

## ABSTRACT

Title of dissertation: SEMIPARAMETRIC THRESHOLD  
REGRESSION ANALYSIS FOR  
TIME-TO-EVENT DATA

Ran Ji, Doctor of Philosophy, 2014

Dissertation directed by: Professor Xin He  
Department of Epidemiology and Biostatistics  
Professor Abram M. Kagan  
Department of Mathematics

Threshold regression is a relatively new alternative approach to the Cox proportional hazards model when the proportional hazards assumption is violated. It is based on first-hitting-time models, where the time-to-event data can be modeled as the time at which the stochastic process of interest first hits a boundary or threshold state. In this dissertation, we develop a semiparametric threshold regression model with flexible covariate effects. Specifically, we propose a B-spline approximation method to estimate nonlinear covariate effects on both the initial state and the rate of the process. We show that the spline based estimators are consistent and achieve the possible optimal rate of convergence under the smooth assumption. Simulation studies are conducted for practical situations, and the methodology is applied to a study of osteoporotic fractures that motivated this investigation.

To check the validity of threshold regression model with parametric link functions, we propose two supremum-type test processes: one is based on cumulative

sums of martingale residuals; the other one is based on censoring consistent residuals. The realizations of these test stochastic processes under the assumed model can be easily generated by computer simulation. We show that both tests are consistent against model misspecification. Both model checking methods have been applied to a kidney dialysis data set.

*Keywords:* B-spline; Censoring consistent residuals; Empirical process; Goodness-of-fit; Martingale residuals; Sieve maximum likelihood; Spline approximation; Survival analysis; Threshold regression; Wiener process.

SEMIPARAMETRIC THRESHOLD REGRESSION ANALYSIS  
FOR TIME-TO-EVENT DATA

by

Ran Ji

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2014

Advisory Committee:

Dr. Xin He, Co-Advisor

Dr. Abram M. Kagan, Co-Advisor

Dr. Armand M. Makowski, Dean's Representative

Dr. Mei-Ling Ting Lee

Dr. Paul J. Smith

© Copyright by  
Ran Ji  
2014

## Dedication

This dissertation is dedicated to my parents

嵇成新 查莎娜

## Acknowledgments

I am grateful for all the people who offered their help and support during my doctoral study.

I would like to express my sincere gratitude to my two advisors, Dr. Xin He and Dr. Abram M. Kagan for their invaluable guidance for my research. This dissertation would never have existed without their consistent support, encouragement, patience and immense knowledge. I would like to thank Dr. He for bringing me these interesting research topics to study. I very much appreciate for his patient guidance, advice he has provided on both my academic and personal development throughout these years. I am also very grateful to Dr. Kagan for his insightful discussions about the research. He has been supportive and has given me great freedom to pursue research topics I am interested in. I thank them for everything they have done during the course of supervision of this dissertation.

I would like to extend my gratitude to my committee members, Drs. Paul Smith, Mei-Ling Ting Lee and Armand Makowski for their constructive suggestions and comments. I would like to thank Dr. Mei-Ling Ting Lee for her support and helpful discussions on related topics. Special thanks to the professors in the mathematical statistics program: Dr. Paul Smith, Dr. Abram Kagan, Dr. Eric Slud, Dr. Grace Yang, Dr. Leonid Korolov and Dr. Benjamin Kedem for their great courses and help during my graduate study. In particular, I thank Dr. Slud for referring me to related literature on sieve maximum likelihood.

I am indebted to my parents for everything. You made me who I am. I am

also grateful to my girlfriend Jiehong Lou for her constant love, understanding and support on this journey.

My friends, who were sources of joy and support, also deserve my sincere thanks.

## Table of Contents

List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 Background	1
1.2 Literature Review: Threshold Regression	8
1.3 TR Model Based on One Dimensional Wiener Process	12
1.4 Overview of the Dissertation	17
2 Threshold Regression with Flexible Covariate Effects	19
2.1 Introduction	19
2.1.1 Background	19
2.1.2 Review of B-spline Approximation	20
2.2 Model and Likelihood	25
2.3 Estimation Procedure	28
2.4 Asymptotic Property of Estimator	33
2.4.1 Introduction	33
2.4.2 Consistency	38
2.4.3 Convergence Rate	59
2.5 Simulation Study	66
2.6 An Application	69
2.7 Summary	72
3 Model Checking Techniques for TR Models	79
3.1 Model Checking Based on Martingale Residuals	80
3.1.1 Test Statistic	81
3.1.2 Consistency	88
3.2 Model Checking Based on Censoring Consistent Residuals	90
3.2.1 Test Statistic	91
3.2.2 Consistency	94
3.3 Simulation Study	96



3.4	An Application . . . . .	99
3.5	Summary . . . . .	101
4	Future Study	102
A	Appendix	104
A.1	R Functions for Model Checking Based on Martingale Residuals . . .	104
A.2	R Functions for Model Checking Based on Censoring Consistent Residuals . . . . .	109
	Bibliography	112

## List of Tables

2.2	Predictor variables in the analysis of osteoporotic fracture data set. . .	70
2.1	Simulation results for the parametric components of link functions based on 100 replications. . . . .	73
2.3	Constant regression coefficient estimates in the osteoporotic fracture data set. . . . .	77
3.1	Empirical sizes of test statistics at the $\alpha = 0.05$ significance level. . .	99
3.2	Empirical powers of test statistics at the $\alpha = 0.05$ significance level. .	99
3.3	Estimated regression coefficients in the analysis of kidney dialysis data set. . . . .	100

## List of Figures

1.1	Two illustrative sample paths of health status: path 1 reaches the threshold boundary zero before the end of a follow-up time of 600 days and event time is observed; path 2 has not reached the threshold boundary by the end of the follow-up time and event time is censored.	16
2.1	Calculation steps for $B_{1,3,t}(t)$ . Here, $B_{3,1,t}(t) = I[t_3, t_4)$ , $B_{2,1,t}(t) = I[t_2, t_3)$ and $B_{1,1,t}(t) = I[t_1, t_2)$ based on formula (1.1.1).	23
2.2	Calculation steps for $B_{7,3,t}(t)$ . Here, $B_{9,1,t}(t) = I[t_9, t_{10})$ , $B_{8,1,t}(t) = I[t_8, t_9)$ and $B_{7,1,t}(t) = I[t_7, t_8)$ based on formula (1.1.1).	23
2.3	Simulation results based on sample size 200 and 100 runs for covariate functions $\varphi(W_\mu)$ and $\eta(W_\delta)$ . Solid line: true function; dotted line: estimated function; dashed line: 95% point-wise CI.	74
2.4	Simulation results based on sample size 400 and 100 runs for covariate functions $\varphi(W_\mu)$ and $\eta(W_\delta)$ . Solid line: true function; dotted line: estimated function; dashed line: 95% point-wise CI.	75
2.5	Simulation results based on sample size 600 and 100 runs for covariate functions $\varphi(W_\mu)$ and $\eta(W_\delta)$ . Solid line: true function; dotted line: estimated function; dashed line: 95% point-wise CI.	76
2.6	Estimated functional coefficient $\hat{\eta}(age)$ .	78
3.1	Predicted survival curve from TR model versus Kaplan-Meier curve for the kidney dialysis data set.	101

## Chapter 1: Introduction

### 1.1 Background

Analysis of time-to-event data has become a very important and active area in statistical research in the past several decades. One major issue in the study of time-to-event data is how to deal with censored data. Censoring means that the time to the event of interest is not available for all individuals due to loss to follow-up or non-occurrence of event before the end of study. It often occurs in clinical trials and reliability studies. For example, in a two year clinical study evaluating treatment effect of a new drug for lung cancer, death from the disease is the event of interest. Those patients who drop out of the study before the end and those who have not died by the end of the study are censored. For those patients, we only know that their event times are beyond some time points. There are different types of censoring, such as right censoring, left censoring and interval censoring. Right censoring occurs if the event of interest occurs at a time after a right bound but the actual time of event is unknown. Left censoring is when the event has already occurred before the observation time and the time of event is not known precisely. Interval censoring occurs when an interval bounding the event time is known but the exact event time is not known.

Traditional statistical methods can not be used in analyzing time-to-event data since there is often a mixture of complete and incomplete observations because of censoring. Many methods including both parametric and semiparametric approaches have been proposed to model time-to-event data in the past several decades. Survival analysis which focuses on time-to-event data has become one of the hottest topics in statistical research.

*Survival function* and *hazard function* are two basic concepts in survival analysis. Let  $T$  denote the random variable representing time to event of interest with the cumulative distribution function  $F(t) = \Pr(T \leq t)$ . The survival function  $S(t)$  is defined as

$$S(t) = \Pr(T > t) = 1 - F(t), \tag{1.1.1}$$

which is the probability that the event has not happen by time  $t$ . The hazard function  $\lambda(t)$  gives the instantaneous rate of occurrence of the event at time  $t$ . It is defined formally as

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt | T \geq t)}{dt}. \tag{1.1.2}$$

Let

$$\Lambda(t) = \int_0^t \lambda(s) ds \tag{1.1.3}$$

denote the cumulative hazard function. Let  $f(t)$  be the probability density function

of  $T$  if  $F(t)$  is absolutely continuous. By some simple probabilistic arguments, one can get following basic relations between survival function, hazard function and probability density function:

$$\begin{cases} \lambda(t) = -\frac{d}{dt}(\log S(t)) \\ S(t) = \exp(-\Lambda(t)) \\ f(t) = \lambda(t)S(t). \end{cases}$$

As shown above, specifying one of the three functions  $S(t)$ ,  $\lambda(t)$  and  $f(t)$  specifies the other two functions. The survival function  $S(t)$  and the cumulative hazard function  $\Lambda(t)$  can be estimated by *Kaplan-Meier estimator* and *Nelson-Aalen estimator* (Chapter 1, Kalbfleisch and Prentice 2002), respectively. Both of them are constructed based on nonparametric methods. Because of the flexibility of nonparametric statistics they are commonly adopted by researchers. In some special cases, parametric distributions (e.g., Weibull, Exponential, Gompertz-Makeham, Gamma, etc.) can also be used to model statistical characteristics of the failure time  $T$ . One important part of modern survival analysis is to explore effects of covariates on individual's survival experience. Next, we will briefly review some popular survival analysis models that can be used to deal with covariate effects in analyzing time-to-event data.

**Cox model:** The Cox model (Cox, 1972) is often referred to as the proportional hazards model. Let  $\lambda(t|\mathbf{Z})$  denote the hazard function for an individual with

a  $p \times 1$  vector  $\mathbf{Z}$  of covariates. Under the Cox model, the hazard function  $\lambda(t|\mathbf{Z})$  can be written as the product of an unspecified baseline hazard function  $\lambda_0(t)$  and  $\exp(\boldsymbol{\beta}'_0\mathbf{Z})$ , i.e.,

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) \exp(\boldsymbol{\beta}'_0\mathbf{Z})$$

where  $\boldsymbol{\beta}_0$  is a  $p \times 1$  vector of unknown regression parameters. One can estimate  $\boldsymbol{\beta}_0$  through the partial likelihood method. Let  $\hat{\boldsymbol{\beta}}$  denote the resulting estimate of  $\boldsymbol{\beta}_0$ . Andersen and Gill (1982) and Tsiatis (1981) discussed the asymptotic behavior of  $\hat{\boldsymbol{\beta}}$ .

As shown in the above functional form of  $\lambda(t|\mathbf{Z})$ , the covariates have a linear effect on the log hazard function in Cox model. This assumption is relatively strong and may not always be guaranteed. To relax this assumption and enhance the accuracy of the model, nonlinear covariate effects have been considered by many authors. Hastie and Tibshirani (1986) proposed a local scoring algorithm for estimating smooth covariate functions (potentially nonlinear) in a class of generalized additive models (which is a set of likelihood based regression models). Its application on the Cox model was also discussed. O'Sullivan (1988) adopted smoothing splines to estimate nonlinear covariate effects in the Cox model through penalized partial likelihood. Sleeper and Harrington (1990) discussed using splines to approximate additive nonlinear covariate effects in the Cox model. Strawderman and Tsiatis (1996) and Huang (1999) explored consistency and convergence rate for the spline based estimators in the Cox model with unknown functional forms of covariates by various approaches.

In the application of survival analysis, some prognostic factors may vary over time or other prognostic factors. For example, a drug may gradually lose its efficacy on treating certain diseases because of virus mutation. Several authors have studied Cox models with varying coefficients. Murphy and Sen (1991) proposed sieve method to estimate varying coefficient functions in the Cox model by modeling them as piecewise constants. Cai and Sun (2003) adopted the kernel smoothing approach to estimate time-dependent regression coefficients in the Cox model locally, based on partial likelihood within an interval around each time point. They also established the pointwise consistency and asymptotic normality of their estimators. This method was further discussed by Tian *et al.* (2005). Our model development in Chapter 2 is inspired by discussions of these authors.

**Accelerated failure time model:** The accelerated failure time model (Chapter 7, Kalbfleisch and Prentice 2002) describes the logarithm of the event time  $T$  as a linear function of covariates  $\mathbf{Z}$ . Because of the way it links event time with covariates, the accelerated failure time model has a direct physical interpretation. The model can be written in the following general form

$$\log(T) = \boldsymbol{\beta}_0' \mathbf{Z} + \varepsilon,$$

where  $\boldsymbol{\beta}_0$  is a  $p \times 1$  vector of unknown regression parameters and  $\varepsilon$  is an independent error term under some distribution. Let  $S_0(\cdot)$  denote the survival function of  $\varepsilon$  and



let  $T_0 = \exp(\varepsilon)$  be the baseline survival time. By some simple arguments, the survival function of  $T$  given  $\mathbf{Z}$  can be written as

$$S(t|\mathbf{Z}) = S_1(t \exp(-\boldsymbol{\beta}'_0 \mathbf{Z})),$$

where  $S_1(\cdot)$  is the survival function of  $T_0$  and  $S_1(\cdot) = S_0(\log(\cdot))$ . This can be interpreted as a reduction of the survival time  $T$  by a factor  $\exp(-\boldsymbol{\beta}'_0 \mathbf{Z})$  as compared to the baseline survival time. If the distribution of  $\varepsilon$  is known,  $\boldsymbol{\beta}_0$  can be estimated based on the maximum likelihood method.

When the distribution of  $\varepsilon$  is unknown, Buckley and James (1979) developed an estimation procedure for  $\boldsymbol{\beta}_0$  by modifying the least squares normal equations to accommodate censored observations. Tsiatis (1990) derived a linear rank test statistics based estimating equations for regression parameters  $\boldsymbol{\beta}_0$ . Asymptotic properties of the rank estimator were studied by Tsiatis (1990), Wei *et al.* (1990), Lai and Ying (1991), and Wei (1993), among others. Ritov (1990) discussed the relation between two classes of estimating equations and showed that they are asymptotically equivalent.

The Cox model has been the most commonly used approach among others, since one does not need to specify a particular baseline hazard function. However, the proportional hazards assumption may not always be satisfied in real applications and one needs to carefully check this assumption when using the standard Cox

model.

The threshold regression (TR) model can be one alternative approach when the proportional hazards assumption is violated. The TR model, which is based on first-hitting-time (FHT) models, describes the time-to-event as the first time a stochastic process reaches a boundary. FHT models have been widely used in many fields, such as sociology, engineering, medicine, etc. Lee *et al.* (2006) pointed out that a FHT model has two basic components:

1. a parent stochastic process  $\{X(t), t \in \mathcal{T}, x \in \mathcal{X}\}$  with an initial value  $X(0) = x_0$ , where  $\mathcal{T}$  is the time space and  $\mathcal{X}$  is the state space of the process.
2. a boundary set  $\mathcal{B}$ , where  $\mathcal{B} \in \mathcal{X}$ .

Then the event time  $T$  can be interpreted as the first time the parent stochastic process  $X(t)$  starting at  $X(0)$  reaching the threshold  $\mathcal{B}$ , that is,  $T = \inf\{t : X(t) \in \mathcal{B}\}$ .

Lee *et al.* (2006) also mentioned many possible choices for the parent stochastic process  $X(t)$ , such as Gamma process, Ornstein-Uhlenbeck process, Wiener process, etc. Because of its feasibility in its assumption and explicit distribution of first hitting time  $T$ , Wiener process has been chosen by many researchers as the parent stochastic process in their studies. It can be shown that the first hitting time  $T$  of a Wiener process with a fixed threshold  $\mathcal{B}$  follows an inverse Gaussian distribution. For the statistical properties of the inverse Gaussian distribution one can refer to Tweedie (1957). Lancaster (1972) adopted a FHT model based on the Wiener

process to describe the duration of a strike and he found that the fit of the model to the observations was very close. Eaton and Whitmore (1977) considered the length of hospital stay as a Wiener process based FHT model and got a better fitted result compared with results of other models. Chhikara and Folks (1977) suggested using the inverse Gaussian distribution to model the lifetime of a industrial product because of its advantages in statistical characterization. Whitmore (1979) adopted a FHT model based on the Wiener process for describing the employee service time. Doksum and Hoyland (1992) developed a time-scale changed Wiener process to model accumulated decay of a product under accelerated life testing trial where time-transformed inverse Gaussian distribution were used for parameter estimation. However, in these applications, the covariate effects were not taken into account for the heterogeneity in populations.

## 1.2 Literature Review: Threshold Regression

As pointed out by Lee *et al.* (2006), to be truly valuable in applications, FHT models must be capable of extension to include regression structures since regression structures allow the effects of covariates to explain the inherent dispersion of the data. Usually, FHT models with regression structures are referred to as threshold regression.

Whitmore (1983) proposed a regression method for analyzing censored time-to-event data. He adopted an FHT model based on the Wiener process starting from

origin with a drift parameter  $\delta$  and a volatility parameter  $v$  for modeling censored failure time data and set the absorbing threshold as a constant (one unit from origin). The drift parameter  $\delta$  was modeled as a linear function of covariates. An EM algorithm was developed to find the maximum likelihood estimates of the regression parameters  $\beta$  and volatility parameter  $v$  where the first passage time (event time) follows the inverse Gaussian distribution. The similarities and differences of other alternative general regression methodologies (i.e., Cox and log-linear models) from the inverse Gaussian model were also discussed.

Lee and Whitmore (2010) investigated in depth the connections between the Cox model and TR models and showed that the Cox model is a special case of TR models for most purposes. They showed that TR models can yield proportional hazards functions mainly through the following two methods: (1) changing the time scale of TR model; (2) altering the boundary of TR model. They also discussed how to estimate the time scale and boundary with or without the proportional hazards assumption.

Lee *et al.* (2004) used the TR model for analyzing lung cancer risk in railroad workers. They modeled the latent health status process by a Wiener process with an unknown positive initial value  $x_0$  and the occurrence of event as the first time this Wiener process reaching threshold boundary zero. They described the initial value  $x_0$  and drift  $\mu$  of the Wiener process as linear functions of covariates for consideration of the inherent dispersion of data. Instead of using calendar time they

used operational time scale since the rate of disease progression varied for different intervals of life experience.

Lee *et al.* (2009) extended the TR model for analyzing survival data with time-varying covariates using a Markov decomposition technique. They showed that the proposed Markov TR model is consistent with the Cox model with time-varying covariates. The connection between the Markov TR model and the concept of a collapsible survival model was also discussed.

Pennell *et al.* (2010) proposed a Bayesian approach for TR models. They assumed that an individual's health status can be modeled as a Wiener process with subject-specific initial state and drift. In addition to modeling initial process state and drift as linear functions of covariates, they assumed a prior distribution for regression parameters in the process initial state and drift. They argued that this Bayesian approach is able to address the issue of presence of unmeasured covariates. They applied the Bayesian method to a melanoma data set and compared their results with regular TR models.

Whitmore *et al.* (1998) proposed a bivariate Wiener process model in which one process models the marker and the other latent process determines the failure time. The failure occurs when the latent process hits a threshold boundary. They derived the joint probability distribution and density function for event time and marker level where the maximum likelihood method can be used for parameter es-

timination. Lee *et al.* (2000) extended this model by relating covariates to the model parameters through linear regression functions. They applied the model to an AIDS clinical trial data set in which the CD4 count was treated as the bio-marker. Tong *et al.* (2008) further extended this bivariate Wiener process model to analyze current status data (sometimes called case I interval censored data where there is only one observation time for each subject).

Horrocks and Thompson (2004) developed a model for length of stay in hospital (LOS) through a Wiener process reaching two absorbing thresholds, one for the healthy discharge and the other for the death in hospital, where the two thresholds are modeled as linear functions of covariates. They also showed that the density of the first hitting time and its derivatives are absolutely and uniformly convergent under certain regularity conditions.

Yu *et al.* (2009) extended TR models to accommodate nonlinear independent covariate effects. They proposed a spline approximation method for the estimation of unknown functional effect of covariate. Cross-validation method was also discussed for the selection of number of knots and determination of smoothing parameters. However, consistency of the resulting spline based estimator was not shown.

Many researchers have discussed various ways to link covariates' effects in TR models, but most of them only considered linear regression structures of covariates to explain the heterogeneity in the sample data. However, in real applications most

phenomena can not be fully explained by models with linear structures. Analysis results may have misleading conclusions if one forces such a linear relation in TR models. This is the main inspiration for the model development in Chapter 2 and model checking in Chapter 3 of this dissertation.

Recently, Li and Lee (2011) discussed TR models with varying coefficients for prognostic factors. In their article, they considered the case that regression coefficients vary with time dependent variables. However, our model proposed in Chapter 2 is different from their model in the following ways:

1. Li and Lee (2011) modeled drift parameter of Wiener process  $X(t)$  as a function of covariates and assumed each individual had the same initial value (i.e.,  $X(0) = 1$ ). In our model, the variability in different individuals when they first come into the study is also considered through modeling initial value  $X(0)$  as a regression function of covariates with both linear and nonlinear structures.
2. Li and Lee (2011) adopted a local polynomial method to approximate unknown functions; spline approximation technique is used to estimate smooth functions of covariates in our proposed method. Besides, the consistency and convergence rate of the estimator are discussed in a different manner.

### 1.3 TR Model Based on One Dimensional Wiener Process

Although TR model is applicable to many different areas, we focus on its applications in clinical and epidemiological studies. Fluctuations of an individual's

latent health status can be well described by a stochastic process and time to event can be interpreted as the first time this stochastic process hits a threshold boundary. As shown in Section 1.2, many authors have chosen a Wiener process as the parent stochastic process for TR models. We also model latent health status by a Wiener process, mainly because this kind of process has been often found to be a suitable model for many physical processes that exhibit random variation over time (Lee *et al.* 2004). Besides, Wiener process has flexible assumptions and its statistical properties have been well developed in the past several decades.

Specifically, let  $\{X(t)\}$  be a Wiener process that represents the health status of a subject at time  $t$ . At  $t = 0$ , we assume that the subject's initial health status is a positive value  $\delta = X(0)$ , which is an unknown parameter to be estimated. Let  $\mu$  and  $\sigma^2$  denote the drift and infinitesimal variance of  $\{X(t)\}$ , respectively. Then  $X(t)$  can be written as

$$X(t) = \delta + \mu t + \sigma W_t,$$

where  $W_t$  is a standard Brownian motion. The process  $X(t)$  has independent increments with

$$X(t) - X(s) \sim N(\mu(t - s), \sigma^2(t - s))$$

for  $0 \leq s < t$ .

An event time, such as death and disease progression can be interpreted as the first time  $X(t)$  reaches the threshold boundary  $\mathcal{B} = 0$ . Denote the first hitting



time by  $S$ , where

$$S = \inf\{t : X(t) = 0\}.$$

Let  $\Phi(\cdot)$  and  $\phi(\cdot)$  be the standard normal cumulative distribution function and standard normal density function, respectively. Following Lee *et al.* (2004), the first hitting time  $S$  for the Wiener process  $\{X(t)\}$  follows an inverse Gaussian distribution with the following density function :

$$f(s|\boldsymbol{\theta}) = \frac{\delta}{(2\pi\sigma^2s^2)^{1/2}} \exp\left[-\frac{(\delta + \mu s)^2}{2\sigma^2s}\right] \quad s > 0, \quad (1.3.1)$$

for  $-\infty < \mu < +\infty$ ,  $\sigma^2 > 0$  and  $\delta > 0$ ,

where  $\boldsymbol{\theta} = (\delta, \mu, \sigma)'$ . Then the cumulative distribution function of  $S$  is given by

$$F(s|\boldsymbol{\theta}) = 1 - \Phi\left[\frac{\mu s + \delta}{(\sigma^2s)^{1/2}}\right] + \exp\left(\frac{-2\delta\mu}{\sigma^2}\right) \Phi\left[\frac{\mu s - \delta}{(\sigma^2s)^{1/2}}\right]. \quad (1.3.2)$$

If  $\mu > 0$  there is a point mass at infinity:

$$\Pr(S = \infty) = 1 - \exp(-2\delta\mu/\sigma^2).$$

The underlying Wiener process  $X(t)$  may not reach the threshold boundary  $\mathcal{B} = 0$  eventually. This case can be interpreted as a certain proportion of individuals are “cured” or never experience the event of interest. For  $\mu \leq 0$ , the first-hitting time  $S < \infty$  almost surely (Chhikara and Folks, 1989). The expected survival time

(first-hitting-time) is

$$E(s|\boldsymbol{\theta}) = \frac{\delta}{|\mu|}$$

whenever  $\mu < 0$ .

Assume right censored data are available. Let  $Y = S \wedge C$  represent the event time subject to right censoring time  $C$ . Assume that  $S$  and  $C$  are independent. Let  $\Delta = I(S \leq C)$  be the event indicator ( $\Delta = 1$  if  $Y$  is an observed event time and 0 otherwise). Figure 1.3 illustrates two cases: path 1 shows an observed event before the end of a follow-up time of 600 days, while a censored event time is illustrated by path 2 when health process has not reached the threshold boundary  $\mathcal{B} = 0$  by the end of the follow-up time.

Suppose that  $(Y_i, \Delta_i), i = 1, \dots, n$ , are  $n$  i.i.d. copies of  $(Y, \Delta)$ . Then the log-likelihood function of the observed data has the form

$$\begin{aligned} l_n(\boldsymbol{\theta}) &= \sum_{i=1}^n l_i(\boldsymbol{\theta}; Y_i, \Delta_i) \\ &= \sum_{i=1}^n \left\{ \Delta_i \log [f(Y_i|\boldsymbol{\theta})] + (1 + \Delta_i) \log [1 - F(C_i|\boldsymbol{\theta})] \right\} \end{aligned} \quad (1.3.3)$$

It is clear that the proposed log-likelihood function (1.3.3) depends on the following two parameter quantities:  $\delta/\sigma$  and  $\mu/\sigma$ . The health status process is latent so it can be given an arbitrary measurement unit (Lee and Whitmore, 2006). Thus, we assume that  $\sigma = 1$ .

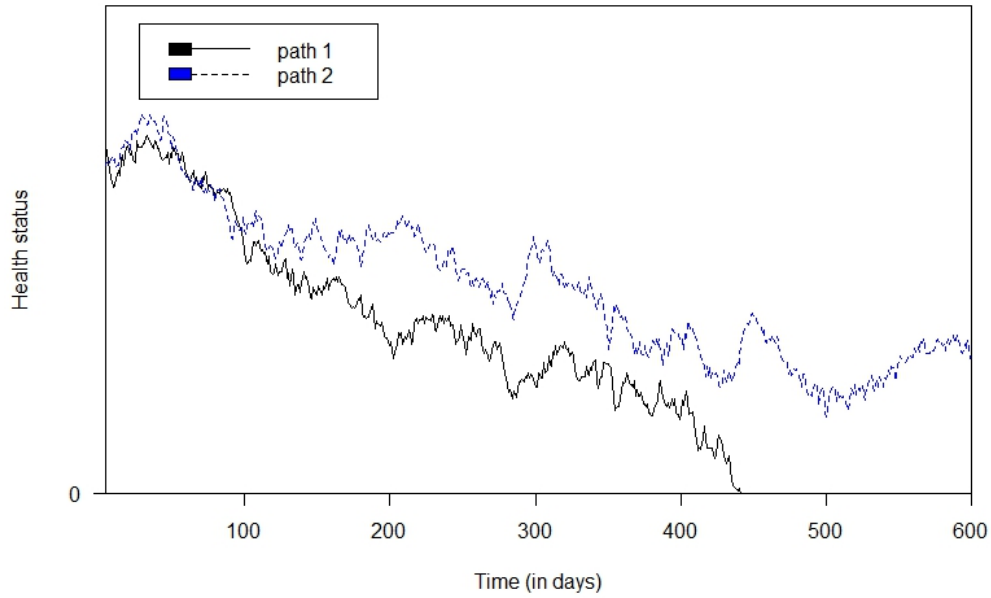


Figure 1.1: Two illustrative sample paths of health status: path 1 reaches the threshold boundary zero before the end of a follow-up time of 600 days and event time is observed; path 2 has not reached the threshold boundary by the end of the follow-up time and event time is censored.

To incorporate the covariate effects several authors have considered the link functions:

$$\ln(\delta) = \mathbf{Z}'_{\delta} \boldsymbol{\beta}_{\delta} ,$$

$$\mu = \mathbf{Z}'_{\mu} \boldsymbol{\beta}_{\mu} ,$$

where  $\mathbf{Z}_{\delta} = (Z_{\delta 1}, \dots, Z_{\delta d_1})'$  and  $\mathbf{Z}_{\mu} = (Z_{\mu 1}, \dots, Z_{\mu d_2})'$  are  $d_1 \times 1$  and  $d_2 \times 1$  vectors of covariates, and  $\boldsymbol{\beta}_{\delta} = (\beta_{\delta 1}, \dots, \beta_{\delta d_1})'$  and  $\boldsymbol{\beta}_{\mu} = (\beta_{\mu 1}, \dots, \beta_{\mu d_2})'$  are the corresponding vectors of unknown regression parameters. Modeling the logarithm of  $\delta$  as a function

of covariates ensures that  $\delta > 0$ . One advantage of TR models with the above link functions is their ability to distinguish two types of covariate effects: the covariate effects on how far the process has advanced prior to the study (i.e., the effects on  $\delta$ ) and the causal effects on the degradation (i.e., the effects on  $\mu$ ) (Pennell *et al.*, 2010). In contrast, the Cox model is only able to estimate covariate effects during the study and explain covariate effects in terms of hazard ratio. However, for example, a smaller hazard rate of treatment group compared to control group could be a result of effective treatment or better health condition at the beginning of the study. This question can not be answered using Cox model. We refer the readers to Section 6 of Lee and Whitmore (2010) for more details on this issue. We will further discuss covariate effects on latent process  $X(t)$  in Chapter 2. Another advantage of TR models is that they can estimate hazard ratios at different time points. This can be seen from the hazard function

$$\begin{aligned}\lambda(s|\boldsymbol{\theta}) &= \frac{f(s|\boldsymbol{\theta})}{1 - F(s|\boldsymbol{\theta})} \\ &= \frac{\delta(2\pi s^2)^{-1/2} \exp[-(\delta + \mu s)^2/2s]}{\Phi[(\mu s + \delta)/(s)^{1/2}] - \exp(-2\delta\mu)\Phi[(\mu s - \delta)/(s)^{1/2}]},\end{aligned}$$

while Cox model can only estimate a constant hazard ratio.

## 1.4 Overview of the Dissertation

This dissertation is organized as follows. In Chapter 2, we present the threshold regression model with flexible covariate effects in which covariates are modeled

as partially linear and partially nonlinear in the link functions. We adopt the B-spline approximation method to estimate nonlinear covariate effects. Regression parameters and coefficients of B-spline basis functions are estimated by the method of maximum likelihood. Asymptotic properties of the spline-based estimator are established through the empirical process. In particular, our proposed estimator attains the optimal convergence rate in nonparametric regression analysis when the number of spline basis functions are chosen properly. Simulation studies show that the proposed estimation procedure works well. The methodology is applied to the data from a study of osteoporotic fractures.

In Chapter 3, we discuss two model checking techniques for assessing the goodness-of-fit of TR models with parametric link functions. One is based on cumulative sums of martingale residuals, which was first proposed by Lin *et al.* (1993) under the Cox model. The other is based on cumulative sums of censoring consistent residuals, which was originally introduced by León and Tsai (2004) for testing the functional form specification of covariates under the Cox model. We show how to calculate  $p$ -values for both test processes through simulation. Large sample properties are established under fixed alternative hypotheses for the two methods. We apply both model checking approaches to a kidney dialysis data set.

In Chapter 4, we provide a summary of this dissertation and discuss several directions for future research.

## Chapter 2: Threshold Regression with Flexible Covariate Effects

### 2.1 Introduction

#### 2.1.1 Background

Linear models are one of the most commonly used statistical methods to explore effects of covariates on a particular outcome. The assumption that regression coefficients are constants is relatively strong and thus may not be satisfied in certain applications. One may get biased estimates and misleading results if this assumption is violated. Varying-coefficient models which allow regression coefficients to vary with certain covariates are a natural alternative to linear models and can improve the accuracy of parameter estimation. There are mainly two approaches to estimate functional coefficients. One is the local polynomial method and the other is the spline approximation technique. Hastie and Tibshirani (1993) discussed the regression and generalized regression models in which the coefficients are allowed to vary over other variables. They adopted a penalized spline method to estimate coefficient functions. Hoover *et al.* (1998) studied varying-coefficient model for longitudinal data, in which both local polynomial and smoothing spline methods were discussed. Chiang *et al.* (2001) considered the component-wise smoothing spline method for

the estimation of functional coefficients. Huang *et al.* (2002) studied the spline approximation for estimating parameters of a varying-coefficient model. Consistency and convergence rates of their resulting estimators were established. Huang *et al.* (2004) further explored the asymptotic normality of the spline approximation based estimator for functional coefficients. Huang and Shen (2004) discussed the spline approximation technique for estimating functional coefficient regression models for non-linear time series. For the estimation of functional coefficients through the local polynomial approach, one can refer to the review paper by Fan and Zhang (2008).

In survival analysis, the varying-coefficient modeling approach has received much attention recently. These works include Murphy and Sen (1991), Cai and Sun (2003), Tian *et al.* (2005), Nan *et al.* (2005), Cai *et al.* (2007), among others. In this chapter, we will discuss TR models with flexible covariate effects in which some of the covariates may have nonlinear coefficients and others retain the linear effects on both the initial health status  $\delta = X(0)$  and the drift parameter  $\mu$  of the latent process  $X(t)$ . We will adopt a spline approximation technique to estimate functional coefficients because it has been well developed and demonstrated to be a powerful tool in the estimation of semiparametric models.

### 2.1.2 Review of B-spline Approximation

In this section, we will briefly review basic concepts and properties of spline approximation. Interested readers can refer to the book by de Boor (2001) for more

detailed information on splines. Using the spline approximation technique to approximate unknown functions has a long history in statistical inference. One can find rich resources in this area from Sleeper and Harrington (1990), Shen and Wong (1994), Stone (1994), Huang (1996), and Lu *et al.* (2007), among others.

A polynomial function of order  $k$  has the functional form

$$f(x) = a_1 + a_2x + \dots + a_kx^{k-1} = \sum_{i=1}^k a_i x^{i-1} .$$

A spline function is constructed based on piecewise polynomial functions with same order  $k$ , where the degree of spline with order  $k$  is  $k - 1$ . The points at which the piecewise polynomial functions meet are called *knots*. In this section we focus on a class of splines which are named B-splines. The term ‘‘B-spline’’ is short for basis-spline. One important property of B-splines of order  $k$  is that its  $(k - 2)$ th derivative is continuous at the knots if the knots are distinct.

We will first review how B-spline basis functions are generated when their order and knot sequence of support are given. Let  $\mathbf{t}$  be the knot sequence for closed interval  $[a, b]$  where  $\mathbf{t} = (t_i)_1^{n+k}$  with

$$t_1 = \dots = t_k = a < t_{k+1} \leq \dots < b = t_{n+1} = \dots = t_{n+k} .$$

Let  $B_{j,k,\mathbf{t}}$  denote the  $j$ th B-spline basis function of order  $k$  for the knot sequence



$\mathbf{t}$ . The number of B-spline basis functions is  $n$ , that is, the number of the knots  $n + k$  minus the order of B-spline basis  $k$ . The B-spline basis functions  $B_{j,k,\mathbf{t}}$  for  $j = 1, 2, \dots, n$  are defined through their order  $k$  and the number of knots  $n + k$  and can be recursively constructed through the following formulas (Chapter IX, De Boor, 2001):

$$B_{j,1,\mathbf{t}}(t) = \begin{cases} 1, & \text{if } t_j \leq t < t_{j+1}; \\ 0, & \text{otherwise} \end{cases} \quad (2.1.1)$$

$$B_{j,k,\mathbf{t}}(t) = w_{j,k}(t)B_{j,k-1,\mathbf{t}}(t) + [1 - w_{j+1,k}(t)]B_{j+1,k-1,\mathbf{t}}(t), \text{ for } k > 1, \quad (2.1.2)$$

where

$$w_{j,k}(t) = \frac{t - t_j}{t_{j+k-1} - t_j}.$$

In the above formulas,  $0/0$  is defined as 0.

For example, given a knot sequence  $\mathbf{t}$  where  $t_1 = t_2 = t_3 = 0, t_4 = 0.3, t_5 = 0.5, t_6 = 0.5, t_7 = 0.6$  and  $t_8 = t_9 = t_{10} = 1$ , we get

$$B_{1,3,\mathbf{t}}(t) = \left(1 - \frac{10}{3}t\right)^2, 0 \leq t < 0.3;$$

$$B_{7,3,\mathbf{t}}(t) = 2.5(5t - 3)^2, 0.6 \leq t < 1$$

based on the above recursive formula (2.1.2). Figures 2.1 and 2.2 show the steps to compute  $B_{1,3,\mathbf{t}}(t)$  and  $B_{7,3,\mathbf{t}}(t)$ , respectively.

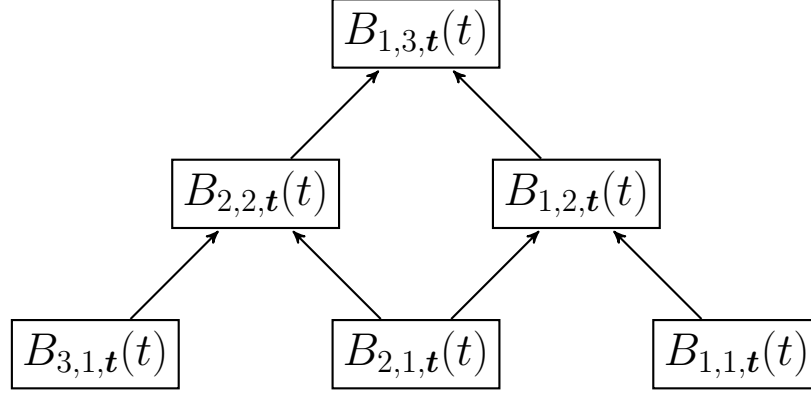


Figure 2.1: Calculation steps for  $B_{1,3,t}(t)$ . Here,  $B_{3,1,t}(t) = I[t_3, t_4)$ ,  $B_{2,1,t}(t) = I[t_2, t_3)$  and  $B_{1,1,t}(t) = I[t_1, t_2)$  based on formula (1.1.1).

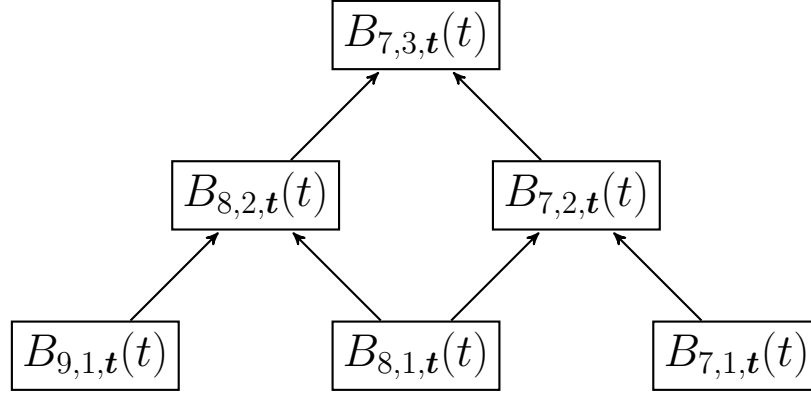


Figure 2.2: Calculation steps for  $B_{7,3,t}(t)$ . Here,  $B_{9,1,t}(t) = I[t_9, t_{10})$ ,  $B_{8,1,t}(t) = I[t_8, t_9)$  and  $B_{7,1,t}(t) = I[t_7, t_8)$  based on formula (1.1.1).

As pointed out by de boor (2001), any spline function  $S(x)$  of order  $k$  with knot sequence  $\mathbf{t}$  can be written as a linear combination of B-spline basis functions with the same order and knot sequence  $\mathbf{t}$ . Let  $\mathcal{S}_{k,\mathbf{t}}$  denote the set of all such functions. Then, we have

$$\mathcal{S}_{k,\mathbf{t}} = \left\{ \sum_i \alpha_i B_{i,k,\mathbf{t}} : \alpha_i \text{ real, all } i \right\}.$$

Because of their ability to approximate functions with unknown functional form, splines are widely used in semiparametric and nonparametric analysis. The function

is approximated by B-spline basis (polynomial function) within consecutive knots. The convergence of spline approximation for a smooth function  $g$  is guaranteed by Theorem (6) (p. 149) of de Boor (2001), as stated below.

**Lemma 2.1.1** (de Boor 2001). *For  $j = 0, \dots, k - 1$ , there exists  $const_{k,j}$  so that, for all  $\mathbf{t} = (t_i)_1^{n+k}$  with*

$$t_1 = \dots = t_k = a < t_{k+1} \leq \dots < b = t_{n+1} = \dots = t_{n+k}$$

and for all  $g \in C^{(j)}[a, b]$ ,

$$dist(g, \mathcal{S}_{k,\mathbf{t}}) \leq const_{k,j} |\mathbf{t}|^j \omega(D^j g; |\mathbf{t}|).$$

In particular, for  $j = k - 1$ , we get

$$dist(g, \mathcal{S}_{k,\mathbf{t}}) \leq const_k |\mathbf{t}|^k \|D^k g\|$$

in case  $g$  has  $k$  continuous derivatives (since then  $\omega(D^{k-1}g; h) \leq h \|D^k g\|$ ).

Notation in the above lemma is specified below:

1.  $D^j f := j$ th derivative of  $f$ .
2.  $dist(g, S) := \inf\{\|g - f\| : f \in S\}$ , the distance of  $g \in C[a, b]$  from the subset  $S$  of  $C[a, b]$ .
3.  $C^{(n)}[a, b] := \{f : [a, b] \rightarrow \mathbb{R} : f \text{ is } n \text{ times continuously differentiable}\}$ .

4.  $\|f\| := \max\{|f(x)| : a \leq x \leq b\}$ , the uniform norm of  $f \in C[a, b]$ .
5.  $\Delta\tau_i := \tau_{i+1} - \tau_i$ , the forward difference.
6.  $|\boldsymbol{\tau}| := \max_i \Delta\tau_i$ , the mesh size.
7.  $\omega(f; h) := \max\{|f(x) - f(y)| : x, y \in [a, b], |x - y| \leq h\}$ , the modulus of continuity for  $f \in C[a, b]$ .

The above lemma will be used in Section 2.5 for proving the consistency and convergence rate of the resulting estimator.

## 2.2 Model and Likelihood

In Section 1.3 we briefly introduced TR models with linear covariate effects. However, this assumption is not always guaranteed, which can lead to misleading conclusions. To relax this assumption, Yu *et al.* (2009) incorporated penalized and regression splines into TR models to accommodate nonlinear covariates' effects. Li and Lee (2011) generalized their models and discussed TR models with varying coefficients. They adopted a local polynomial technique to approximate the unknown functional form of coefficients. Both articles assume that the initial health status  $\delta = X(0)$  to be constant (one unit away from threshold boundary  $\mathcal{B} = 0$ ) and describe the drift parameter  $\mu$  of latent health process  $X(t)$  as a function of covariates. However, artificially forcing the initial health for all subjects to be the same is not appropriate since different individuals with different values of covariates may not share the same initial health status. To better understand covariate effects on  $X(t)$ ,

we propose a more general approach in which both  $\delta$  and  $\mu$  are considered as functions of covariates with possibly varying coefficients to accommodate the potential nonlinear interactions between covariates. We also assume  $\sigma^2$ , the infinitesimal variance of  $X(t)$ , to be 1 for which the reason is explained in Section 1.3, and rewrite the parameters as  $\boldsymbol{\theta} = (\delta, \mu)'$ . Then from (1.3.1), the density function for the first hitting time  $S$  of the latent Wiener process  $\{X(t)\}$  is given by

$$f(s|\boldsymbol{\theta}) = \frac{\delta}{(2\pi s^2)^{1/2}} \exp\left[-\frac{(\delta + \mu s)^2}{2s}\right] \quad s > 0, \quad (2.2.1)$$

for  $-\infty < \mu < +\infty$  and  $\delta > 0$ .

The corresponding cumulative distribution function is

$$F(s|\boldsymbol{\theta}) = 1 - \Phi\left[\frac{\mu s + \delta}{(s)^{1/2}}\right] + \exp(-2\delta\mu)\Phi\left[\frac{\mu s - \delta}{(s)^{1/2}}\right]. \quad (2.2.2)$$

For each subject, the observed data are denoted by  $(Y, \Delta, \mathbf{Z}'_{\delta}, \mathbf{Z}'_{\mu}, \mathbf{V}'_{\delta}, \mathbf{W}'_{\delta}, \mathbf{V}'_{\mu}, \mathbf{W}'_{\mu})$ ; where  $Y$  and  $\Delta$  follow the same definition as in Section 1.3;  $\mathbf{Z}_{\delta} = (Z_{\delta 1}, \dots, Z_{\delta d_1})'$  and  $\mathbf{Z}_{\mu} = (Z_{\mu 1}, \dots, Z_{\mu d_2})'$  are  $d_1$ -dimensional and  $d_2$ -dimensional vectors of covariates, respectively;  $\mathbf{V}_{\delta} = (V_{\delta 1}, \dots, V_{\delta p_1})'$  and  $\mathbf{W}_{\delta} = (W_{\delta 1}, \dots, W_{\delta p_1})'$  are  $p_1$ -dimensional vectors of covariates;  $\mathbf{V}_{\mu} = (V_{\mu 1}, \dots, V_{\mu p_2})'$  and  $\mathbf{W}_{\mu} = (W_{\mu 1}, \dots, W_{\mu p_2})'$  are  $p_2$ -dimensional vectors of covariates. We assume that the censoring time  $C$  is independent with the stochastic process  $\{X(t)\}$  given the covariates  $\mathbf{Z}_{\delta}, \mathbf{Z}_{\mu}, \mathbf{V}_{\delta}, \mathbf{W}_{\delta},$

$\mathbf{V}_\mu$  and  $\mathbf{W}_\mu$ .

To incorporate the covariate effects into the stochastic process  $X(t)$ , we propose a class of semiparametric partially linear varying-coefficient TR models. Suppose that the covariate effects can be described by

$$\begin{cases} \ln(\delta) = \mathbf{Z}_\delta^T \boldsymbol{\beta}_\delta + \sum_{m=1}^{p_1} V_{\delta m} \eta_m(W_{\delta m}) \\ \mu = \mathbf{Z}_\mu^T \boldsymbol{\beta}_\mu + \sum_{n=1}^{p_2} V_{\mu n} \varphi_n(W_{\mu n}) \end{cases}$$

where  $\boldsymbol{\beta}_\delta$  and  $\boldsymbol{\beta}_\mu$  are  $d_1$ -dimensional and  $d_2$ -dimensional vectors of unknown regression parameters and  $\eta_m(\cdot)$  ( $m = 1, \dots, p_1$ ) and  $\varphi_n(\cdot)$  ( $n = 1, \dots, p_2$ ) are completely unspecified smooth functions. When  $V_{\delta m} = 0$  ( $m = 1, \dots, p_1$ ) and  $V_{\mu n} = 0$  ( $n = 1, \dots, p_2$ ), the model reduces to the regular TR model, which has been well studied by Lee *et al.* (2000, 2004), among others. When  $V_{\delta m} = 1$  ( $m = 1, \dots, p_1$ ) and  $V_{\mu n} = 1$  ( $n = 1, \dots, p_2$ ), it reduces to the additive semiparametric TR model which proposed by Yu *et al.* (2009). In this dissertation, for simplicity of notation, we only consider the case with  $p_1 = p_2 = 1$ , that is,

$$\begin{cases} \ln(\delta) = \mathbf{Z}_\delta^T \boldsymbol{\beta}_\delta + V_\delta \eta(W_\delta) \\ \mu = \mathbf{Z}_\mu^T \boldsymbol{\beta}_\mu + V_\mu \varphi(W_\mu). \end{cases}$$

The proposed models can be easily generalized to the cases with  $p_1 > 1$  and  $p_2 > 1$ .

Suppose the observed data for the  $i$ th subject is  $\mathbf{U}_i = (Y_i, \Delta_i, \mathbf{Z}'_{\delta i}, \mathbf{Z}'_{\mu i}, V'_{\delta i}, W'_{\delta i}, V'_{\mu i}, W'_{\mu i})$  for  $i = 1, \dots, n$ . Then the log-likelihood function has the form

$$\begin{aligned} l_n(\boldsymbol{\tau}^*) &= \sum_{i=1}^n l_i(\boldsymbol{\tau}^*; \mathbf{U}_i) \\ &= \sum_{i=1}^n \left\{ \Delta_i \log[f(Y_i | \mathbf{Z}_{\delta i}, \mathbf{Z}_{\mu i}, V_{\delta i}, W_{\delta i}, V_{\mu i}, W_{\mu i}, \boldsymbol{\tau}^*)] \right. \\ &\quad \left. + (1 - \Delta_i) \log[1 - F(C_i | \mathbf{Z}_{\delta i}, \mathbf{Z}_{\mu i}, V_{\delta i}, W_{\delta i}, V_{\mu i}, W_{\mu i}, \boldsymbol{\tau}^*)] \right\}, \quad (2.2.3) \end{aligned}$$

where  $\boldsymbol{\tau}^* = (\boldsymbol{\beta}'_{\delta}, \boldsymbol{\beta}'_{\mu}, \eta(\cdot), \varphi(\cdot))'$ .

### 2.3 Estimation Procedure

The functions  $\eta(\cdot)$  and  $\varphi(\cdot)$  can not be estimated directly from the log-likelihood function (2.2.3) since their functional forms are completely unspecified. We use the B-spline function approximation technique as briefly reviewed in Section 2.1 for the estimation of smooth functions  $\eta(\cdot)$  and  $\varphi(\cdot)$ .

Without loss of generality, assume that  $W_{\delta}$  has support on  $[a_{\eta}, b_{\eta}]$ . Let

$$\mathbf{t}_{\eta} = \{t_{\eta i}\}_1^{k_n+m},$$

with

$$t_{\eta 1} = \dots = t_{\eta m} = a_{\eta} < t_{\eta m+1} \leq \dots < b_{\eta} = t_{\eta k_n+1} = \dots = t_{\eta k_n+m},$$

be a sequence of knots that partitions  $[a_{\eta}, b_{\eta}]$ , the support of  $W_{\delta}$ , into  $k_n + 1 - m$

subintervals  $I_{\eta i} = [s_{\eta m+i}, s_{\eta m+i+1}]$ , for  $i = 0, 1, \dots, k_n - m$ . Let

$$\mathbf{f}_\eta(\cdot) = (f_{\eta 1}(\cdot), \dots, f_{\eta k_n}(\cdot))'$$

the B-spline basis function for the space of  $m$ -th order polynomial splines with knots  $\mathbf{t}_\eta$ .  $f_{\eta i}$ ,  $i = 1, \dots, k_n$ , can be constructed recursively by (2.1.1) and (2.1.2). Then we can approximate  $\eta(W_\delta)$  by

$$\eta_m(W_\delta) = \boldsymbol{\alpha}'_\eta \mathbf{f}_\eta(W_\delta) ,$$

where  $\boldsymbol{\alpha}_\eta = (\alpha_{\eta 1}, \dots, \alpha_{\eta k_n})'$  is a  $k_n$ -dimensional vector of unknown coefficients.

Similarly, we can approximate  $\varphi(\cdot)$  by a spline function. Assume that  $W_\mu$  has support on  $[a_\varphi, b_\varphi]$ . Let

$$\mathbf{t}_\varphi = \{t_{\varphi i}\}_1^{k_n+m} ,$$

with

$$t_{\varphi 1} = \dots = t_{\varphi m} = a_\varphi < t_{\varphi m+1} \leq \dots < b_\varphi = t_{\varphi k_n+1} = \dots = t_{\varphi k_n+m}$$

be a sequence of knots that partitions  $[a_\varphi, b_\varphi]$ , the support of  $W_\varphi$ , into  $k_n + 1 - m$  subintervals  $I_{\mu i} = [s_{\mu m+i}, s_{\mu m+i+1}]$ , for  $i = 0, 1, \dots, k_n - m$ . Let

$$\mathbf{f}_\varphi(\cdot) = (f_{\varphi 1}(\cdot), \dots, f_{\varphi k_n}(\cdot))'$$

the B-spline basis function for the space of  $m$ -th order polynomial splines with knots



$\mathbf{t}_\varphi$ . Then  $f_{\varphi i}$  for  $i = 1, \dots, k_n$  can be constructed recursively by (2.1.1) and (2.1.2).

Thus  $\varphi(W_\mu)$  can be approximated by

$$\varphi_n(W_\mu) = \boldsymbol{\alpha}'_\varphi \mathbf{f}_\varphi(W_\mu),$$

where  $\boldsymbol{\alpha}_\varphi = (\alpha_{\varphi 1}, \dots, \alpha_{\varphi k_n})'$  is a  $k_n$ -dimensional vector of unknown coefficients.

A spline function of order  $m$  with knot sequence  $\mathbf{t}$  is any linear combination of B-splines of order  $m$  for the knot sequence  $\mathbf{t}$  (de Boor 2001, p. 93). Let

$$\mathcal{S}_{m, \mathbf{t}_\eta} = \left\{ \eta_n(W_\delta) : \eta_n(W_\delta) = \sum_{i=1}^{k_n} \alpha_{\eta i} f_{\eta i}(W_\delta), \alpha_{\eta i} \text{ real, all } i, W_\delta \in [a_\eta, b_\eta] \right\}$$

be the collection of spline functions of order  $m$  with knot sequence  $\mathbf{t}_\eta$  and

$$\mathcal{S}_{m, \mathbf{t}_\varphi} = \left\{ \varphi_n(W_\mu) : \varphi_n(W_\mu) = \sum_{i=1}^{k_n} \alpha_{\varphi i} f_{\varphi i}(W_\mu), \alpha_{\varphi i} \text{ real, all } i, W_\mu \in [a_\varphi, b_\varphi] \right\}$$

denote the collection of spline functions of order  $m$  with knot sequence  $\mathbf{t}_\varphi$ . Let  $\mathcal{A}_\eta$  be the support of  $\boldsymbol{\alpha}_\eta$  and  $\mathcal{A}_\varphi$  be the support of  $\boldsymbol{\alpha}_\varphi$ . After replacing  $\eta(W_\delta)$  and  $\varphi(W_\mu)$  in the log-likelihood function (2.2.3) by their spline approximation functions

$\eta_n(W_\delta)$  and  $\varphi_n(W_\mu)$ , respectively, the log-likelihood function has the following form

$$\begin{aligned}
l_n(\boldsymbol{\tau}) &= \sum_{i=1}^n l_i(\boldsymbol{\tau}; \mathbf{U}_i) \\
&= \sum_{i=1}^n \left\{ \Delta_i \log[f(Y_i | \mathbf{Z}_{\delta i}, \mathbf{Z}_{\mu i}, V_{\delta i}, W_{\delta i}, V_{\mu i}, W_{\mu i}, \boldsymbol{\tau})] \right. \\
&\quad \left. + (1 - \Delta_i) \log[1 - F(C_i | \mathbf{Z}_{\delta i}, \mathbf{Z}_{\mu i}, V_{\delta i}, W_{\delta i}, V_{\mu i}, W_{\mu i}, \boldsymbol{\tau})] \right\}, \quad (2.3.1)
\end{aligned}$$

where  $\boldsymbol{\tau} = (\boldsymbol{\beta}'_\delta, \boldsymbol{\beta}'_\mu, \boldsymbol{\alpha}'_\eta, \boldsymbol{\alpha}'_\varphi)'$ . Let  $\Theta \subset \mathbb{R}^{d_1+d_2}$  denote the domain of the regression parameters  $(\boldsymbol{\beta}'_\delta, \boldsymbol{\beta}'_\mu)$ . To estimate the unknown parameters  $\boldsymbol{\tau}$  one can maximize the log-likelihood function  $l_n(\boldsymbol{\tau})$  over its parameter space  $\Theta \times \mathcal{A}_\eta \times \mathcal{A}_\varphi$  which is equivalent to maximizing  $l_n(\boldsymbol{\tau}^*)$  over  $\Theta \times \mathcal{S}_{m, t_\eta} \times \mathcal{S}_{m, t_\varphi}$ .

The parameter estimator which maximizes  $l_n(\boldsymbol{\tau})$  over its parameter space  $\Theta \times \mathcal{A}_\eta \times \mathcal{A}_\varphi$  is denoted by  $(\hat{\boldsymbol{\beta}}'_\delta, \hat{\boldsymbol{\beta}}'_\mu, \hat{\boldsymbol{\alpha}}'_\eta, \hat{\boldsymbol{\alpha}}'_\varphi)'$ . The regression spline estimators can be obtained by

$$\hat{\eta}_n(W_\delta) = \mathbf{f}_\eta(W_\delta)' \hat{\boldsymbol{\alpha}}_\eta$$

and

$$\hat{\varphi}_n(W_\mu) = \mathbf{f}_\varphi(W_\mu)' \hat{\boldsymbol{\alpha}}_\varphi,$$

respectively. Let

$$H(\hat{\boldsymbol{\beta}}'_\delta, \hat{\boldsymbol{\beta}}'_\mu, \hat{\boldsymbol{\alpha}}'_\eta, \hat{\boldsymbol{\alpha}}'_\varphi) = \left[ \frac{\partial^2 l(\boldsymbol{\tau})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'} \right]_{(d_1+d_2+2k_n) \times (d_1+d_2+2k_n); \boldsymbol{\tau}=(\hat{\boldsymbol{\beta}}'_\delta, \hat{\boldsymbol{\beta}}'_\mu, \hat{\boldsymbol{\alpha}}'_\eta, \hat{\boldsymbol{\alpha}}'_\varphi)'}$$

be the Hessian matrix of the log-likelihood function  $l_n(\boldsymbol{\tau})|_{\boldsymbol{\tau}=(\hat{\boldsymbol{\beta}}'_\delta, \hat{\boldsymbol{\beta}}'_\mu, \hat{\boldsymbol{\alpha}}'_\eta, \hat{\boldsymbol{\alpha}}'_\varphi)'}$ . Then the

variance estimators for  $\hat{\eta}_n$  and  $\hat{\varphi}_n$  are given by

$$\mathbf{f}_\eta(W_\delta)' \left[ -H(\hat{\beta}'_\delta, \hat{\beta}'_\mu, \hat{\alpha}'_\eta, \hat{\alpha}'_\varphi) \right]_{\alpha_\eta, \alpha_\eta}^{-1} \mathbf{f}_\eta(W_\delta)$$

and

$$\mathbf{f}_\varphi(W_\mu)' \left[ -H(\hat{\beta}'_\delta, \hat{\beta}'_\mu, \hat{\alpha}'_\eta, \hat{\alpha}'_\varphi) \right]_{\alpha_\varphi, \alpha_\varphi}^{-1} \mathbf{f}_\varphi(W_\mu),$$

respectively, where  $[-H(\hat{\beta}'_\delta, \hat{\beta}'_\mu, \hat{\alpha}'_\eta, \hat{\alpha}'_\varphi)]_{\alpha_\eta, \alpha_\eta}^{-1}$  and  $[-H(\hat{\beta}'_\delta, \hat{\beta}'_\mu, \hat{\alpha}'_\eta, \hat{\alpha}'_\varphi)]_{\alpha_\varphi, \alpha_\varphi}^{-1}$  are the  $k_n \times k_n$  submatrices of  $[-H(\hat{\beta}'_\delta, \hat{\beta}'_\mu, \hat{\alpha}'_\eta, \hat{\alpha}'_\varphi)]^{-1}$  corresponding to  $\hat{\alpha}_\eta$  and  $\hat{\alpha}_\varphi$ , respectively.

To check whether the functional coefficient is a parametric function, one can adopt the likelihood ratio test. For example, to examine if  $\eta(W_\delta)$  in  $\ln(\delta)$  is linear over  $W_\delta$ , one can fit the nested TR model under the null hypothesis  $H_0 : \eta(W_\delta) = \alpha_0 + \alpha_1 W_\delta$  and then conduct the likelihood ratio test against the full model.

We will discuss the asymptotic behavior of spline-based maximum likelihood estimator in the next section.

## 2.4 Asymptotic Property of Estimator

### 2.4.1 Introduction

Empirical process theory is a powerful tool for studying asymptotic properties of estimators from complex statistical models. In this section we will briefly review some basic concepts and important theorems of empirical processes. These theorems will be used to prove the main results in this section. For a comprehensive treatment and for references to the extensive literature on the subject one may refer to the books by van der Vaart and Wellner (1996) and van der Vaart (1998).

We mainly follow the notation in van der Vaart (1998). For a probability measure  $P$  on measurable space  $(\mathcal{X}, \mathcal{B})$  and a measurable function  $f : \mathcal{X} \mapsto \mathbb{R}$ , let

$$Pf = \int f dP .$$

That is,  $Pf = E_P(f)$ . Let  $X_1, \dots, X_n$  be a random sample from  $P$ . We denote

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(X_i) .$$

The empirical process  $\mathbb{G}_n f$  is defined as

$$\mathbb{G}_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{f(X_i) - Pf\} .$$

Research on empirical processes focuses on studying empirical processes as random functions over the associated index set. Let  $\mathcal{F}$  denote a class of measurable functions  $f : \mathcal{X} \mapsto \mathbb{R}$ . If

$$\|\mathbb{P}_n f - P f\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - P f| \xrightarrow{a.s.} 0 ,$$

we say that  $\mathcal{F}$  is *P-Glivenko-Cantelli*. Let  $\mathbb{G}$  be a mean zero Gaussian process which is indexed by  $\mathcal{F}$ . Assume that the covariance of  $\mathbb{G}$  is  $P(fg) - P f P g$  for all  $f, g \in \mathcal{F}$ .

Then  $\mathcal{F}$  is called *P-Donsker* if

$$\mathbb{G}_n \xrightarrow{D} \mathbb{G}$$

in  $l^\infty(\mathcal{F})$ , where  $l^\infty(\mathcal{F})$  is the collection of all bounded functions in  $\mathcal{F}$ . One can interpret P-Glivenko-Cantelli and P-Donsker as stronger versions of the law of large numbers and central limit theorem, respectively.

The *Bracketing number* which is a way of measuring the size of a function class  $\mathcal{F}$ . It can be used to determine if  $\mathcal{F}$  is a Glivenko-Cantelli or Donsker class. The *bracket*  $[l, u]$  is the set of all functions  $f$  with  $l \leq f \leq u$  where  $l$  and  $u$  are two given functions. The  $L_r(P)$ -norm is defined as

$$\begin{aligned} \|f\|_{P,r} &= \left( \int_{\mathcal{X}} |f(x)|^r dP(x) \right)^{1/r} \\ &= (P|f|^r)^{1/r} . \end{aligned}$$

An  $\varepsilon$ -*bracket* in  $L_r(P)$  is a bracket  $[l, u]$  with  $P(u - l)^r < \varepsilon^r$ . The bracketing number  $N_{[\cdot]}(\varepsilon, \mathcal{F}, L_r(P))$  is the minimum number of  $\varepsilon$ -brackets needed to cover  $\mathcal{F}$ . (The

bracketing functions  $l$  and  $u$  must have finite  $L_r(P)$ -norms but need not belong to  $\mathcal{F}$ .) The *entropy with bracketing* is the logarithm of the bracketing number. The *covering number*, denoted by  $N(\varepsilon, \mathcal{F}, L_r(Q))$ , is another way of measuring the size of a function class  $\mathcal{F}$ . The covering number  $N(\varepsilon, \mathcal{F}, L_r(Q))$  is the minimum number of  $\varepsilon$ -balls need to cover  $\mathcal{F}$ , where an  $\varepsilon$ -ball around a function  $f$  is the set  $\{h : \|h - f\|_{Q,r} < \varepsilon\}$ . The *entropy* is the logarithm of the covering number. Bracketing number and covering number are closely related. They have the following relations (Kosorok 2008, p. 160-161):

1. Bracketing numbers are in general larger than covering numbers. For any norm  $\|\cdot\|$  on  $\mathcal{F}$  one has

$$N(\varepsilon, \mathcal{F}, \|\cdot\|) \leq N_{[\cdot]}(\varepsilon, \mathcal{F}, \|\cdot\|) .$$

2. For any norm  $\|\cdot\|$  dominated by  $\|\cdot\|_\infty$ ,

$$\log N_{[\cdot]}(2\varepsilon, \mathcal{F}, \|\cdot\|) \leq \log N(\varepsilon, \mathcal{F}, \|\cdot\|_\infty) .$$

The following theorem answers what kind of function class is P-Glivenko-Cantelli.

**Lemma 2.4.1** (van der Vaart 2002, Theorem 19.4). *Every class  $\mathcal{F}$  of measurable functions such that  $N_{[\cdot]}(\varepsilon, \mathcal{F}, L_1(P)) < \infty$  for every  $\varepsilon$  is P-Glivenko-Cantelli.*

In most cases, the bracketing numbers  $N_{[\cdot]}(\varepsilon, \mathcal{F}, L_r(P))$  are decreasing functions of  $\varepsilon$  and grow to infinity when  $\varepsilon \downarrow 0$ . One way to tell whether a function class

$\mathcal{F}$  is P-Donsker is to check how fast  $N_{[\cdot]}(\varepsilon, \mathcal{F}, L_r(P))$  goes to infinity as  $\varepsilon \downarrow 0$ . The bracketing integral

$$J_{[\cdot]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{F}, L_2(P))} d\varepsilon$$

can be used to measure the speed. Then we have the following theorem.

**Lemma 2.4.2** (van der Vaart 2002, Theorem 19.5). *Every class  $\mathcal{F}$  of measurable functions with  $J_{[\cdot]}(\delta, \mathcal{F}, L_2(P)) < \infty$  is P-Donsker.*

The above lemma requires more stringent condition on the number of brackets needed to cover  $\mathcal{F}$ . The value of  $J_{[\cdot]}(\delta, \mathcal{F}, L_2(P))$  depends only on the size of the bracketing numbers for  $\varepsilon \downarrow 0$ . Since we know that  $\int_0^1 (\frac{1}{\varepsilon})^r d\varepsilon$  converges given  $r < 1$ ,  $J_{[\cdot]}(\delta, \mathcal{F}, L_2(P))$  is bounded if the entropy with bracketing grows with an order slower than  $(1/\varepsilon)^2$ .

The following result provides other equivalent descriptions of a P-Donsker class  $\mathcal{F}$  and they can be used as properties of  $\mathcal{F}$ , as stated below.

**Lemma 2.4.3** (van der Vaart and Wellner 1996, Corollary 2.3.12). *Let  $\mathcal{F}$  be a class of measurable functions. Then the following are equivalent:*

- (1)  $\mathcal{F}$  is P-Donsker;
- (2)  $(\mathcal{F}, \rho_P)$  is totally bounded and

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} P \left( \sup_{\rho_P(f-g) < \delta} |\mathbb{G}_n(f-g)| > \varepsilon \right) = 0$$

where  $\rho_P(f) = (P(f - Pf)^2)^{1/2}$  for  $f \in \mathcal{F}$ .

(3)  $(\mathcal{F}, \rho_P)$  is totally bounded and

$$E\sqrt{n}\|\mathbb{P}_n - P\|_{\mathcal{F}_{\delta_n}} \rightarrow 0, \text{ for every } \delta_n \rightarrow 0$$

where  $\|\mathbb{P}_n - P\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf|$  and  $\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, \rho_P(f - g) < \delta\}$ .

Lemma 2.4.3 will be used to show consistency of our proposed spline based estimator. To get exact convergence rate of our estimator, we will mainly check conditions in the next lemma.

**Lemma 2.4.4** (van der Vaart and Wellner 1996, Theorem 3.4.1). *Let  $\mathbb{M}_n$  be stochastic processes indexed by a semimetric space  $\Theta$  and  $\mathbb{M} : \Theta \mapsto \mathbb{R}$  a deterministic function, such that for every  $\theta$  in a neighborhood of  $\theta_0$ ,*

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta, \theta_0).$$

*Suppose that, for every  $n$  and sufficiently small  $\delta$ , the centered process  $\mathbb{M}_n - \mathbb{M}$  satisfies*

$$E \sup_{d(\theta, \theta_0) < \delta} |(\mathbb{M}_n - \mathbb{M})(\theta) - (\mathbb{M}_n - \mathbb{M})(\theta_0)| \lesssim \frac{\phi_n(\delta)}{\sqrt{n}},$$

*for functions  $\phi_n$  such that  $\delta \mapsto \phi_n(\delta)/\delta^\alpha$  is decreasing for some  $\alpha < 2$  (not depending on  $n$ ). Let*

$$r_n^2 \phi_n\left(\frac{1}{r_n}\right) \leq \sqrt{n}, \quad \text{for every } n.$$



If the sequence  $\hat{\theta}_n$  satisfies  $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_0) - O_P(r_n^{-2})$  and converges in probability to  $\theta_0$ , then  $r_n d(\hat{\theta}_n, \theta_0) = O_P(1)$ . If the displayed conditions are valid for every  $\theta$  and  $\delta$ , then the condition that  $\hat{\theta}_n$  is consistent is unnecessary.

The condition on the centered process  $\mathbb{M}_n - \mathbb{M}$  in Lemma 2.4.4 is hard to check in most situations. One alternative is to use Lemma 3.4.2 in van der Vaart and Wellner (1996). We define the bracketing integral of a function class  $\mathcal{F}$  by

$$\tilde{J}_{[\cdot]}(\delta, \mathcal{F}, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[\cdot]}(\varepsilon, \mathcal{F}, \|\cdot\|)} d\varepsilon,$$

where  $\|\cdot\|$  is a given norm.

**Lemma 2.4.5** (van der Vaart and Wellner 1996, Lemma 3.4.2). *Let  $\mathcal{F}$  be class a of measurable functions such that  $Pf^2 < \delta^2$  and  $\|f\|_\infty \leq M$  for every  $f$  in  $\mathcal{F}$ . Then*

$$E_P \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \tilde{J}_{[\cdot]}(\delta, \mathcal{F}, L_2(P)) \left( 1 + \frac{\tilde{J}_{[\cdot]}(\delta, \mathcal{F}, L_2(P))}{\delta^2 \sqrt{n}} M \right).$$

## 2.4.2 Consistency

We will show the consistency of the spline based estimator for TR models with flexible covariate effects which was discussed in Section 2.3. We will mainly use the empirical process techniques that reviewed in Section 2.4.1 to prove the main result in this section.

Analogous to Wellner and Zhang (2007), the following regularity conditions

are needed to derive consistency of spline based maximum likelihood estimator

$$\hat{\boldsymbol{\tau}}_n = (\hat{\boldsymbol{\beta}}_\delta, \hat{\boldsymbol{\beta}}_\mu, \hat{\eta}_n(W_\delta), \hat{\varphi}_n(W_\mu))'.$$

**Assumption 2.4.1.**

(1)  $\mathbf{Z}_\delta$  and  $\mathbf{Z}_\mu$  are bounded, i.e., there exists  $Z_0 > 0$  such that  $P(\|\mathbf{Z}_\delta\| \leq Z_0) = 1$  and  $P(\|\mathbf{Z}_\mu\| \leq Z_0) = 1$ .

(2)  $\Theta$  is a compact subset of  $\mathbb{R}^{d_1+d_2}$ .

(3)  $\eta(W_\delta) \in C^1[a_\eta, b_\eta]$  and  $\varphi(W_\mu) \in C^1[a_\varphi, b_\varphi]$ , where  $C^i[a, b]$  denotes a class of functions with continuous  $i$ -th derivatives in  $[a, b]$ .

(4) Censoring time  $C$  is bounded,  $C < T < \infty$ .

**Lemma 2.4.6.** For  $f \in C^j[a, b]$ , there exists a  $f_n \in \mathcal{S}_{m, \mathbf{t}}$  with order  $m \geq j+1$ , where  $\mathbf{t} = (t_i)_{i=1}^{k_n+m}$  is the knot sequence for  $[a, b]$  and  $\max_{i \in \{1, 2, \dots, k_n+m-1\}} (t_{i+1} - t_i) = O(k_n^{-1})$  such that

$$\|f - f_n\|_\infty = O(k_n^{-j}) .$$

*Proof.* This is a direct conclusion of Lemma 2.1.1. Since  $f^{(j)} \in C^{(j)}[a, b] \subset C^{(j-1)}[a, b]$ , we have

$$\text{dist}(f, \mathcal{S}_{m, \mathbf{t}}) \leq \text{const}_{k, j-1} |\mathbf{t}|^{j-1} \omega(D^{j-1} f; |\mathbf{t}|) .$$

By the mean value theorem,  $\omega(D^{j-1} f; |\mathbf{t}|) \leq |\mathbf{t}| \|D^j f\|_\infty$ . This implies that

$$\text{dist}(f, \mathcal{S}_{m, \mathbf{t}}) \leq \text{const}_{k, j-1} |\mathbf{t}|^j \|D^j f\|_\infty .$$

The quantity  $D^j f$  is bounded by some constant  $C$  since  $D^j f$  is continuous on  $[a, b]$  ( $f \in C^j[a, b]$ ). This leads to

$$\text{dist}(f, \mathbb{S}_{m,t}) \leq \text{const}_{k,j-1} |\mathbf{t}|^j C ;$$

that is,  $\|f - f_n\|_\infty = O(k_n^{-j})$  for some  $f_n \in \mathbb{S}_{m,t}$ . □

Let  $\mathbf{t}_\eta$  and  $\mathbf{t}_\varphi$  be partitions of  $[a_\eta, b_\eta]$  and  $[a_\varphi, b_\varphi]$  as defined in Section 2.2. For each  $n$ , we further assume that  $\mathbf{t}_\eta$  and  $\mathbf{t}_\varphi$  have meshes which satisfy

$$\begin{cases} \bar{\Delta}_{\eta n} = \max_i \{t_{\eta i} - t_{\eta i-1}\} = O(k_n^{-1}) \\ \bar{\Delta}_{\varphi n} = \max_i \{t_{\varphi i} - t_{\varphi i-1}\} = O(k_n^{-1}) \end{cases} \quad (2.4.1)$$

where  $k_n = O(n^v)$  for  $v \in (0, 1)$ . Based on Lemma 2.4.6 and Assumption 2.4.1 (3), we have the following facts:

1.  $\exists \boldsymbol{\alpha}_\eta^0 = (\alpha_{\eta 1}^0, \dots, \alpha_{\eta k_n}^0)' \in \mathcal{A}_\eta$  such that  $\|\eta(W_\delta) - \eta_n^0(W_\delta)\|_\infty \leq O(k_n^{-1})$  where  $\eta_n^0(W_\delta) = \boldsymbol{\alpha}_\eta^{0'} \mathbf{f}_\eta(W_\delta) \in \mathbb{S}_{m,t_\eta}$ .
2.  $\exists \boldsymbol{\alpha}_\varphi^0 = (\alpha_{\varphi 1}^0, \dots, \alpha_{\varphi k_n}^0)' \in \mathcal{A}_\varphi$  such that  $\|\varphi(W_\mu) - \varphi_n^0(W_\mu)\|_\infty \leq O(k_n^{-1})$  where  $\varphi_n^0(W_\mu) = \boldsymbol{\alpha}_\varphi^{0'} \mathbf{f}_\varphi(W_\mu) \in \mathbb{S}_{m,t_\varphi}$ .

The functions  $\eta(W_\delta)$  and  $\varphi(W_\mu)$  are bounded under Assumption 2.4.1 (3). We

further assume that

$$\mathcal{A}_\eta = \{\boldsymbol{\alpha}_\eta = (\alpha_{\eta 1}, \dots, \alpha_{\eta k_n})' : \max_{i=1, \dots, k_n} |\alpha_{\eta i}| \leq M\}, \quad (2.4.2)$$

$$\mathcal{A}_\varphi = \{\boldsymbol{\alpha}_\varphi = (\alpha_{\varphi 1}, \dots, \alpha_{\varphi k_n})' : \max_{i=1, \dots, k_n} |\alpha_{\varphi i}| \leq M\} \quad (2.4.3)$$

for some constant  $M$  such that  $\boldsymbol{\alpha}_\eta^0 \in \mathcal{A}_\eta$  and  $\boldsymbol{\alpha}_\varphi^0 \in \mathcal{A}_\varphi$ .

Let  $\|\cdot\|$  denote the Euclidean distance; that is,  $\|\mathbf{p}\| = \sqrt{p_1^2 + \dots + p_n^2}$  where  $\mathbf{p}$  is a vector in  $\mathbb{R}^n$ . For any probability measure  $P$ , the  $L^2$ -norm is defined as

$$\|f\|_2 = \left( \int f^2 dP \right)^{1/2}.$$

For  $\boldsymbol{\tau}_1 = (\boldsymbol{\beta}_{\delta 1}, \boldsymbol{\beta}_{\mu 1}, \eta_1(W_\delta), \varphi_1(W_\mu))$  and  $\boldsymbol{\tau}_2 = (\boldsymbol{\beta}_{\delta 2}, \boldsymbol{\beta}_{\mu 2}, \eta_2(W_\delta), \varphi_2(W_\mu))$ , define the following  $L^2$ -metric:

$$\begin{aligned} d(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) &= \left\{ \|\boldsymbol{\beta}_{\delta 1} - \boldsymbol{\beta}_{\delta 2}\|^2 + \|\boldsymbol{\beta}_{\mu 1} - \boldsymbol{\beta}_{\mu 2}\|^2 \right. \\ &\quad \left. + \|\eta_1(W_\delta) - \eta_2(W_\delta)\|_2^2 + \|\varphi_1(W_\mu) - \varphi_2(W_\mu)\|_2^2 \right\}^{1/2} \\ &= \left\{ \|\boldsymbol{\beta}_{\delta 1} - \boldsymbol{\beta}_{\delta 2}\|^2 + \|\boldsymbol{\beta}_{\mu 1} - \boldsymbol{\beta}_{\mu 2}\|^2 \right. \\ &\quad \left. + \int [(\eta_1 - \eta_2)(W_\delta)]^2 dP(W_\delta) + \int [(\varphi_1 - \varphi_2)(W_\mu)]^2 dP(W_\mu) \right\}^{1/2}. \end{aligned}$$

The main result of this section is given in the following theorem.

**Theorem 2.4.7.** *Suppose (1)–(4) in Assumption 2.4.1 hold. As  $n \rightarrow \infty$  and for*

$k_n = O(n^v)$  where  $v \in (0, 1)$ ,

$$d(\hat{\boldsymbol{\tau}}_n, \boldsymbol{\tau}_0) \rightarrow 0,$$

where  $\hat{\boldsymbol{\tau}}_n = (\hat{\boldsymbol{\beta}}_\delta, \hat{\boldsymbol{\beta}}_\mu, \hat{\eta}_n(W_\delta), \hat{\varphi}_n(W_\mu))$  is the spline based maximum likelihood estimator and  $\boldsymbol{\tau}_0 = (\boldsymbol{\beta}_\delta^0, \boldsymbol{\beta}_\mu^0, \eta^0(W_\delta), \varphi^0(W_\mu))$  denotes the true parameter value.

We will prove Theorem 2.4.7 by checking conditions in Theorem 5.7 of van der Vaart (1998), which is an important tool for showing consistency of estimators based on estimating equations, as stated below.

**Lemma 2.4.8** (van der Vaart 1998, Theorem 5.7). *Let  $M_n$  be random functions and let  $M$  be a fixed function of  $\theta$  such that for every  $\varepsilon > 0$*

$$\begin{aligned} \sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| &\xrightarrow{p} 0, \\ \sup_{\theta: d(\theta, \theta_0) \geq \varepsilon} M(\theta) &< M(\theta_0) \end{aligned} \tag{2.4.4}$$

*Then any sequence of estimators  $\hat{\theta}_n$  with  $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_p(1)$  converges in probability to  $\theta_0$ .*

We will prove each condition in Lemma 2.4.8 as a lemma for our spline based maximum likelihood estimator  $\hat{\boldsymbol{\tau}}_n = (\hat{\boldsymbol{\beta}}_\delta, \hat{\boldsymbol{\beta}}_\mu, \hat{\eta}_n(W_\delta), \hat{\varphi}_n(W_\mu))'$ . In the process of proving these conditions, we will summarize other useful conclusions as lemmas.

**Lemma 2.4.9.** Let  $\mathcal{S}_{m,t} = \left\{ \sum_{i=1}^{k_n} \alpha_i B_{i,m,t}, \max_{i=1, \dots, k_n} |\alpha_i| \leq M \right\}$ , where  $B_{i,m,t}$  are B-spline basis functions of order  $m$  on  $[a, b]$ . Then the bracketing number  $N_{[\cdot]}(\varepsilon, \mathcal{S}_{m,t}, \|\cdot\|_\infty)$  under the  $\|\cdot\|_\infty$ -norm and  $N_{[\cdot]}(\varepsilon, \mathcal{S}_{m,t}, L_2(P))$  under  $L_2(P)$ -norm are bounded by  $c(M/\varepsilon)^{k_n}$  and  $c(\sqrt{k_n}M/\varepsilon)^{k_n}$  for some constants  $c$ , respectively.

By a calculation of Shen and Wong (1994) (p. 597) we get  $N_{[\cdot]}(\varepsilon, \mathcal{S}_{m,t}, \|\cdot\|_\infty) \leq c(M/\varepsilon)^{k_n}$ . For  $N_{[\cdot]}(\varepsilon, \mathcal{S}_{m,t}, L_2(P))$  one just needs to change the  $\delta$ -sphere and the cube with diameter  $\varepsilon/n$  in Lemma 5 of Shen and Wong (1994) into a cube in  $\mathbb{R}^{k_n}$  with diameter  $M$  and a small cube with diameter  $\varepsilon/\sqrt{k_n}$ , respectively.

We denote

$$l(\boldsymbol{\tau}; \mathbf{U}) = \Delta \log[f(S|\mathbf{Z}_\delta, \mathbf{Z}_\mu, V_\delta, W_\delta, V_\mu, W_\mu, \boldsymbol{\tau})] + (1 - \Delta) \\ \times \log[1 - F(C|\mathbf{Z}_\delta, \mathbf{Z}_\mu, V_\delta, W_\delta, V_\mu, W_\mu, \boldsymbol{\tau})] .$$

Let  $M(\boldsymbol{\tau}) = Pl(\boldsymbol{\tau}; \mathbf{U})$  and  $M_n(\boldsymbol{\tau}) = \mathbb{P}_n l(\boldsymbol{\tau}; \mathbf{U})$ . In this dissertation, we do not consider the cure rate model and assume that the drift parameter  $\mu \leq 0$ . The first condition in Lemma 2.4.8 is given by the following lemma.

**Lemma 2.4.10.** Let  $\mathcal{T}_n = \Theta \times \mathcal{A}_\eta \times \mathcal{A}_\varphi$ , where  $\mathcal{A}_\eta$  and  $\mathcal{A}_\varphi$  are defined in (2.4.2) and (2.4.3), respectively. Under Assumption 2.4.1 we have:

$$\sup_{\boldsymbol{\tau} \in \mathcal{T}_n} |M_n(\boldsymbol{\tau}) - M(\boldsymbol{\tau})| \xrightarrow{P} 0 . \quad (2.4.5)$$

*Proof.* Let  $\mathcal{F}_1 = \{l(\boldsymbol{\tau}; \mathbf{U}) : \boldsymbol{\tau} \in \mathcal{T}_n\}$ . From Lemma 2.4.9, for all  $\varepsilon > 0$ , there exists a set of brackets

$$\left\{ [\eta_i^l(W_\delta), \eta_i^u(W_\delta)] : \|\eta_i^u(W_\delta) - \eta_i^l(W_\delta)\|_\infty \leq \varepsilon \text{ for } i = 1, 2, \dots, (\sqrt{k_n}M/\varepsilon)^{k_n} \right\}$$

such that for any  $\eta_n(W_\delta) \in \mathcal{S}_{m, \mathbf{t}_\eta}$ ,

$$\eta_i^l(W_\delta) \leq \eta_n(W_\delta) \leq \eta_i^u(W_\delta)$$

for some  $1 \leq i \leq (\sqrt{k_n}M/\varepsilon)^{k_n}$  and all  $W_\delta \in [a_\eta, b_\eta]$ . Similarly, for all  $\varepsilon > 0$ , there exists a set of brackets

$$\left\{ [\varphi_j^l(W_\mu), \varphi_j^u(W_\mu)] : \|\varphi_j^u(W_\mu) - \varphi_j^l(W_\mu)\|_\infty \leq \varepsilon \text{ for } j = 1, 2, \dots, (\sqrt{k_n}M/\varepsilon)^{k_n} \right\}$$

such that for any  $\varphi_n(W_\mu) \in \mathcal{S}_{m, \mathbf{t}_\varphi}$ ,

$$\varphi_j^l(W_\mu) \leq \varphi_n(W_\mu) \leq \varphi_j^u(W_\mu)$$

for some  $1 \leq j \leq (\sqrt{k_n}M/\varepsilon)^{k_n}$  and all  $W_\mu \in [a_\varphi, b_\varphi]$ . From arguments in the proof of Lemma 7.2 of Huang (1991),  $\Theta$  can be covered by  $[c(1/\varepsilon)^{d_1+d_2}]$  balls with radius  $\varepsilon$  where  $c$  is some constant since  $\Theta \in \mathbb{R}^{d_1+d_2}$  is compact. So for any  $\boldsymbol{\theta} \in \Theta$  there

exists an  $1 \leq s \leq [c(1/\varepsilon)^{d_1+d_2}]$  such that  $\|\boldsymbol{\theta} - \boldsymbol{\theta}_s\| < \varepsilon$ . Then we have:

$$\begin{cases} \|\boldsymbol{\beta}_\delta - \boldsymbol{\beta}_{\delta s}\| < \varepsilon \text{ for some } 1 \leq m \leq [c(1/\varepsilon)^{d_1+d_2}] \\ \|\boldsymbol{\beta}_\mu - \boldsymbol{\beta}_{\mu s}\| < \varepsilon \text{ for some } 1 \leq n \leq [c(1/\varepsilon)^{d_1+d_2}]. \end{cases}$$

Since  $\mathbf{Z}_\delta$  and  $\mathbf{Z}_\mu$  are bounded, we have

$$\begin{cases} |\boldsymbol{\beta}'_\delta \mathbf{Z}_\delta - \boldsymbol{\beta}'_{\delta m} \mathbf{Z}_\delta| < c\varepsilon \\ |\boldsymbol{\beta}'_\mu \mathbf{Z}_\mu - \boldsymbol{\beta}'_{\mu n} \mathbf{Z}_\mu| < c\varepsilon \end{cases}$$

for some constant  $c$ . The above inequalities are equivalent to:

$$\begin{cases} \boldsymbol{\beta}'_\delta \mathbf{Z}_\delta \in [\boldsymbol{\beta}'_{\delta m} \mathbf{Z}_\delta - c\varepsilon, \boldsymbol{\beta}'_{\delta m} \mathbf{Z}_\delta + c\varepsilon] \\ \boldsymbol{\beta}'_\mu \mathbf{Z}_\mu \in [\boldsymbol{\beta}'_{\mu n} \mathbf{Z}_\mu - c\varepsilon, \boldsymbol{\beta}'_{\mu n} \mathbf{Z}_\mu + c\varepsilon] \end{cases}$$

for any  $\mathbf{Z}_\delta, \mathbf{Z}_\mu \in \Theta$ .

Based on above facts, we can build a set of brackets

$$\left\{ [l_{i,j,m,n}^l(\mathbf{U}), l_{i,j,m,n}^u(\mathbf{U})] : i, j = 1, 2, \dots, [c(\frac{1}{\varepsilon})^{d_1+d_2}]; \right. \\ \left. m, n = 1, 2, \dots, [(\frac{\sqrt{k_n}M}{\varepsilon})^{k_n}] \right\}$$

such that for any function  $l(\boldsymbol{\tau}; \mathbf{U}) \in \mathcal{F}_1$  and any sample point  $\mathbf{U}$

$$l(\boldsymbol{\tau}; \mathbf{U}) \in [l_{i,j,m,n}^l(\mathbf{U}), l_{i,j,m,n}^u(\mathbf{U})]$$



for some  $1 \leq i, j \leq [c(\frac{1}{\varepsilon})^{d_1+d_2}]$  and  $1 \leq m, n \leq [(\frac{\sqrt{k_n M}}{\varepsilon})^{k_n}]$ . The brackets  $l_{i,j,m,n}^l(\mathbf{U})$  and  $l_{i,j,m,n}^u(\mathbf{U})$  are constructed as follows:

$$l_{i,j,m,n}^l(\mathbf{U}) = \Delta \log \left\{ \frac{\delta_{i,j,m,n}^l}{(2\pi S^3)^{1/2}} \exp \left[ -\frac{(\delta_{i,j,m,n}^l + \mu_{i,j,m,n}^l S)^2}{2S} \right] \right\} \\ + (1 - \Delta) \log \left\{ \Phi \left[ \frac{\mu_{i,j,m,n}^l C + \delta_{i,j,m,n}^l}{C^{1/2}} \right] \right. \\ \left. - \exp(-2\delta_{i,j,m,n}^l \mu_{i,j,m,n}^l) \Phi \left[ \frac{\mu_{i,j,m,n}^l C - \delta_{i,j,m,n}^l}{C^{1/2}} \right] \right\}$$

and

$$l_{i,j,m,n}^u(\mathbf{U}) = \Delta \log \left\{ \frac{\delta_{i,j,m,n}^u}{(2\pi S^3)^{1/2}} \exp \left[ -\frac{(\delta_{i,j,m,n}^u + \mu_{i,j,m,n}^u S)^2}{2S} \right] \right\} \\ + (1 - \Delta) \log \left\{ \Phi \left[ \frac{\mu_{i,j,m,n}^u C + \delta_{i,j,m,n}^u}{C^{1/2}} \right] \right. \\ \left. - \exp(-2\delta_{i,j,m,n}^u \mu_{i,j,m,n}^u) \Phi \left[ \frac{\mu_{i,j,m,n}^u C - \delta_{i,j,m,n}^u}{C^{1/2}} \right] \right\}$$

where

$$\begin{cases} \ln(\delta_{i,j,m,n}^l) = \beta'_{\delta m} \mathbf{Z}_\delta + V \eta_i^l(W_\delta) - c\varepsilon \\ \ln(\delta_{i,j,m,n}^u) = \beta'_{\delta m} \mathbf{Z}_\delta + V \eta_i^u(W_\delta) + c\varepsilon \end{cases}$$

and

$$\begin{cases} \mu_{i,j,m,n}^l = \beta'_{\mu n} \mathbf{Z}_\mu + V \varphi_j^l(W_\mu) - c\varepsilon \\ \mu_{i,j,m,n}^u = \beta'_{\mu n} \mathbf{Z}_\mu + V \varphi_j^u(W_\mu) + c\varepsilon. \end{cases}$$

The above brackets are constructed based on the monotonic properties of the  $\Phi(\cdot)$  function and the exponential function. Without loss of generality, we assume that

$V > 0$ . By using Taylor expansion and Assumption 2.4.1 one can easily get

$$P|l_{i,j,m,n}^u(\mathbf{U}) - l_{i,j,m,n}^l(\mathbf{U})| \leq c\varepsilon$$

for all  $1 \leq i, j \leq [c(1/\varepsilon)^{d_1+d_2}]$  and  $1 \leq m, n \leq [(\sqrt{k_n}M/\varepsilon)^{k_n}]$  where  $c$  is some constant. This means that the  $\varepsilon$ -bracketing number for  $\mathcal{F}_1$  with  $L_1(P)$ -norm is bounded by  $[c(1/\varepsilon)^{d_1+d_2}] \times [(\sqrt{k_n}M/\varepsilon)^{2k_n}]$ , which is less than infinity. So  $\mathcal{F}_1$  is P-Glivenko-Cantelli by Lemma 2.4.1. This leads to

$$\sup_{\tau \in \mathcal{T}_n} |M_n(\tau) - M(\tau)| \xrightarrow{P} 0 .$$

□

Let

$$\begin{aligned} l_{\beta_\delta}(\tau; \mathbf{U}) &= l_{\beta_\delta}(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U}) = \frac{\partial l(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U})}{\partial \beta_\delta} , \\ l_{\beta_\mu}(\tau; \mathbf{U}) &= l_{\beta_\mu}(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U}) = \frac{\partial l(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U})}{\partial \beta_\mu} , \\ l_{\beta_\delta \beta_\delta}(\tau; \mathbf{U}) &= l_{\beta_\delta \beta_\delta}(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U}) = \frac{\partial l_{\beta_\delta}(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U})}{\partial \beta_\delta} , \\ l_{\beta_\delta \beta_\mu}(\tau; \mathbf{U}) &= l_{\beta_\delta \beta_\mu}(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U}) = \frac{\partial l_{\beta_\delta}(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U})}{\partial \beta_\mu} , \\ l_{\beta_\mu \beta_\delta}(\tau; \mathbf{U}) &= l_{\beta_\mu \beta_\delta}(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U}) = \frac{\partial l_{\beta_\mu}(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U})}{\partial \beta_\delta} , \\ l_{\beta_\mu \beta_\mu}(\tau; \mathbf{U}) &= l_{\beta_\mu \beta_\mu}(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U}) = \frac{\partial l_{\beta_\mu}(\beta_\delta, \beta_\mu, \eta, \varphi; \mathbf{U})}{\partial \beta_\mu} . \end{aligned}$$

For any fixed  $\eta(W_\delta) \in C^1[a_\eta, b_\eta]$ , suppose that  $\{\eta_{\alpha_1} : \alpha_1 \text{ in a neighborhood of } 0 \in$

$\mathbb{R}$  is a smooth curve in  $C^1[a_\eta, b_\eta]$  running through  $\eta$  at  $\alpha_1 = 0$ , i.e.,  $\eta_{\alpha_1}|_{\alpha_1=0} = \eta$ .

Let  $\{\varphi_{\alpha_2} : \alpha_2 \text{ in a neighborhood of } 0 \in \mathbb{R}\}$  denote a smooth curve in  $C^1[a_\varphi, b_\varphi]$  such that  $\varphi_{\alpha_2}|_{\alpha_2=0} = \varphi$ . Let

$$\mathcal{H}_1 = \left\{ h_1 : h_1 = \frac{\partial \eta_{\alpha_1}}{\partial \alpha_1} \Big|_{\alpha_1=0} \right\},$$

$$\mathcal{H}_2 = \left\{ h_2 : h_2 = \frac{\partial \varphi_{\alpha_2}}{\partial \alpha_2} \Big|_{\alpha_2=0} \right\}.$$

For any  $h_1 \in \mathcal{H}_1$  and any  $h_2 \in \mathcal{H}_2$ , we define

$$\begin{aligned}
l_\eta(\boldsymbol{\tau}; \mathbf{U})[h_1] &= l_\eta(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_1] = \frac{\partial l(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta_{\alpha_1}, \varphi; \mathbf{U})}{\partial \alpha_1} \Big|_{\alpha_1=0}, \\
l_\varphi(\boldsymbol{\tau}; \mathbf{U})[h_2] &= l_\varphi(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_2] = \frac{\partial l(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi_{\alpha_2}; \mathbf{U})}{\partial \alpha_2} \Big|_{\alpha_2=0}, \\
l_{\beta_\delta \eta}(\boldsymbol{\tau}; \mathbf{U})[h_1] &= l_{\beta_\delta \eta}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_1] = \frac{\partial l_{\beta_\delta}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta_{\alpha_1}, \varphi; \mathbf{U})}{\partial \alpha_1} \Big|_{\alpha_1=0}, \\
l_{\beta_\mu \eta}(\boldsymbol{\tau}; \mathbf{U})[h_1] &= l_{\beta_\mu \eta}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_1] = \frac{\partial l_{\beta_\mu}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta_{\alpha_1}, \varphi; \mathbf{U})}{\partial \alpha_1} \Big|_{\alpha_1=0}, \\
l_{\beta_\delta \varphi}(\boldsymbol{\tau}; \mathbf{U})[h_2] &= l_{\beta_\delta \varphi}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_2] = \frac{\partial l_{\beta_\delta}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi_{\alpha_2}; \mathbf{U})}{\partial \alpha_2} \Big|_{\alpha_2=0}, \\
l_{\beta_\mu \varphi}(\boldsymbol{\tau}; \mathbf{U})[h_2] &= l_{\beta_\mu \varphi}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_2] = \frac{\partial l_{\beta_\mu}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi_{\alpha_2}; \mathbf{U})}{\partial \alpha_2} \Big|_{\alpha_2=0}, \\
l_{\eta \beta_\delta}(\boldsymbol{\tau}; \mathbf{U})[h_1] &= l_{\eta \beta_\delta}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_1] = \frac{\partial l_\eta(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_1]}{\partial \beta_\delta}, \\
l_{\eta \beta_\mu}(\boldsymbol{\tau}; \mathbf{U})[h_1] &= l_{\eta \beta_\mu}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_1] = \frac{\partial l_\eta(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_1]}{\partial \beta_\mu}, \\
l_{\varphi \beta_\delta}(\boldsymbol{\tau}; \mathbf{U})[h_2] &= l_{\varphi \beta_\delta}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_2] = \frac{\partial l_\varphi(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_2]}{\partial \beta_\delta}, \\
l_{\varphi \beta_\mu}(\boldsymbol{\tau}; \mathbf{U})[h_2] &= l_{\varphi \beta_\mu}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_2] = \frac{\partial l_\varphi(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_2]}{\partial \beta_\mu}, \\
l_{\eta \eta}(\boldsymbol{\tau}; \mathbf{U})[h_1, h_1] &= l_{\eta \eta}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_1, h_1] = \frac{\partial l_\eta(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta_{\alpha_1}, \varphi; \mathbf{U})[h_1]}{\partial \alpha_1} \Big|_{\alpha_1=0}, \\
l_{\eta \varphi}(\boldsymbol{\tau}; \mathbf{U})[h_1, h_2] &= l_{\eta \varphi}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_1, h_2] = \frac{\partial l_\eta(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi_{\alpha_2}; \mathbf{U})[h_1]}{\partial \alpha_2} \Big|_{\alpha_2=0}, \\
l_{\varphi \eta}(\boldsymbol{\tau}; \mathbf{U})[h_2, h_1] &= l_{\varphi \eta}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_2, h_1] = \frac{\partial l_\varphi(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta_{\alpha_1}, \varphi; \mathbf{U})[h_2]}{\partial \alpha_1} \Big|_{\alpha_1=0}, \\
l_{\varphi \varphi}(\boldsymbol{\tau}; \mathbf{U})[h_2, h_2] &= l_{\varphi \varphi}(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi; \mathbf{U})[h_2, h_2] = \frac{\partial l_\varphi(\boldsymbol{\beta}_\delta, \boldsymbol{\beta}_\mu, \eta, \varphi_{\alpha_2}; \mathbf{U})[h_2]}{\partial \alpha_2} \Big|_{\alpha_2=0}.
\end{aligned}$$

The  $h_1$  and  $h_2$  in brackets are functions denoting the direction of the functional derivative with respect to  $\eta$  and  $\varphi$ , respectively.

To show the second condition in Lemma 2.4.8 for our proposed spline based estimator  $\hat{\tau}_n$ , Lemma 8.8 from van der Vaart (2002) is used and stated below.

**Lemma 2.4.11** (van der Vaart 2002). *Let  $h$ ,  $g_1$  and  $g_2$  be measurable functions such that  $c_1 \leq h \leq c_2$  and  $(Pg_1g_2)^2 \leq cPg_1^2Pg_2^2$  for a constant  $c < 1$  and constants  $c_1 < 1 < c_2$  close to 1. Then*

$$P(hg_1 + g_2)^2 \geq C(Pg_1^2 + Pg_2^2),$$

for a constant  $C$  depending on  $c$ ,  $c_1$  and  $c_2$  that approaches  $1 - \sqrt{c}$  as  $c_1 \uparrow 1$  and  $c_2 \downarrow 1$ .

**Lemma 2.4.12.** *Suppose Assumption 2.4.1 holds. Then*

$$\sup_{\tau: d(\tau, \tau_0) \geq \varepsilon} M(\tau) \leq M(\tau_0) .$$

*Proof.*

$$\begin{aligned} M(\tau) - M(\tau_0) &= Pl(\tau; \mathbf{U}) - Pl(\tau_0; \mathbf{U}) \\ &= P[l(\tau; \mathbf{U}) - l(\tau_0; \mathbf{U})] \\ &= E_{\tau_0}[l(\tau; \mathbf{U}) - l(\tau_0; \mathbf{U})] . \end{aligned} \tag{2.4.6}$$

By using Taylor expansion of  $l(\boldsymbol{\tau}; \mathbf{U})$  around  $\boldsymbol{\tau}_0$  we have:

$$\begin{aligned} l(\boldsymbol{\tau}; \mathbf{U}) &\cong l(\boldsymbol{\tau}_0; \mathbf{U}) + [\mathbf{U}_0(\boldsymbol{\tau}_0)]'(\boldsymbol{\beta}'_\delta - \boldsymbol{\beta}^{0'}_\delta, \boldsymbol{\beta}'_\mu - \boldsymbol{\beta}^{0'}_\mu, 1, 1)' \\ &\quad + \frac{1}{2}(\boldsymbol{\beta}'_\delta - \boldsymbol{\beta}^{0'}_\delta, \boldsymbol{\beta}'_\mu - \boldsymbol{\beta}^{0'}_\mu, 1, 1)' \mathbf{H}_0(\boldsymbol{\tau}_0)(\boldsymbol{\beta}'_\delta - \boldsymbol{\beta}^{0'}_\delta, \boldsymbol{\beta}'_\mu - \boldsymbol{\beta}^{0'}_\mu, 1, 1) \end{aligned}$$

where

$$[\mathbf{U}_0(\boldsymbol{\tau}_0)]' = \left( l_{\beta_\delta}(\boldsymbol{\tau}; \mathbf{U}), l_{\beta_\mu}(\boldsymbol{\tau}; \mathbf{U}), l_\eta(\boldsymbol{\tau}; \mathbf{U})[h_1], l_\varphi(\boldsymbol{\tau}; \mathbf{U})[h_2] \right)_{\boldsymbol{\tau}=\boldsymbol{\tau}_0}$$

and

$$\begin{aligned} &\mathbf{H}_0(\boldsymbol{\tau}_0) \\ &= \begin{pmatrix} l_{\beta_\delta \beta_\delta}(\boldsymbol{\tau}; \mathbf{U}) & l_{\beta_\delta \beta_\mu}(\boldsymbol{\tau}; \mathbf{U}) & l_{\beta_\delta \eta}(\boldsymbol{\tau}; \mathbf{U})[h_1] & l_{\beta_\delta \varphi}(\boldsymbol{\tau}; \mathbf{U})[h_2] \\ l_{\beta_\mu \beta_\delta}(\boldsymbol{\tau}; \mathbf{U}) & l_{\beta_\mu \beta_\mu}(\boldsymbol{\tau}; \mathbf{U}) & l_{\beta_\mu \eta}(\boldsymbol{\tau}; \mathbf{U})[h_1] & l_{\beta_\mu \varphi}(\boldsymbol{\tau}; \mathbf{U})[h_2] \\ l_{\eta \beta_\delta}(\boldsymbol{\tau}; \mathbf{U})[h_1] & l_{\eta \beta_\mu}(\boldsymbol{\tau}; \mathbf{U})[h_1] & l_{\eta \eta}(\boldsymbol{\tau}; \mathbf{U})[h_1, h_1] & l_{\eta \varphi}(\boldsymbol{\tau}; \mathbf{U})[h_1, h_2] \\ l_{\varphi \beta_\delta}(\boldsymbol{\tau}; \mathbf{U})[h_2] & l_{\varphi \beta_\mu}(\boldsymbol{\tau}; \mathbf{U})[h_2] & l_{\varphi \eta}(\boldsymbol{\tau}; \mathbf{U})[h_2, h_1] & l_{\varphi \varphi}(\boldsymbol{\tau}; \mathbf{U})[h_2, h_2] \end{pmatrix}_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \end{aligned}$$

Without loss of generality, we take derivative of  $l(\boldsymbol{\tau}; \mathbf{U})$  with respect to  $\eta$  and  $\varphi$  with direction  $\eta - \eta^0$  and  $\varphi - \varphi^0$ , respectively, i.e.,  $h_1 = \eta - \eta^0$  and  $h_2 = \varphi - \varphi^0$ .

By the chain rule, this leads to:

$$\begin{aligned} l(\boldsymbol{\tau}; \mathbf{U}) &\cong l(\boldsymbol{\tau}_0; \mathbf{U}) + [\mathbf{U}(\boldsymbol{\tau}_0)]'(\boldsymbol{\beta}'_\delta - \boldsymbol{\beta}^{0'}_\delta, \boldsymbol{\beta}'_\mu - \boldsymbol{\beta}^{0'}_\mu, \eta - \eta^0, \varphi - \varphi^0)' \\ &\quad + \frac{1}{2}(\boldsymbol{\beta}'_\delta - \boldsymbol{\beta}^{0'}_\delta, \boldsymbol{\beta}'_\mu - \boldsymbol{\beta}^{0'}_\mu, \eta - \eta^0, \varphi - \varphi^0)' \mathbf{H}(\boldsymbol{\tau}_0) \\ &\quad \times (\boldsymbol{\beta}'_\delta - \boldsymbol{\beta}^{0'}_\delta, \boldsymbol{\beta}'_\mu - \boldsymbol{\beta}^{0'}_\mu, \eta - \eta^0, \varphi - \varphi^0) \end{aligned}$$

where

$$[\mathbf{U}(\tau_0)]' = \left( \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \beta_\delta}, \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \beta_\mu}, \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta}, \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi} \right)_{\boldsymbol{\tau}=\boldsymbol{\tau}_0}$$

and

$$\mathbf{H}(\boldsymbol{\tau}_0) = \begin{pmatrix} l_{\beta_\delta \beta_\delta}(\boldsymbol{\tau}; \mathbf{U}) & l_{\beta_\delta \beta_\mu}(\boldsymbol{\tau}; \mathbf{U}) & \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \beta_\delta \partial \eta} & \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \beta_\delta \partial \varphi} \\ l_{\beta_\mu \beta_\delta}(\boldsymbol{\tau}; \mathbf{U}) & l_{\beta_\mu \beta_\mu}(\boldsymbol{\tau}; \mathbf{U}) & \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \beta_\mu \partial \eta} & \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \beta_\mu \partial \varphi} \\ \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta \partial \beta_\delta} & \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta \partial \beta_\mu} & \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta \partial \eta} & \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta \partial \varphi} \\ \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi \partial \beta_\delta} & \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi \partial \beta_\mu} & \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi \partial \eta} & \frac{\partial^2 l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi \partial \varphi} \end{pmatrix}_{\boldsymbol{\tau}=\boldsymbol{\tau}_0}.$$

Plugging the above expression for  $l(\boldsymbol{\tau}; \mathbf{U})$  into equation (2.4.6) we get:

$$\begin{aligned} M(\boldsymbol{\tau}) - M(\boldsymbol{\tau}_0) &= E_{\boldsymbol{\tau}_0} \left\{ [\mathbf{U}(\boldsymbol{\tau}_0)]' (\boldsymbol{\beta}'_\delta - \boldsymbol{\beta}^{0'}_\delta, \boldsymbol{\beta}'_\mu - \boldsymbol{\beta}^{0'}_\mu, \eta - \eta^0, \varphi - \varphi^0)' \right\} \\ &\quad + E_{\boldsymbol{\tau}_0} \left\{ \frac{1}{2} (\boldsymbol{\beta}'_\delta - \boldsymbol{\beta}^{0'}_\delta, \boldsymbol{\beta}'_\mu - \boldsymbol{\beta}^{0'}_\mu, \eta - \eta^0, \varphi - \varphi^0)' \mathbf{H}(\boldsymbol{\tau}_0) \right. \\ &\quad \left. \times (\boldsymbol{\beta}'_\delta - \boldsymbol{\beta}^{0'}_\delta, \boldsymbol{\beta}'_\mu - \boldsymbol{\beta}^{0'}_\mu, \eta - \eta^0, \varphi - \varphi^0) \right\}. \end{aligned} \tag{2.4.7}$$

By changing the order of derivative and integral one can derive that the first term  $E_{\boldsymbol{\tau}_0} \left\{ [\mathbf{U}(\boldsymbol{\tau}_0)]' (\boldsymbol{\beta}'_\delta - \boldsymbol{\beta}^{0'}_\delta, \boldsymbol{\beta}'_\mu - \boldsymbol{\beta}^{0'}_\mu, \eta - \eta^0, \varphi - \varphi^0)' \right\}$  in equation (2.4.7) equals 0. Similarly by changing the order of derivative and integral and doing some calcu-

lation for the second term in equation (2.4.7) we have:

$$\begin{aligned}
& M(\boldsymbol{\tau}) - M(\boldsymbol{\tau}_0) \\
&= -E_{\boldsymbol{\tau}_0} \left\{ \frac{1}{2} (\boldsymbol{\beta}'_{\delta} - \boldsymbol{\beta}^{0'}_{\delta}, \boldsymbol{\beta}'_{\mu} - \boldsymbol{\beta}^{0'}_{\mu}, \eta - \eta^0, \varphi - \varphi^0)' [\mathbf{U}(\boldsymbol{\tau}_0)] [\mathbf{U}(\boldsymbol{\tau}_0)]' \right. \\
&\quad \left. \times (\boldsymbol{\beta}'_{\delta} - \boldsymbol{\beta}^{0'}_{\delta}, \boldsymbol{\beta}'_{\mu} - \boldsymbol{\beta}^{0'}_{\mu}, \eta - \eta^0, \varphi - \varphi^0) \right\} \\
&= -\frac{1}{2} E \left\{ E_{\boldsymbol{\tau}_0} \left\{ \left[ (\boldsymbol{\beta}'_{\delta} - \boldsymbol{\beta}^{0'}_{\delta}) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \boldsymbol{\beta}^0_{\delta}} + (\boldsymbol{\beta}'_{\mu} - \boldsymbol{\beta}^{0'}_{\mu}) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \boldsymbol{\beta}^0_{\mu}} \right. \right. \right. \\
&\quad \left. \left. \left. + (\eta - \eta^0) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \eta^0} + (\varphi - \varphi^0) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \varphi^0} \right]^2 \middle| \mathbf{Z}_{\delta}, \mathbf{Z}_{\mu}, W_{\delta}, W_{\mu}, V \right\} \right\}.
\end{aligned} \tag{2.4.8}$$

Let

$$\begin{cases} g_1(\mathbf{U}) = (\boldsymbol{\beta}'_{\delta} - \boldsymbol{\beta}^{0'}_{\delta}) (\partial / \partial \boldsymbol{\beta}^0_{\delta}) l(\boldsymbol{\tau}_0; \mathbf{U}) + (\eta - \eta^0) (\partial / \partial \eta^0) l(\boldsymbol{\tau}_0; \mathbf{U}) \\ g_2(\mathbf{U}) = (\boldsymbol{\beta}'_{\mu} - \boldsymbol{\beta}^{0'}_{\mu}) (\partial / \partial \boldsymbol{\beta}^0_{\mu}) l(\boldsymbol{\tau}_0; \mathbf{U}) + (\varphi - \varphi^0) (\partial / \partial \varphi^0) l(\boldsymbol{\tau}_0; \mathbf{U}) . \end{cases}$$

By the Cauchy-Schwarz inequality one can easily get

$$(Pg_1g_2)^2 < P(g_1^2)P(g_2^2)$$

since  $g_1$  and  $g_2$  are not linearly dependent. Then based on Lemma 2.4.11 we have

$$P(g_1 + g_2)^2 \geq C(Pg_1^2 + Pg_2^2)$$



for some constant  $C$  with  $h = 1$ . This leads to

$$\begin{aligned}
& M(\boldsymbol{\tau}) - M(\boldsymbol{\tau}_0) \\
& \leq -\frac{1}{2}CE \left\{ E_{\boldsymbol{\tau}_0} \left\{ \left[ (\boldsymbol{\beta}'_{\delta} - \boldsymbol{\beta}^{0'}_{\delta}) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \boldsymbol{\beta}^0_{\delta}} + (\eta - \eta^0) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \eta^0} \right]^2 \right. \right. \\
& \quad \left. \left. + \left[ (\boldsymbol{\beta}'_{\mu} - \boldsymbol{\beta}^{0'}_{\mu}) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \boldsymbol{\beta}^0_{\mu}} + (\varphi - \varphi^0) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \varphi^0} \right]^2 \middle| \mathbf{Z}_{\delta}, \mathbf{Z}_{\mu}, W_{\delta}, W_{\mu}, V \right\} \right\}.
\end{aligned}$$

Let

$$\begin{cases} g_{11} = (\boldsymbol{\beta}'_{\delta} - \boldsymbol{\beta}^{0'}_{\delta})(\partial/\partial \boldsymbol{\beta}^0_{\delta})l(\boldsymbol{\tau}_0; \mathbf{U}) \\ g_{12} = (\eta - \eta^0)(\partial/\partial \eta^0)l(\boldsymbol{\tau}_0; \mathbf{U}) \end{cases}$$

and

$$\begin{cases} g_{21} = (\boldsymbol{\beta}'_{\mu} - \boldsymbol{\beta}^{0'}_{\mu})(\partial/\partial \boldsymbol{\beta}^0_{\mu})l(\boldsymbol{\tau}_0; \mathbf{U}) \\ g_{22} = (\varphi - \varphi^0)(\partial/\partial \varphi^0)l(\boldsymbol{\tau}_0; \mathbf{U}) . \end{cases}$$

We have the following inequalities:

$$\begin{cases} (Pg_{11}g_{12})^2 < P(g_{11}^2)P(g_{12}^2) \\ (Pg_{21}g_{22})^2 < P(g_{21}^2)P(g_{22}^2) . \end{cases}$$

by the Cauchy-Schwarz inequality. After applying Lemma 2.4.11 with  $h = 1$  again,

we get:

$$\begin{cases} P(g_{11} + g_{12})^2 \geq C(Pg_{11}^2 + Pg_{12}^2) \\ P(g_{21} + g_{22})^2 \geq C(Pg_{21}^2 + Pg_{22}^2) \end{cases}$$

for some constant  $C$ . Hence

$$\begin{aligned}
& M(\boldsymbol{\tau}) - M(\boldsymbol{\tau}_0) \\
& \leq -\frac{1}{2}E_{\boldsymbol{\tau}_0} \{c_1 [(\boldsymbol{\beta}'_{\boldsymbol{\delta}} - \boldsymbol{\beta}^0_{\boldsymbol{\delta}})^{\otimes 2} + (\boldsymbol{\beta}'_{\boldsymbol{\mu}} - \boldsymbol{\beta}^0_{\boldsymbol{\mu}})^{\otimes 2} + (\eta - \eta^0)^2 + (\varphi - \varphi^0)^2]\} \\
& \leq -Cd^2(\boldsymbol{\tau}, \boldsymbol{\tau}_0)
\end{aligned} \tag{2.4.9}$$

for some constant  $C$ . Inequality (2.4.9) implies that:

$$\sup_{\boldsymbol{\tau}:d(\boldsymbol{\tau},\boldsymbol{\tau}_0)\geq\varepsilon} M(\boldsymbol{\tau}) \leq M(\boldsymbol{\tau}_0) - C\varepsilon^2 < M(\boldsymbol{\tau}_0) .$$

□

The third condition of Lemma 2.4.8 for our proposed spline based estimator  $\hat{\boldsymbol{\tau}}_n$  is stated as following Lemma 2.4.13.

**Lemma 2.4.13.** *Under Assumption 2.4.1 we have:*

$$M_n(\hat{\boldsymbol{\tau}}_n) \geq M_n(\boldsymbol{\tau}_0) - o_p(1) . \tag{2.4.10}$$

*Proof.* Let  $\boldsymbol{\tau}_{0,n} = (\boldsymbol{\beta}_\delta^0, \boldsymbol{\beta}_\mu^0, \boldsymbol{\alpha}_\eta^0, \boldsymbol{\alpha}_\varphi^0)'$ . We have:

$$\begin{aligned}
& M_n(\hat{\boldsymbol{\tau}}_n) - M_n(\boldsymbol{\tau}_0) \\
&= M_n(\hat{\boldsymbol{\tau}}_n) - M_n(\boldsymbol{\tau}_{0,n}) + M_n(\boldsymbol{\tau}_{0,n}) - M_n(\boldsymbol{\tau}_0) \\
&\geq M_n(\boldsymbol{\tau}_{0,n}) - M_n(\boldsymbol{\tau}_0) \\
&= \mathbb{P}_n[l(\boldsymbol{\tau}_{0,n}; \mathbf{U}) - l(\boldsymbol{\tau}_0; \mathbf{U})] \\
&= (\mathbb{P}_n - P)[l(\boldsymbol{\tau}_{0,n}; \mathbf{U}) - l(\boldsymbol{\tau}_0; \mathbf{U})] + P[l(\boldsymbol{\tau}_{0,n}; \mathbf{U}) - l(\boldsymbol{\tau}_0; \mathbf{U})] \tag{2.4.11}
\end{aligned}$$

where the first inequality follows from the fact that  $\hat{\boldsymbol{\tau}}_n$  is MLE. Let

$$\mathcal{F}_2 = \left\{ l_c[(\eta, \varphi); \mathbf{U}] = l[(\boldsymbol{\beta}_\delta^0, \boldsymbol{\beta}_\mu^0, \eta, \varphi); \mathbf{U}] - l[\boldsymbol{\tau}^0; \mathbf{U}] : \eta \in \mathcal{S}_{m,t_\eta}, \varphi \in \mathcal{S}_{m,t_\varphi}, \right. \\
\left. \|\eta - \eta^0\|_\infty = O(k_n^{-1}) \text{ and } \|\varphi - \varphi^0\|_\infty = O(k_n^{-1}) \right\}.$$

Based on facts 1 and 2 (p. 40), we can construct a set of brackets

$$\left\{ \left[ l_{c;m,n}^l(\mathbf{U}), l_{c;m,n}^u(\mathbf{U}) \right] : m, n = 1, 2, \dots, \left[ \left( \frac{\sqrt{k_n} M}{\varepsilon} \right)^{k_n} \right] \right\}$$

such that for any function  $l_c[(\eta, \varphi); \mathbf{U}] \in \mathcal{F}_2$  and any sample point  $\mathbf{U}$

$$l_c[(\eta, \varphi); \mathbf{U}] \in \left[ l_{c;m,n}^l(\mathbf{U}), l_{c;m,n}^u(\mathbf{U}) \right]$$

for some  $1 \leq m, n \leq [(\sqrt{k_n} M / \varepsilon)^{k_n}]$ . The brackets  $l_{c;m,n}^l(\mathbf{U})$  and  $l_{c;m,n}^u(\mathbf{U})$  are

constructed as follows:

$$\begin{aligned}
l_{c;m,n}^l(\mathbf{U}) = & \Delta \log \left\{ \frac{\delta_{c;m,n}^l}{(2\pi S^3)^{1/2}} \exp \left[ -\frac{(\delta_{c;m,n}^u + \mu_{c;m,n}^u S)^2}{2S} \right] \right\} \\
& + (1 - \Delta) \log \left\{ \Phi \left[ \frac{\mu_{c;m,n}^l C + \delta_{c;m,n}^l}{C^{1/2}} \right] \right. \\
& \quad \left. - \exp(-2\delta_{c;m,n}^l \mu_{c;m,n}^l) \Phi \left[ \frac{\mu_{c;m,n}^l C - \delta_{c;m,n}^l}{C^{1/2}} \right] \right\} - l[\boldsymbol{\tau}^0; \mathbf{U}]
\end{aligned}$$

and

$$\begin{aligned}
l_{c;m,n}^u(\mathbf{U}) = & \Delta \log \left\{ \frac{\delta_{c;m,n}^u}{(2\pi S^3)^{1/2}} \exp \left[ -\frac{(\delta_{c;m,n}^l + \mu_{c;m,n}^l S)^2}{2S} \right] \right\} \\
& + (1 - \Delta) \log \left\{ \Phi \left[ \frac{\mu_{c;m,n}^u C + \delta_{c;m,n}^u}{C^{1/2}} \right] \right. \\
& \quad \left. - \exp(-2\delta_{c;m,n}^u \mu_{c;m,n}^u) \Phi \left[ \frac{\mu_{c;m,n}^u C - \delta_{c;m,n}^u}{C^{1/2}} \right] \right\} - l[\boldsymbol{\tau}^0; \mathbf{U}]
\end{aligned}$$

where

$$\begin{cases} \ln(\delta_{c;m,n}^l) = \boldsymbol{\beta}_\delta^{0'} \mathbf{Z}_\delta + V \eta_i^l(W_\delta) - c\varepsilon \\ \ln(\delta_{c;m,n}^u) = \boldsymbol{\beta}_\delta^{0'} \mathbf{Z}_\delta + V \eta_i^u(W_\delta) + c\varepsilon \end{cases}$$

and

$$\begin{cases} \mu_{c;m,n}^l = \boldsymbol{\beta}_\mu^{0'} \mathbf{Z}_\mu + V \varphi_j^l(W_\mu) - c\varepsilon \\ \mu_{c;m,n}^u = \boldsymbol{\beta}_\mu^{0'} \mathbf{Z}_\mu + V \varphi_j^u(W_\mu) + c\varepsilon. \end{cases}$$

The above brackets are constructed based on the monotonic properties of  $\Phi(\cdot)$  function and exponential function. Without loss of generality, we also assume that:

$V > 0$ . By Taylor expansion and Assumption 2.4.1 one can get

$$P|l_{c;m,n}^u(\mathbf{U}) - l_{c;m,n}^l(\mathbf{U})|^2 \leq c\varepsilon$$

for all  $1 \leq m, n \leq [(\sqrt{k_n}M/\varepsilon)^{k_n}]$  where  $c$  is some constant. This means that the  $\varepsilon$ -bracketing number for  $\mathcal{F}_1$  with  $L_2(P)$ - norm is bounded by  $[(\sqrt{k_n}M/\varepsilon)^{2k_n}]$ , i.e.,

$$N_{[\cdot]}(\varepsilon, \mathcal{F}_2, L_2(P)) \leq c \left[ \frac{k_n^{1/2}M}{\varepsilon} \right]^{2k_n}.$$

This leads to:

$$\begin{aligned} J_{[\cdot]}(1, \mathcal{F}_2, L_2(P)) &= \int_0^1 \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{F}_2, L_2(P))} d\varepsilon \\ &= \int_0^1 2ck_n \log \left[ \frac{k_n^{1/2}M}{\varepsilon} \right] d\varepsilon. \end{aligned}$$

We have  $J_{[\cdot]}(1, \mathcal{F}_2, L_2(P)) < \infty$  since the integrand grows of slower order than  $(1/\varepsilon)^2$  which can be seen from the functional form of the integrand in the integral of  $J_{[\cdot]}(1, \mathcal{F}_2, L_2(P))$ . Therefore,  $\mathcal{F}_2$  is  $P$ -Donsker. For any function  $l_c[(\eta, \varphi); \mathbf{U}]$  in class  $\mathcal{F}_2$  we find that

$$\begin{aligned} P \left[ l_c[(\eta, \varphi); \mathbf{U}] \right]^2 &= P \left[ l[(\boldsymbol{\beta}_\delta^0, \boldsymbol{\beta}_\mu^0, \eta, \varphi); \mathbf{U}] - l[(\boldsymbol{\beta}_\delta^0, \boldsymbol{\beta}_\mu^0, \eta^0, \varphi^0); \mathbf{U}] \right]^2 \\ &\rightarrow 0 \end{aligned}$$

by the Dominated Convergence Theorem. Then by Lemma 2.4.3 we have

$$\begin{aligned} E\sqrt{n}\|\mathbb{P}_n - P\|_{\mathcal{F}_2} &\rightarrow 0 \\ \Rightarrow (\mathbb{P}_n - P)[l(\boldsymbol{\tau}_{\mathbf{0},n}; \mathbf{U}) - l(\boldsymbol{\tau}_{\mathbf{0}}; \mathbf{U})] &= o_p(n^{-\frac{1}{2}}). \end{aligned} \quad (2.4.12)$$

By the Dominated Convergence Theorem, it follows that

$$P[l(\boldsymbol{\tau}_{\mathbf{0},n}; \mathbf{U}) - l(\boldsymbol{\tau}_{\mathbf{0}}; \mathbf{U})] = o_p(1). \quad (2.4.13)$$

By plugging equations (2.4.12) and (2.4.13) into (2.4.11) we find that

$$M_n(\hat{\boldsymbol{\tau}}_n) - M_n(\boldsymbol{\tau}_{\mathbf{0}}) \geq -o_p(1).$$

This completes the proof of Lemma 2.4.13. □

### 2.4.3 Convergence Rate

In this section, we further explore the convergence rate of the spline based estimator  $\hat{\boldsymbol{\tau}}_n = (\hat{\boldsymbol{\beta}}_{\delta}'', \hat{\boldsymbol{\beta}}_{\mu}'', \hat{\eta}_n(W_{\delta}), \hat{\varphi}_n(W_{\mu}))'$  to the true parameter vector  $\boldsymbol{\tau}_{\mathbf{0}} = (\boldsymbol{\beta}_{\delta}^{0\prime}, \boldsymbol{\beta}_{\mu}^{0\prime}, \eta^0(W_{\delta}), \varphi^0(W_{\mu}))'$ , when sample size  $n$  goes to infinity.

**Assumption 2.4.2.**  $\eta(W_{\delta}) \in C^j[a_{\eta}, b_{\eta}]$  and  $\varphi(W_{\mu}) \in C^j[a_{\varphi}, b_{\varphi}]$  where  $j \geq 1$ .

**Theorem 2.4.14.** Let  $K_n = O(n^v)$  where  $v$  satisfies  $[2(1+j)]^{-1} < v < (2j)^{-1}$ .

Suppose that Assumption 2.4.1 (1), (2), (4) and Assumption 2.4.2 hold. Then

$$d(\hat{\boldsymbol{\tau}}_n, \boldsymbol{\tau}_0) = O_p\left(n^{-\min(jv, (1-v)/2)}\right).$$

**Remark.** We want to point out that when  $v = 1/(2j + 1)$

$$d(\hat{\boldsymbol{\tau}}_n, \boldsymbol{\tau}_0) = O_p\left(n^{-j/(1+2j)}\right)$$

which is the optimal convergence rate in nonparametric regression analysis.

*Proof.* We will use Lemma 2.4.4 to show the result of Theorem 2.4.14. From equation (2.4.9) we have

$$M(\boldsymbol{\tau}_0) - M(\boldsymbol{\tau}) \geq Cd^2(\boldsymbol{\tau}, \boldsymbol{\tau}_0)$$

where  $C$  is some constant. From equation (2.4.11) we have

$$\begin{aligned} & M_n(\hat{\boldsymbol{\tau}}_n) - M_n(\boldsymbol{\tau}_0) \\ & \geq (\mathbb{P}_n - P)[l(\boldsymbol{\tau}_{0,n}; \mathbf{U}) - l(\boldsymbol{\tau}_0; \mathbf{U})] + P[l(\boldsymbol{\tau}_{0,n}; \mathbf{U}) - l(\boldsymbol{\tau}_0; \mathbf{U})]. \end{aligned} \quad (2.4.14)$$

By using Taylor expansion of  $l(\boldsymbol{\tau}_{0,n}; \mathbf{U})$  around  $\boldsymbol{\tau}_0$  we get

$$(\mathbb{P}_n - P)[l(\boldsymbol{\tau}_{0,n}; \mathbf{U}) - l(\boldsymbol{\tau}_0; \mathbf{U})] \cong (\mathbb{P}_n - P)\left\{[\mathbf{U}_1(\boldsymbol{\tau}_0)]'(1, 1)'\right\}$$

where  $\mathbf{U}_1(\boldsymbol{\tau}_0) = (l_\eta(\boldsymbol{\tau}; \mathbf{U})[h_1], l_\varphi(\boldsymbol{\tau}; \mathbf{U})[h_2])'_{\boldsymbol{\tau}=\boldsymbol{\tau}_0}$ . Without loss of generality, we take the derivative of  $l(\boldsymbol{\tau}; \mathbf{U})$  with respect to  $\eta$  and  $\varphi$  with direction  $\eta_n^0 - \eta^0$  and  $\varphi_n^0 - \varphi^0$ ,

respectively, i.e.,  $h_1 = \eta_n^0 - \eta^0$  and  $h_2 = \varphi_n^0 - \varphi^0$ . By the chain rule, this leads to

$$\begin{aligned} & (\mathbb{P}_n - P)[l(\boldsymbol{\tau}_{0,n}; \mathbf{U}) - l(\boldsymbol{\tau}_0; \mathbf{U})] \\ &= (\mathbb{P}_n - P) \left[ \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} (\eta_n^0 - \eta^0) + \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} (\varphi_n^0 - \varphi^0) \right] \\ &= n^{-jv+\varepsilon} (\mathbb{P}_n - P) \left[ \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \frac{(\eta_n^0 - \eta^0)}{n^{-jv+\varepsilon}} + \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \frac{(\varphi_n^0 - \varphi^0)}{n^{-jv+\varepsilon}} \right] \end{aligned}$$

in which  $0 < \varepsilon < \frac{1}{2} + jv$ . Let

$$\begin{aligned} \mathcal{F}_3 = \left\{ l_c[(\eta, \varphi); \mathbf{U}] = l[(\boldsymbol{\beta}_\delta^0, \boldsymbol{\beta}_\mu^0, \eta, \varphi); \mathbf{U}] - l[\boldsymbol{\tau}_0; \mathbf{U}] : \eta \in \mathcal{S}_{m,t_\eta}, \varphi \in \mathcal{S}_{m,t_\varphi}, \right. \\ \left. \|\eta - \eta^0\|_\infty = O(k_n^{-1}) \text{ and } \|\varphi - \varphi^0\|_\infty = O(k_n^{-1}) \right\}. \end{aligned}$$

By a similar argument as that in the proof of Lemma 2.4.13,  $\mathcal{F}_3$  is  $P$ -Donsker. Note that

$$\begin{aligned} & P \left[ \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \frac{(\eta_n^0 - \eta^0)}{n^{-jv+\varepsilon}} + \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \frac{(\varphi_n^0 - \varphi^0)}{n^{-jv+\varepsilon}} \right]^2 \\ & \leq 2 \left\{ P \left[ \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \frac{(\eta_n^0 - \eta^0)}{n^{-jv+\varepsilon}} \right]^2 + P \left[ \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \frac{(\varphi_n^0 - \varphi^0)}{n^{-jv+\varepsilon}} \right]^2 \right\}. \end{aligned}$$

Under Assumption 2.4.1 (1), (2), (4) and Assumption 2.4.2,  $\frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0}$  and  $\frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0}$  are uniformly bounded. From Lemma 2.4.6,

$$\begin{cases} \|\eta_n^0 - \eta^0\|_\infty = O(k_n^{-j}) = O(n^{-jv}) \\ \|\varphi_n^0 - \varphi^0\|_\infty = O(k_n^{-j}) = O(n^{-jv}) \end{cases}.$$



Therefore,

$$P \left[ \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \frac{(\eta_n^0 - \eta^0)}{n^{-jv+\varepsilon}} + \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \frac{(\varphi_n^0 - \varphi^0)}{n^{-jv+\varepsilon}} \right]^2 \rightarrow 0 \text{ as } n \rightarrow \infty .$$

Since  $\mathcal{F}_3$  is  $P$ -Donsker, from Lemma 2.4.3 we get

$$(\mathbb{P}_n - P) \left[ \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \eta} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \frac{(\eta_n^0 - \eta^0)}{n^{-jv+\varepsilon}} + \frac{\partial l(\boldsymbol{\tau}; \mathbf{U})}{\partial \varphi} \Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \frac{(\varphi_n^0 - \varphi^0)}{n^{-jv+\varepsilon}} \right] = o_p(n^{-1/2}) .$$

Then

$$(\mathbb{P}_n - P) [l(\boldsymbol{\tau}_{0,n}; \mathbf{U}) - l(\boldsymbol{\tau}_0; \mathbf{U})] = o_p(n^{-jv+\varepsilon} \cdot n^{-1/2}) = o_p(n^{-2jv}) \quad (2.4.15)$$

because  $2jv < jv - \varepsilon + \frac{1}{2} < jv + \frac{1}{2}$ . From equation ( 2.4.8), we have

$$\begin{aligned} & P[l(\boldsymbol{\tau}_{0,n}; \mathbf{U}) - l(\boldsymbol{\tau}_0; \mathbf{U})] \\ &= -\frac{1}{2} E \left[ E_{\boldsymbol{\tau}_0} \left\{ \left[ (\eta_n^0 - \eta^0) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \eta^0} + (\varphi_n^0 - \varphi^0) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \varphi^0} \right]^2 \Big| \mathbf{Z}_\delta, \mathbf{Z}_\mu, W_\delta, W_\mu, V \right\} \right] \\ &\geq -E \left[ E_{\boldsymbol{\tau}_0} \left\{ \left[ (\eta_n^0 - \eta^0) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \eta^0} \right]^2 + \left[ (\varphi_n^0 - \varphi^0) \frac{\partial l(\boldsymbol{\tau}_0; \mathbf{U})}{\partial \varphi^0} \right]^2 \Big| \mathbf{Z}_\delta, \mathbf{Z}_\mu, W_\delta, W_\mu, V \right\} \right] \\ &= -O(n^{-2jv}) \end{aligned} \quad (2.4.16)$$

Plugging equation (2.4.15) and equation (2.4.16) into equation (2.4.14), we have

$$\begin{aligned} M_n(\hat{\boldsymbol{\tau}}_n) - M_n(\boldsymbol{\tau}_0) &\geq -O_p(n^{-2jv}) \\ &\geq -O_p\left(n^{-2\min(jv, (1-v)/2)}\right). \end{aligned} \quad (2.4.17)$$

Let

$$\mathcal{F}_4 = \left\{ l(\boldsymbol{\tau}; \mathbf{U}) - l(\boldsymbol{\tau}_0; \mathbf{U}) : \boldsymbol{\tau} \in \mathcal{T}_n \text{ and } d(\boldsymbol{\tau}, \boldsymbol{\tau}_0) \leq \delta \right\}.$$

By a similar construction of brackets as that for  $\mathcal{F}_1$  along with Lemma 2.4.9, the  $\varepsilon$ - bracketing number of  $\mathcal{F}_4$  with  $L_2(P)$ -norm,  $N_{[\cdot]}(\varepsilon, \mathcal{F}_4, L_2(P))$  is bounded by  $(M/\varepsilon)^{ck_n}$  where  $c$  is some constant. Then

$$\begin{aligned} \tilde{J}_{[\cdot]}(\delta, \mathcal{F}_4, L_2(P)) &= \int_0^\delta \sqrt{1 + \log N_{[\cdot]}(\varepsilon, \mathcal{F}_4, L_2(P))} d\varepsilon \\ &= \int_0^\delta \sqrt{1 + \log \left(\frac{M}{\varepsilon}\right)^{ck_n}} d\varepsilon \\ &= \int_0^\delta \sqrt{1 + ck_n \log \left(\frac{M}{\varepsilon}\right)} d\varepsilon \\ &\leq \int_0^\delta \sqrt{ck_n + ck_n \log \left(\frac{M}{\varepsilon}\right)} d\varepsilon \\ &= \sqrt{ck_n} \int_0^\delta \sqrt{1 + \log \left(\frac{M}{\varepsilon}\right)} d\varepsilon. \end{aligned}$$

Let  $X = \sqrt{1 + \log(M/\varepsilon)}$ , then

$$\varepsilon = M \exp(1 - X^2)$$

$$d\varepsilon = -2M \exp(1 - X^2) dX.$$

We get

$$\begin{aligned}
\int \sqrt{1 + \log \frac{M}{\varepsilon}} d\varepsilon &= \int X(-2M) \exp(1 - X^2) dX \\
&= M \int \exp(1 - X^2) d(1 - X^2) \\
&= M \exp(1 - X^2) \\
&= M \exp \left[ 1 - \left( \sqrt{1 + \log \frac{M}{\varepsilon}} \right)^2 \right] \\
&= \varepsilon .
\end{aligned}$$

Therefore,

$$\begin{aligned}
\tilde{J}_{[\cdot]}(\delta, \mathcal{F}_4, L_2(P)) &\leq \sqrt{ck_n} \int_0^\delta \sqrt{1 + \log \left( \frac{M}{\varepsilon} \right)} d\varepsilon \\
&= \sqrt{ck_n} \delta .
\end{aligned}$$

Under Assupmtion 2.4.1 (1), (2), (3) and Assumption 2.4.2,  $l(\boldsymbol{\tau}; \mathbf{U})$  is uniformly bounded. By Lemma 2.4.5, we know that

$$\begin{aligned}
E \|\mathbb{G}_n\|_{\mathcal{F}_4} &\lesssim \sqrt{ck_n} \delta \left( 1 + \frac{\sqrt{ck_n} \delta}{\delta^2 \sqrt{n}} M \right) \\
&= \sqrt{c} \sqrt{k_n} \delta + \frac{Mck_n}{\sqrt{n}} .
\end{aligned}$$

That is,

$$\begin{aligned} E \sup_{d(\boldsymbol{\tau}, \boldsymbol{\tau}_0) < \delta} |(M_n - M)(\boldsymbol{\tau}) - (M_n - M)(\boldsymbol{\tau}_0)| &\lesssim \frac{1}{\sqrt{n}} \left( \sqrt{c} \sqrt{k_n} \delta + \frac{Mck_n}{\sqrt{n}} \right) \\ &= \frac{1}{\sqrt{n}} \phi_n(\delta), \end{aligned}$$

where  $\phi_n(\delta) = \sqrt{c} \sqrt{k_n} \delta + \frac{Mck_n}{\sqrt{n}}$ . Since

$$\frac{\phi_n(\delta)}{\delta^\alpha} = \sqrt{c} \sqrt{k_n} \delta^{1-\alpha} + \frac{Mck_n}{\sqrt{n}} \delta^{-\alpha},$$

it is obvious that  $\phi_n(\delta)/\delta^\alpha$  is decreasing for some  $\alpha < 2$  (e.g.,  $\alpha$  can be 1.5.). We note that

$$\begin{aligned} n^{2jv} \phi_n \left( \frac{1}{n^{jv}} \right) &= n^{2jv} \sqrt{c} \sqrt{k_n} \frac{1}{n^{jv}} + n^{2jv} \frac{Mck_n}{\sqrt{n}} \\ &= O \left( n^{jv} n^{\frac{v}{2}} + n^{2jv} n^v \frac{1}{\sqrt{n}} \right) \\ &= O \left( n^{\frac{1}{2}} [n^{jv - \frac{1-v}{2}} + n^{2(jv - \frac{1-v}{2})}] \right) \end{aligned} \quad (2.4.18)$$

and

$$\begin{aligned} n^{2 \times (1-v)/2} \phi_n \left( \frac{1}{n^{(1-v)/2}} \right) &= n^{2 \times (1-v)/2} \sqrt{c} \sqrt{k_n} \frac{1}{n^{(1-v)/2}} + n^{2 \times (1-v)/2} \frac{Mck_n}{\sqrt{n}} \\ &= O \left( n^{(1-v)/2} n^{(v)/2} + n^{1-v} n^v n^{-1/2} \right) \\ &= O(n^{1/2}) \end{aligned} \quad (2.4.19)$$

since  $k_n = O(n^v)$ . Expressions (2.4.18) and (2.4.19) imply that

$$r_n^2 \phi_n \left( \frac{1}{r_n} \right) \leq n^{1/2}$$

where  $r_n = \min(jv, (1-v)/2)$ . From (2.4.17) we get

$$M_n(\hat{\boldsymbol{\tau}}_n) - M_n(\boldsymbol{\tau}_0) \geq -O_p(r_n^{-2}).$$

Therefore

$$r_n d(\hat{\boldsymbol{\tau}}_n, \boldsymbol{\tau}_0) = O_p(1)$$

by Lemma 2.4.4. □

## 2.5 Simulation Study

In this section, we conduct simulation studies to evaluate the finite sample performance of the proposed methodology under different scenarios. Observed failure times are generated following the idea of Tong *et al.* (2008). We first generate a follow-up time  $C$ . The latent process  $X(t)$  is simulated by accumulating normally distributed independent increments  $\Delta_X = X(t + \Delta_t) - X(t)$  over time increments of length  $\Delta_t$ . Specifically, for subject  $i$ , we first generate a set of increments  $\{\Delta_{Xj} : j = 1, \dots, n_i\}$  from a normal distribution with mean  $\Delta_t \mu$  and variance 1, where  $n_i = \lceil C_i / \Delta_t \rceil$ . Let  $X_{ij} = \delta + \sum_{k=1}^j \Delta_{Xk}$ ,  $j = 1, \dots, n_i$ . If  $X_{ij} > 0$  for all  $j$ , then subject  $i$  has not failed by the end of follow-up time  $C_i$ . If there exists  $j$  such that  $X_{il} > 0$  for  $l = 1, \dots, j-1$  and  $X_{ij} \leq 0$ , then subject  $i$  has failed before

follow-up time  $C_i$  and the corresponding failure time is calculated as  $j \times \Delta_t$  ( $\Delta_t$  is chosen to be very small).

For given covariates, we simulate event times based on the following link functions

$$\begin{cases} \ln(\delta) = \mathbf{Z}_\delta^T \boldsymbol{\beta}_\delta + V_\delta \eta(W_\delta) \\ \mu = \mathbf{Z}_\mu^T \boldsymbol{\beta}_\mu + V_\mu \varphi(W_\mu) \end{cases},$$

where  $\mathbf{Z}_\delta = (1, Z_{\delta 2})^T$ ,  $\mathbf{Z}_\mu = (1, Z_{\mu 2})^T$ ,  $\eta(W_\delta) = -0.8 \sin[\pi(W_\delta + 1)] + 1.2W_\delta$  and  $\varphi(W_\mu) = -\frac{4}{\sqrt{\pi}} \exp[-(3W_\mu - 4)^2] - 3.9$ . The covariates are simulated by  $Z_{\delta 2} \sim \text{Bernoulli}(1, 0.5)$ ,  $Z_{\mu 2} \sim N(0, 1)$ ,  $V_\delta \sim N(0.5, 0.2)$ ,  $V_\mu \sim \text{Uniform}(0, 1)$ ,  $W_\delta \sim \text{Uniform}(0, 1)$  and  $W_\mu \sim \text{Uniform}(1, 2)$ . The time increment  $\Delta_t$  is chosen to be 0.001. We consider the following cases of regression parameters:

Case 1.  $\boldsymbol{\beta}_\delta = (1, -0.5)^T$ ,  $\boldsymbol{\beta}_\mu = (-0.4, -0.3)^T$ ,

Case 2.  $\boldsymbol{\beta}_\delta = (1, -0.6)^T$ ,  $\boldsymbol{\beta}_\mu = (-0.4, 1)^T$ ,

Case 3.  $\boldsymbol{\beta}_\delta = (1, 0)^T$ ,  $\boldsymbol{\beta}_\mu = (-0.4, 0)^T$ ,

Case 4.  $\boldsymbol{\beta}_\delta = (1, 0.4)^T$ ,  $\boldsymbol{\beta}_\mu = (-0.4, -0.4)^T$ ,

Case 5.  $\boldsymbol{\beta}_\delta = (1, 0.6)^T$ ,  $\boldsymbol{\beta}_\mu = (-0.4, -1)^T$ .

Let the censoring time  $C \sim N(2.8, 0.2)$ , which is independent of event times. The event rates range from 65% to 80% roughly for the above cases. Sample sizes of 200, 400 and 600 are considered for each case. For the estimation of smooth functions  $\eta(\cdot)$  and  $\varphi(\cdot)$ , we place the inner knots of the B-spline basis functions at the 1st

quartile, 2nd quartile and 3rd quartile of  $W_\delta$  and  $W_\mu$ , respectively.

Table 2.1 presents the simulation results for the parametric components of link functions based on 100 replications. The table includes the estimated bias (BIAS), the averages of estimated standard errors (Estimated SE) and the sample standard deviations of the estimates (Empirical SE). We display the true functions, estimated functions and 95% point-wise confidence intervals for all cases in Figures 2.3 to 2.5 corresponding to different sample sizes.

These simulation results indicate that our proposed B-spline based estimation procedure works fairly well. Specifically, it can be seen from Table 2.1 that the estimates of parametric coefficients  $\hat{\beta}_\delta$  and  $\hat{\beta}_\mu$  are approximately unbiased for all cases and all sample sizes. Their estimated standard errors increase as sample size increases and they are all very close to the corresponding empirical standard deviations. For the functional coefficients, the estimated function  $\hat{\varphi}(W_\mu)$  does not match the true function  $\varphi(W_\mu)$  perfectly when the sample size is 200, but it can capture the trend of  $\varphi(W_\mu)$ . However, when the sample size is 400 or more, the estimated covariate functions  $\hat{\varphi}(W_\mu)$  and  $\hat{\eta}(W_\delta)$  match the true functions  $\varphi(W_\mu)$  and  $\eta(W_\delta)$  very well, as can be seen from Figures 2.4 and 2.5.

## 2.6 An Application

In this section we illustrate the proposed threshold regression models with flexible covariate effects to a study of osteoporotic fractures.

The primary purpose of this study is to identify risk factors for osteoporotic fractures. Briefly, 9,704 primarily Caucasian women aged 65 or older were recruited in the Study of Osteoporotic Fractures (SOF) from four metropolitan areas in the United States, Baltimore, Pittsburgh, Minneapolis and Portland, from 1986 through 1989. They have been continuously tracked with clinical visits approximately every other year. In 1997, 662 African-American women were added into this cohort. The second visit was set to be the baseline in our analysis. The final analytic sample consists of 6,869 individuals after removing missing data, among which 3,075 had experienced any type of fracture by the end of the study. The predictor variables are listed in Table 2.2.



Table 2.2: Predictor variables in the analysis of osteoporotic fracture data set.

Variable Name	Label
<i>estrk</i>	Doctor ever told you have had a stroke?
<i>earth</i>	Doctor ever told you have arthritis?
<i>ediab</i>	Doctor ever told you have diabetes?
<i>lndpth</i>	Low near depth perception
<i>estcur</i>	Currently taking estrogen pills
<i>bmi</i>	Body mass index (kg/m <sup>2</sup> )
<i>thd</i>	Total hip BMD (g/cm <sup>2</sup> )

We consider *ediab*, *earth*, *estrk* in the initial status  $\ln(\delta)$  and *estcur* in the degradation rate  $\mu$  based on past literature. Since *bmi* would affect both  $\ln(\delta)$  and  $\mu$ , it is included in both of these link functions. Based on our preliminary data analysis, interaction effect of *age* and *thd* is significant in  $\ln(\delta)$ , and thus we suspect that the effect for *thd* may change with *age*. Therefore *age* is modeled as a varying coefficient of *thd*. Specifically, covariate effects are described in the following link functions

$$\left\{ \begin{array}{l} \ln(\delta) = \beta_{\delta 1} + \textit{estrk}\beta_{\delta 2} + \textit{earth}\beta_{\delta 3} + \textit{ediab}\beta_{\delta 4} + \textit{bmi}\beta_{\delta 5} \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad + \textit{lndpth}\beta_{\delta 6} + \eta(\textit{age})\textit{thd} \quad \cdot \\ \mu = \beta_{\mu 1} + \textit{estcur}\beta_{\mu 2} + \textit{bmi}\beta_{\mu 3} \end{array} \right. .$$

To estimate the unknown functional coefficient  $\eta(\textit{age})$ , we place the inner knots of the B-spline basis functions at 1st quartile, 2nd quartile and 3rd quartile of *age*.

The estimate  $\hat{\eta}(\textit{age})$  along with 95% point-wise confidence interval are displayed in

Figure 2.6. Table 2.3 presents the constant regression coefficient estimates.

As shown in Table 2.3, we note that the coefficient of estrogen pills is positive (0.00190) with a  $p$ -value of 0.01 in the drift parameter  $\mu$ , indicating that the bone health status of those individuals who are currently taking estrogen pills tends to decline slower as compared to that of those without taking estrogen pills. The bone health degradation of individuals with a higher body mass index tends to be faster (-0.00013,  $p$ -value=0.07). The results for the initial status  $\ln(\delta)$  indicate that individuals who ever had stroke or diabetes have a worse initial bone health status. Positive coefficient estimates of arthritis and body mass index in  $\ln(\delta)$  indicate that individuals who ever had arthritis or higher body mass index have a better initial bone health status.

The estimated functional coefficient  $\hat{\eta}(age)$  of the total hip BMD stays positive, suggesting that the effect of the total hip BMD is positive. That is, the higher total hip BMD, the lower is the chance of developing any type of fractures. The curve of age first increases from 65 to 73 and then starts to decrease for older individuals, indicating that the protective effect of total hip BMD on the risk of osteoporotic fractures reaches its maximum around age 73. Further studies may be needed to confirm this finding.

## 2.7 Summary

In this chapter, we have explored a general approach to incorporate covariates in TR models in which both the drift parameter  $\mu$  and the natural logarithm of the initial status parameter  $\delta$  are functions of given covariates with partially constant coefficients and partially varying coefficients. A spline approximation method was proposed to estimate varying coefficients. One main advantage of the proposed methodology is that it leaves functional forms of certain covariates completely unspecified. We have shown that our spline based estimator is consistent under the  $L^2$ -norm. Besides, the optimal rate of convergence under nonparametric regression setting can be achieved if the number of spline basis functions are chosen properly. Our simulation results suggest that the proposed estimation procedure performs well for practical situations. The analysis of the osteoporotic fracture data set indicates that the total hip BMD has protective effect on the risk of osteoporotic fractures and that the effect varies with age.

Table 2.1: Simulation results for the parametric components of link functions based on 100 replications.

Sample size: $n = 200$				
	True	Bias	Estimated SE	Empirical SE
Case 1	$\beta_{\delta}^T = (1, -0.5)$ $\beta_{\mu}^T = (-0.4, -0.3)$	(0.0492, 0.0047) (-0.0069, -0.0230)	(0.0772, 0.0401) (0.1441, 0.0710)	(0.0834, 0.0434) (0.1518, 0.0806)
Case 2	$\beta_{\delta}^T = (1, -0.6)$ $\beta_{\mu}^T = (-0.4, 1)$	(0.0616, 0.0029) (-0.0430, 0.0791)	(0.0775, 0.0408) (0.1543, 0.0957)	(0.0891, 0.0398) (0.1679, 0.0980)
Case 3	$\beta_{\delta}^T = (1, 0)$ $\beta_{\mu}^T = (-0.4, 0)$	(0.0501, 0.0054) (-0.0496, 0.0008)	(0.0769, 0.0361) (0.1526, 0.0662)	(0.0869, 0.03883) (0.1670, 0.0645)
Case 4	$\beta_{\delta}^T = (1, 0.4)$ $\beta_{\mu}^T = (-0.4, 0.4)$	(0.0702, 0.0025) (0.0153, 0.0295)	(0.0793, 0.0345) (0.1742, 0.0726)	(0.0830, 0.0359) (0.1637, 0.0781)
Case 5	$\beta_{\delta}^T = (1, 0.6)$ $\beta_{\mu}^T = (-0.4, -1)$	(0.0670, -0.0026) (-0.0660, -0.0756)	(0.0818, 0.0348) (0.1905, 0.0989)	(0.0908, 0.0388) (0.2124, 0.1121)
Sample size: $n = 400$				
	True	Bias	Estimated SE	Empirical SE
Case 1	$\beta_{\delta}^T = (1, -0.5)$ $\beta_{\mu}^T = (-0.4, -0.3)$	(0.0322, 0.0010) (-0.0161, -0.0022)	(0.0542, 0.0283) (0.1004, 0.0488)	(0.0551, 0.0298) (0.1025, 0.0565)
Case 2	$\beta_{\delta}^T = (1, -0.6)$ $\beta_{\mu}^T = (-0.4, 1)$	(0.0345, -0.0013) (-0.0231, 0.0278)	(0.0548, 0.0289) (0.1068, 0.0646)	(0.0576, 0.0310) (0.1105, 0.0656)
Case 3	$\beta_{\delta}^T = (1, 0)$ $\beta_{\mu}^T = (-0.4, 0)$	(0.0297, -0.0003) (-0.0083, -0.0033)	(0.0537, 0.0254) (0.1051, 0.0448)	(0.0535, 0.0263) (0.1073, 0.0483)
Case 4	$\beta_{\delta}^T = (1, 0.4)$ $\beta_{\mu}^T = (-0.4, 0.4)$	(0.0418, -0.0002) (-0.0337, 0.0062)	(0.0557, 0.0244) (0.1196, 0.0491)	(0.0634, 0.0240) (0.1099, 0.0530)
Case 5	$\beta_{\delta}^T = (1, 0.6)$ $\beta_{\mu}^T = (-0.4, -1)$	(0.0284, 0.0000) (-0.0163, -0.0299)	(0.0577, 0.0247) (0.1294, 0.0670)	(0.0601, 0.0235) (0.1306, 0.0677)
Sample size: $n = 600$				
	True	Bias	Estimated SE	Empirical SE
Case 1	$\beta_{\delta}^T = (1, -0.5)$ $\beta_{\mu}^T = (-0.4, -0.3)$	(0.0220, 0.0034) (-0.0151, -0.0040)	(0.0442, 0.0230) (0.0811, 0.0395)	(0.0446, 0.0253) (0.0825, 0.0391)
Case 2	$\beta_{\delta}^T = (1, -0.6)$ $\beta_{\mu}^T = (-0.4, 1)$	(0.0191, -0.0013) (-0.0014, 0.0145)	(0.0446, 0.0235) (0.0865, 0.0523)	(0.0430, 0.0241) (0.0834, 0.0537)
Case 3	$\beta_{\delta}^T = (1, 0)$ $\beta_{\mu}^T = (-0.4, 0)$	(0.0245, -0.0005) (-0.0157, 0.0024)	(0.0439, 0.0207) (0.0858, 0.0367)	(0.0458, 0.0214) (0.0839, 0.0414)
Case 4	$\beta_{\delta}^T = (1, 0.4)$ $\beta_{\mu}^T = (-0.4, 0.4)$	(0.0252, -0.0040) (-0.0297, 0.0043)	(0.0456, 0.0199) (0.0960, 0.0396)	(0.0416, 0.0196) (0.0878, 0.0405)
Case 5	$\beta_{\delta}^T = (1, 0.6)$ $\beta_{\mu}^T = (-0.4, -1)$	(0.0257, 0.0004) (-0.0076, -0.0244)	(0.0468, 0.0199) (0.1057, 0.0546)	(0.0461, 0.0203) (0.1021, 0.0613)

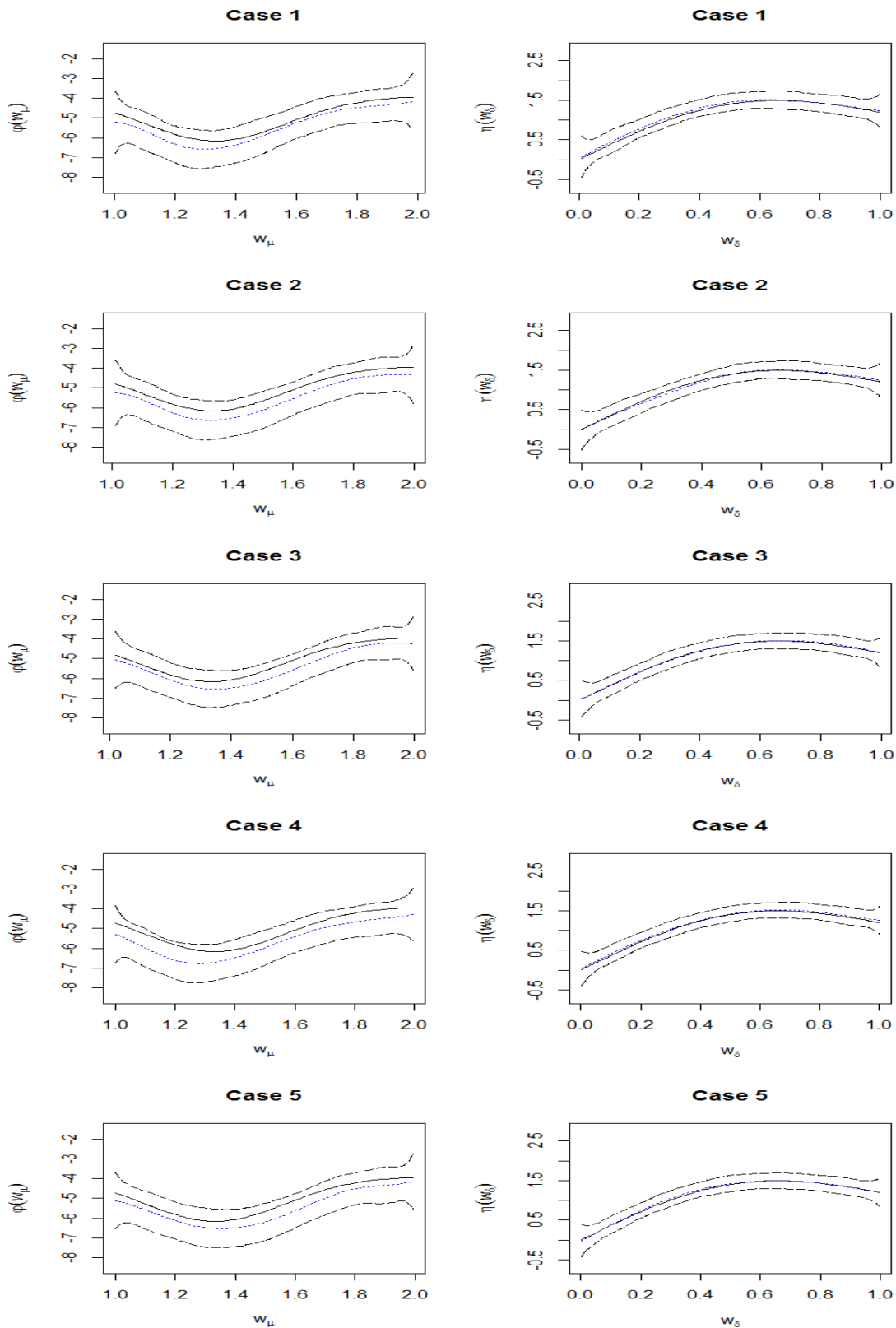


Figure 2.3: Simulation results based on sample size 200 and 100 runs for covariate functions  $\varphi(W_\mu)$  and  $\eta(W_\delta)$ . Solid line: true function; dotted line: estimated function; dashed line: 95% point-wise CI.

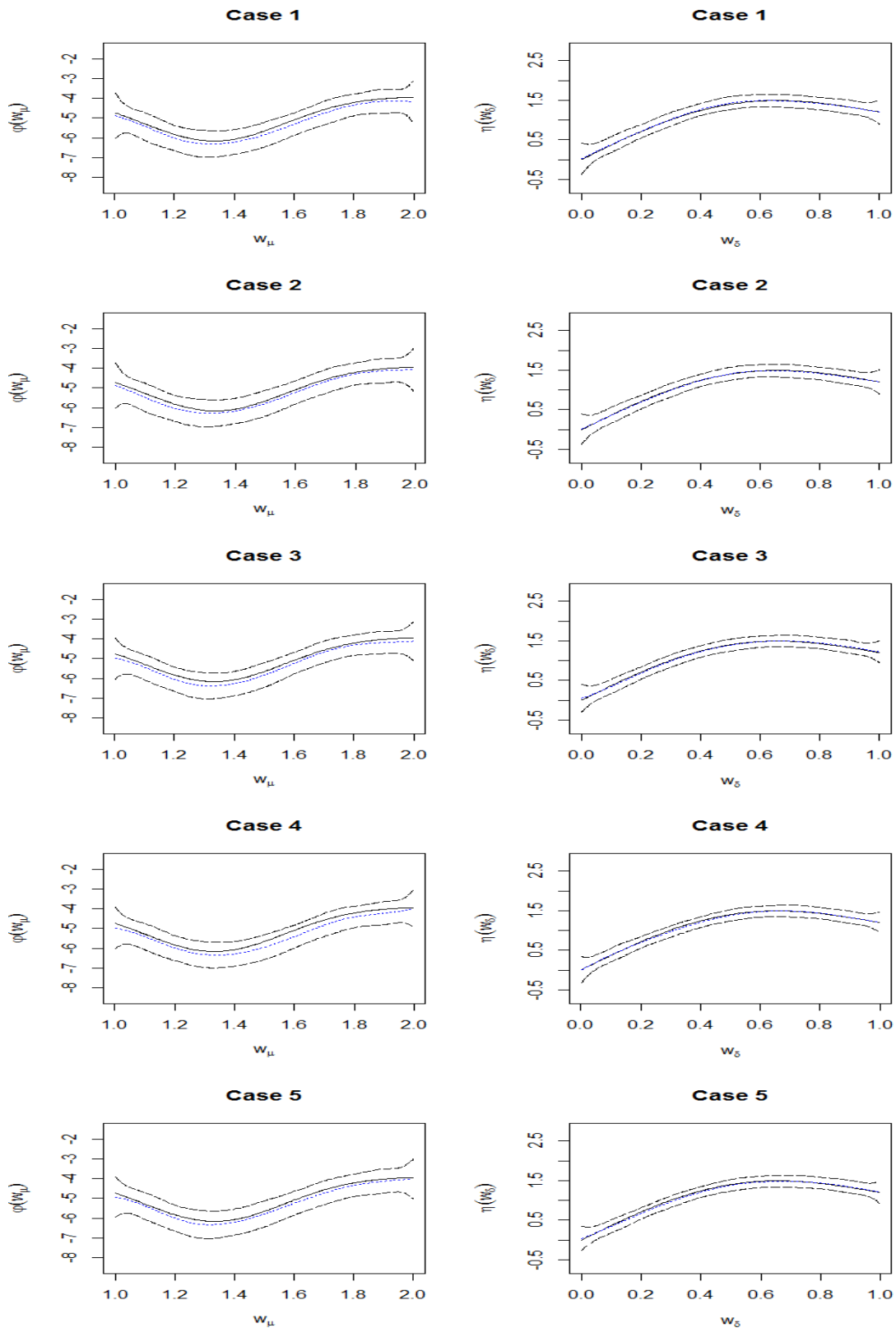


Figure 2.4: Simulation results based on sample size 400 and 100 runs for covariate functions  $\varphi(W_\mu)$  and  $\eta(W_\delta)$ . Solid line: true function; dotted line: estimated function; dashed line: 95% point-wise CI.

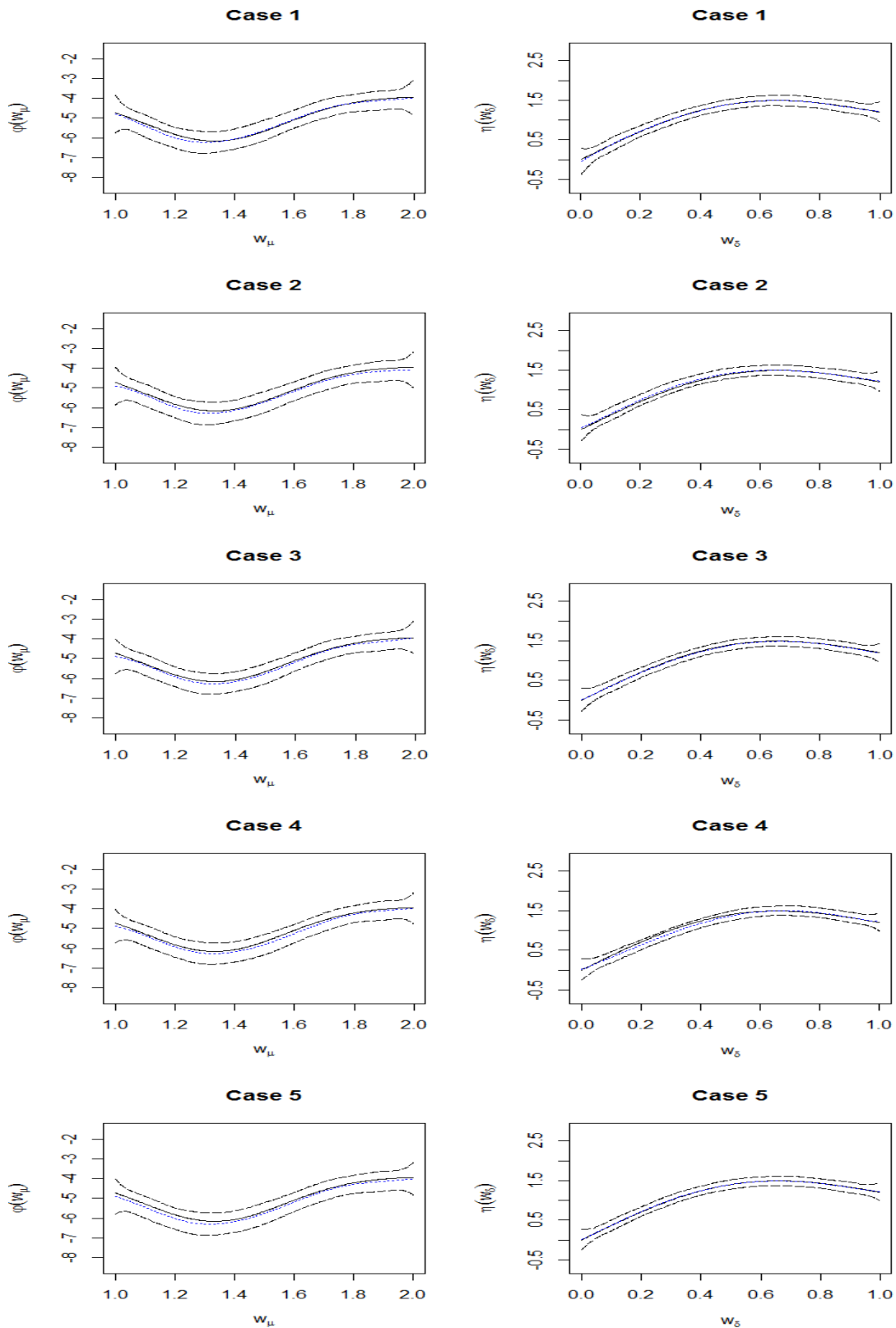


Figure 2.5: Simulation results based on sample size 600 and 100 runs for covariate functions  $\varphi(W_\mu)$  and  $\eta(W_\delta)$ . Solid line: true function; dotted line: estimated function; dashed line: 95% point-wise CI.

Table 2.3: Constant regression coefficient estimates in the osteoporotic fracture data set.

Regression for degradation rate $\mu$ :			
	Coefficient Estimate	Standard Error	$p$ -value
estcur	0.00190	0.00070	0.01
bmi	-0.00014	0.00007	0.07
intercept	0.00980	0.00191	< 0.01
Regression for initial status $\ln(\delta)$ :			
	Coefficient Estimate	Standard Error	$p$ -value
estrk	-0.33444	0.05946	< 0.01
earth	0.14756	0.02000	< 0.01
ediab	-0.55546	0.05865	< 0.01
bmi	0.03475	0.00279	< 0.01
lndpth	-0.23780	0.02785	< 0.01
intercept	2.26417	0.08035	< 0.01



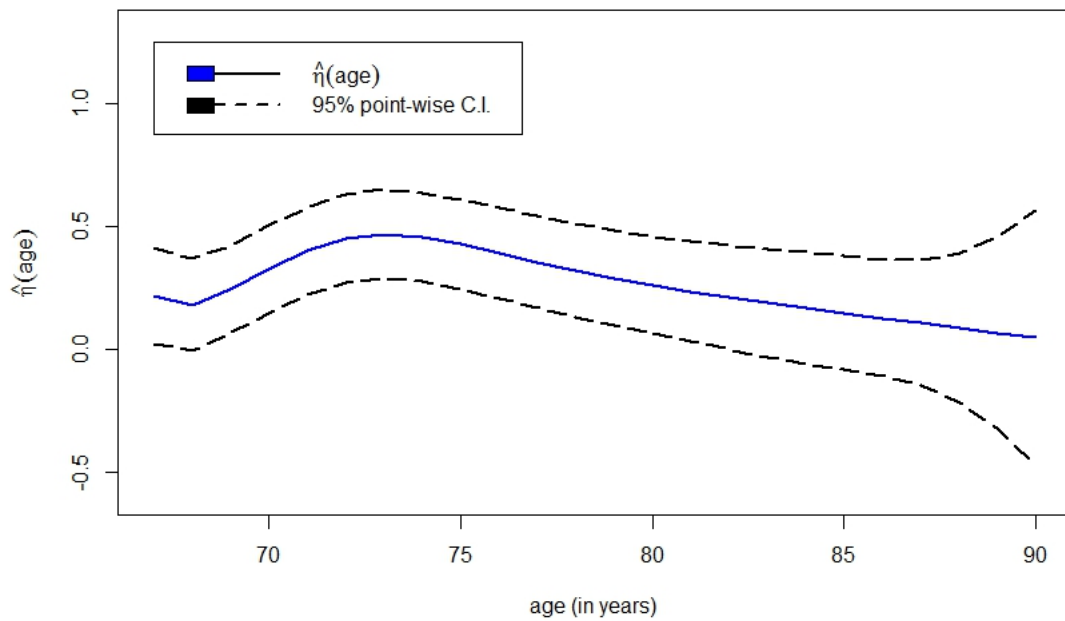


Figure 2.6: Estimated functional coefficient  $\hat{\eta}(age)$ .

## Chapter 3: Model Checking Techniques for TR Models

In this chapter, we discuss two supremum type goodness-of-fit techniques for regular TR models (with parametric link functions). TR models assume that (1) latent health status follows a Wiener process  $X(t)$ ; (2) initial health status  $X(0)$  and drift parameter  $\mu$  depend on covariates. Violation of these assumptions may seriously affect validity and efficiency of statistical inference. Thus, ensuring the correctness of model assumptions is an important issue. Since the application of TR models in survival analysis is relatively new, little has been done for model checking. It is imperative to develop model checking techniques for TR models.

This chapter is organized as follows. First, Section 3.1 describes goodness-of-fit procedures based on cumulative sums of martingale residuals. Specifically, we derive a test statistic and discuss how to calculate its  $p$ -value in Section 3.1.1. Section 3.1.2 discusses the consistency of the test. Second, we explore model checking techniques through censoring consistent residuals for TR models in Section 3.2. Section 3.2.1 discusses test statistic and how to evaluate the  $p$ -value using the bootstrap method. We show the consistency of the test in Section 3.2.2. Simulation studies are conducted in Section 3.3 to evaluate the performance of the proposed

tests based on martingale residuals and censoring consistent residuals. In Section 3.4, we apply both model checking procedures to a kidney dialysis data set. We conclude in Section 3.5 with a discussion.

### 3.1 Model Checking Based on Martingale Residuals

Lin *et al.* (1993) proposed model checking techniques for the Cox model using cumulative sums of martingale residuals. It is supremum-type test and can be used to check the functional form of a covariate, the form of the link function, the proportional hazards assumption, and the overall fit of the Cox model. Under model assumptions, the observed test process converges to a zero-mean Gaussian process which can be easily simulated through the Monte Carlo method.

This model checking approach has received a considerable attention in the literature recently. Lin *et al.* (2000) discussed this test procedure for recurrent event data, while Lin *et al.* (2002) explored this model checking method for generalized linear models and marginal models. Li and Sun (2003) adopted this idea and proposed a simulation-based supremum-type test for Aalen's multiplicative intensity counting process model. Furthermore, Li (2003) developed a nonparametric likelihood ratio goodness-of-fit test to measure the discrepancy between a parametric family and the observed data by a similar idea.

Next, we discuss in detail how this method can be applied to TR models with

parametric link functions.

### 3.1.1 Test Statistic

Recall for TR models, the first hitting time  $S$  of a latent Wiener process  $\{X(t)\}$  has the density function

$$f(s|\boldsymbol{\theta}) = \frac{\delta}{(2\pi s^2)^{1/2}} \exp\left[-\frac{(\delta + \mu s)^2}{2s}\right] \quad s > 0, \quad (3.1.1)$$

for  $-\infty < \mu < +\infty$  and  $\delta > 0$

and the cumulative distribution function

$$F(s|\boldsymbol{\theta}) = 1 - \Phi\left[\frac{\mu s + \delta}{(s)^{1/2}}\right] + \exp(-2\delta\mu)\Phi\left[\frac{\mu s - \delta}{(s)^{1/2}}\right], \quad (3.1.2)$$

where  $\boldsymbol{\theta} = (\delta, \mu)'$ . Let  $C$  denote the censoring time. For each subject, the observed data are denoted by  $(Y, \Delta, \mathbf{Z}'_{\delta}, \mathbf{Z}'_{\mu})$ , where  $Y = S \wedge C$ ,  $\Delta = I(S \leq C)$ ,  $\mathbf{Z}_{\delta} = (Z_{\delta 1}, \dots, Z_{\delta d_1})'$  and  $\mathbf{Z}_{\mu} = (Z_{\mu 1}, \dots, Z_{\mu d_2})'$ . We assume that the censoring time  $C$  is independent with  $S$  given the covariates  $\mathbf{Z}_{\delta}$  and  $\mathbf{Z}_{\mu}$ .

Suppose that the covariate effects on the drift parameter  $\mu$  and the nature logarithm of the initial health status  $\delta$  are linear, i.e.,

$$\begin{cases} \ln(\delta) = \mathbf{Z}'_{\delta}\boldsymbol{\beta}_{\delta} \\ \mu = \mathbf{Z}'_{\mu}\boldsymbol{\beta}_{\mu} \end{cases}. \quad (3.1.3)$$

where  $\boldsymbol{\beta}_\delta = (\beta_{\delta 1}, \dots, \beta_{\delta d_1})'$  and  $\boldsymbol{\beta}_\mu = (\beta_{\mu 1}, \dots, \beta_{\mu d_2})'$  are  $d_1$ -dimensional and  $d_2$ -dimensional vectors of regression parameters, respectively. Let  $\mathbf{Z} = (\mathbf{Z}'_\delta, \mathbf{Z}'_\mu)'$  and  $\boldsymbol{\beta} = (\boldsymbol{\beta}'_\delta, \boldsymbol{\beta}'_\mu)'$ . Suppose that the observed data consist of  $n$  independent replicates of  $(Y, \Delta, \mathbf{Z})'$ . Then the log-likelihood function for  $\boldsymbol{\beta}$  takes the form

$$\begin{aligned} l_n(\boldsymbol{\beta}) &= \sum_{i=1}^n l(\boldsymbol{\beta}; Y_i, \Delta_i, \mathbf{Z}_i) \\ &= \sum_{i=1}^n \left\{ \Delta_i \log f(Y_i | \mathbf{Z}_i, \boldsymbol{\beta}) + (1 - \Delta_i) \log [1 - F(C_i | \mathbf{Z}_i, \boldsymbol{\beta})] \right\}. \end{aligned} \quad (3.1.4)$$

The maximum likelihood estimator  $\hat{\boldsymbol{\beta}}$  is the solution to  $U(\boldsymbol{\beta}) = 0$  where  $U(\boldsymbol{\beta}) = \partial l_n(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$ . By Taylor expansion and some simple probabilistic arguments, one can easily get the following fact

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}} I^{-1}(\boldsymbol{\beta}_0) \sum_{i=1}^n \frac{\partial l(\boldsymbol{\beta}; Y_i, \Delta_i, \mathbf{Z}_i)}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + o_p(1) \quad (3.1.5)$$

where  $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_\delta^0, \boldsymbol{\beta}_\mu^0)'$  is the vector of true values of regression parameters and

$$I(\boldsymbol{\beta}_0) = E_{\boldsymbol{\beta}_0} \left[ \left( \frac{\partial l(\boldsymbol{\beta}; Y, \Delta, \mathbf{Z})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right) \left( \frac{\partial l(\boldsymbol{\beta}; Y, \Delta, \mathbf{Z})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right)' \right]$$

denotes the Fisher information matrix, which can be consistently estimated by

$$I(\hat{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^n \left( \frac{\partial l(\boldsymbol{\beta}; Y_i, \Delta_i, \mathbf{Z}_i)}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right) \left( \frac{\partial l(\boldsymbol{\beta}; Y_i, \Delta_i, \mathbf{Z}_i)}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right)'. \quad (3.1.6)$$

Let  $\Theta$  be the domain of  $\boldsymbol{\beta}$ . We assume that  $\Theta$  is a convex compact subset of Euclidean space  $\mathbb{R}^{d_1+d_2}$ . To check the overall fit of TR models with linear link

functions (3.1.3), one may consider testing the null hypothesis  $H_0$  that  $F(s|\mathbf{Z}, \boldsymbol{\beta})$  belongs to a parametric family

$$\mathcal{D} = \{F(s|\mathbf{Z}, \boldsymbol{\beta}) : \boldsymbol{\beta} \in \Theta\} ;$$

that is,

$$H_0 : F(s|\mathbf{Z}, \boldsymbol{\beta}) \in \mathcal{D}, \quad \text{v.s.} \quad H_1 : F(s|\mathbf{Z}, \boldsymbol{\beta}) = F_1(s|\mathbf{Z}, \boldsymbol{\beta}) \notin \mathcal{D}.$$

For  $i = 1, \dots, n$ , define the counting process  $N_i(t) = \Delta_i I(Y_i \leq t)$ . Then the intensity function of  $N_i(t)$  is given by  $R_i(t)\lambda_i(t|\mathbf{Z}, \boldsymbol{\beta})$  where  $R_i(t) = I(Y_i \geq t)$  is the at-risk function and  $\lambda(t|\mathbf{Z}_i, \boldsymbol{\beta})$  represents the hazard function of  $S$  for the  $i$ th subject. Based on the definition of the hazard function given in Chapter 1,

$$\lambda(t|\mathbf{Z}, \boldsymbol{\beta}) = \frac{f(t|\mathbf{Z}, \boldsymbol{\beta})}{1 - F(t|\mathbf{Z}, \boldsymbol{\beta})}.$$

We define

$$M_i(t) = N_i(t) - \int_0^t R_i(s)\lambda(s|\mathbf{Z}_i, \boldsymbol{\beta}_0)ds .$$

Then  $\{M_i(t) : 0 \leq t \leq \tau\}$  are martingales with respect to the filtration

$$\sigma\{N_i(s), \mathbf{Z}_i, R_i(s), 0 \leq s \leq t\}$$

where  $\tau$  represents the study duration, i.e.,  $0 \leq C \leq \tau$ .

Following Lin *et al.* (1993), we check the adequacy of the proposed model by using the following cumulative sums of  $\hat{M}_i(t)$  which is given by

$$\mathcal{F}(t, \mathbf{z}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z}) \hat{M}_i(t) \quad (3.1.7)$$

where  $I(\mathbf{Z}_i \leq \mathbf{z})$  means that each of the  $d_1 + d_2$  components of  $\mathbf{Z}_i$  is no larger than the corresponding components of  $\mathbf{z}$  and

$$\hat{M}_i(t) = N_i(t) - \int_0^t R_i(s) \lambda(s | \mathbf{Z}_i, \hat{\boldsymbol{\beta}}) ds .$$

By using the Taylor series expansion around  $\beta_0$  and equation (3.1.5), we have

$$\begin{aligned}
& \mathcal{F}(t, \mathbf{z}) \\
& \approx \frac{1}{\sqrt{n}} \sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z}) \left\{ N_i(t) - \int_0^t R_i(s) \left[ \lambda(s|\mathbf{Z}_i, \beta_0) \right. \right. \\
& \qquad \qquad \qquad \left. \left. + (\hat{\beta} - \beta_0) \frac{\partial \lambda(s|\mathbf{Z}_i, \beta)}{\partial \beta} \Big|_{\beta=\beta_0} \right] ds \right\} \\
& = \frac{1}{\sqrt{n}} \sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z}) \left\{ \left[ N_i(t) - \int_0^t R_i(s) \lambda(s|\mathbf{Z}_i, \beta_0) ds \right] \right. \\
& \qquad \qquad \qquad \left. - (\hat{\beta} - \beta_0) \int_0^t R_i(s) \frac{\partial \lambda(s|\mathbf{Z}_i, \beta)}{\partial \beta} \Big|_{\beta=\beta_0} ds \right\} \\
& = \frac{1}{\sqrt{n}} \sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z}) M_i(t) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ I(\mathbf{Z}_i \leq \mathbf{z}) \frac{1}{n} I^{-1}(\beta_0) \right. \\
& \qquad \qquad \qquad \times \left[ \sum_{j=1}^n \frac{\partial l(\beta; Y_j, \Delta_j, \mathbf{Z}_j)}{\partial \beta} \Big|_{\beta=\beta_0} \right] \int_0^t R_i(s) \frac{\partial \lambda(s|\mathbf{Z}_i, \beta)}{\partial \beta} \Big|_{\beta=\beta_0} ds \left. \right\} \\
& = \frac{1}{\sqrt{n}} \sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z}) M_i(t) - \frac{1}{\sqrt{n}} \left[ \sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z}) \frac{1}{n} I^{-1}(\beta_0) \right. \\
& \qquad \qquad \qquad \times \int_0^t R_i(s) \frac{\partial \lambda(s|\mathbf{Z}_i, \beta)}{\partial \beta} \Big|_{\beta=\beta_0} ds \left. \right] \left[ \sum_{j=1}^n \frac{\partial l(\beta; Y_j, \Delta_j, \mathbf{Z}_j)}{\partial \beta} \Big|_{\beta=\beta_0} \right] \\
& = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ I(\mathbf{Z}_i \leq \mathbf{z}) \left[ M_i(t) - \frac{1}{n} I^{-1}(\beta_0) \int_0^t R_i(s) \frac{\partial \lambda(s|\mathbf{Z}_i, \beta)}{\partial \beta} \Big|_{\beta=\beta_0} ds \right. \right. \\
& \qquad \qquad \qquad \left. \left. \times \left( \sum_{j=1}^n \frac{\partial l(\beta; Y_j, \Delta_j, \mathbf{Z}_j)}{\partial \beta} \Big|_{\beta=\beta_0} \right) \right] \right\} \\
& = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ I(\mathbf{Z}_i \leq \mathbf{z}) \left[ M_i(t) - \frac{1}{n} I^{-1}(\beta_0) \int_0^t R_i(s) \frac{\partial \lambda(s|\mathbf{Z}_i, \beta)}{\partial \beta} \Big|_{\beta=\beta_0} ds \cdot \mathbf{U}(\beta_0) \right] \right\}.
\end{aligned}$$



Let

$$\begin{aligned} & \tilde{\mathcal{F}}(t, \mathbf{z}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ I(\mathbf{Z}_i \leq \mathbf{z}) \left[ M_i(t) - \frac{1}{n} I^{-1}(\boldsymbol{\beta}_0) \int_0^t R_i(s) \frac{\partial \lambda(s|\mathbf{Z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} ds \cdot U(\boldsymbol{\beta}_0) \right] \right\}. \end{aligned}$$

Then the processes  $\mathcal{F}$  and  $\tilde{\mathcal{F}}$  are asymptotically equivalent. The score function  $U(\boldsymbol{\beta})$  has the following martingale representation (Kalbfleisch and Prentice 2002, p. 179):

$$U(\boldsymbol{\beta}_0) = \sum_{i=1}^n \int_0^\tau \left[ \frac{\partial \log \lambda(s|\mathbf{Z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right] dM_i(s).$$

This leads to

$$\begin{aligned} & \tilde{\mathcal{F}}(t, \mathbf{z}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ I(\mathbf{Z}_i \leq \mathbf{z}) \left[ M_i(t) - \frac{1}{n} I^{-1}(\boldsymbol{\beta}_0) \int_0^t R_i(s) \frac{\partial \lambda(s|\mathbf{Z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} ds \right. \right. \\ & \quad \left. \left. \times \sum_{i=1}^n \int_0^\tau \left[ \frac{\partial \log \lambda(s|\mathbf{Z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right] dM_i(s) \right] \right\}. \end{aligned} \tag{3.1.8}$$

It is difficult to evaluate the above distribution analytically, because the limiting process of  $\tilde{\mathcal{F}}$  does not have an independent increment structure. Lin *et al.* (1993) proposed to substitute  $N_i(t)G_i$  for  $M_i(t)$  where  $\{G_i; i = 1, \dots, n\}$  denotes a random sample of standard normal variables. The idea behind using  $N_i(t)G_i$  is that the variance function of  $M_i(t)$  is  $E(N_i(t))$  (Fleming and Harrington 1991, Theorem 2.5.3). Replacing  $\boldsymbol{\beta}_0$  and  $M_i(t)$  in (3.1.8) with  $\hat{\boldsymbol{\beta}}$  and  $\{N_i(t)G_i\}$ , respectively, we

have

$$\begin{aligned}
& \hat{\mathcal{F}}(t, \mathbf{z}) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ I(\mathbf{Z}_i \leq \mathbf{z}) \left[ N_i(t) G_i - \frac{1}{n} I^{-1}(\hat{\boldsymbol{\beta}}) \int_0^t R_i(s) \frac{\partial \lambda(s | \mathbf{Z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} ds \right. \right. \\
&\quad \left. \left. \times \sum_{i=1}^n \int_0^\tau \left[ \frac{\partial \log \lambda(s | \mathbf{Z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right] dN_i(s) G_i \right] \right\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ I(\mathbf{Z}_i \leq \mathbf{z}) \left[ \Delta_i I(Y_i \leq t) G_i - \frac{1}{n} I^{-1}(\hat{\boldsymbol{\beta}}) \int_0^t R_i(s) \frac{\partial \lambda(s | \mathbf{Z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} ds \right. \right. \\
&\quad \left. \left. \times \sum_{i=1}^n \Delta_i \left[ \frac{\partial \log \lambda(s | \mathbf{Z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right] G_i \right] \right\}. \quad (3.1.9)
\end{aligned}$$

Following the same line of Lin *et al.* (1993)'s proof, we know that the process  $\hat{\mathcal{F}}(t, \mathbf{z})$  given the observed data  $\{(Y_i, \Delta_i, \mathbf{Z}_i) : i = 1, \dots, n\}$  and the process  $\mathcal{F}(t, \mathbf{z})$  converge in distribution to the same zero-mean Gaussian random field.

To approximate the distribution of  $\mathcal{F}(t, \mathbf{z})$ , one can obtain a large number of realizations from  $\hat{\mathcal{F}}(t, \mathbf{z})$  by repeatedly generating the standard normal random sample  $\{G_i : i = 1, \dots, n\}$  while fixing  $\{(Y_i, \Delta_i, \mathbf{Z}_i) : i = 1, \dots, n\}$  at their observed values. More formally, we apply the supremum test statistic  $\sup_{0 \leq t \leq \tau, \mathbf{z}} |\mathcal{F}(t, \mathbf{z})|$ , for which the  $p$ -value can be obtained by comparing the observed value of  $\sup_{0 \leq t \leq \tau, \mathbf{z}} |\tilde{\mathcal{F}}(t, \mathbf{z})|$  to a large number of realizations from  $\sup_{0 \leq t \leq \tau, \mathbf{z}} |\hat{\mathcal{F}}(t, \mathbf{z})|$ . Let  $L$  denote the number of independent samples of  $\{G_i; i = 1, \dots, n\}$ . Let  $p_n(\alpha)$  be the  $100(1 - \alpha)$ th percentile of

$$\sup_{0 \leq t \leq \tau, \mathbf{z}} |\hat{\mathcal{F}}_1(t, \mathbf{z})|, \dots, \sup_{0 \leq t \leq \tau, \mathbf{z}} |\hat{\mathcal{F}}_L(t, \mathbf{z})|,$$

where each  $\hat{\mathcal{F}}_l(t, \mathbf{z})$  is calculated from (3.1.9) for the  $l$ -th sample set  $\{G_i; i = 1, \dots, n\}$ ,  $l = 1, \dots, L$ . The null hypothesis  $H_0$  is rejected if

$$\sup_{0 \leq t \leq \tau, \mathbf{z}} |\tilde{\mathcal{F}}(t, \mathbf{z})| > p_n(\alpha) .$$

### 3.1.2 Consistency

Under TR models with link functions (3.1.3), we show that the asymptotic power of the test discussed in Section 3.1.1 under the alternative

$$H_1 : F(s|\mathbf{Z}, \boldsymbol{\beta}) = F_1(s|\mathbf{Z}, \boldsymbol{\beta}) \notin \mathcal{D} \tag{3.1.10}$$

(i.e., the probability of rejecting the null hypothesis  $H_0$  under  $H_1$ ) is 1.

Let  $M_n(\boldsymbol{\beta}) = \frac{1}{n}l_n(\boldsymbol{\beta})$  and let

$$M(\boldsymbol{\beta}) = E_{F_1} [M_n(\boldsymbol{\beta})] = E_{F_1} [l(\boldsymbol{\beta}; Y_i, \Delta_i, \mathbf{Z}_i)] .$$

To derive the asymptotic power of the test, we require the following assumptions.

#### Assumption 3.1.1.

(1)  $\sup_{\boldsymbol{\beta} \in \Theta} |M_n(\boldsymbol{\beta}) - M(\boldsymbol{\beta})| \xrightarrow{p} 0$ .

(2) *There exists  $\tilde{\boldsymbol{\beta}}$  such that  $\sup_{\boldsymbol{\beta}: \|\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}\| \geq \varepsilon} M(\boldsymbol{\beta}) \leq M(\tilde{\boldsymbol{\beta}})$  for any  $\varepsilon > 0$ .*

**Theorem 3.1.1.** *Suppose (1) and (2) in Assumption 3.1.1 hold. Then under the fixed alternative hypothesis  $H_1$ ,*

$$P\left(\sup_{0 \leq t \leq \tau, \mathbf{z}} |\tilde{\mathcal{F}}(t, \mathbf{z})| > p_n(\alpha)\right) \rightarrow 1, \text{ as } n \rightarrow \infty .$$

*Proof.* By law of large numbers, we have

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sup_{0 \leq t \leq \tau, \mathbf{z}} |\hat{\mathcal{F}}(t, \mathbf{z})| = 0 .$$

Hence, we only need to show that  $\sup_{0 \leq t \leq \tau, \mathbf{z}} \frac{1}{\sqrt{n}} |\mathcal{F}(t, \mathbf{z})|$  does not converge to zero under the alternative hypothesis  $H_1$  since  $\mathcal{F}(t, \mathbf{z})$  and  $\tilde{\mathcal{F}}(t, \mathbf{z})$  are asymptotically equivalent. Assume that the cumulative distribution function of  $\mathbf{Z}$  is  $F_z(\mathbf{Z})$ . Let  $\lambda_{H_1}(t|\mathbf{Z}, \beta_1)$  be the hazard function under the alternative hypothesis  $H_1$ . Suppose that  $H_1$  holds. We note that

$$\hat{\beta} \xrightarrow{P} \tilde{\beta}$$

by Lemma 2.4.8. Then,

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \mathcal{F}(t, \mathbf{z}) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z}) \hat{M}_i(t) \\
&= \frac{1}{n} \sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z}) \left[ N_i(t) - \int_0^t R_i(s) \lambda(s | \mathbf{Z}_i, \hat{\boldsymbol{\beta}}) ds \right] \\
&= \frac{1}{n} \sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z}) \left[ N_i(t) - \int_0^t R_i(s) \lambda(s | \mathbf{Z}_i, \tilde{\boldsymbol{\beta}}) ds \right] + o_p(1) \\
&\xrightarrow{a.s.} E_{F_z} \left\{ I(\mathbf{Z}_i \leq \mathbf{z}) E_{H_1} \left[ \left( N_i(t) - \int_0^t R_i(s) \lambda(s | \mathbf{Z}_i, \tilde{\boldsymbol{\beta}}) ds \right) | \mathbf{Z}_i \right] \right\} \\
&= \int_{\mathbf{Z}_i \leq \mathbf{z}} \left\{ E_{H_1}(N_i(t) | \mathbf{Z}_i) - E_{H_1} \left[ \int_0^t R_i(s) \lambda(s | \mathbf{Z}_i, \tilde{\boldsymbol{\beta}}) ds | \mathbf{Z}_i \right] \right\} dF_z(\mathbf{Z}_i) \\
&= \int_{\mathbf{Z}_i \leq \mathbf{z}} \left\{ E_{H_1} \left[ \int_0^t R_i(s) \lambda_{H_1}(s | \mathbf{Z}_i, \boldsymbol{\beta}_1) ds | \mathbf{Z}_i \right] \right. \\
&\quad \left. - \int_0^t E_{H_1} [R_i(s) | \mathbf{Z}_i] \lambda(s | \mathbf{Z}_i, \tilde{\boldsymbol{\beta}}) ds \right\} dF_z(\mathbf{Z}_i) \\
&= \int_{\mathbf{Z}_i \leq \mathbf{z}} \left\{ \int_0^t E_{H_1} [R_i(s) | \mathbf{Z}_i] \lambda_{H_1}(s | \mathbf{Z}_i, \boldsymbol{\beta}_1) ds \right. \\
&\quad \left. - \int_0^t E_{H_1} [R_i(s) | \mathbf{Z}_i] \lambda(s | \mathbf{Z}_i, \tilde{\boldsymbol{\beta}}) ds \right\} dF_z(\mathbf{Z}_i) \\
&= \int_{\mathbf{Z}_i \leq \mathbf{z}} \left\{ \int_0^t E_{H_1} [R_i(s) | \mathbf{Z}_i] [\lambda_{H_1}(s | \mathbf{Z}_i, \boldsymbol{\beta}_1) - \lambda(s | \mathbf{Z}_i, \tilde{\boldsymbol{\beta}})] ds \right\} dF_z(\mathbf{Z}_i) .
\end{aligned}$$

The above expression is not zero for some  $t, \mathbf{z}$  (i.e.,  $\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq \tau, \mathbf{z}} n^{-1/2} |\mathcal{F}(t, \mathbf{z})| \neq 0$ ) . □

### 3.2 Model Checking Based on Censoring Consistent Residuals

In this section, we discuss the goodness-of-fit test based on censoring consistent residuals for checking the validity of TR models with linear link functions.

León and Tsai (2004) proposed censoring consistent residuals based test statistics for identifying the functional forms of covariates in the Cox model. In contrast to the martingale residuals, which can be interpreted as the difference between the observed and expected number of events for the  $i$ th subject, the censoring consistent residuals can be interpreted, asymptotically, as the difference between the observed and expected survival time for the  $i$ th subject (León and Tsai, 2004).

### 3.2.1 Test Statistic

Consider the following hypotheses

$$H_0 : F(s|\mathbf{Z}, \boldsymbol{\beta}) \in \mathcal{D}, \quad \text{v.s.} \quad H_1 : F(s|\mathbf{Z}, \boldsymbol{\beta}) = F_1(s|\mathbf{Z}, \boldsymbol{\beta}) \notin \mathcal{D}$$

where  $F(s|\mathbf{Z}, \boldsymbol{\beta})$  denotes the cumulative distribution function of the first hitting time  $S$  and  $\mathcal{D}$  is the parametric family defined in Section 3.1.1.

We follow the notation in Section 3.1. The expectations of observed survival times  $Y_i$  are not the same as those of true survival times  $S_i$  (i.e.,  $E(Y_i|\mathbf{Z}_i) \neq E(S_i|\mathbf{Z}_i)$ ) because of censoring. Fan and Gijbels (1996) discussed a censoring unbiased transformation such that the expectations of the transformed observed survival times are equivalent to those of the survival times without censoring. We now briefly explain their approach.

Let  $G(t) = P(C > t)$  be the survival function of censoring time. Define

$$\theta = \min_{\{i:\Delta_i=1\}} \frac{\int_0^{Y_i} [1/G(t)] dt - Y_i}{Y_i/G(Y_i) - \int_0^{Y_i} [1/G(t)] dt} .$$

Let  $\varphi(\cdot)$  denote the transformation function. Then the transformed observed survival time is given by

$$\begin{aligned} \varphi(Y_i) &= \begin{cases} \varphi_1(Y_i) & \text{if uncensored} \\ \varphi_2(Y_i) & \text{if censored} \end{cases} \\ &= \Delta_i \varphi_1(Y_i) + (1 - \Delta_i) \varphi_2(Y_i) , \end{aligned} \tag{3.2.1}$$

where

$$\varphi_1(t) = (1 + \theta) \int_0^t \frac{1}{G(s)} ds - \theta \frac{t}{G(t)}$$

and

$$\varphi_2(t) = (1 + \theta) \int_0^t \frac{1}{G(s)} ds .$$

Then we have

$$E(\varphi(Y_i) | \mathbf{Z}_i) = E(S_i | \mathbf{Z}_i) .$$

One can estimate  $G(t)$  by the Nelson-Aalen estimator  $\hat{G}(t)$ . Let  $\hat{\varphi}(\cdot)$  denote the estimated transformation function upon replacing  $G(\cdot)$  in (3.2.1) by  $\hat{G}(\cdot)$ . Let  $m(\mathbf{Z}_i) = E(S_i | \mathbf{Z}_i)$ . Then  $m(\mathbf{Z}_i)$  can be estimated by

$$\hat{m}(\mathbf{Z}_i) = \int_0^\infty \left( 1 - F(s | \hat{\beta}, \mathbf{Z}_i) \right) ds .$$

León and Tsai (2004) referred to

$$\hat{R}_i = \hat{\varphi}(Y_i) - \hat{m}(\mathbf{Z}_i)$$

as the censoring consistent residual.

To check the overall fit of TR models with linear link functions, we adopt the following cumulative sums of censoring consistent residuals

$$\mathcal{R}(\mathbf{z}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z}) \left[ \hat{\varphi}(Y_i) - \hat{m}(\mathbf{Z}_i) \right].$$

We approximate the above process  $\mathcal{R}(\mathbf{z})$  by the following bootstrap method as discussed in León and Tsai (2004). Let  $(Y_i^*, \Delta_i^*, \mathbf{Z}_i^*)$  denote the bootstrap data. Specifically, for  $i = 1, \dots, n$ ,

1. Let  $\mathbf{Z}_i^* = \mathbf{Z}_i$ .
2. Generate  $u_i^* \sim U(0, 1)$ .
3. If  $u_i^* \leq F(\max_{j \in \{1, \dots, n\}} \{Y_j\} | \hat{\boldsymbol{\beta}}, \mathbf{Z}_i)$ , then  $S_i^* = \min_{j \in \{1, \dots, n\}} \{S_j : F(S_j | \hat{\boldsymbol{\beta}}, \mathbf{Z}_i)\}$ ; otherwise,  $S_i^* = \max_{j \in \{1, \dots, n\}} \{Y_j\}$ .
4. If  $\Delta_i = 0$ , let  $C_i^* = Y_i$ ; if  $\Delta_i = 1$ , by analogy to step 2, generate  $C_i^*$  from  $F_Y(t) = [\hat{G}(t) - \hat{G}(Y_i)] / \hat{G}(Y_i)$  for  $t \geq Y_i$ .

Then  $\Delta_i^* = I(S_i^* \leq C_i^*)$  and  $Y_i^* = S_i^* \wedge C_i^*$ . We can approximate the distribution of



$\mathcal{R}(\mathbf{z})$  by generating  $L$  bootstrap realizations

$$\mathcal{R}_k^*(\mathbf{z}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n 1(\mathbf{Z}_i^* \leq \mathbf{z}) \left[ \hat{\varphi}(Y_i^*) - \hat{m}(\mathbf{Z}_i^*) \right], \quad k = 1, \dots, L .$$

We apply the supremum-type test statistic  $\sup_{\mathbf{z}} |\mathcal{R}(\mathbf{z})|$  to check the adequacy of TR models with linear link functions. Let  $p^*(\alpha)$  denote the  $100(1-\alpha)$ th percentile of

$$\sup_{\mathbf{z}} |\mathcal{R}_1^*(\mathbf{z})|, \dots, \sup_{\mathbf{z}} |\mathcal{R}_L^*(\mathbf{z})| .$$

The null hypothesis  $H_0$  is rejected if

$$\sup_{\mathbf{z}} |\mathcal{R}(\mathbf{z})| \geq p^*(\alpha) .$$

### 3.2.2 Consistency

In this section, we investigate the asymptotic power of the proposed test above.

**Theorem 3.2.1.** *Suppose Assumption 3.1.1 holds. Then under the fixed alternative hypothesis  $H_1$ ,*

$$\mathbb{P} \left( \sup_{\mathbf{z}} |\mathcal{R}(\mathbf{z})| \geq p^*(\alpha) \right) \rightarrow 1, \quad \text{as } n \rightarrow \infty .$$

*Proof.* It can be shown that

$$\frac{1}{\sqrt{n}} \mathcal{R}_k^*(\mathbf{z}) \xrightarrow{a.s.} 0, \quad \text{by law of large numbers .}$$

To establish the conclusion of Theorem 3.2.1, we only need to show that

$$\frac{1}{\sqrt{n}}\mathcal{R}(\mathbf{z}) \xrightarrow{p} a \neq 0$$

under  $H_1$ . We have

$$\hat{\beta} \xrightarrow{p} \tilde{\beta}$$

under Assumption 3.1.1 by Lemma 2.4.8. Then,

$$\begin{aligned} \frac{1}{\sqrt{n}}\mathcal{R}(\mathbf{z}) &= \frac{1}{n} \sum_{i=1}^n I[\mathbf{Z}_i \leq \mathbf{z}] \left\{ \hat{\varphi}(Y_i) - \hat{m}(\mathbf{Z}_i) \right\} \\ &= \frac{1}{n} \sum_{i=1}^n I[\mathbf{Z}_i \leq \mathbf{z}] \left\{ \hat{\varphi}(Y_i) - \tilde{m}(\mathbf{Z}_i) \right\} + o_p(1), \end{aligned}$$

where  $\tilde{m}(\mathbf{Z}_i) = E_{\tilde{\beta}}(S_i|\mathbf{Z}_i)$ . This leads to

$$\begin{aligned} \frac{1}{\sqrt{n}}\mathcal{R}(\mathbf{z}) &\xrightarrow{a.s.} E \left\{ I[\mathbf{Z}_i \leq \mathbf{z}] E \left[ \hat{\varphi}(Y_i) - \tilde{m}(\mathbf{Z}_i) | \mathbf{Z}_i \right] \right\} \\ &= \int_{\mathbf{Z}_i \leq \mathbf{z}} E \left[ \hat{\varphi}(Y_i) - \tilde{m}(\mathbf{Z}_i) | \mathbf{Z}_i \right] dF_z(\mathbf{Z}_i) \\ &= \int_{\mathbf{Z}_i \leq \mathbf{z}} E_{H_1}[S_i | \mathbf{Z}_i] dF_z(\mathbf{Z}_i) - \int_{\mathbf{Z}_i \leq \mathbf{z}} \tilde{m}(\mathbf{Z}_i) dF_z(\mathbf{Z}_i) \\ &= \int_{\mathbf{Z}_i \leq \mathbf{z}} \int_0^\infty (1 - F_1(S_i | \mathbf{Z}_i)) dS_i dF_z(\mathbf{Z}_i) - \\ &\quad \int_{\mathbf{Z}_i \leq \mathbf{z}} \int_0^\infty (1 - F(S_i | \mathbf{Z}_i, \tilde{\beta})) dS_i dF_z(\mathbf{Z}_i) \\ &= \int_{\mathbf{Z}_i \leq \mathbf{z}} \int_0^\infty (F_1(S_i | \mathbf{Z}_i) - F(S_i | \mathbf{Z}_i, \tilde{\beta})) dS_i dF_z(\mathbf{Z}_i). \end{aligned}$$

The above expression is not zero for some  $\mathbf{z}$ . □

### 3.3 Simulation Study

To assess the performance of the tests based on martingale residuals and censoring consistent residuals, we conduct simulation studies under various situations. We simulate an event time  $S$  (first hitting time) using the method described in Section 2.5. Specifically, we consider the following four different cases:

**Case 1** (Covariate-dependent initial status with correct link functions in the estimation):

- Event time  $S$  is generated from the following link functions

$$\begin{cases} \ln(\delta) = \beta_{\delta 1} + \beta_{\delta 2} Z_{\delta 2} \\ \mu = \beta_{\mu 1} + \beta_{\mu 2} Z_{\mu 2} , \end{cases}$$

where  $(\beta_{\delta 1}, \beta_{\delta 2})' = (1, 1.8)'$ ,  $(\beta_{\mu 1}, \beta_{\mu 2})' = (-1.3, -1.6)'$ , and  $Z_{\delta 2} = Z_{\mu 2} \sim \text{Uniform}(0, 1)$ .

- We use the following correctly specified link functions in estimation

$$\begin{cases} \ln(\delta) = \beta_{\delta 1} + \beta_{\delta 2} Z_{\delta 2} \\ \mu = \beta_{\mu 1} + \beta_{\mu 2} Z_{\mu 2} . \end{cases}$$

**Case 2** (Covariate-dependent initial status with wrong link functions in the estimation):

- Event time  $S$  is generated from the following link functions

$$\begin{cases} \ln(\delta) = \beta_{\delta 1} + \beta_{\delta 2} Z_{\delta 2} \\ \mu = \beta_{\mu 1} + \beta_{\mu 2} Z_{\mu 2} + \beta_{\mu 3} Z_{\mu 3} , \end{cases}$$

where  $(\beta_{\delta 1}, \beta_{\delta 2})' = (1, 1.8)'$ ,  $(\beta_{\mu 1}, \beta_{\mu 2}, \beta_{\mu 3})' = (-1.3, -1.6, -1)'$ ,  $Z_{\delta 2} = Z_{\mu 2} \sim \text{Uniform}(0, 1)$ , and  $Z_{\mu 3} \sim N(1, 1.4)$ .

- We use the following wrong link functions in the estimation

$$\begin{cases} \ln(\delta) = \beta_{\delta 1} + \beta_{\delta 2} Z_{\delta 2} \\ \mu = \beta_{\mu 1} + \beta_{\mu 2} Z_{\mu 2} . \end{cases}$$

**Case 3** (Covariate-independent initial status with correct link functions in the estimation):

- Event time  $S$  is generated from the following link functions

$$\begin{cases} \ln(\delta) = \beta_{\delta 1} \\ \mu = \beta_{\mu 1} + \beta_{\mu 2} \ln(Z_{\mu 2}) , \end{cases}$$

where  $\beta_{\delta 1} = 2.3$ ,  $(\beta_{\mu 1}, \beta_{\mu 2})' = (-1.3, 1.6)'$ , and  $Z_{\mu 2} \sim \text{Uniform}(0, 1)$ .

- We use the following correctly specified link functions in the estimation

$$\begin{cases} \ln(\delta) = \beta_{\delta 1} \\ \mu = \beta_{\mu 1} + \beta_{\mu 2} \ln(Z_{\mu 2}) . \end{cases}$$

**Case 4** (Covariate-independent initial status with wrong link functions in the estimation):

- Event time  $S$  is generated from the following link functions

$$\begin{cases} \ln(\delta) = \beta_{\delta 1} \\ \mu = \beta_{\mu 1} + \beta_{\mu 2} \ln(Z_{\mu 2}) , \end{cases}$$

where  $\beta_{\delta 1} = 2.3$ ,  $(\beta_{\mu 1}, \beta_{\mu 2})' = (-1.3, 1.6)'$ , and  $Z_{\mu 2} \sim \text{Uniform}(0, 1)$ .

- We use the following wrong link functions in the estimation

$$\begin{cases} \ln(\delta) = \beta_{\delta 1} \\ \mu = \beta_{\mu 1} + \beta_{\mu 2} Z_{\mu 2} . \end{cases}$$

The censoring time  $C$  follows a chi-squared distribution with 7 degrees of freedom. The censoring rates for Case 1 and Case 2 are around 15%. The censoring rates for Case 3 and Case 4 are around 27%. Simulation results are summarized in Table 3.1 and Table 3.2. For each simulation case, we simulate 100 data sets with sample size  $n = 200$ . The  $p$ -values are calculated from  $L = 100$  realizations of  $\hat{\mathcal{F}}(t, \mathbf{z})$  and  $\mathcal{R}^*(\mathbf{z})$ . The significance level  $\alpha$  is 0.05.

Table 3.1: Empirical sizes of test statistics at the  $\alpha = 0.05$  significance level.

Using correct link functions in the estimation		
	Case 1	Case 3
Martingale residual test	0.04	0.07
Censoring consistent residual test	0	0.04

Table 3.2: Empirical powers of test statistics at the  $\alpha = 0.05$  significance level.

Using wrong link functions in the estimation		
	Case 2	Case 4
Martingale residual test	0.95	0.90
Censoring consistent residual test	0.63	0.73

Results in Table 3.1 indicate that both test procedures maintain the null size when link functions are correctly specified. Results in Table 3.2 suggest that the rejection powers are reasonable for both tests if the model is misspecified.

### 3.4 An Application

In this section, we apply our proposed model checking methods to the kidney dialysis data set in Nahman *et al.* (1992). The data contains 119 patients with renal insufficiency in which 43 patients used a surgically placed catheter (group 1) and 76 patients utilized a percutaneous placement of their catheter (group 2). Time (in months) to first exit-site infection in patients was recorded. Of the 43 patients in group 1, 28 were censored; while 65 among 76 patients in group 2 were censored.

The proportional hazards assumption does not hold for this data set based previous studies (Klein and Moeschberger, 2003; Xiao *et al.*, 2012). Xiao *et al.* (2012) also noted that regular TR model fits this data set fairly well since the predicted survival curves from the TR model match the Kaplan-Meier curves. We apply our model checking procedures to test the overall fit of the TR model using this data set. The test based on cumulative sums of martingale residuals yields a  $p$ -value of 0.319 using  $L = 1,000$  realizations from  $\hat{\mathcal{F}}(t, \mathbf{z})$  under the null hypothesis. We obtain a  $p$ -value of 0.381 from the test based on censoring consistent residuals using  $L = 1,000$  bootstrap samples. Therefore,  $p$ -values from both tests show no evidence against the TR model. For the reader's convenience, we also report the estimated parameters in Table 3.3 and plot the predicted survival curves from the TR model along with Kaplan-Meier survival curves for the two groups in Figure 3.1. We code the *group* variable as  $group = 1$  if percutaneous placed catheter, 0 if surgically placed catheter, in the analysis.

Table 3.3: Estimated regression coefficients in the analysis of kidney dialysis data set.

Regression for degradation rate $\mu$ :			
	Coefficient Estimate	Standard Error	$p$ -value
group	0.6377	0.1280	< 0.01
intercept	-0.0959	0.0765	0.21
Regression for $\ln(\delta)$ :			
	Coefficient Estimate	Standard Error	$p$ -value
group	-1.0731	0.1891	< 0.01
intercept	1.4113	0.1434	< 0.01

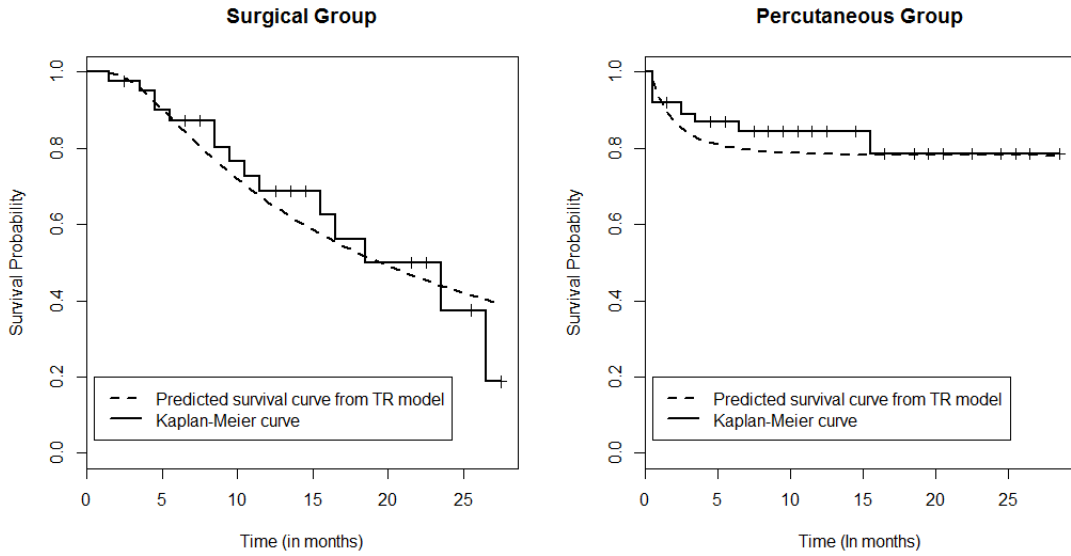


Figure 3.1: Predicted survival curve from TR model versus Kaplan-Meier curve for the kidney dialysis data set.

### 3.5 Summary

Lin *et al.* (1993) proposed cumulative sums of martingale residuals for the model checking of Cox model. León and Tsai (2004) introduced functional form diagnostics for the Cox model based on censoring consistent residuals. We have adopted their ideas and proposed two model checking procedures to check the overall fit of the TR model in this dissertation. We have showed that both tests are asymptotically consistent when the model is misspecified. Simulation studies indicate that the performances of the two tests are quite satisfactory. We illustrate the practical application of the proposed two tests using data from a kidney dialysis.



## Chapter 4: Future Study

In this chapter, we discuss several potential directions for future research.

In Chapter 2, we investigated TR models with flexible covariate effects for right censored data. This modeling approach can be easily extended to accommodate left and interval censored or truncated observations. We explored consistency and convergence rate of the spline based estimator for TR models with flexible covariate effects. It would be useful to derive the asymptotic distribution of the spline based estimator. One can use its asymptotic distribution to construct simultaneous variability bands. The interested reader is referred to Strawderman and Tsiatis (1996) and Huang *et al.* (2004) for further information.

Knot selection is a main research direction in regression splines. The performance of spline smoothing depends on the location of the knots and number of knots, especially for curves with varying shapes. It would be meaningful to develop knot selection techniques for TR models with flexible covariate effects to obtain an efficient estimator. To conduct further research in this direction, one may consider the cross-validation method. For more details we refer the reader to Rice and Sil-

verman (1991) and Yu *et al.* (2009).

We have noticed that local polynomial approach is another popular tool in statistical models with varying coefficients. It is of interest to compare the performance between the proposed spline based estimator and the estimator based on local polynomial regression under different circumstances.

We discussed model checking techniques for TR models with linear link functions in Chapter 3. Extending these model checking approaches for TR models with flexible covariate effects is worth exploration.

## Chapter A: Appendix

### A.1 R Functions for Model Checking Based on Martingale Residuals

```
#####  
# Function name: Ftx #  
#-----#  
# Input variables: #  
# delta: vector of length n which records estimated delta for all #  
# subjects #  
# mu: vector of length n which records estimated mu for all subjects #  
# cov: n by (d_1+d_2) matrix which records values of all covariates #  
# cov_com: matrix with (d_1+d_2+1) columns which records all #  
# different combinations of values of covariates and event #  
# time; covariates order is same as cov and the last column #  
# is for event time #  
# dt: vector of length n which records the event times for all #  
# subjects #  
# ft: numerical valued (0,1) vector which records event indicators #  
# for all subjects #  
#-----#  
# Output of Ftx: #  
# Return the maximum of absolute values of the true process #  
#####
```

```
Ftx= function(delta, mu, cov, cov_com, dt, ft) {
```

```
  ind= length(dt)  
  ncounti= rep(0, ind)  
  secteri= rep(0, ind)  
  Mit= rep(0, ind)
```

```
  # Distribution function of the event time S  
  F= function(t) {1-pnorm((mu*t+delta)/(1*t)^0.5)+  
    exp((-2*delta*mu)/(1))*pnorm((mu*t- delta)/(1*t)^0.5)}
```

```

# Density function of the event time S
f=function(t) (delta)/(2*pi*1*t^3)^.5*exp(-(delta+mu*t )^2/(2*1*t ))
# Function of N_i(t)
ncount= function(t) ft*as.numeric(dt<=t)
# Hazard function;
# I used the relationship of F(t) and cumulative hazard function H(t)
secter= function(t) -log(1- F(pmin(t, dt)))
# I(X_i<=x)
indi= function(x,y) as.numeric(x<=y)

last_col_no= dim(cov_com)[2]
indi_m= matrix(0, ind, (last_col_no-1))
rol_no= dim(cov_com)[1]
# each_value records values of process at each point
each_value= rep(0, rol_no)

for(w in 1:rol_no){
  for(i in 1:(last_col_no-1)) {indi_m[ ,i]= indi(cov[,i], cov_com[w, i])}
  indic= rep(0, ind)
  rowProds= function(X) {apply(X, 1, FUN="prod")}
  indic= rowProds(indi_m)
  secteri= secter(cov_com[w, last_col_no])
  ncounti= ncount(cov_com[w, last_col_no])
  # Mit records values of M_i(t) at each point
  Mit= ncounti- secteri
  each_value[w]=sum(indic*Mit)*1/sqrt(ind)
}
max(abs(each_value))
}

#=====
# Function name: Fhat #
#-----#
# Input variables: #
# delta: vector of length n which records estimated delta for all #
# subjects #
# mu: vector of length n which records estimated mu for all subjects #
# cov: n by (p+q) matrix which records values of all covariates #
# cov_com: matrix with (p+q+1) columns which records all different #
# combinations of values of covariates and event time; #
# covariates order is same as cov and the last column is #
# for event time #
# cov_mu: n by (1+d_2) matrix which records all covariates used in #
# mu with first column all 1 representing intercept #

```

```

# cov_delta: n by (1+d_1) matrix which records all covariates used #
#           in delta with first column all 1 representing intercept #
# dt: vector of length n which records the event times for all #
#       subjects #
# ft: numerical valued (0,1) vector which records event indicators #
#     for all subjects #
# number: number of approximated processes #
#-----#
# Output of Fhatx: #
# Return a vector of length 'number' which records all maximums of #
# absolute values of the simulated processes #
#=====#

Fhat= function(delta, mu, cov_mu, cov_delta, cov, cov_com, dt, ft, number) {

  ind= length(dt)
  # Distribution function of the event time S
  F= function(t) {1-pnorm((mu*t+delta)/(1*t)^0.5)+
                 exp((-2*delta*mu)/(1))*pnorm((mu*t- delta)/(1*t)^0.5)}
  # Density function of the event time S
  f=function(t) (delta)/(2*pi*1*t^3)^.5*exp(-(delta+mu*t )^2/(2*1*t ))
  # Derivative of f w.r.t mu
  f1= function(t) (1/(2*pi*t^3)^.5)*exp(-(delta+mu*t)^2/(2*t))*
                (-delta^2-delta*mu*t)
  # Derivative of f w.r.t delta
  f2= function(t) (1/(2*pi*t^3)^.5)*exp(-(delta+mu*t)^2/(2*t))*
                (1-delta^2/t-delta*mu)
  # Derivative of F w.r.t mu
  F1= function(t) -dnorm( (mu*t+delta)/t^.5 )*t^.5+
                 exp(-2*delta*mu)*(-2*delta)*pnorm((mu*t-delta)/t^.5 )+
                 exp(-2*delta*mu)*dnorm((mu*t-delta)/t^.5 )*t^.5
  # Derivative of F w.r.t delta
  F2= function(t) -dnorm( (mu*t+delta)/t^.5 )*t^-.5+
                 exp(-2*delta*mu)*(-2*mu)*pnorm((mu*t-delta)/t^.5 )+
                 exp(-2*delta*mu)*dnorm((mu*t-delta)/t^.5 )*(-t^-.5)

  # Estimated fishier information matrix
  mu_col_no= dim(cov_mu)[2]
  delta_col_no= dim(cov_delta)[2]
  li_mu= matrix(0, ind, mu_col_no)
  li_delta= matrix(0, ind, delta_col_no)
  i=1
  while (mu_col_no>=1) {
    # take the derivative w.r.t coeff of cov_mu[ ,mu_col_no] in mu
    li_mu[ ,i]= (ft*f1(dt)/f(dt)- (1-ft)*F1(dt)/(1-F(dt)))*cov_mu[ ,mu_col_no]
  }
}

```

```

i= i+1
mu_col_no= mu_col_no-1
}
i=1
while (delta_col_no>=1) {
# take the derivative w.r.t coeff of cov_delta[ ,delta_col_no] in delta
li_delta[ ,i]= (ft*f2(dt)/f(dt)- (1-ft)*F2(dt)/(1-F(dt)))*
cov_delta[ ,delta_col_no]
i= i+1
delta_col_no= delta_col_no-1
}
infisher= solve(1/ind*t(cbind(li_mu, li_delta))%*%cbind(li_mu, li_delta) )

# Function lambda
lambd= function(t) f(t)/(1-F(t))
# derivative of lambd w.r.t mu
lambdp1= function(t) 1/(1-F(t))*f1(t)+f(t)*F1(t)/(1-F(t))^2
# derivative of lambd w.r.t delta
lambdp2= function(t) 1/(1-F(t))*f2(t)+f(t)*F2(t)/(1-F(t))^2

#Derivative of function lambda
mu_col_no= dim(cov_mu)[2]
delta_col_no= dim(cov_delta)[2]
dlambd_mu= matrix(0, ind, mu_col_no)
dlambd_delta= matrix(0, ind, delta_col_no)
i=1
while(mu_col_no>=1) {
# derivative of log(lambd(yi)) w.r.t coeff of cov_mu[ ,mu_cov_no] in mu
dlambd_mu[ ,i]= lambd(dt)^(-1)*lambdp1(dt)*cov_mu[ ,mu_col_no]
i= i+1
mu_col_no= mu_col_no-1
}
i=1
while(delta_col_no>=1) {
# derivative of log(lambd(yi)) wrt coef of cov_delta[ ,delta_cov_no] in delta
dlambd_delta[ ,i]= lambd(dt)^(-1)*lambdp2(dt)*cov_delta[ ,delta_col_no]
i= i+1
delta_col_no= delta_col_no-1
}
dlambd= rbind(t(dlambd_mu), t(dlambd_delta))

long_no= dim(cov_com)[1]
cov_no= dim(cov_com)[2]
# Function N_i(t)
ncount= function(t) ft*as.numeric(dt<=t)

```

```

# Function I(X_i<=x)
indi= function(x,y) as.numeric(x<=y)
indi_m= matrix(0, ind, (cov_no-1))
# Store all values of 1st term besides G_i
ecount= matrix(0, long_no, ind)
for(w in 1:long_no) {
  for(i in 1:(cov_no-1)) {indi_m[ ,i]= indi(cov[,i], cov_com[w, i])}
  indic= rep(0, ind)
  rowProds= function(X) {apply(X, 1, FUN="prod")}
  indic= rowProds(indi_m)
  ecount[w, ]= 1/sqrt(ind)*indic*ncount(cov_com[w, cov_no])
}

# 2nd term
mu_col_no= dim(cov_mu)[2]
delta_col_no= dim(cov_delta)[2]
# store all values of Bhat
ebhat= matrix(0, long_no, (mu_col_no+delta_col_no))
## Record integral part of B_hat(t,x)
Bint_mu= matrix(0, mu_col_no, ind)
Bint_delta= matrix(0, delta_col_no, ind)
Bint= matrix(0, (mu_col_no+delta_col_no), ind)
# derivative of the integral w.r.t mu;
# I used the relationship btw survival function and cumulative hazard function
integr1= function(t) 1/(1- F(pmin(t, dt)))*F1(pmin(t, dt))
# derivative of the integral w.r.t delta;
# I used the relationship btw survival function and cumulative hazard function
integr2= function(t) 1/(1- F(pmin(t, dt)))*F2(pmin(t, dt))
# R_i(t)
rit= function(t) 1-as.numeric(dt<t)
indi_m= matrix(0, ind, (cov_no-1))
for(w in 1:long_no) {
  for(i in 1:mu_col_no) {
    Bint_mu[i, ]= integr1(cov_com[w , cov_no])*cov_mu[ , mu_col_no]
    mu_col_no= mu_col_no-1
  }

  for(i in 1:delta_col_no) {
    Bint_delta[i, ]= integr2(cov_com[w , cov_no])*cov_delta[ , delta_col_no]
    delta_col_no= delta_col_no-1
  }
}
Bint= rbind(Bint_mu, Bint_delta)
for(i in 1:(cov_no-1)) {indi_m[ ,i]= indi(cov[,i], cov_com[w, i])}
indic= rep(0, ind)
rowProds= function(X) {apply(X, 1, FUN="prod")}

```

```

indic= rowProds(indi_m)
mu_col_no= dim(cov_mu)[2]
delta_col_no= dim(cov_delta)[2]
for (i in 1:(mu_col_no+delta_col_no)) {
  ebhat[w, i]= 1/ind*sum(indic*Bint[i, ])
}
}

# Simulate G_i
g= matrix(rnorm(number*ind), number, ind)
# msimulated stores the maximum for each simulated data
msimulated= rep(0, number)
# install foreach package first
library(foreach)
msilulated= foreach(i= 1:number, .combine='c') %dopar% {
max(abs(ecount%*%g[i,]- 1/sqrt(ind)*ebhat%*%infisher%*%dlambd%*%(g[i,]*ft) ))
}
msilulated
}

```

## A.2 R Functions for Model Checking Based on Censoring Consistent Residuals

```

#=====#
# Function name: R_n #
#-----#
# Input variables: #
# delta: vector of length n which records estimated delta for all #
# subjects #
# mu: vector of length n which records estimated mu for all subjects #
# cov: n by (d_1+d_2) matrix which records values of all covariates #
# cov_com: matrix with (d_1+d_2) columns which records all different #
# combinations of values of covariates #
# covariates order is same as cov #
# dt: vector of length n which records event times for all subjects #
# ft: numerical valued (0,1) vector which records event indicators #
# for all subjects #
#-----#
# Output of R_n: #
# Return the maximum of absolute values of the process #

```



```

#####
R_n= function(delta, mu, cov ,cov_com ,dt, ft) {

# No. of individuals
ind= length(dt)
# survival function for the censor time
library(survival)
temp= survfit(coxph(Surv(dt,as.numeric(!ft))~1),type="aalen")
# f=0 makes it right continuous
survivalfunc1= stepfun(temp$time,c(1,temp$sur) ,f=0)
# next to calculate theta
# integrate of 1/G(t) where 1-G(t) is the censoring distribution
int_G= function(t) integrate(function(t) 1/survivalfunc1(t),0, t,
                             subdivisions=100000)$value
min_theta= function(x) (int_G(x)-x)/(x/survivalfunc1(x)-int_G(x))
sub_set= dt[(ft==1)]
thetam= rep(0, length(sub_set))
for(i in 1:length(sub_set)) {thetam[i]=min_theta(sub_set[i])}
# omit those NaN in thetam which is a result from zero in denominator
# of min_theta for some dt
theta=min(na.omit(thetam))
# transformed death time for uncensored obs.
fi1= function(x) (1+theta)*int_G(x)- theta*x/survivalfunc1(x)
# transformed death time for censored obs.
fi2= function(x) (1+theta)*int_G(x)
# new_survivaltime records transformed survival time
new_survivaltime= rep(0, ind)
new_survivaltime[ft==1]= mapply(fi1, dt[ft==1])
new_survivaltime[ft==0]= mapply(fi2, dt[ft==0])

# value of test statistics
col_no= dim(cov)[2]
indi_m= matrix(0, ind, col_no)
# I(z_i<=z)
indi= function(x,y) as.numeric(x<=y)
rol_no= dim(cov_com)[1]
# each_value records values of process at each point
each_value= rep(0, rol_no)
for(w in 1:rol_no){
  for(i in 1:col_no) {indi_m[ ,i]= indi(cov[,i], cov_com[w, i])}
  indic= rep(0, ind)
  rowProds= function(X) {apply(X, 1, FUN="prod")}
  indic= rowProds(indi_m)
  each_value[w]=sum(indic*(new_survivaltime- delta/abs(mu)))*1/sqrt(ind)
}
}

```

```
}  
max(abs(each_value))  
}
```

## Bibliography

- [1] Aalen, O. O., Borgan, A. and Gjessing, H. K. (2008). *Survival and Event History Analysis: A Process Point of View*. Springer, New York.
- [2] Aalen, O. O. and Gjessing, H. K. (2001). Understanding the Shape of the Hazard Rate: A Process Point of View. *Statistical Science*. **16** (1), 1-22.
- [3] Kagan, A. M. and Nagaev, S. (2001). How Many Moments Can Be Estimated from a Large Sample? *Statistics and Probability Letters*. **55**, 99-105.
- [4] 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.
- [5] Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993) *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- [6] Bickel, P. J. and Wichura, M. J. (1971). Convergence Criteria for Multiparameter Stochastic Processes and Some Applications. *The Annals of Mathematical Statistics*. **42** (5), 1656-1670.
- [7] Buckley, J. and James, I. (1979). Linear Regression with Censored Data. *Biometrika*. **66**, 429-436.
- [8] Cai, Z. and Sun, Y. (2003). Local Linear Estimation for Time-Dependent Coefficients in Cox's Regression Models. *Scandinavian Journal of Statistics*. **30**, 93-111.
- [9] Chhikara, R. S. and Folks, J. H. (1977). The Inverse Gaussian Distribution as a lifetime model. *Technometrics*. **19**, 461-468.
- [10] Chhikara, R. S. and Folks, J. L. (1989). *The inverse Gaussian distribution: theory, methods and applications* Marcel Dekker: New-York.

- [11] Cox, D. R. (1972). Regression Models and Life Tables (with Discussion). *Journal of Royal Statistical Society, Series B.* **34**, 187-220.
- [12] de Boor, C. (2001). *A Practical Guide to Splines*. Revised Edition, Springer, New York.
- [13] Doksum, K. and Hyland, A. (1992). Models for Variable-Stress Accelerated Test Experiments Based on Wiener Processes and the Inverse Gaussian Distribution. *Technometrics.* **34**, 74-82.
- [14] Eaton, W. W. and Whitmore, G. A. (1977). Length of Stay as A Stochastic Process: A General Approach and Application to hospitalization for Schizophrenia. *Journal of Mathematical Sociology.* **5**, 273-292.
- [15] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman and Hall.
- [16] Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.
- [17] Folks, J. L. and Chhikara, R. S. (1978). The Inverse Gaussian Distribution and Its Statistical Application- A Review. *Journal of Royal Statistical Society, Series B.* **40**, 263-289.
- [18] Hastie, T. and Tibshirani, R. (1986). Generalized Additive Models. *Statistical Sciece.* **1** (3), 297-318.
- [19] Hjort, N. L. (1990). Goodness of Fit Tests in Models for Life History Data Based on Cumulative Hazard Rates. *The Annals of Statistics.* **18**, 1221-1258.
- [20] Horrocks, J. and Thompson, M. (2004). Modelling Event Times with Multiple Outcomes Using the Wiener Process with Drift. *Lifetime Data Analysis.* **10**, 29-49.
- [21] Huang, J. (1996). Efficient Estimation for the Proportional Hazards Model with Interval Censoring. *The Annals of Statistics.* **24**, 540-68.
- [22] Huang, J. Z., Wu, C. O. and Zhou L. (2004). Polynomial Spline Estimation and Inference for Varying Coefficient Models with Longitudinal Data. *Statistica Sinica.* **14**, 763-788.

- [23] Huang, C. Y., Luo, X. and Follmann, D. A. (2011). A Model Checking Method for the Proportional Hazards Model with Recurrent Gap Time Data. *Biostatistics*. **12** (3), 535-547.
- [24] Huber, P. J. (1967). The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions. *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley: University of California Press, 221-233.
- [25] Jennrich, R. I. (1969). Asymptotic Properties of Non-Linear Least Squares Estimators. *The Annals of Mathematical Statistics*. **40** (2), 633-643.
- [26] Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data* (2nd ed.), Hoboken, NJ: Wiley.
- [27] Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*, Springer-New York.
- [28] Lai, T. L. and Ying, Z. (1991). Rank Regression Methods for Left Truncated and Right Censored Data. *The Annals of Statistics*. **19**, 531-554.
- [29] Lancaster, T. (1972). A Stochastic Model for the Duration of A Strike. *Journal of Royal Statistical Society, Series A*. **135**, 257-271.
- [30] Lee, M.-L. T. and Whitmore, G. A. (2006). Threshold Regression for Survival Analysis: Modeling Event Times by a Stochastic Process Reaching a Boundary. *Statistical Science*. **21**, 501-513.
- [31] Lee, M.-L. T., DeGruttola, V and Schoenfeld, D. A. (2000). A Model for Markers and Latent Health Status. *Journal of Royal Statistical Society, Series B*. **62**, 747-762.
- [32] Lee, M.-L. T. and Whitmore, G. A. (2010). Proportional Hazards and Threshold Regression: Their Theoretical and Practical Connections. *Lifetime Data Analysis*. **16**, 196-214.
- [33] Lee, M.-L. T., Whitmore, G. A. and Rosner, B. A. (2010). Threshold Regression for Survival Data with Time-Varying Covariates. *Statistics in Medicine*. **29**, 896-905.
- [34] Lee, M.-L. T., Whitmore, G. A., Laden F., Hart, J. E. and Garshick, E. (2004). Assessing Lung Cancer Risk in Railroad Workers Using a First Hitting Time Regression Model. *Environmetrics*. **15** (5), 501-512.

- [35] León, L. F. and Cai, T. (2012). Model Checking Techniques for Assessing Functional Form Specifications in Censored Linear Regression Models. *Statistica Sinica*. **22**, 509-530.
- [36] León, L. F. and Tsai, C. L. (2004). Functional Form Diagnostics for Cox's Proportional Hazards Model. *Biometrics*. **60**, 75-84.
- [37] Li, G. (2003). Nonparametric Likelihood Ratio Goodness-of-Fit Tests for Survival Data. *Journal of Multivariate Analysis*. **86**, 166-182.
- [38] Li, J. and Lee, M.-L. T. (2011). Analysis of Failure Time Using Threshold Regression with Semi-Parametric Varying Coefficients. *Statistica Neerlandica*. **65**, 164-182.
- [39] Lin, D. Y. (1991). Goodness-of-Fit Analysis for the Cox Regression Model Based on a Class of Parameter Estimators. *Journal of American Statistical Association*. **86** (415), 725-728.
- [40] Lin, D. Y. and Spiekerman, C. F. (1996). Model Checking Techniques for Parametric Regression with Censored Data. *Scandinavian Journal of Statistics*. **23**, 157-177.
- [41] Lin, D. Y. and Wei, L. J. (1991). Goodness-of-Fit Tests for the General Cox Regression Model. *Statistica Sinica*. **1**, 1-17.
- [42] Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric Regression for the mean and Rate Functions of Recurrent Events. *Journal of Royal Statistical Society, Series B*. (62), Part 4, 711-730.
- [43] Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox Model with Cumulative Sums of Martingale-based Residuals. *Biometrika*. **80** (3), 557-572.
- [44] Lin, D. Y., Wei, L. J. and Ying, Z. (2002). Model-Checking Techniques Based on Cumulative Residuals. *Biometrics*. **58**, 1-12.
- [45] Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the Mean Function with Panel Count Data Using Monotone Polynomial Splines. *Biometrika*. **94**, 705-718.
- [46] Murphy, S. A. and Sen, P. K. (1991). Time-dependent Coefficients in a Cox-type Regression Model. *Stochastic Processes and their Applications*. **39**, 153-180.

- [47] Nahman, N. S., Middendorf, D. F., Bay, W. H., McElligott, R., Powell, S. and Anderson, J. (1992). Modification of the Percutaneous Approach to Peritoneal Dialysis Catheter Placement Under Peritoneoscopic Visualization: Clinical Results in 78 Patients. *Journal of the American Society of Nephrology*. **3**, 103-107.
- [48] O'Sullivan, F. (1988). Nonparametric Estimation of Relative Risk Using Splines and Cross-Validation. *SIAM Journal on Scientific and Statistical Computing*. **9**, 531-542.
- [49] Pennell, M. L., Whitmore, G. A. and Lee, M.-L. T. (2010). Bayesian Random-Effects Threshold Regression with Application to Survival Data with Nonproportional Hazards. *Biostatistics*. **11(1)**, 111-126.
- [50] Rice, J. A. and Silverman, B. W. (1991). Estimating the Mean and Covariance Structure Nonparametrically When the Data Are Curves. *Journal of Royal Statistical Society, Series B*. **53**, 233-243.
- [51] Ritov, Y. (1990). Estimation in a Linear Regression Model with Censored Data. *The Annals of Statistics*. **18**, 303-328.
- [52] Shen, X. and Wong, W. (1994). Convergence Rate of Sieve Estimates. *The Annals of Statistics*. **22**, 580-615.
- [53] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.
- [54] Sleeper, L. A. and Harrington, D. P. (1990). Regression Splines in the Cox Model with Application to Covariate Effects in Liver Disease. *Journal of the American Statistical Association*. **85** (412), 941-949.
- [55] Stone, C. J. (1994). The Use of Polynomial Splines and Their Tensor Products in Multivariate Function Estimation (with discussion). *The Annals of Statistics*. **22**, 118-184.
- [56] Strawderman, R. L and Tsiatis, A. A. (1996). On the Asymptotic Properties of a Flexible Hazard Estimator. *The Annals of Statistics*. **24** (1), 41-63.
- [57] Strawderman, R. L. and Tsiatis, A. A. (1996). On Consistency in Parameter Spaces of Expanding Dimension: An Application of the Inverse Function Theorem. *Statistica Sinica*. **6** 917-923.
- [58] Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified Proportional Hazard Models. *Biometrika*. **73** (2), 363-369.

- [59] Sun, Y. (1997). Weak Convergence of the Generalized Parametric Empirical Processes and Goodness-of-Fit Tests for Parametric Models. *Communications in Statistics- theory and Methods*. **26** (10), 2393-2413.
- [60] Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990). Martingale-Based Residuals for Survival Models. *Biometrika*. **77**, 147-160.
- [61] Tian, L. , Zucker, D. and Wei, L. J. (2005). On the Cox Model with Time-Varying Regression Coefficients. *Journal of the American Statistical Association*. **100** (469), 172-183.
- [62] Tong, X., He, X., Sun, J. and Lee, M.-L. T. (2008). Joint Analysis of Current Status and Marker Data: An Extension of a Bivariate Threshold Model. *The International Journal of Biostatistics*. **4**(1), Article 21.
- [63] Tsiatis, A. A. (1981). A Large Sample Study of Cox's Regression Model. *The Annals of Statistics*. **9**, 93-108.
- [64] Tsiatis, A. A. (1990). Estimating Regression Parameters Using Linear Rank Tests for Censored Data. *The Annals of Statistics*. **18**, 354-372.
- [65] Tweedie, M. C. K. (1957). Statistical Properties of Inverse Gaussian Distribution. *The Annals of Mathematical Statistics*. **28**(2), 362-377.
- [66] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- [67] van der Vaart, A. W. (2002). Semiparametric Statistics. *Lectures on Probability Theory and Statistics. Lecture Notes in Math.* **1781** 331-457. Springer, New York.
- [68] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1915446
- [69] 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** (408), 1065-1073.
- [70] Wei, L. J., Ying, Z. and Lin, D. Y. (1990). Linear Regression Analysis of Censored Survival Data Based on Rank Tests. *Biometrika*. **77**, 845-851.



- [71] Wellner, J. and Zhang, Y. (2007). Two Likelihood-based Semiparametric Estimation Methods for Panel Count Data with Covariates. *The Annals of Statistics*. **35**, 2106-2142.
- [72] White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*. **50**, 1-26.
- [73] Whitmore, G. A. (1983). A Regression Method for Censored Inverse-Gaussian Data. *Canadian Journal of Statistics*. **11**, 305-315.
- [74] Whitmore, G. A. (1979). An Inverse Gaussian Model for Labour Turnover. *Journal of Royal Statistical Society, Series A*. **142**, 468-478.
- [75] Whitmore, G. A., Crowder, M. J. and Lawless, J. F. (1998). Failure Inference from a Marker Process Based on Bivariate Wiener Model. *Lifetime Data Analysis*. **4**, 229-251.
- [76] Xiao, T., Whitmore, G. A., He, X. and Lee, M.-L. T. (2012). Threshold Regression for Time-To-Event Analysis: The Stthreg Package. *The Stata Journal*. **12** (2), 1-27.
- [77] Ying, Z. (1993). A large Sample Study of Rank Estimation for Censored Regression Data. *The Annals of Statistics*. **21**, 76-99.
- [78] Yu, Z., Tu, W. and Lee, M.-L. T. (2009). A Semi-parametric Threshold Regression Analysis of Sexually Transmitted Infections in Adolescent Women. *Statistics in Medicine*. **28**, 3029-3042.
- [79] Zhang, Y., Hua, L. and Huang, J. (2010). A Spline-Based Semiparametric Maximum Likelihood Estimation Method for Cox Model with Interval-Censored Data. *Scandinavian Journal of Statistics*. **37**, 338-354.