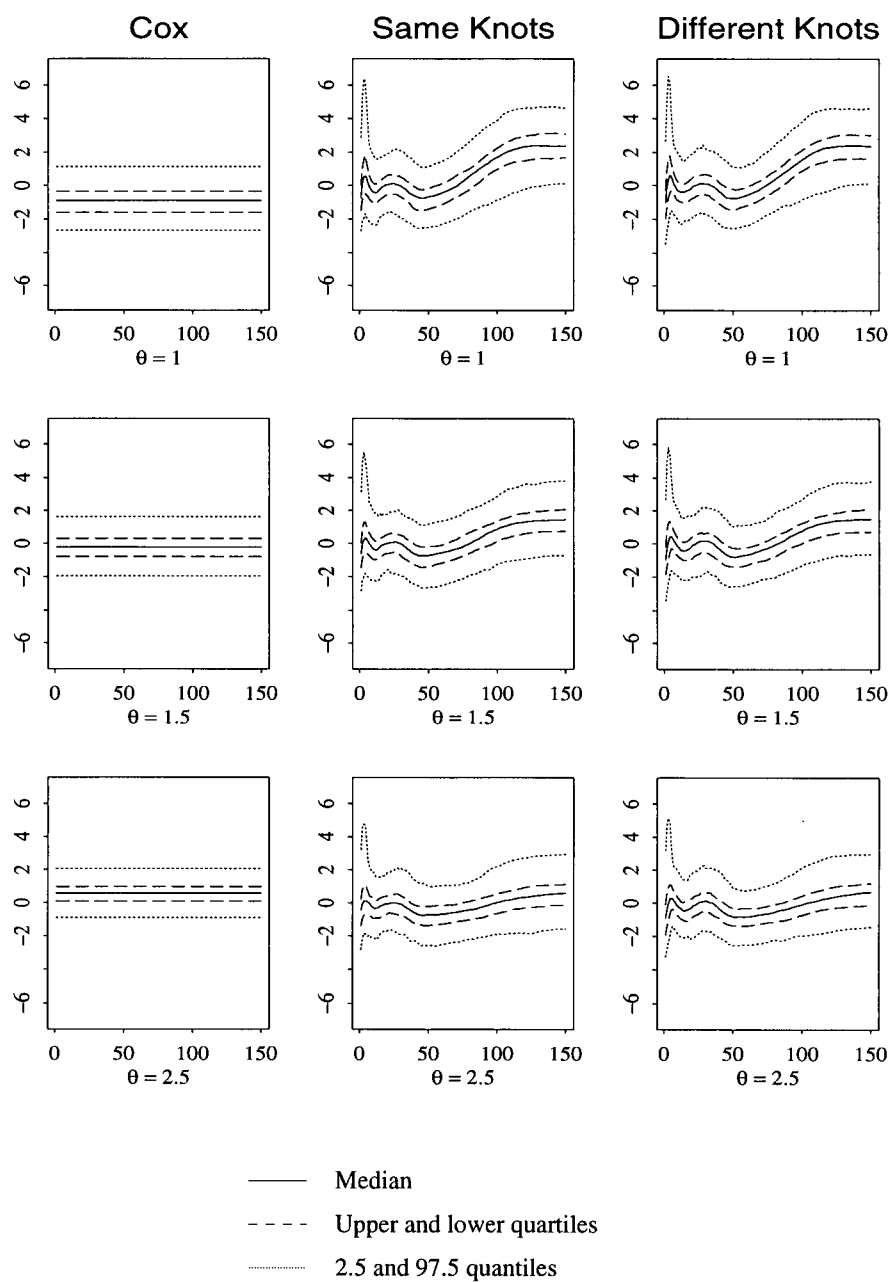


Figure 5.24: The quantiles of the pointwise  $z$  values of the estimated log hazards ratio for the data generated according to the proportional hazards model. The censoring rate is 25% for the “treatment”.

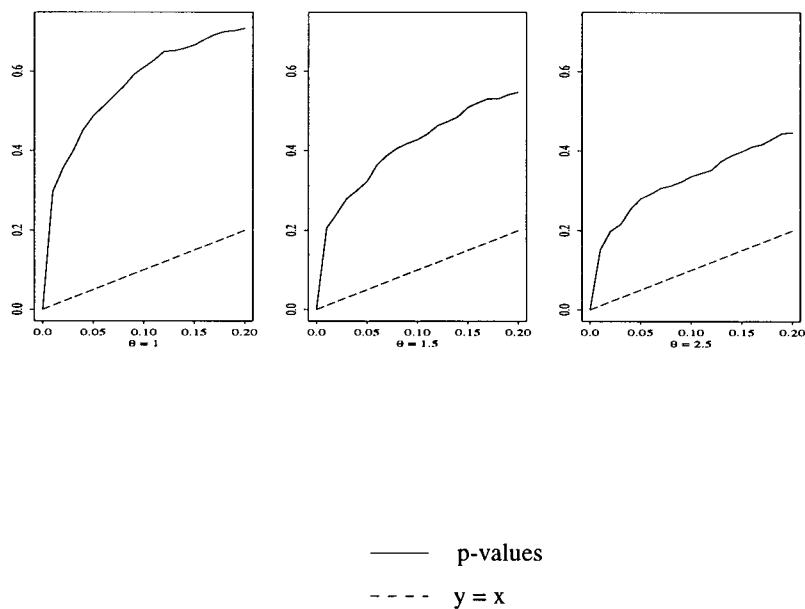


Figure 5.25: Empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the proportional hazards model in Section 5.1 have proportional hazards.

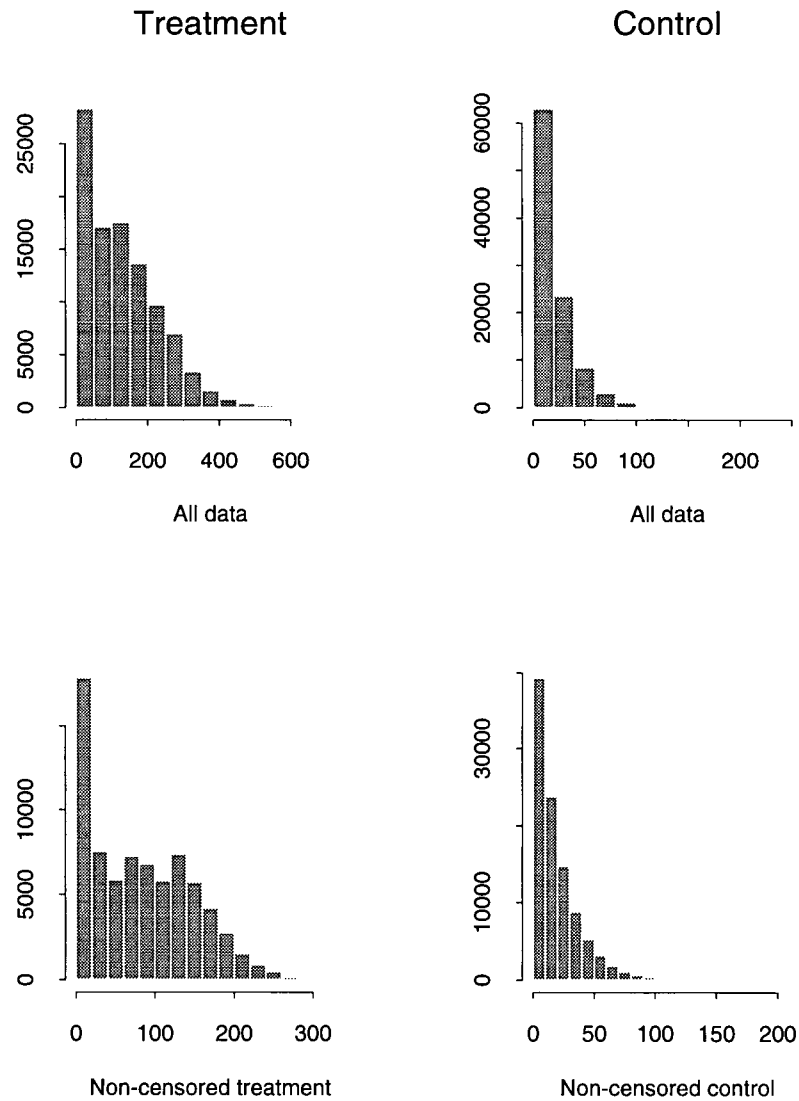


Figure 5.26: The histograms of the marginal failure times, and the non-censored marginal failure times for the data from the non-proportional hazards model in Section 5.1. The censoring rate for the “treatment” is 25%.

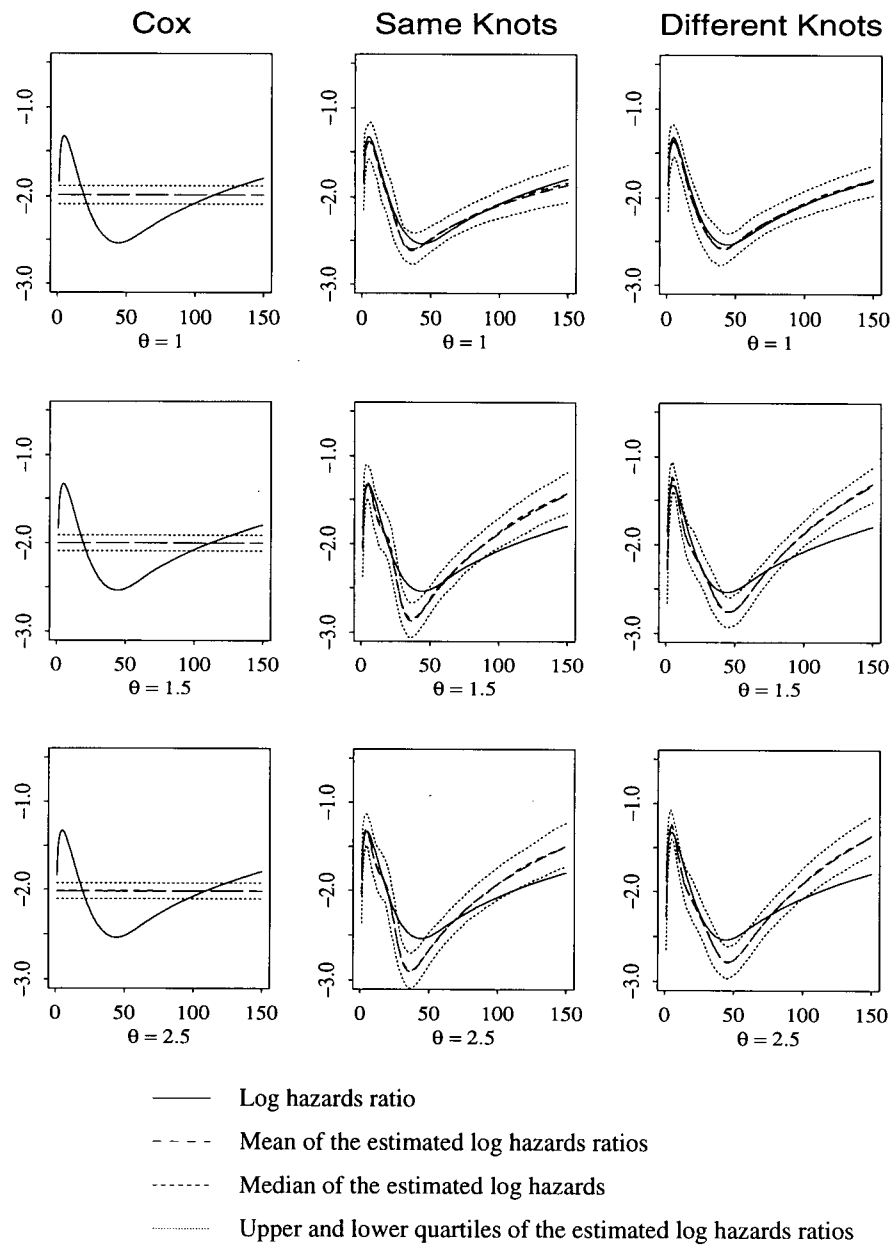


Figure 5.27: The true log hazards ratio of the data from the non-proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the estimated log hazards ratios.

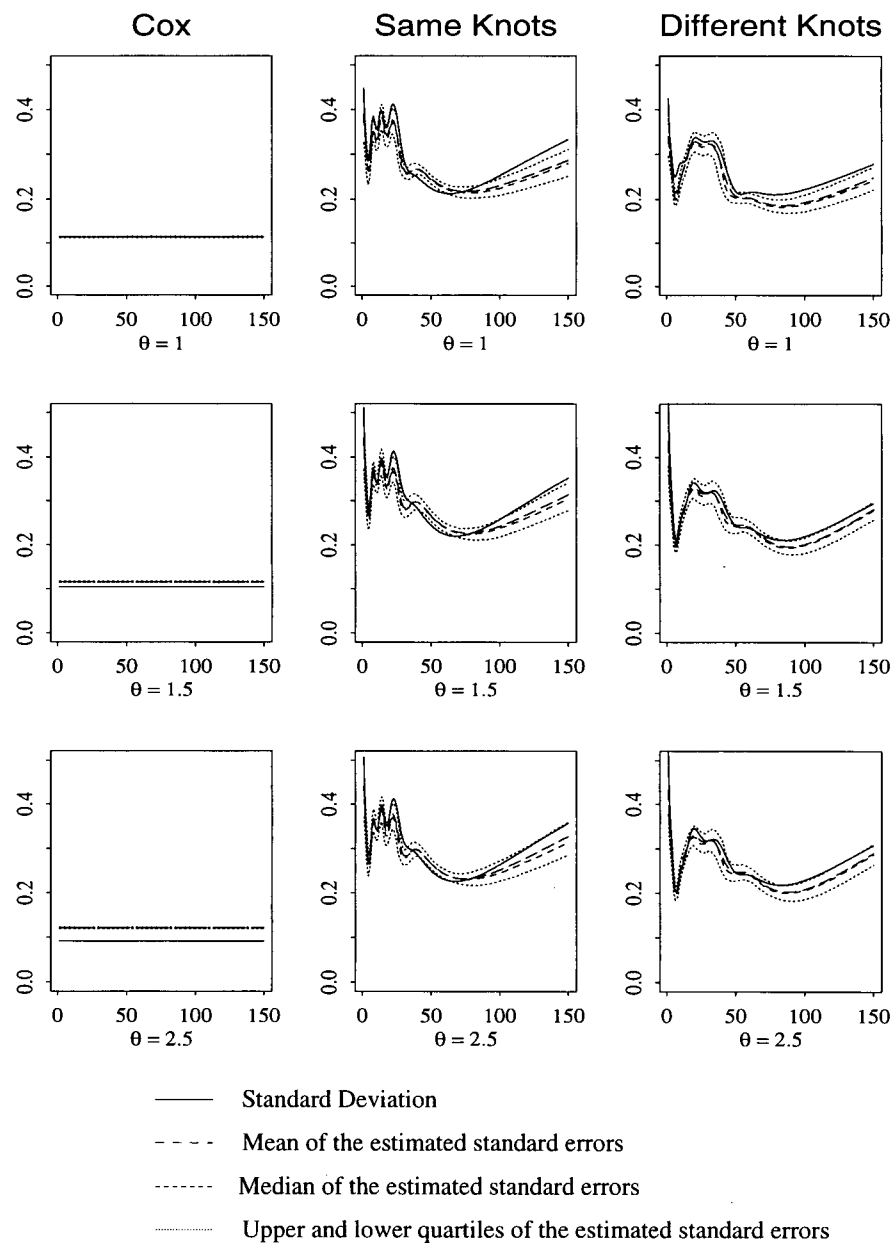


Figure 5.28: The standard deviations of the estimated log hazards ratio for data generated according to the non-proportional hazards model. The censoring rate is 25% for the “treatment”.

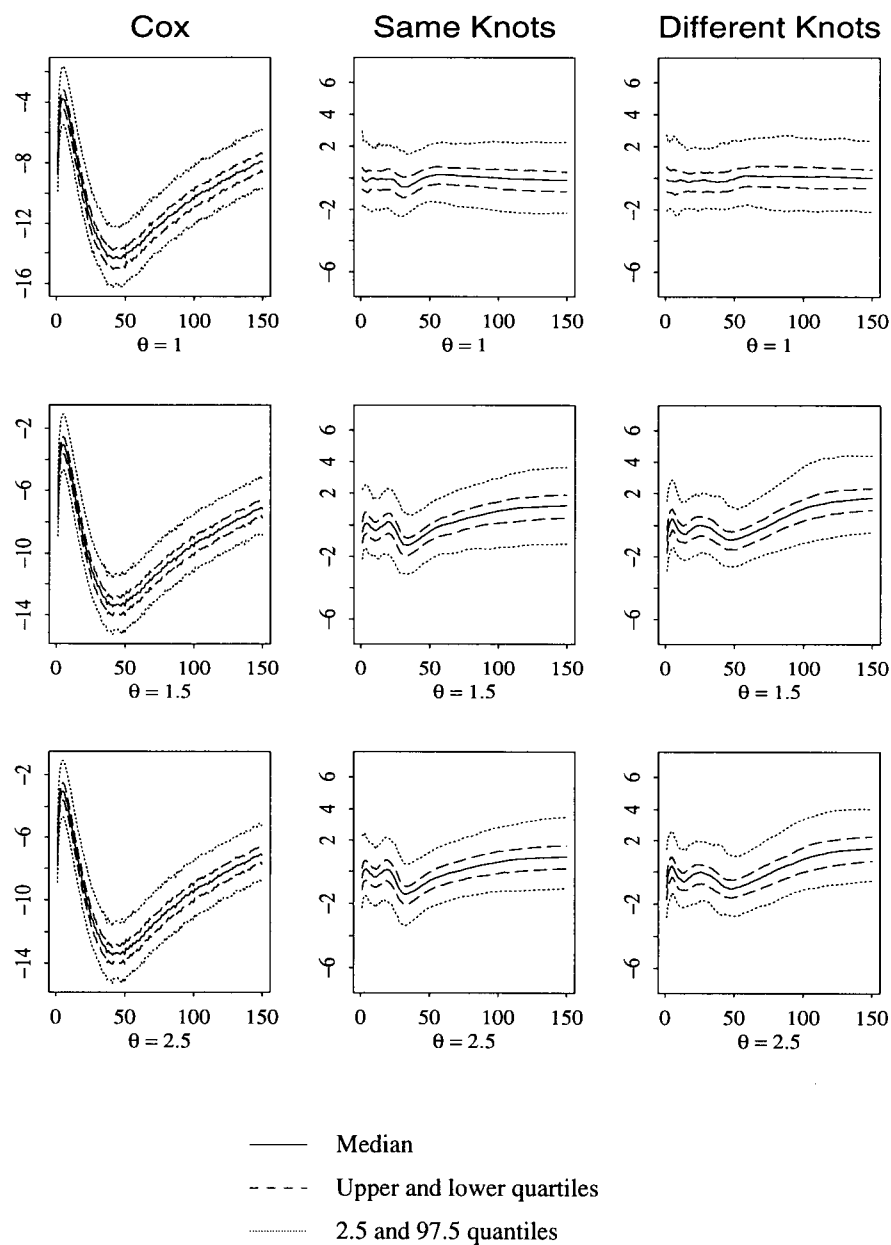


Figure 5.29: The quantiles of the pointwise  $z$  values of the estimated log hazards ratio for the data generated according to the non-proportional hazards model. The censoring rate is 25% for the “treatment”.

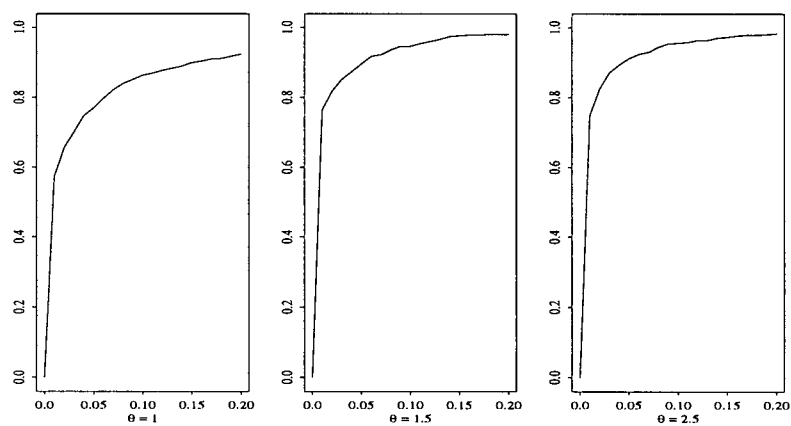


Figure 5.30: Empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the non-proportional hazards model in Section 5.1 have proportional hazards.



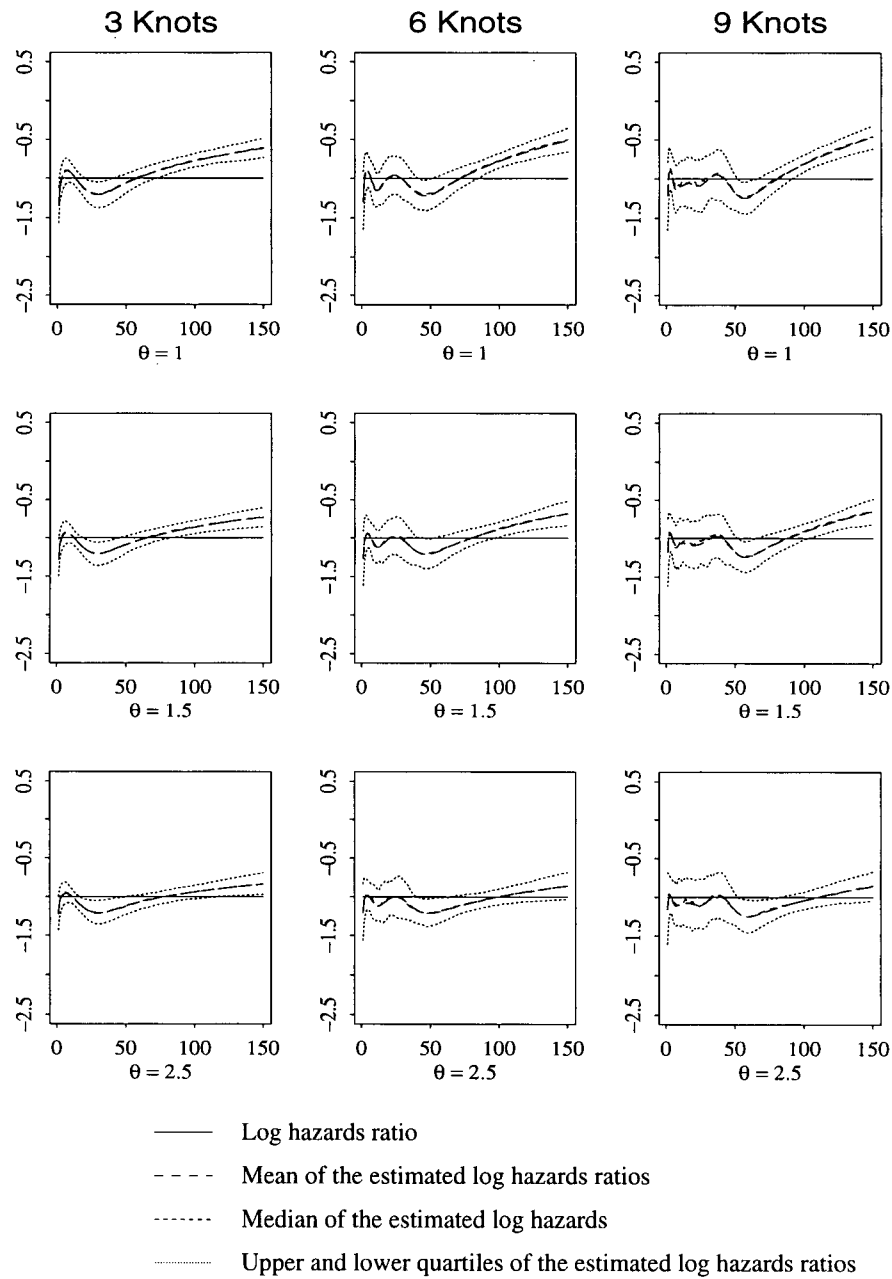


Figure 5.31: The true log hazards ratio of the data from the proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the estimated log hazards ratios by the Same Knots method.

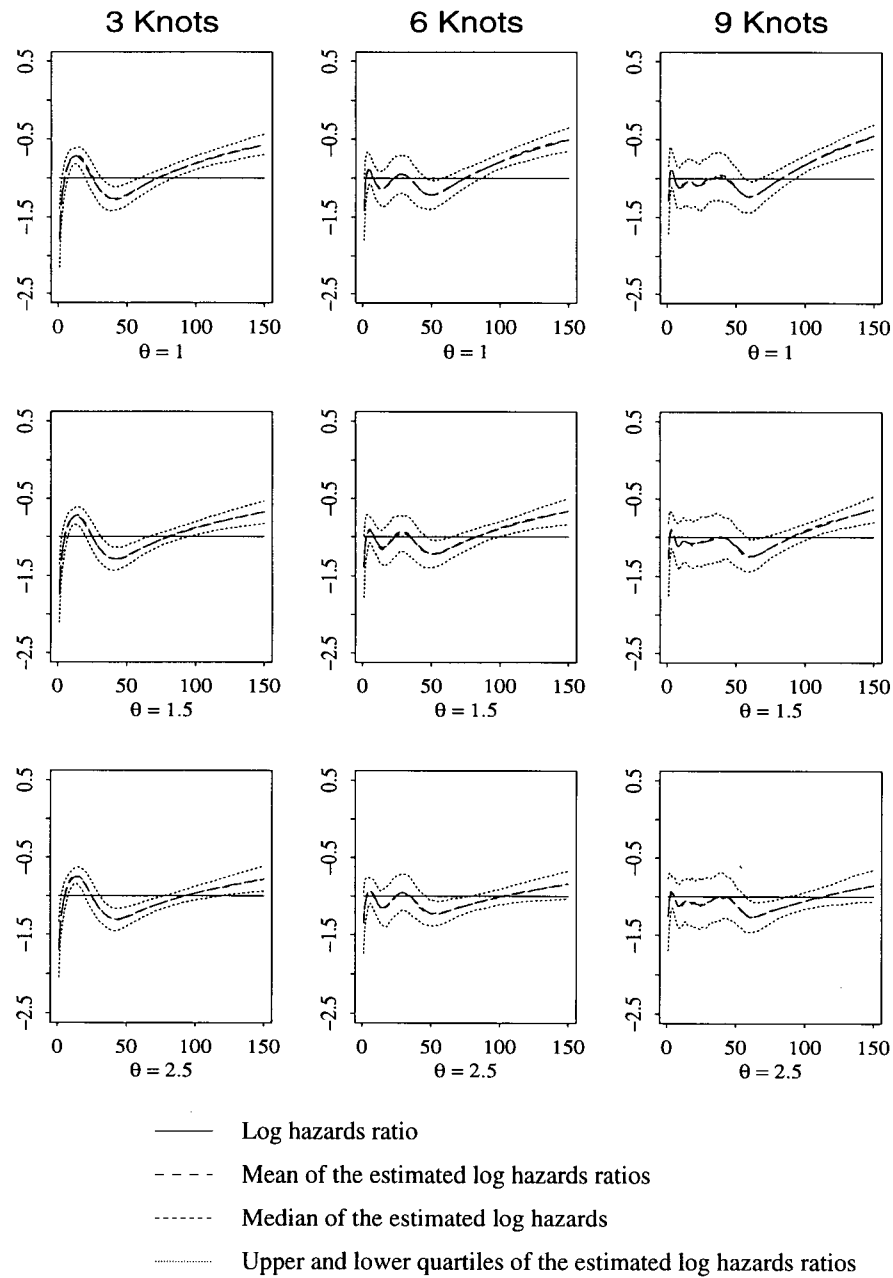


Figure 5.32: The true log hazards ratio of the data from the proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the estimated log hazards ratios by the Different Knots method.

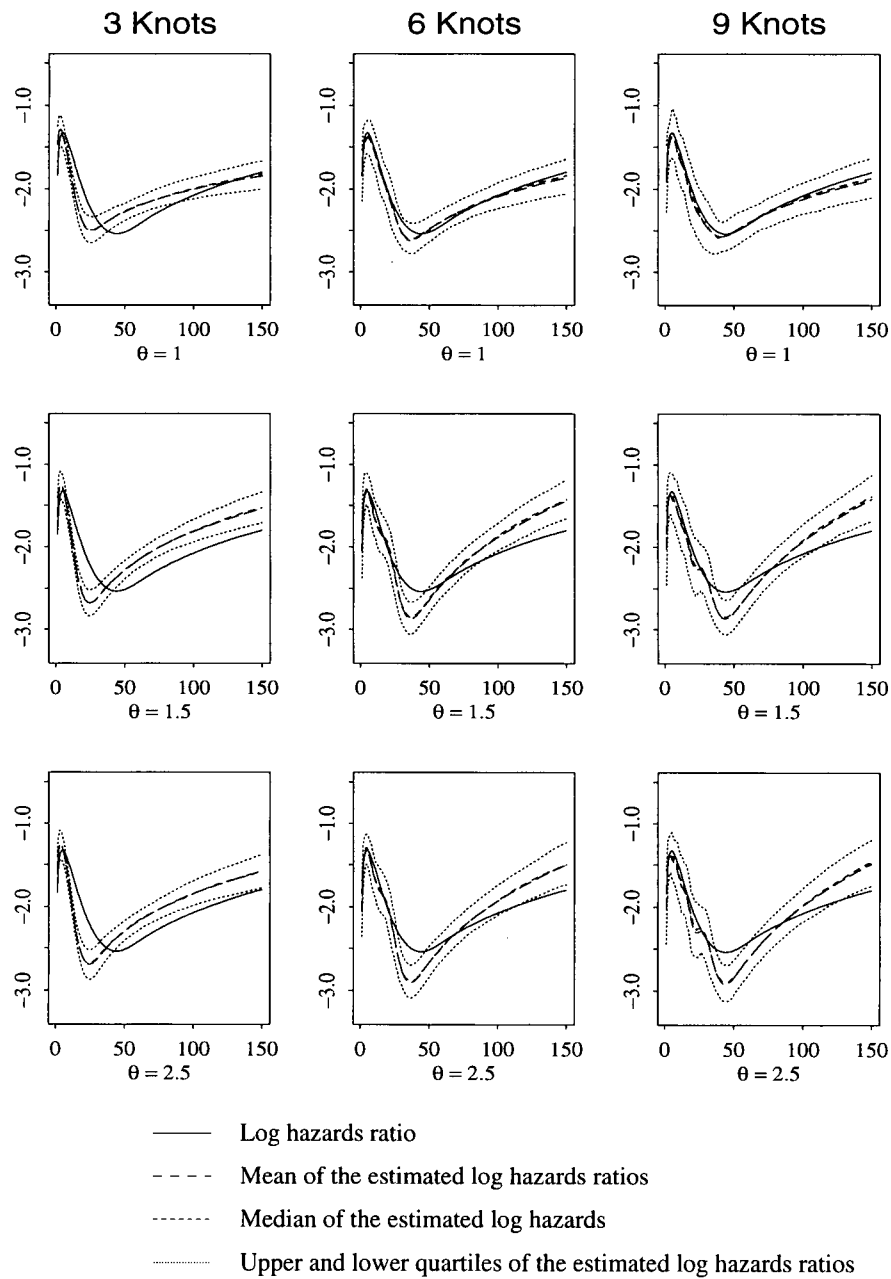


Figure 5.33: The true log hazards ratio of the data from the non-proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the estimated log hazards ratios by the Same Knots method.

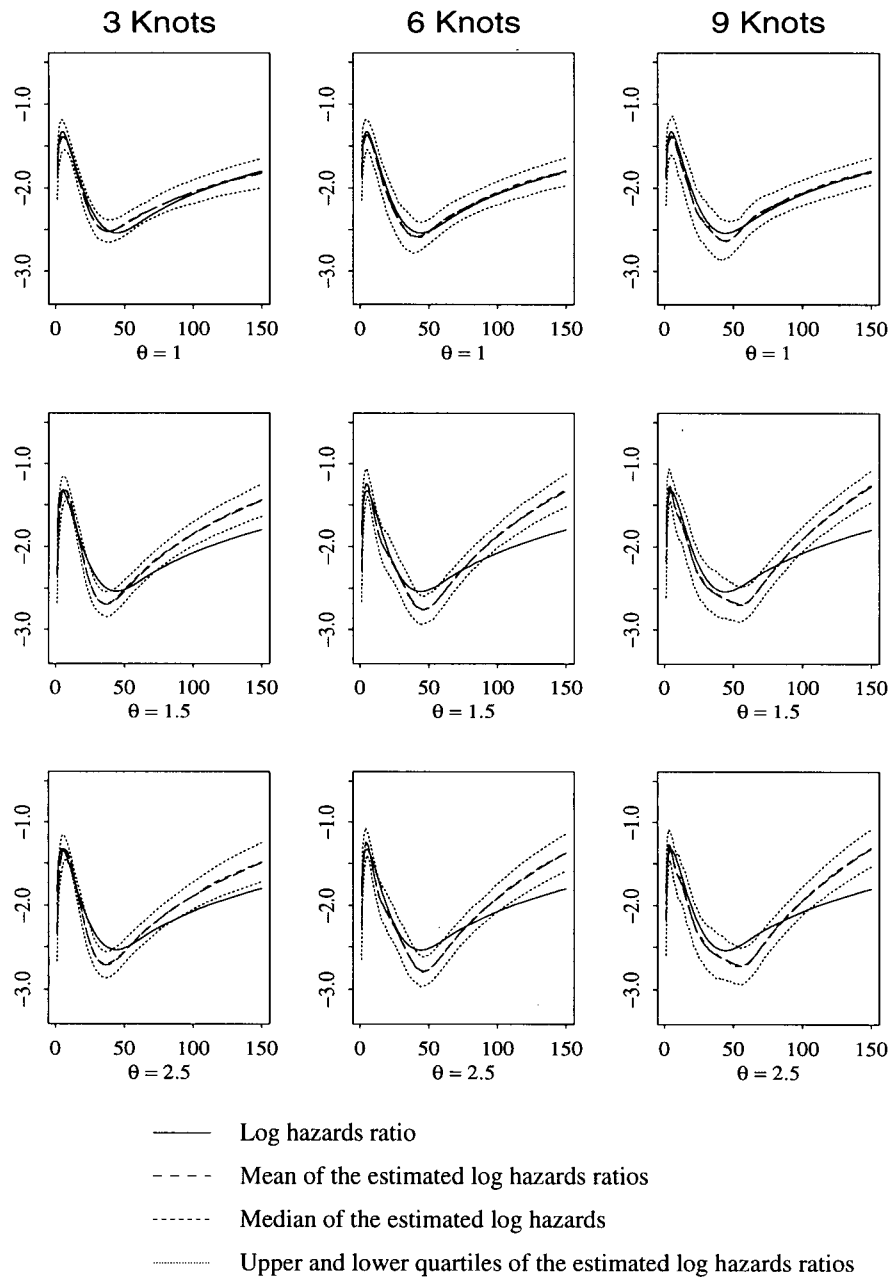


Figure 5.34: The true log hazards ratio of the data from the non-proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the estimated log hazards ratios by the Different Knots method.

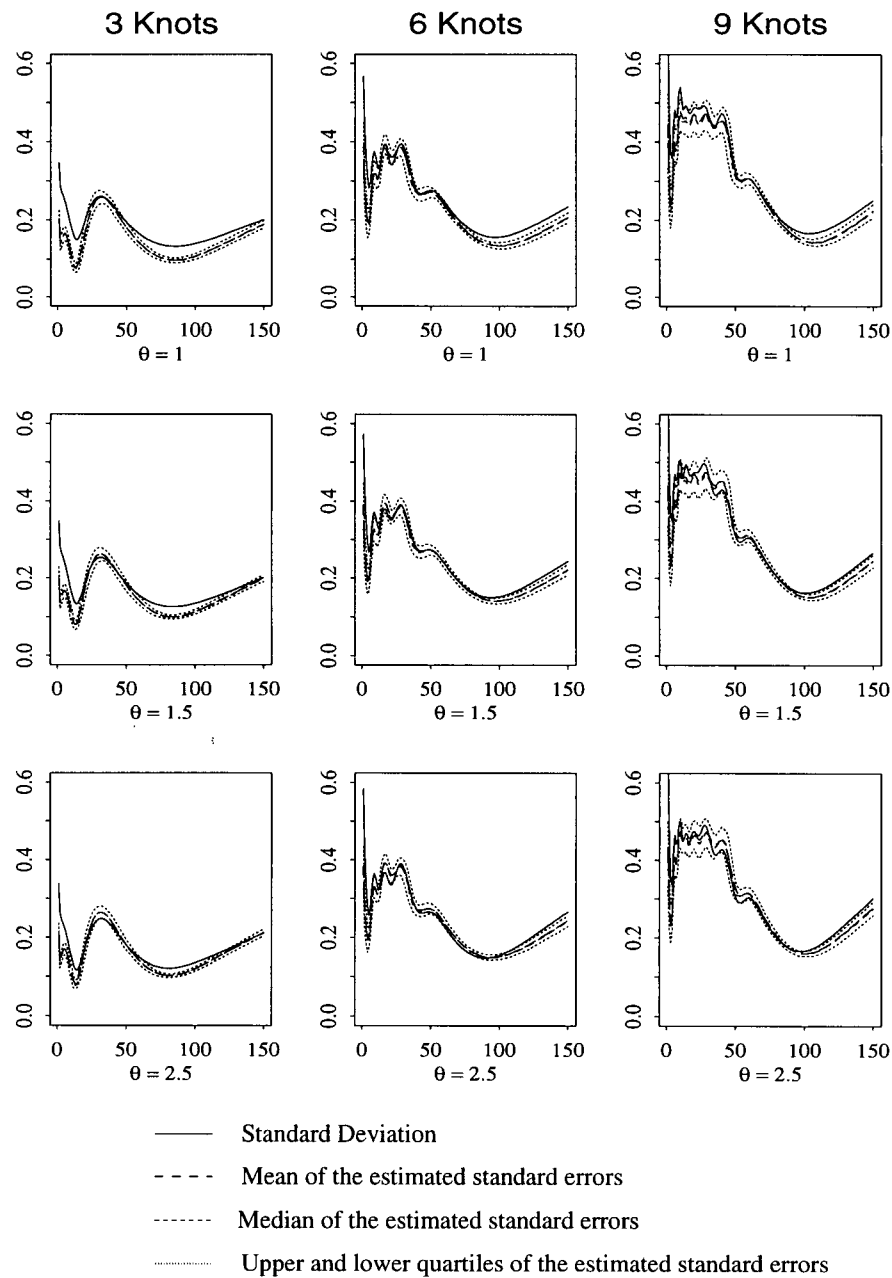


Figure 5.35: The standard deviations of the estimated log hazards ratio by the Same Knots method for data generated according to the proportional hazards model. The censoring rate is 25% for the “treatment”.

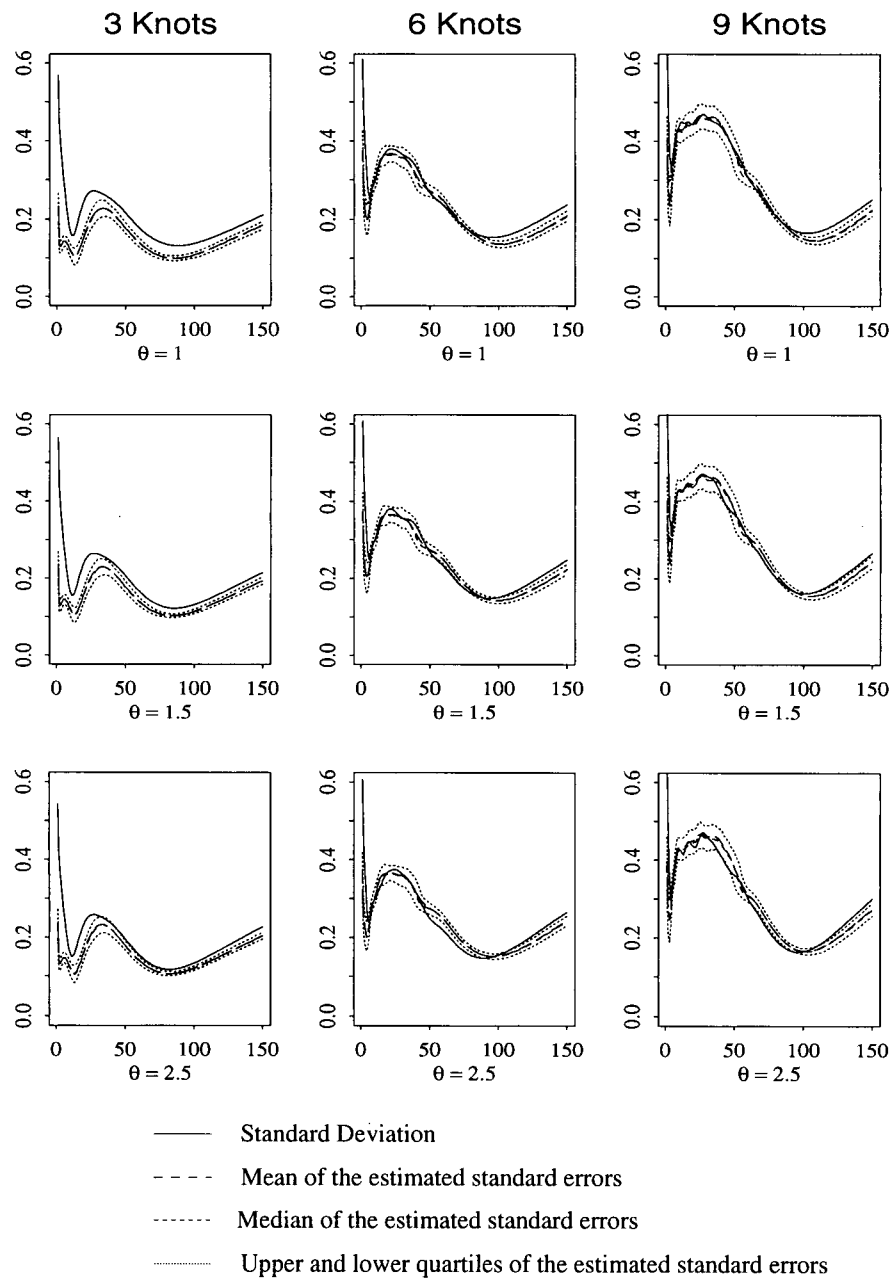


Figure 5.36: The standard deviations of the estimated log hazards ratio by the Different Knots method for data generated according to the proportional hazards model. The censoring rate is 25% for the “treatment”.

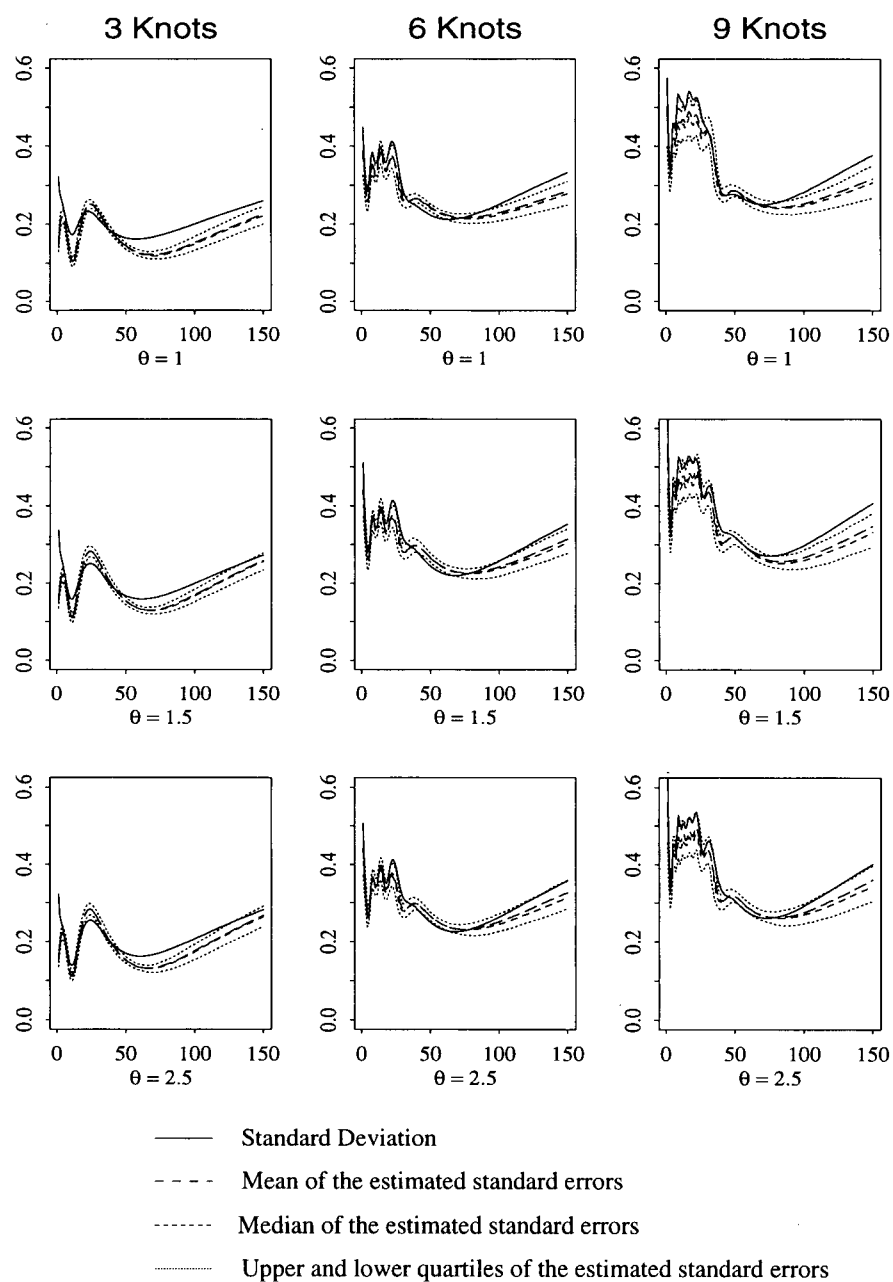


Figure 5.37: The standard deviations of the estimated log hazards ratio by the Same Knots method for data generated according to the non-proportional hazards model. The censoring rate is 25% for the “treatment”.

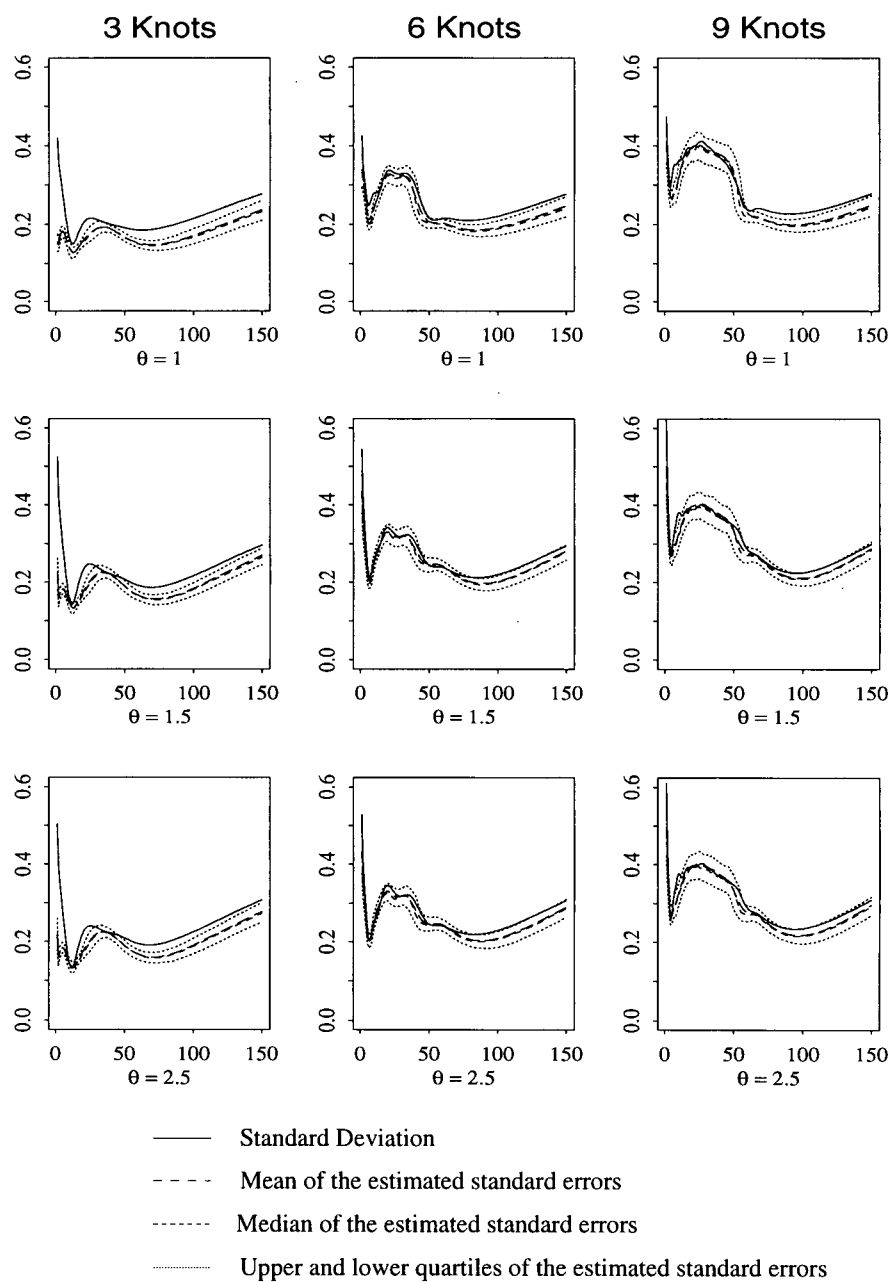


Figure 5.38: The standard deviations of the estimated log hazards ratio by the Different Knots method for data generated according to the non-proportional hazards model. The censoring rate is 25% for the “treatment”.



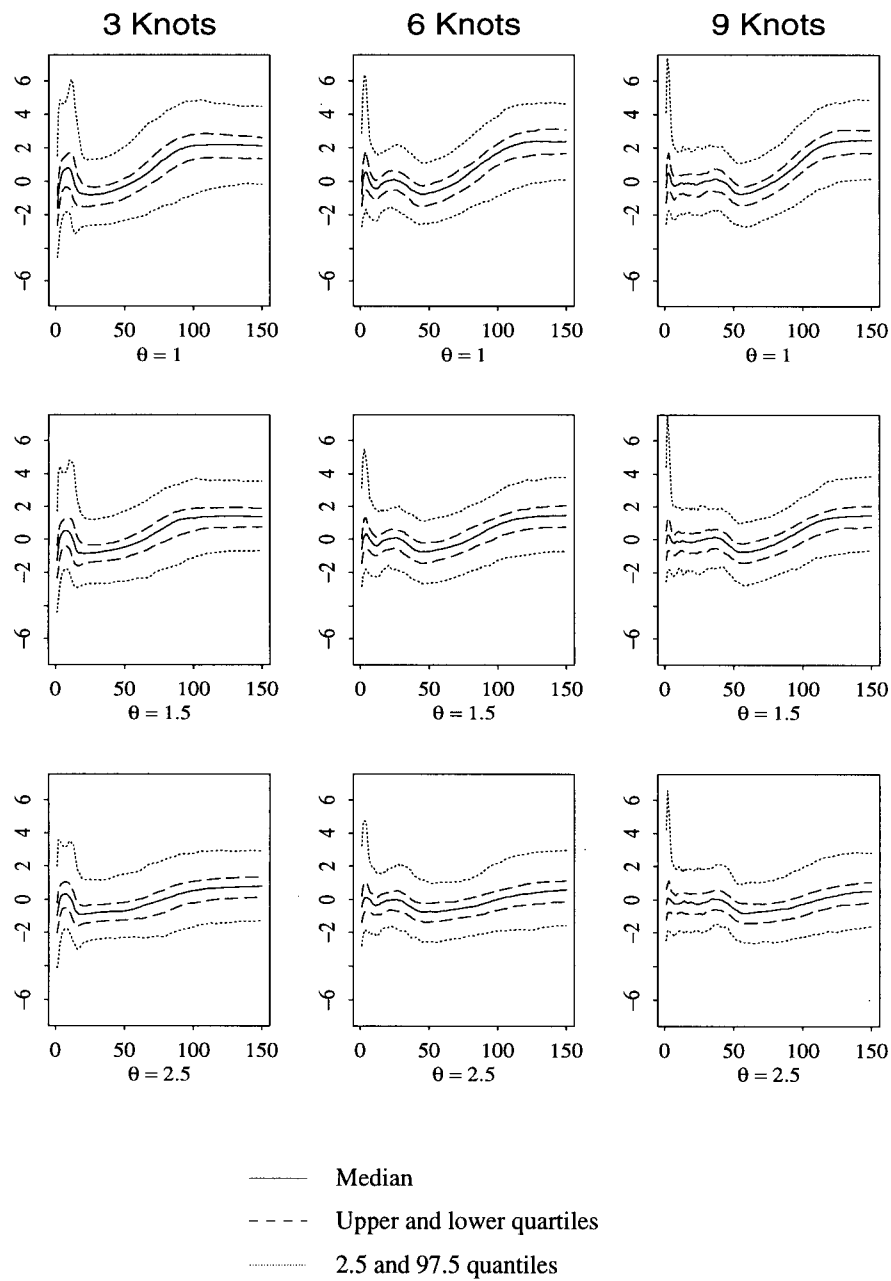


Figure 5.39: The quantiles of the pointwise  $z$  values of the estimated log hazards ratio by the Same Knots method for the data generated according to the proportional hazards model. The censoring rate is 25% for the “treatment”.

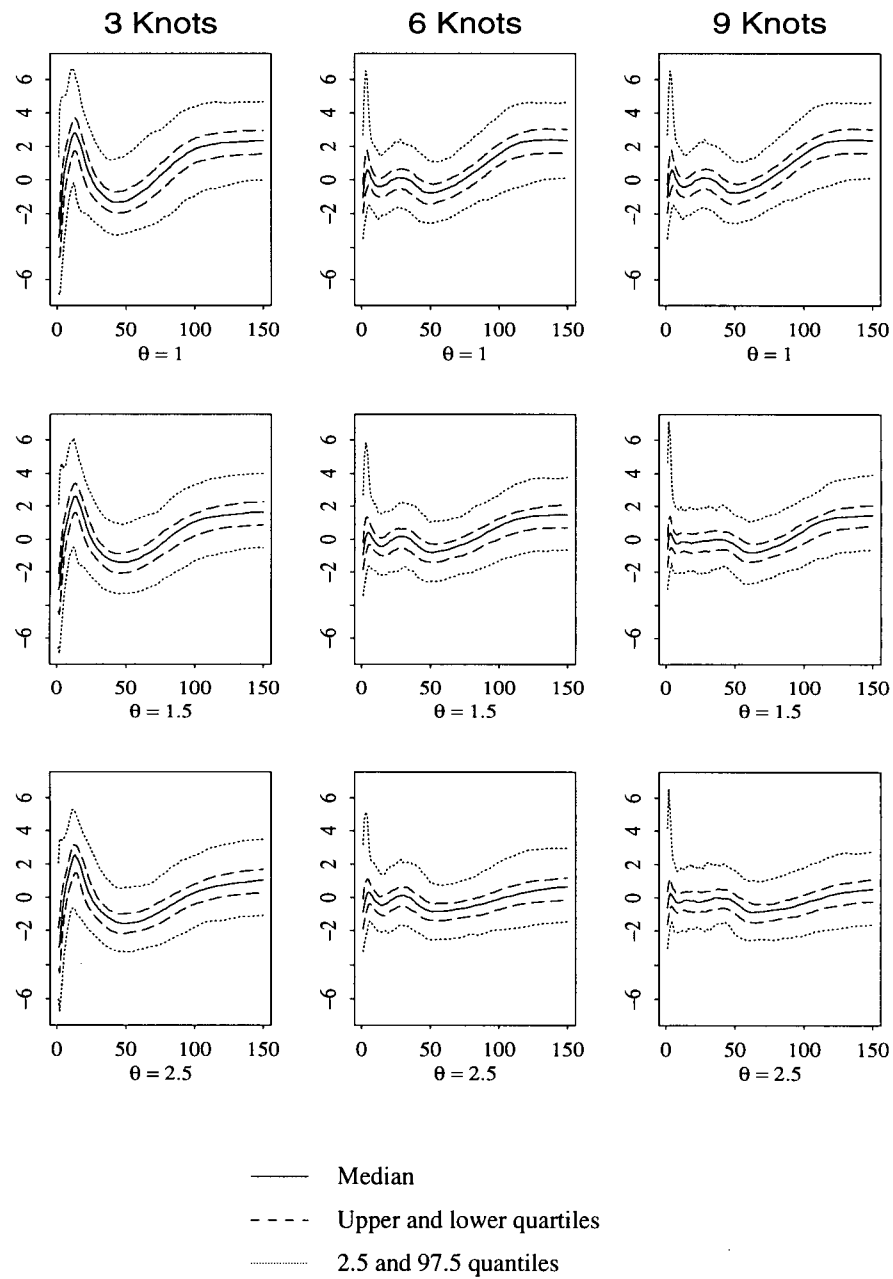


Figure 5.40: The quantiles of the pointwise  $z$  values of the estimated log hazards ratio by the Different Knots method for the data generated according to the proportional hazards model. The censoring rate is 25% for the “treatment”.

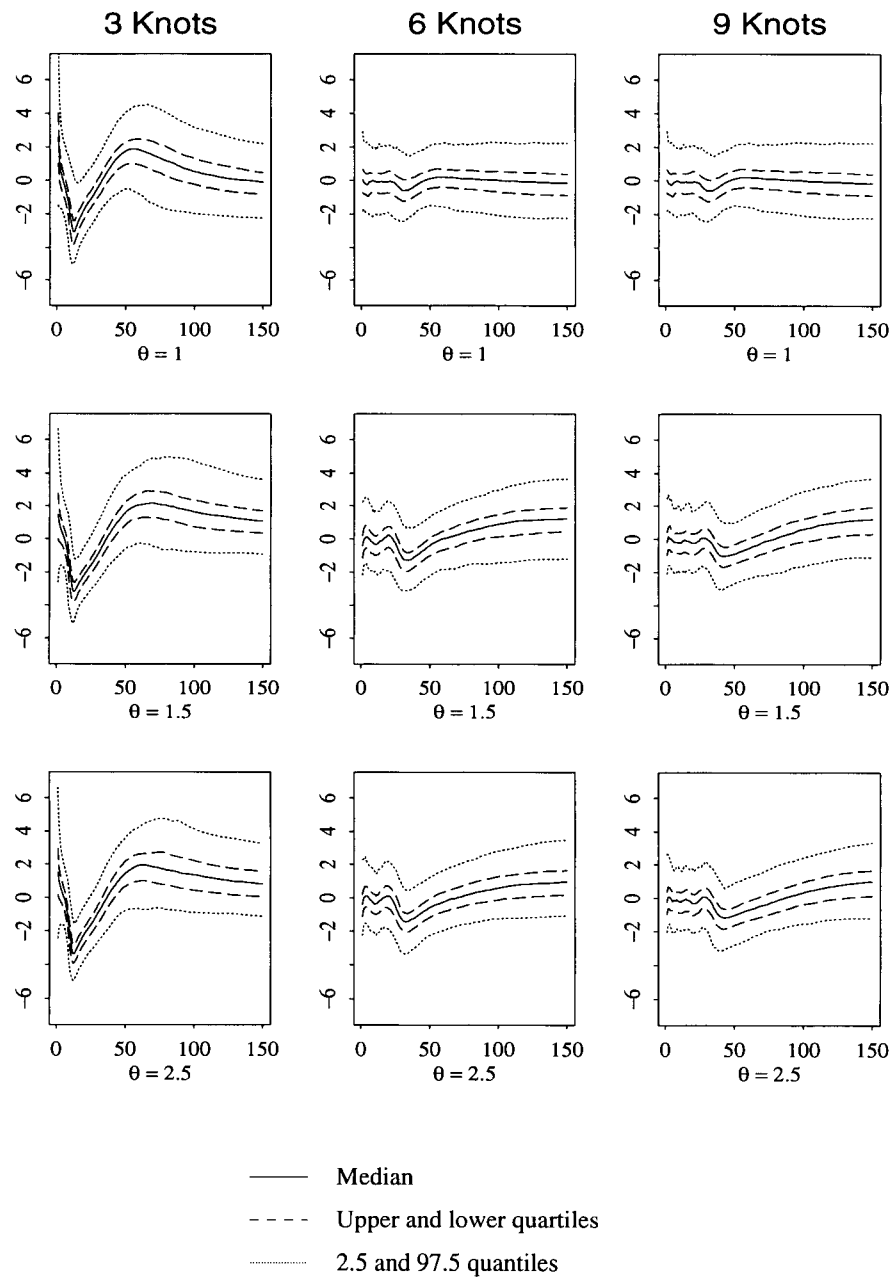


Figure 5.41: The quantiles of the pointwise  $z$  values of the estimated log hazards ratio by the Same Knots method for the data generated according to the non-proportional hazards model. The censoring rate is 25% for the “treatment”.

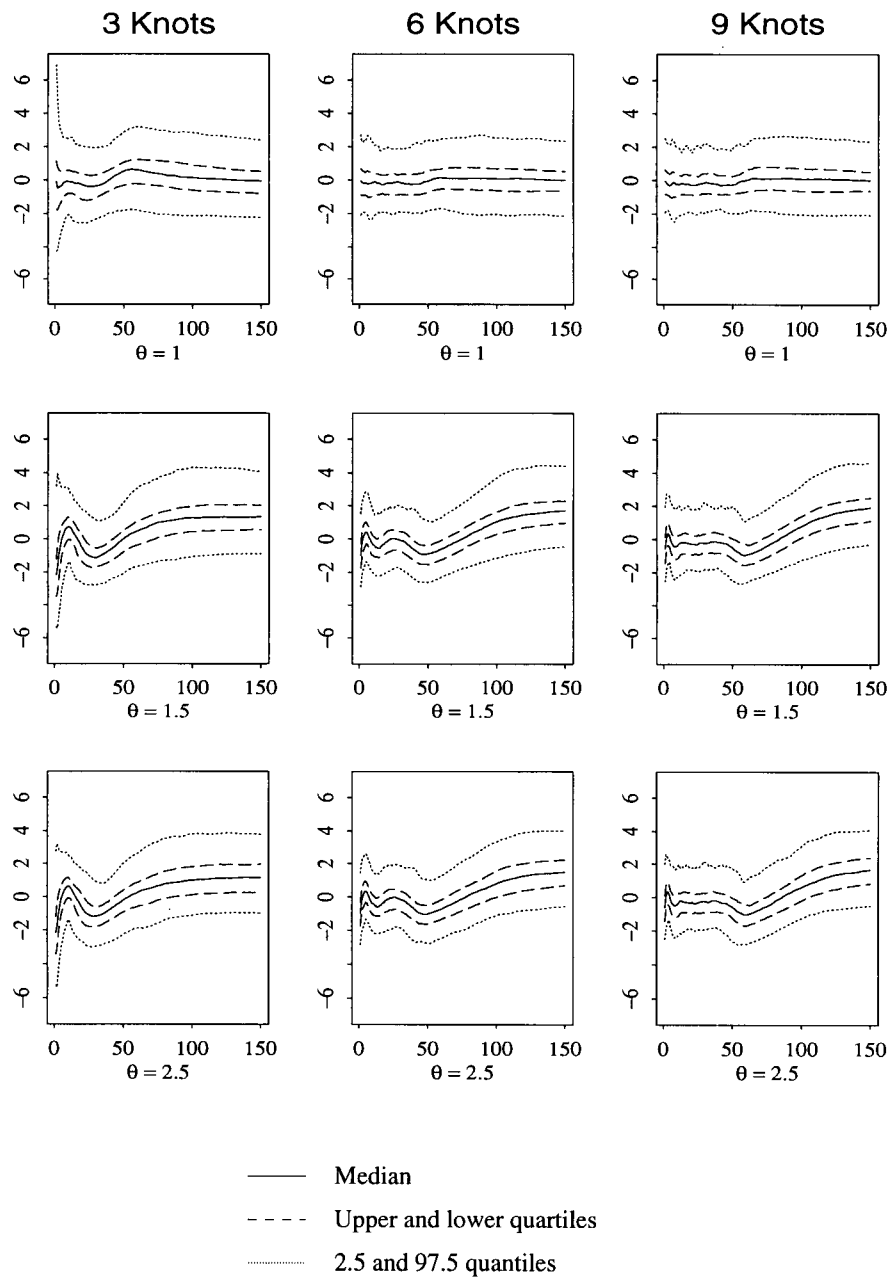


Figure 5.42: The quantiles of the pointwise z values of the estimated log hazards ratio by the Different Knots method for the data generated according to the non-proportional hazards model. The censoring rate is 25% for the “treatment”.

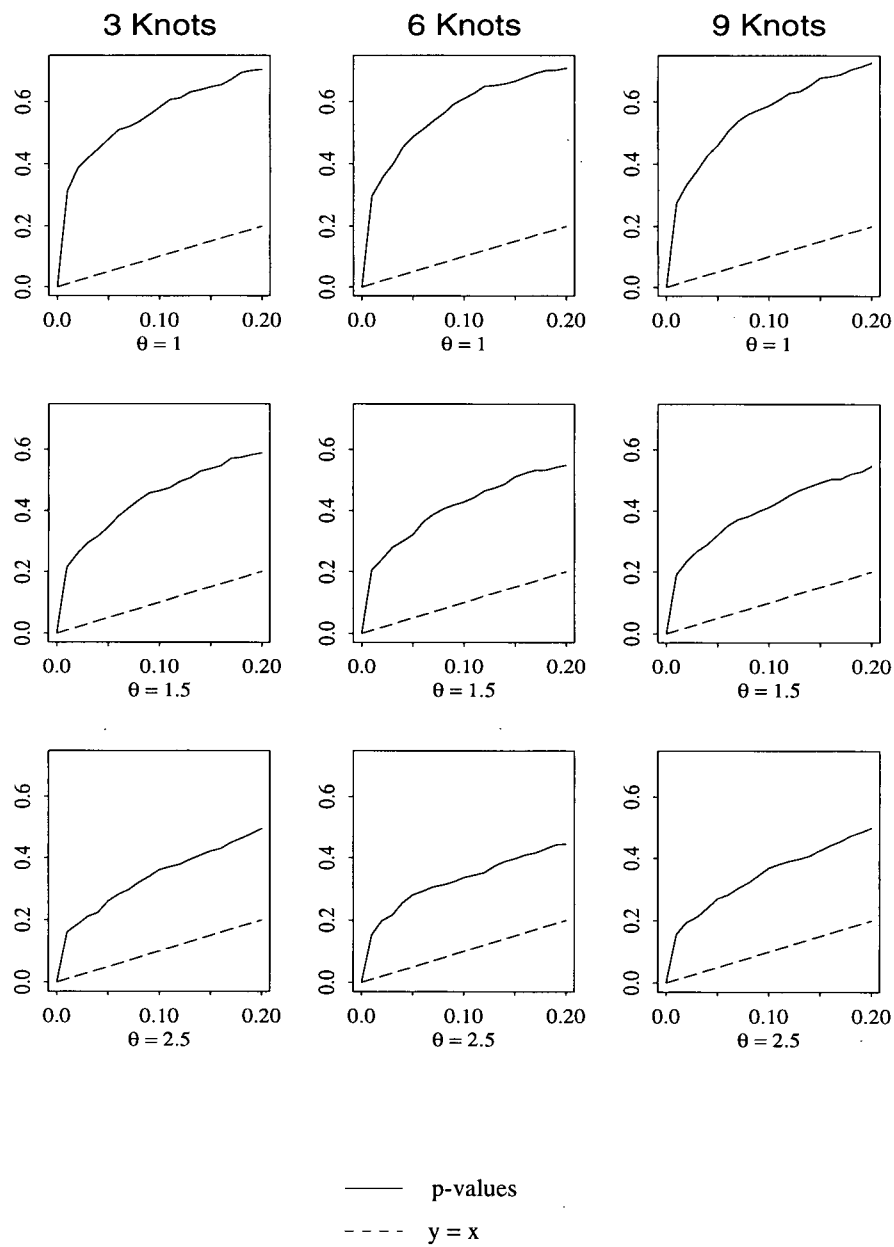


Figure 5.43: By the Same Knots method empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the proportional hazards model in Section 5.1 have proportional hazards.

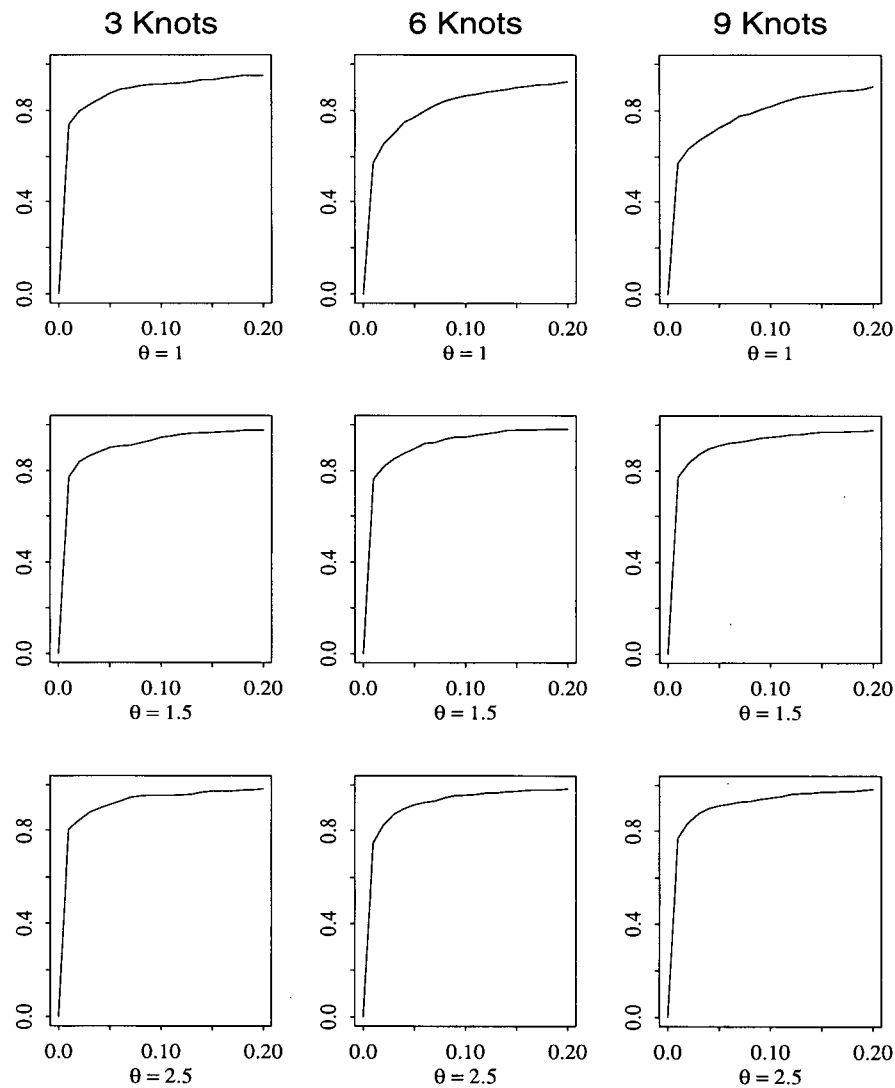


Figure 5.44: By the Same Knots method empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the non-proportional hazards model in Section 5.1 have proportional hazards.

## Chapter 6

### Conclusion

In this thesis a flexible parametric model, log hazard regression model for paired censored failure times, is proposed. In this model B-splines are used to estimate the log hazards of marginal failure times and the log hazards ratio of the two failure times. Consistent estimates of the standard errors for the estimated marginal log hazards and the log hazards ratio are presented as well. Based on this model we propose test procedures for the four hypotheses that the marginals follow an exponential or Weibull distributions and that the two failure times have the same distribution or have proportional hazards.

A simulation study indicates that when the censoring rate is not too high, the estimates of the log hazards and the log hazards ratio perform well within the range of the knots used and the estimates of standard errors for the estimated log hazards ratio are good, but a little bit small. This underestimation is bigger at time zero.

The simulation study also shows that the tests that the marginal failure times follow an exponential or Weibull distribution perform very well. But the test for the hypotheses that the two failure times have proportional hazards or have the same distribution tend to over-reject the null hypothesis when the null hypothesis is true. Also, the test is not

powerful.

The model was applied to the Diabetic Retinopathy Study data to assess the effectiveness of laser photocoagulation in delaying the onset of blindness for diabetic retinopathy patients. The conclusion of the analysis with the log hazard regression model agrees with those from the standard parametric models fitted by Huster et. al (1989), that is, there is a significant laser photocoagulation treatment effect. We use the log hazard model to test the hypothesis that the failure times of the treated eye and untreated eye have an exponential or a Weibull distribution. The test results indicate that there is significant evidence to reject the hypotheses that the failure time of the treated eye follows either exponential or Weibull distribution. The test results also show that there is significant evidence to reject that the failure time of the untreated eye follows an exponential distribution. But the test that the failure time of the untreated eye follows a Weibull yields a marginal result. The p-value is 0.053. The p-value for testing the hypothesis that the failure times of the treated eye and the untreated are proportional hazards is very high, 0.25. This, along with the tendency of the test to over-reject (see Section 5.3), indicates that it is reasonable to assume that the failure times of the treated eye and the untreated eye have proportional hazards. This result supports the analysis of Huster et. al. (1989).

We note that knot selection is a basic step to use the log hazard regression model. The simulation study shows that our Quantile Knots procedure performs well when the censoring rate is not too high. Even when we restrict the range of the knots, the estimates are fairly reliable within the range of the knots used.

We can not always choose knots using Kooperberg et. al's stepwise addition and stepwise deletion method, since the chosen knots often cause numerical problems (see Section 5.2). However, the knots they selected will give us some indication of the number and the location of the knots that should be used. From the experience of our simulation,



we give the following suggestions for using the log hazard regression models to analyze paired data:

1. Use the *heft* algorithm to choose a set of knots for each failure time as we mentioned in Section 3.4. If there is no warning message from the *heft* code and the calculation of estimated standard errors can be performed, then use the Different Knots method to estimate the log hazards ratio with the two sets of knots. If there is any warning message or the calculation of the estimated standard errors is impossible for one or both failure times, choose the knots for the one or both failure times by the quantiles method as we discussed in Section 3.4. Use the number of knots used by *heft*.
2. Test for exponential or Weibull. Fit the data with the log hazards model decided by the selected knots above. Then use the methods we mentioned in Section 3.5, to test if the marginals have exponential or Weibull distributions. If they do have those standard parametric distributions, use those distributions for modeling the marginal failure times.
3. To test proportional hazards, choose one set of knots for the two failure times. Use the quantiles of the non-censored failure times as discussed in Section 3.4. The number of knots should be between  $\max\{K_1, K_2\}$  and  $K_1 + K_2$ , where  $K_1$  and  $K_2$  are the numbers of knots selected by *heft* for  $T_1$  and  $T_2$ , respectively. Then use the method given in Section 3.5 to test whether the two failure times have proportional hazards.

From the definition of our log hazard regression model, we can generalize the model to multivariate censored survival data easily. We also can include covariates in this model. If we assume that the effects of covariates are independent of time, then the calculation of the estimates is straightforward based on the *heft* algorithm. We can model the

log hazard as a function of time plus a function of the covariate. If the effects of the covariates are dependent on time, then the log hazard would be modeled as a bivariate function of time and the covariate. However, the calculations of the estimates would be more complicated. To simplify calculations somewhat, Kooperberg et. al. modeled this bivariate function using linear B-spline and their tensor products.

There are some unanswered questions. First, why is the variability of the estimate of log hazard or log hazards ratio big at time zero? We might reduce the problem if we restrict the B-spline functions to be constant between zero and the smallest knot. Second, how do we choose the range of knots? As we see from our simulations, the estimates perform better within the range of the knots than outside of the range. We need more simulations to find a better way to determine the range of knots. Third, why do the test procedures for testing that the failure times have proportional hazard or have the same distribution work poorly, while the test procedures for testing that marginal failure times follows an exponential or Weibull distribution work well? We expect that a big sample size might improve the behavior of the test procedure. Finally, why does a high correlation level yield a better test for testing that failure times have proportional hazards? We need more simulations to understand the effect of the correlation level.

# Bibliography

- [1] Abrahamowicz, M. Ciampi, A. and Ramsay, J. O. (1992). Nonparametric Density Estimation for Censored Survival Data: Regression-Spline Approach. *The Canadian Journal of Statistics* **20**, 171 - 185.
- [2] Andersen, P. K. and Gill, R. D. (1982). Cox's Regression Model for Counting Processes: a Large Sample Study. *The Annals of Statistics* **10**, 1100 - 1120.
- [3] De Boor, C.(1978). *A Practical Guide to Splines. Applied Mathematical Sciences.* **27**, Springer-Verlag, New York.
- [4] Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika* **65**, 141 - 151.
- [5] Dabrowska, D. M. (1988). Kaplan-Meier Estimate on the Plane. *The Annals of Statistics* **16**, 1475 - 1489.
- [6] Diabetic Retinopatý Study Research Group (1981), Diabetic retinopathy study. *Investigative Ophthalmology and Visual Science*, **21**, 149 - 226.
- [7] Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1996). *Analysis of Longitudinal Data.* Clarendon Press, Oxford.

- [8] Eilers, P. H. C. and Marx, B. D. (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science* **11**, 89 - 102.
- [9] Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. M. Dekker, New York.
- [10] Ferguson, T. S. (1996). *A Course in Large Sample Theory*. Chapman and Hall, London/Weinheim/New York/Tokyo/Melbourne/Madras.
- [11] Fleming, T. R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons, Inc., New York.
- [12] Gill, R. D. (1992). Multivariate Survival Analysis. *Theory Probab. Appl.* **37**, 18 - 31.
- [13] Holt, J. D. and Prentice, R. L. (1974). Survival Analysis in Twin Studies and Matched Pair Experiments. *Biometrika* **61**, 17 - 30.
- [14] Huster, W. J., Brookmeyer, R., and Self, S. G. (1988). Modeling Paired Survival Data with Covariates. *Biometrics* **45**, 145 - 156.
- [15] Kooperberg, C., Stone, C. J, and Truong, Y. K. (1995). Hazard Regression. *Journal of the American Statistical Association* **90**, 78 - 94.
- [16] Kooperberg, C. (1998). Bivariate Density Estimation with an Application to Survival Analysis. *Journal of Computational and Graphical Statistics* **7**, 322 -341.
- [17] Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.
- [18] Murphy, R.P. and Patz, A. (1978). New Concepts in Management of Retinal Vascular Disorder. *Ophthalmology Update*, International Congress Series. No. 508, K.Mizuno and Y.Mitsui(eds), 115 - 125, Amsterdam: Excerpta Medica.

- [19] Oakes, D. (1982). A model for Association in Bivariate Survival Data. *Journal of the Royal Statistical Society, Series B* **44**, 414 - 428.
- [20] Sen, P. K and Singer, J. M. (1993). *Larger Sample Methods in Statistics*. Chapman and Hall, New York, London.
- [21] Shikin, E. V.(1995). *Handbook on Splines for the User*. CRC Press, Boca Rato.
- [22] Stone, C. J., Hansen, M. H., Kooperberg, C., and Truong, Y. K. (1997). Polynomial Splines and their Tensor Products in Extended Linear Modeling. *The Annals of Statistics* **25**, 1371 - 1470.
- [23] Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association*, **84**, 1065 - 1073.
- [24] Wei, L. J. and Lachin, J. M. (1984). Two-Sample Asymptotically Distribution-Free Tests for Incomplete Multivariate Observations. *Journal of the American Statistical Association* **79**, 653 - 661.

# **Appendix A**

## **Simulation Results for Bivariate Data with 50% Censoring**

Figures A.1 to A.10 show the simulation results for the proportional and non-proportional data generated as in the Section 5.1 with 50% censoring.

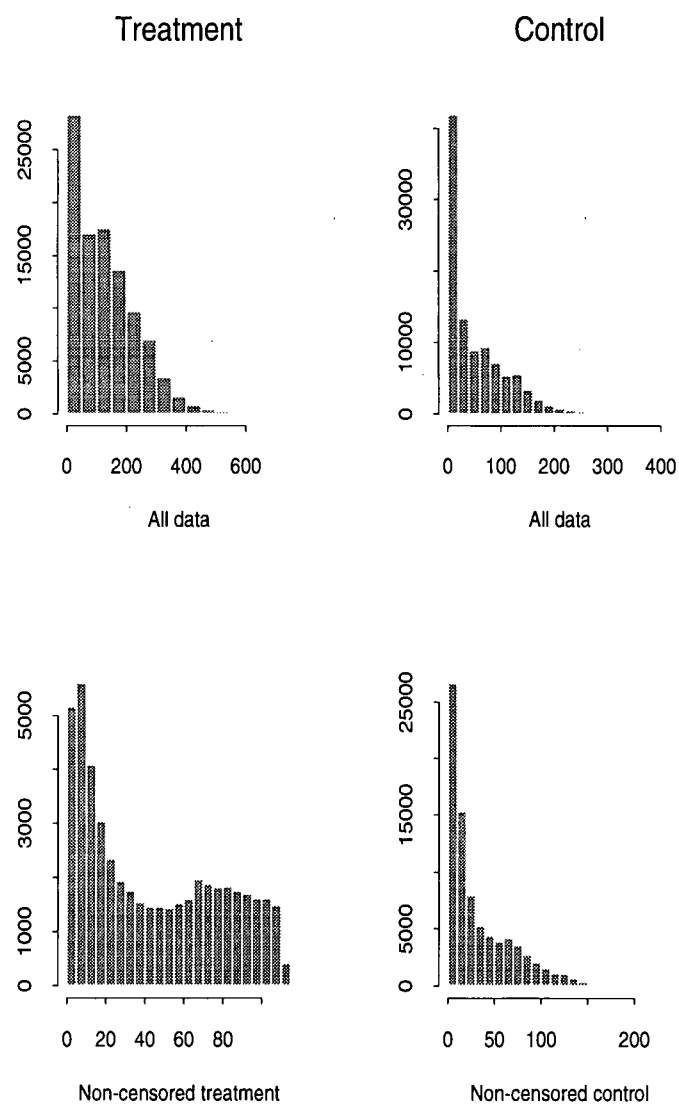


Figure A.1: The histograms of the marginal failure times, and non-censored marginal failure times for the data generated according to the proportional hazards model in Section 5.1. The censoring rate for the “treatment” is 50%.

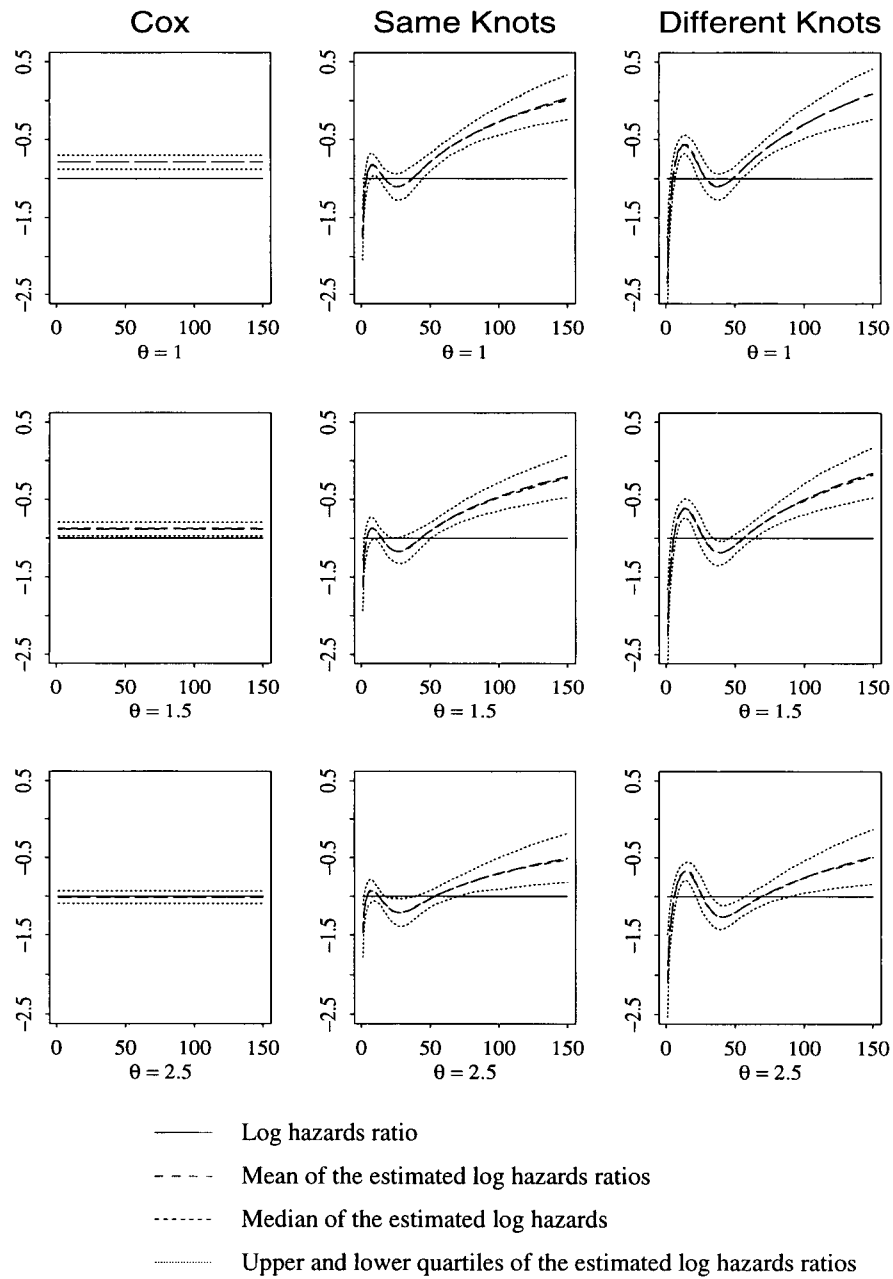


Figure A.2: The true log hazards ratio of the data as in the proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the 500 estimated log hazards ratios.



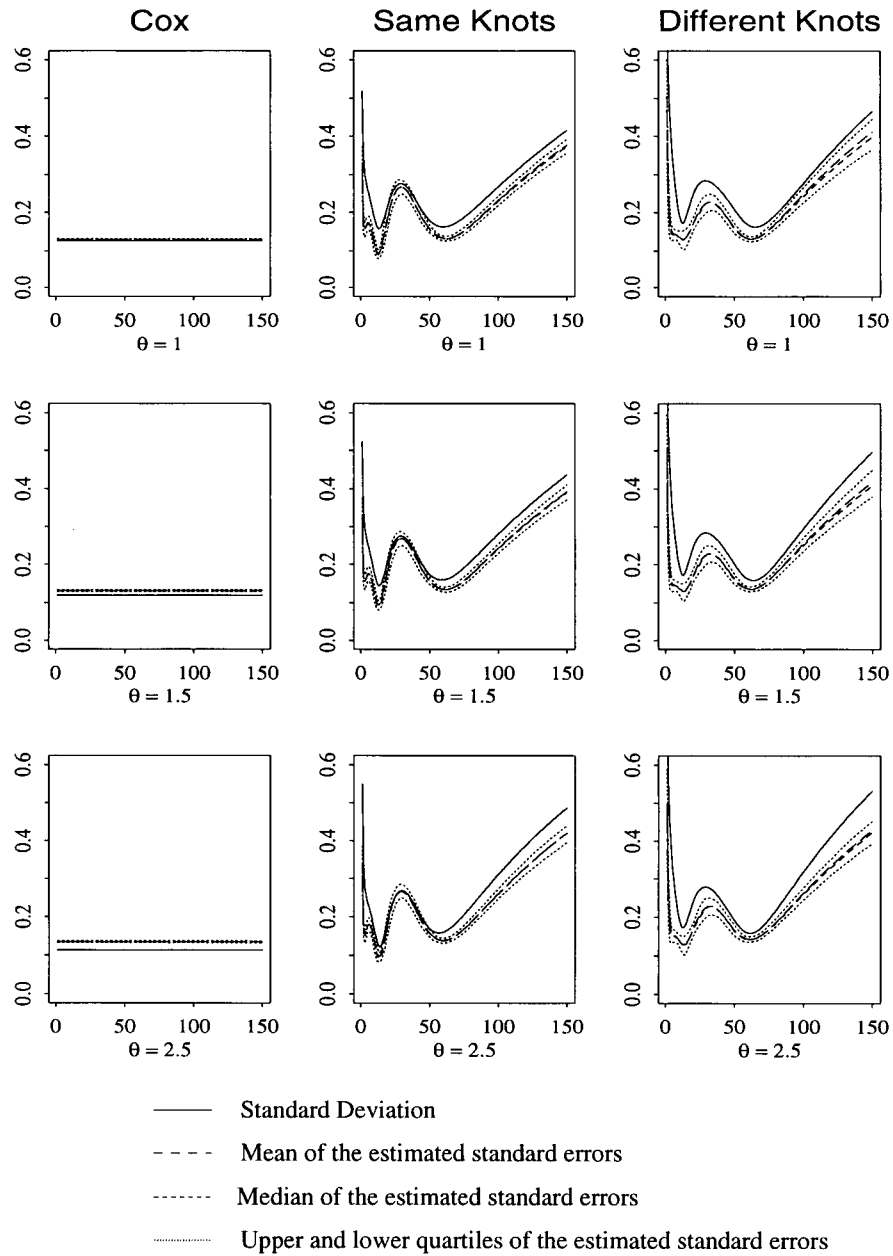


Figure A.3: The standard deviations of the estimated log hazards ratio for data generated according to the proportional hazards model. The censoring rate is 50% for the “treatment”.

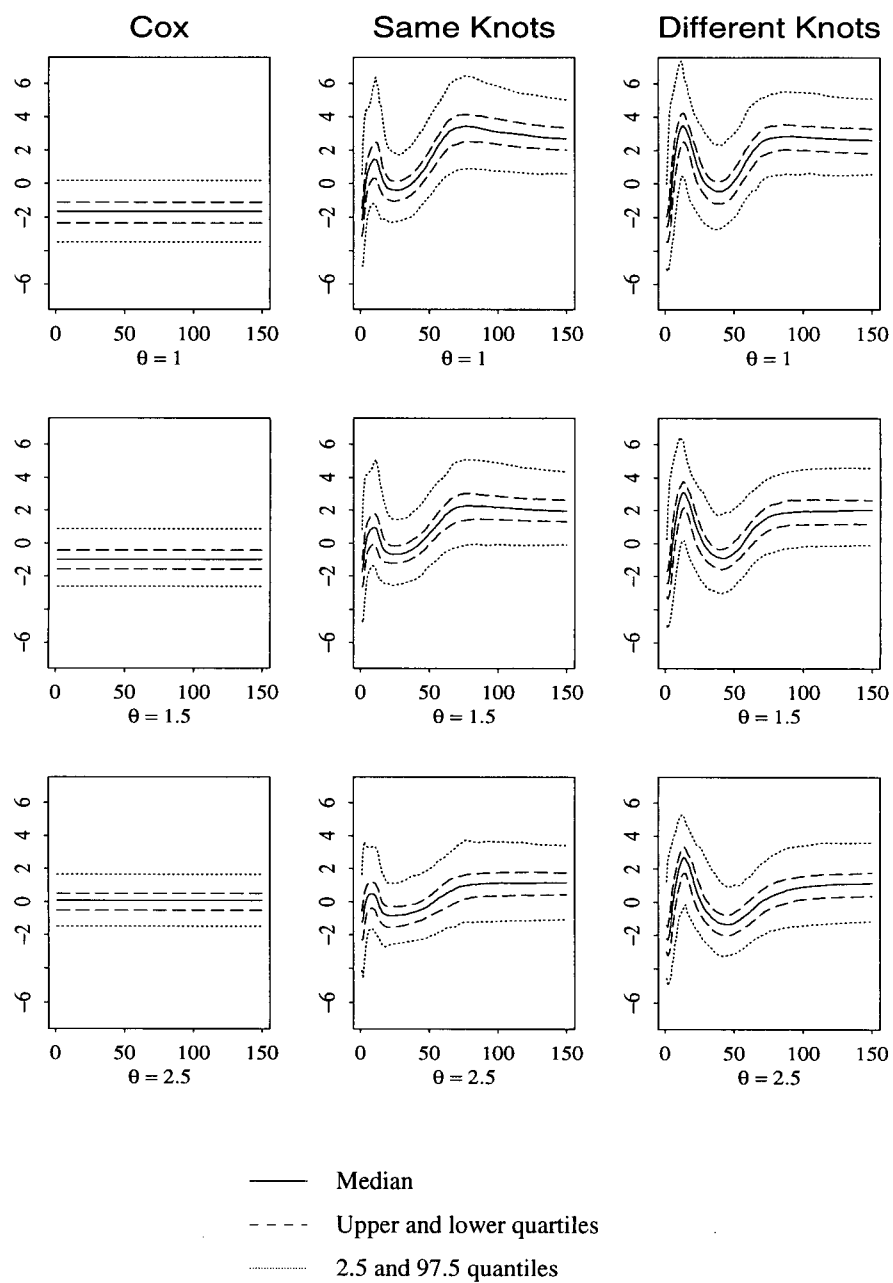


Figure A.4: The quantiles of the pointwise  $z$  values of the estimated log hazards ratio for the data generated according to the proportional hazards model. The censoring rate is 50% for the “treatment”.

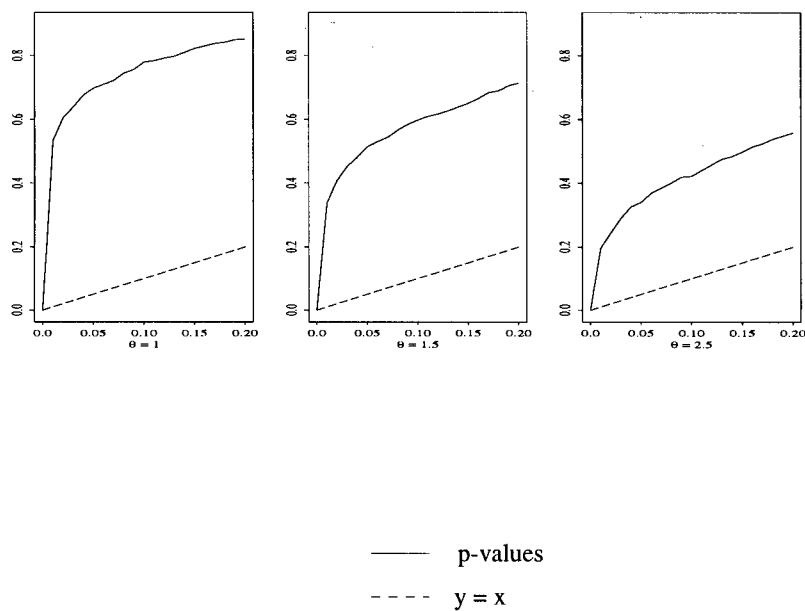


Figure A.5: Empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the proportional hazards model in Section 5.1 have proportional hazards.

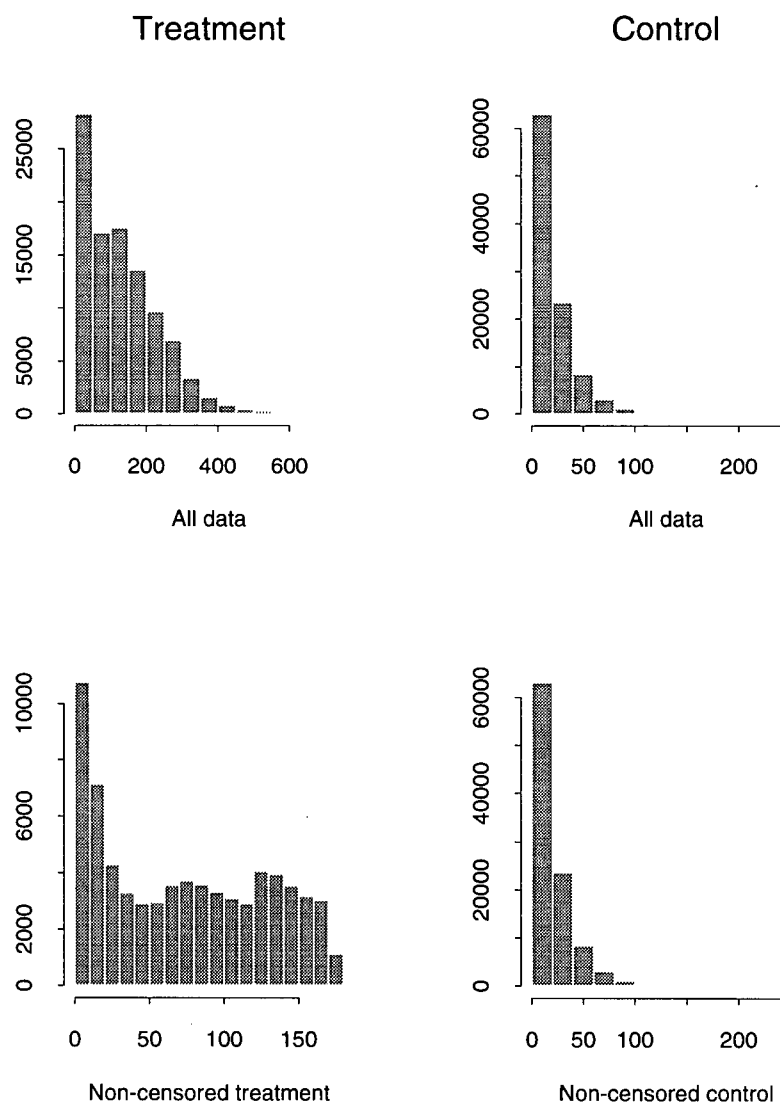


Figure A.6: The histograms of the marginal failure times, and the non-censored marginal failure times for the data from the non-proportional hazards model in Section 5.1. The censoring rate for the “treatment” is 50%.

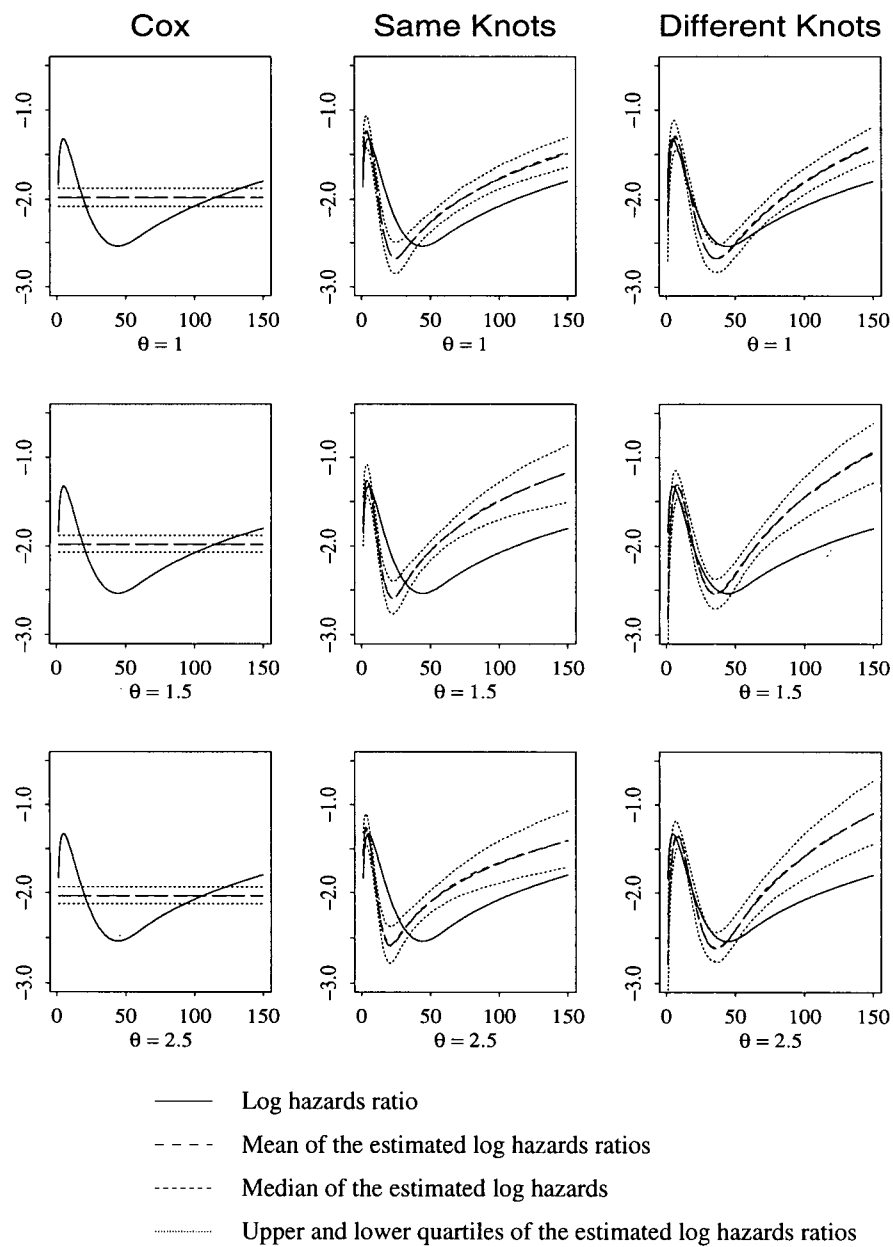


Figure A.7: The true log hazards ratio of the data as in the non-proportional hazards model defined in Section 5.1 and the pointwise quartiles and empirical mean of the 500 estimated log hazards ratios.

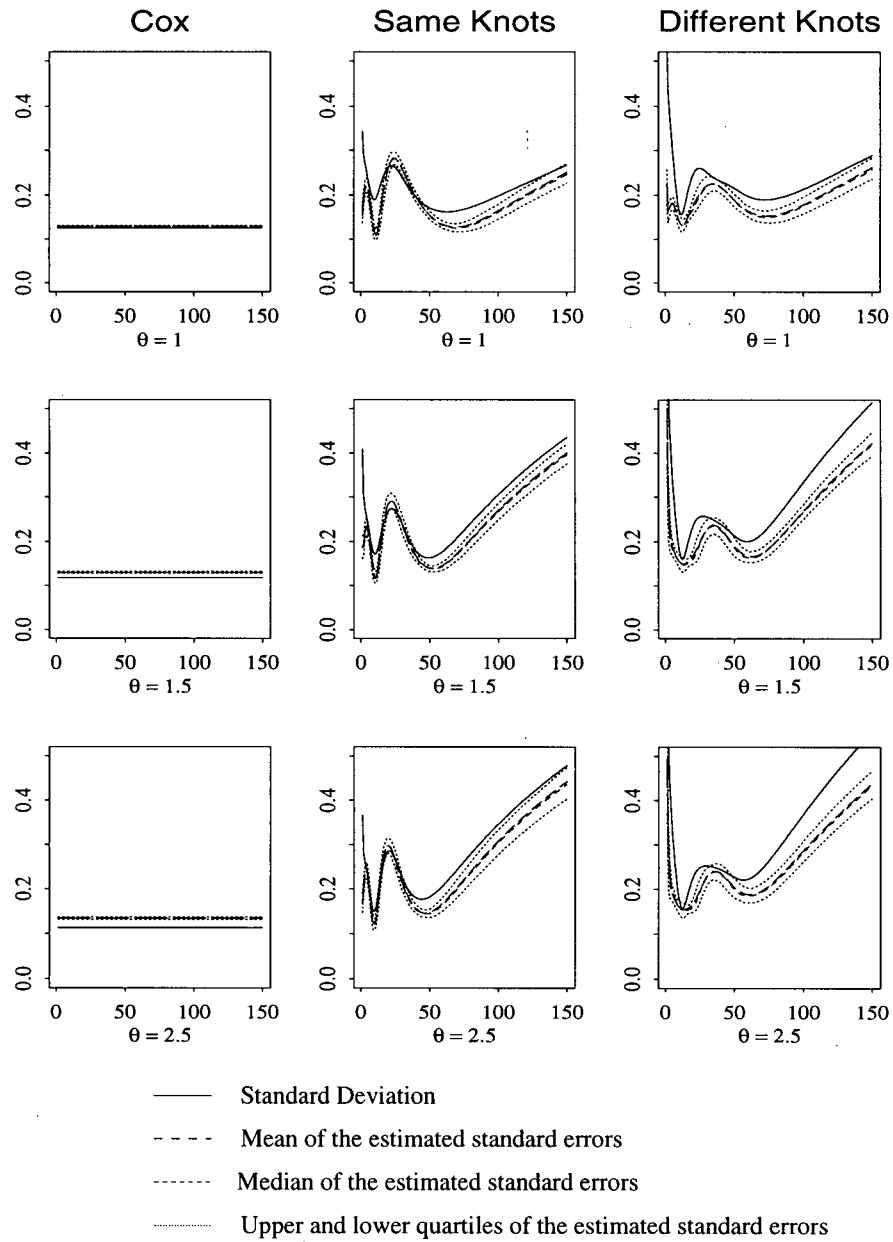


Figure A.8: The standard deviations of the estimated log hazards ratio for data generated according to the non-proportional hazards model. The censoring rate is 50% for the “treatment”.

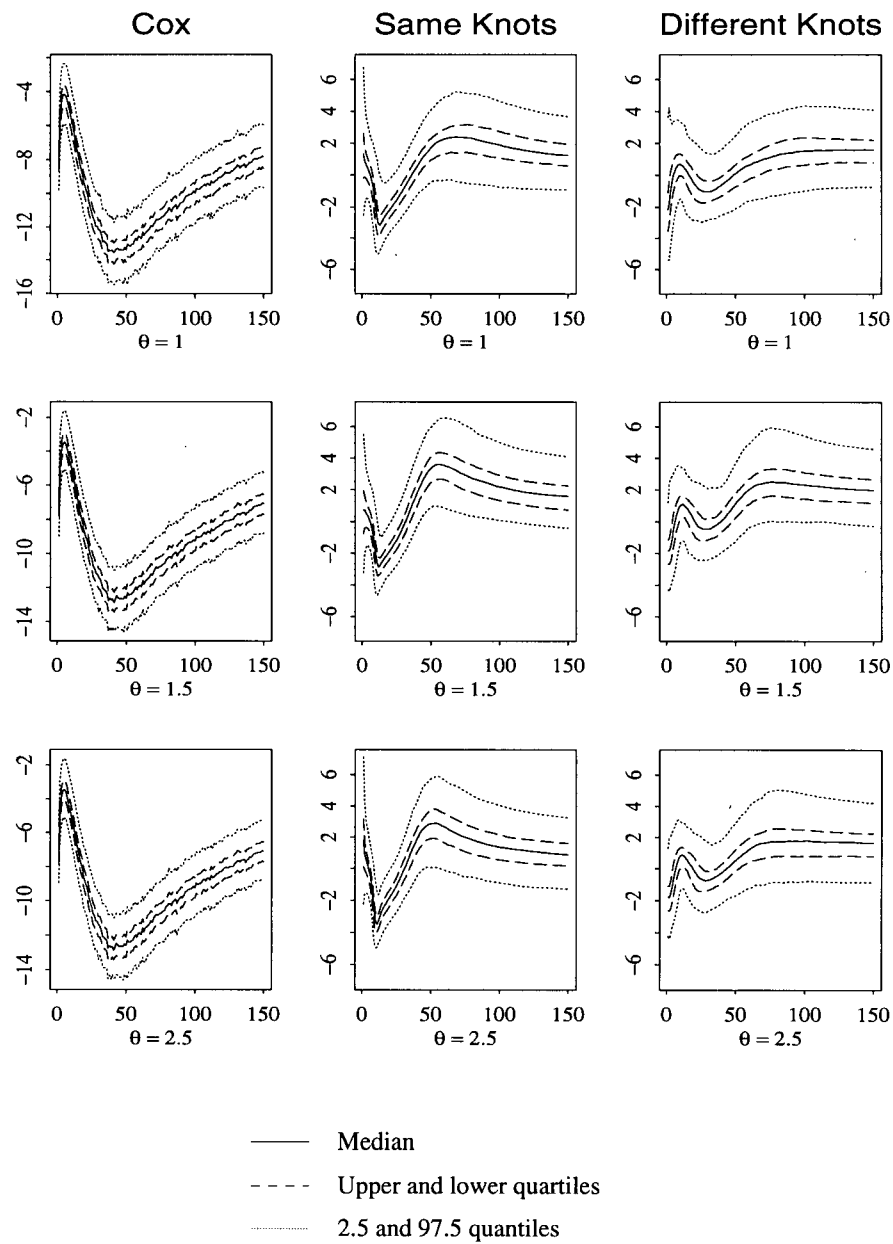


Figure A.9: The quantiles of the pointwise z values of the estimated log hazards ratio for the data generated according to the non-proportional hazards model. The censoring rate is 50% for the “treatment”.

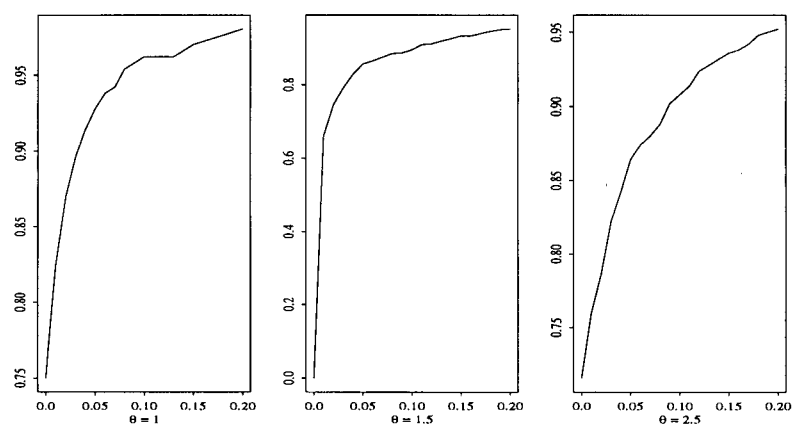


Figure A.10: Empirical distribution functions of the p-values for testing the hypothesis that the two failure times of the data generated as in the non-proportional hazards model in Section 5.1 have proportional hazards.