NONPARAMETRIC STATISTICAL INFERENCE FOR SURVIVAL DATA

MEILING HAO

Ph.D

The Hong Kong Polytechnic University

2016

The Hong Kong Polytechnic University

Department of Applied Mathematics

Nonparametric statistical inference for survival data

Meiling Hao

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

July 2016

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________________ (Signed)

Meiling Hao (Name of student)

Abstract

Censored data, one of the most common data types, arise frequently in many fields of modern science, e.g., health science, reliability, economics, finance, etc. The most prominent feature of this kind of data is that the occurrence of the event could not be observed exactly. Right censored data and interval censored data are among the most popular ones. Over the past decades, there have been numerous state-of-the-art methodologies in survival analysis literature to handle censoring. This thesis would focus on the nonparametric statistical inference of right censored data and interval censored data.

As the first part of this thesis, a penalized nonparametric maximum likelihood estimation of the log-hazard function is introduced in analyzing the right censored data. The smoothing spline is employed for a smooth estimation. The most appealing fact is that a functional Bahadur representation is established, which serves as a key step for nonparametric inference of the unknown parameter/function. Asymptotic properties of the resulting estimate of the unknown log-hazard function are proved. Furthermore, the local confidence interval and simultaneous confidence band of the unknown log-hazard function are provided, along with a local and global likelihood ratio tests. We also investigate issues related to the asymptotic efficiency.

As the second part of this thesis, the aforementioned nonparametric inference approach is extended to handle interval censored data. In particular, we focus on the nonparametric inference of the cumulative hazard function, instead of the log-hazard function of the interval censored data. Similarly, we have derived a functional Bahadur representation and established the asymptotic properties of the resulting estimate of the cumulative function. Particularly, the global asymptotic properties are justified under regularity conditions. A likelihood ratio test is also provided. To the best of our knowledge, there is no report in the literature on the asymptotic properties of a smoothing spline-based nonparametric estimate for the interval censored data.

The theoretical results are validated by extensive simulation studies. Applications are illustrated with some real datasets. A few discussions and closing remarks are given.

Key Words: Functional Bahadur representation; Interval censored data; Likelihood ratio test; Nonparametric inference; Penalized likelihood; Right censored data; Smoothing splines.

Acknowledgements

Although the endeavor of carrying out research is an isolated activity, under the guidance of my supervisors and the company of my classmates, the process is still full of happiness and pleasant surprise. I would like to express my heartfelt appreciation to the several individuals, who have assisted me in various ways during my PhD life and would like to hereby acknowledge their support.

First and foremost, I would love to extend my deepest and sincerest gratitude to my supervisor, Dr. Xingqiu Zhao, for her instructive and enlightening guidance, invaluable discussions and insightful ideas throughout the years. I also benefited from her enthusiasm for life and open mind attitude towards everything. Without her rigorous requirements and vast knowledge, my thesis would not be finished.

Next, I would express my thanks to my supervisor Dr. Yuanyuan Lin, for her strong encouragement and patience during the years of my PhD study. I feel fortune to work with her. In particular, I am grateful to her continuous support and guidance in every aspect before or after her move. What I have benefited most from her is the rigorous and diligent attitude to scientific research, active, open and sharp mind towards any ideas, thirsty and insistence for the purist of truth. With her kind assistance and grave attitude, everything would not have been possible.

Besides, I would express my gratitude to Professor Heung Wong. At the forefront of my PhD experience, he has been a constant source of inspiration and mentorship.

Finally, I would like to express my special thanks to my family, my friends and my classmates for their love, encouragement and support.

Contents

Certificate of Originality iii

Abstract v

Acknowledgements vii

List of Figures xiii

List of Tables xvii

List of Notations xix

1 Introduction 1

1.1 Background 1

1.1.1 Primary biliary cirrhosis datasets 1

1.1.2 Lung Tumor Data 2

1.1.3 Breast cosmesis study 3

1.2 Time origin, censoring and truncation 3

1.2.1 Time origin 4

1.2.2 Censoring and truncation 5

1.3 Distribution, survival time, density, hazard and cumulative hazard 8

1.4 Literature review 10

1.4.1 Right censoring 10

1.4.2 Interval censoring 16

1.5 Research of outline 21

2 Nonparametric Statistical Inference for Right Censored Data Using Smoothing Splines 23

2.1 Introduction 24

2.2 Preliminaries 28

2.2.1 Notation and Methodology. 28

2.2.2 Reproducing Kernel Hilbert Space 30

2.3 Functional Bahadur Representation 34

2.4 Likelihood Ratio Test 39

2.4.1 Local Likelihood Ratio Test 39

2.4.2 Global Likelihood Ratio Test 45

2.5 Simulation Results 47

2.6 Application 52

2.7 Appendix 53

3 Nonparametric Statistical Inference for Case One Interval Censored Data 95

3.1 Introduction 95

3.2 Methodology 97

3.3 Functional Bahadur Representation 102

3.4 Likelihood Ratio Test 105
3.4.1 Local Likelihood Ratio Test 105
3.4.2 Global Likelihood Ratio Test 109
3.5 Simulation 111
3.6 Application 113
3.7 Appendix 114
4 Conclusions and future work **147**
4.1 Conclusions 147
4.2 Future Work 149
Bibliography **151**

List of Figures

2.1 The pictures show the simulation results with censoring rate being 20%. The first panel and fourth panel is the estimation (marked with −.) and the true function (solid line), respectively. The second and fifth panels display the estimate coverage probabilities given by LCP (marked with −.) and BCP (marked with ·), while the third and last panels show the estimate pointwise confidence intervals given by LCI (marked with −.) and BCI (marked with ·). 88

2.2 The pictures show the simulation results with censoring rate being 40%. The first panel and fourth panel is the estimation (marked with −.) and the true function (solid line), respectively. The second and fifth panels display the estimate coverage probabilities given by LCP (marked with −.) and BCP (marked with ·), while the third and last panels show the estimate pointwise confidence intervals given by LCI (marked with −.) and BCI (marked with ·). 89

2.3 The pictures show the simulation results with censoring rate being 20%. The first panel and fourth panel is the estimation (marked with −.) and the true function (solid line), respectively. The second and fifth panels display the estimate coverage probabilities given by LCP (marked with −.) and BCP (marked with ·), while the third and last panels show the estimate pointwise confidence intervals given by LCI (marked with −.) and BCI (marked with ·). 90

2.4 The pictures show the simulation results with censoring rate being 40%. The first panel and fourth panel is the estimation (marked with −.) and the true function (solid line), respectively. The second and fifth panels display the estimate coverage probabilities given by LCP (marked with −.) and BCP (marked with ·), while the third and last panels show the estimate pointwise confidence intervals given by LCI (marked with −.) and BCI (marked with ·). 91

2.5 The pictures display the real data analysis. The first panel displays the cumulative hazard estimation: the kernel methods (solid line), the Kaplan-Meier estimator (marked with −−) and our estimation (the dash line), and the confidence band is given by the Kaplan-Meier estimation marked with −. line. The second panel shows the log-hazard estimation and its confidence band various the three methods: the solid line is the log-hazard estimation, the −− line the simultaneous confidence band, the −. line is the local pointwise confidence interval while the line marked with · is the confidence interval given by Wahba. 92

3.1 The picture gives the estimation and local CI of the example 1 with n=100. It shows the cumulative hazard estimation with censoring rate being 20% and 30% and the coverage probabilities with the censoring rate being 20% and 30%, respectively. Specifically, the solid line is the true cumulative hazard estimation, the −. lines are that according to the censoring rate being 20% and 30%, respectively. The solid lines marked with · are the CI according to the method of Wahba. 141

3.2 The picture gives the estimation and local CI of the example 1 with n=200. It shows the cumulative hazard estimation with censoring rate being 20% and 30% and the coverage probabilities with the censoring rate being 20% and 30%, respectively. Specifically, the solid line is the true cumulative hazard estimation, the −. lines are that according to the censoring rate being 20% and 30%, respectively. The solid lines marked with · are the CI according to the method of Wahba. 142

3.3 The picture gives the estimation and local CI of the example 1 with n=300. It shows the cumulative hazard estimation with censoring rate being 20% and 30% and the coverage probabilities with the censoring rate being 20% and 30%, respectively. Specifically, the solid line is the true cumulative hazard estimation, the −. lines are that according to the censoring rate being 20% and 30%, respectively. The solid lines marked with · are the CI according to the method of Wahba. 143

3.4 The pictures gives the real data analysis. It shows the cumulative hazard estimation and its confidence band various the three methods: the dash line is the cumulative hazard estimation, the −− line the simultaneous confidence band, the −. line is the local pointwise confidence interval while the line marked with · is the confidence interval given by Wahba. 144

List of Tables

2.1 The estimated size and power of PLRT for example 1, where the test function is $g(t) = g_0(t) + ct$ with various c values and the nominal significance level is 95%. 87

2.2 Estimated global coverage probability with the nominal coverage probability being 95% for example 1. 87

2.3 The estimated size and power of PLRT for example 2, where the test function is $g(t) = g_0(t) + ct$ with various c values and the nominal significance level is 95%. 93

2.4 Estimated global coverage probability with the nominal coverage probability being 95% for example 1. 93

2.5 The estimated size and power of PLRT for example 3, where the test function is $g(t) = g_0(t) + ct$ with various c values and the nominal significance level is 95%. 93

3.1 Estimated global coverage probability with the nominal coverage probability being 95% for example 1. 145

List of Notations

$\mathbb{R}$	Set of real numbers
$x^\top$	Transpose of matrix/vector x
$\|\|\|$	Euclidean norm
$\xrightarrow{d}$	Converge in distribution
$tr(A)$	The trace of the matrix A

Chapter 1

Introduction

The chapter briefly introduces the background via some datasets and gives some literature reviews related to the topic. In this thesis, we mainly discuss the nonparametric analysis about the right censored data and interval censored data.

1.1 Background

Three datasets would be introduced first, which would reveal some data structures in the biomedical studies, engineering and many other fields in survival analysis.

1.1.1 Primary biliary cirrhosis datasets

Although the primary biliary cirrhosis (PBC) is very rare, with just about 50-per-million in population, it is a kind of lethal chronic liver disease with ambiguously cause. As types of the events reveal that the disease may be mediated by immunologic mechanism, the Mayo Clinic conducted a trial to explore the drug D-penicillamine (DPCA) on the PBC from January 1974 to May 1984. Specially, 312 patients were double-blinded randomized in the treatment group with DPCA and the control group with a placebo. Many related explanatory factors, such as clinical, biomedical, serological and histological variables were recorded for as many participates as possible. Finally, there were 125 patients died, with 11 deaths were not died of PBC liver. Besides, there were 8 patients lost to follow up, and another 19 patients had chosen to undergo liver transplantation. The data can be found in Appendix 6 in Fleming and Harrington (2011).

1.1.2 Lung Tumor Data

In order to determine the relationship between a suspected agent or environment with the time of tumor onset, some tumorigenicity experiments are often carried out. In these trials, the tumor onset is the occurrence of an interesting event. But most of the time, the death or sacrifice time instead of the time of tumor onset can be observed. Hoel and Walberg (1972) gave a dataset related to a lung tumor trial. In the trail, 96 mice were put into the conventional environment and 48 mice were treated with germ-free environment. Specifically, the death time of 144 male RFM mice were measured in days and whether the lung tumor presence or not at the time of death are recorded by an indicator function. As lung tumor's occurrence in RFM mice will not affect the time of death, the occurrence of tumor onset is only known before or after the death or sacrifice. The data can be found in Table 1.3 in Sun (2007).

1.1.3 Breast cosmesis study

Beadle et al. (1984) conducted a clinical trial to explore the cosmetic different effects of radiotherapy alone and radiotherapy plus adjuvant chemotherapy on women with early breast cancer. In the study, 46 subjects were treated with radiation treatment while the other 48 were assigned to the radiotherapy plus chemotherapy group. Patients were pre-scheduled followed for every 4-6 months, but the time gap between every visit was lengthened with the recovery progressing. Three points of scale (none, moderate, severe) of breast retraction were recorded at each visit. The interest event is the first observation time of the retraction of moderate or severe breast. As the subjects were observed only when they did the examinations, instead of observing the exact time of the retraction of breast, we just know whether it fell in the interval between visits or not. The subjects were followed up to 60 months, and the recorded data of the two groups can be found in Table 1.9 in Klein and Moeschberger (1997).

1.2 Time origin, censoring and truncation

The commonness of the three examples in the previous section is that the response of interest is the time until some events occur. Such events often refer to the occurrence of a disease, death or the onset of certain milestone, the failing of a machine, or learning something. Thereby, the event is often named as failure and the time, usually referring to the death in biological organisms and failure in mechanical systems, is named as failure time or survival time. In order to analyze the data, it is important to have a general knowledge about the data structures. More specifically, in the section, we would introduce the time origin, the censoring mechanism and the similar data structure, truncation.

1.2.1 Time origin

Before analyzing the failure time data, it is crucial to define the time origin clearly and unambiguously. In some instances, it may be the birth time of an individual; in other cases, it can be the occurrence of an event, such as the randomization registered in a trial or the occurrence of a heart attack. According to Kalbfleisch and Prentice (2011), in order to get the time origin, we should make clearly the following:

(1) A clear definition of what constitutes the failure;

(2) An exact definition of the occurrence of the event of interest, namely "response";

After figuring out the above two definitions, we can define the survival or failure time. For example, in PBC data case, failure means a subject died of PBC liver; in lung tumor situation, the response is the tumor onset while in breast cosmesis study, the time of breast retraction is the response. Just like the examples, if the observed time is not at the origin, special data structure arises.

1.2.2 Censoring and truncation

As noted in the PBC data example, lung tumor data and the breast cosmesis study case, we can find that the data often include some individuals whose time could not be observed. The data on these subjects are said to be censored. Specifically, there are three censoring mechanisms: right censoring, left censoring and interval censoring.

In the PBC data example, we can find that some individuals do not fail during the observation period. The data on these individuals are said to be right censoring. Actually, the right censoring include three kinds of cases: (1) the individuals are still alive at the end of the study; (2) the individuals are lost to follow up because of their moving home, or withdrawing from the trial; (3) the other events happened before our interesting event happening, such as the individual died of heart attack while the interesting event is the PBC liver disease. Specifically, right censoring is that you can observe the failure time data exactly or the failure does not occur at the time the censoring happens. In the light of the relationship between the failure time and the censoring time, it consists of different kinds of censoring mechanisms: type one censoring with a random censoring time, type two censoring with the fixed censoring time; independent censoring mechanism with the failure rate applies to the individual at each time is the same as those without censoring, vise versa, it is called dependent censoring. Most of the times in survival analysis, we would assume the failure time is independent or conditional independent of a random censoring time.

Similar to the right censoring mechanism, there also exists the left censoring mechanism. Specifically, left censoring occurs when the failure time could be observed just before some time, otherwise, we could not know whether it fail or not. The analysis of left censoring is very similar to that of the right censoring mechanism, so we do not talk it in detail.

Compared with right censoring, interval censoring is much more complex. As in the example of lung tumor data and the breast cosmesis study, instead of recording the exact failure time, we just know that the event occurred before or after some point or that the event occurred in some interval or not. This kind of data is named as the interval censored data. Specifically, there exist three kinds of interval censored data types: case-one interval censored data, case-two interval censored data and doubly censored failure time data. Generally, case-one interval censoring means that instead of observing the exact of each individual's failure time, one just knows that the failure time is either left before or right behind some time. Case-one interval censoring is also referred as current status data, coming from the demographic studies. The data in example 2 is exactly about case-one interval censoring. According to the sampling schemes, there mainly exist four types of the case-one interval censored data: case-control censored data (Jewell and Van Der Laan (2004)), doubly censored data (De Gruttola and Lagakos (1989)), clustered censored data (Galbraith et al. (2010)) and bivariate case-one interval censored data (Wang and Ding (2000)). From the example 3, we can get that there exists at leat one interval for an individual, and the event failing in an interval or not could be observed. However, the exact failure time could not be recorded. This kind of data is called case-two interval censored data. When two related events are studied in a study and you just know that each event could fail in an interval or not, doubly censored data occurs. Similar to the right censoring mechanism, there also exists independent or dependent censoring mechanism. Specifically, if the failure time is independent of the censoring time, the independent censoring occurs. Vise versa. Under the situation of current status data, independent censoring means that the failure time is independent of the observation time point, while for case-two interval censoring, it means that the failure time is independent of the interval of the observation time. Also, we mainly talk about the case of independent censoring in this thesis.

Truncation data are very similar but different from censored data. Generally, in some cases, whether the subjects could be observed or not depends on their entry of the study. If the individual can be observed if and only if the failure time is larger than some time, we claim that the left truncation or the delay entry occurs. Vise, we claim the right truncation happens. The analysis of left truncation is very similar to the right censored data, and many methods related to the right censoring can be extended easily to the left truncation.

1.3 Distribution, survival time, density, hazard and cumulative hazard

Let T be an arbitrary nonnegative random variable. Denote $F(t) = P(T \leq t)$, which represents the cumulative distribution function of T. In survival analysis, three other statistics related to the distribution function of T can also reveal the structure of the failure time: the survival function, denoted as $S(t)$, the probability density, denoted as $f(t)$ and the hazard denoted as $\lambda(t)$.

The survival function measures the probability of T larger than a fixed value t, namely an individual lives longer than time t. That is:

$$S(t) = P(T > t) = 1 - F(t), 0 < t < \infty. \tag{1.1}$$

It is easy to check that $S(t)$ is right-continuous at time t, and $S(0) = 1$ while $S(\infty) = 0$.

If T is continuous, then it has the probability density function, defined as:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}. \tag{1.2}$$

The hazard function gives an instantaneous rate at which failure occurs for an individual or item that is surviving at time t, which is defined as:

$$\begin{aligned}\lambda(t) &= \lim_{dt\downarrow 0} \frac{1}{dt} P\{t \leq T < t + dt | T \geq t\} \\ &= -\frac{dS(t)}{S(t)} \\ &= \frac{f(t)}{S(t)} \\ &= -d\log\{S(t)\}. \end{aligned} \tag{1.3}$$

Actually, as Spierdijk (2008) said, an increasing hazard rate reflects the positive duration dependence while a decreasing hazard rate means negative duration dependence. Besides, following from (1.3), it is easy to check that:

$$S(t) = \exp\{-\int_0^t \lambda(s)\, ds\} = \exp\{-\Lambda(t)\}, \tag{1.4}$$

where $\Lambda(t) = \int_0^t \lambda(s)\, ds$, which is called the cumulative hazard function.

From (1.1), (1.2), (1.3) and (1.4), it is known that there is a one-to-one mapping relationship between $F(t), S(t), f(t)$ and $\lambda(t)$. So in order to get the feature of the failure time, anyone of the four can be used. we can extend the previous discussion to the discrete case via some slight changes .

1.4 Literature review

In the fields of biometry, reliability, clinical trials and medical follow-up studies, estimating the distribution of the failure or survivor time is of fundamental importance. As the occurrence of censoring would complicate the analysis of survival time, there exists abundant literature that focuses on this topic. This section would give an overview about the analysis of the right censored data and interval censored data in survival analysis.

1.4.1 Right censoring

The random right censoring arises frequently in fields of survival analysis in biomedical studies and of reliability analysis in engineering. In these situations, life-time table estimation is firstly proposed to estimate the survival function (Berkson and Gage (1952), Cutler and Ederer (1958), and Gehan (1969)). Resulting from the grouping of the data structure, the life-time table estimator is slightly biased. Thereby, Böhmer (1912) extended the life-time table to the product limit estimator or the Kaplan-Meier estimator. After Kaplan and Meier (1958) showed that the estimator was a nonparametric maximum likelihood estimation, the method was widely applied in statistics under the scenario of right censoring. Efron (1967) proved the estimator enjoyed the self-consistency property while Breslow and Crowley (1974) gave the asymptotic normality properties.

There also exists abundant literature that talked about the exponential bound for the Kaplan-Meier estimator, see Dabrowska (1989), Bitouze et al. (1999), Wellner (2007). As the similarity of the right censoring and left censoring scheme, many statisticians extended the Kaplan-Meier estimator to the left censored data case, named as Left Kaplan-Meier estimator (LKM). This can refer to Ware and Demets (1976), Csörgö and Horváth (1980), Gomez et al. (1992) and Gómez et al. (1994). The self consistency and asymptotic properties about the LKM are also well established. Although the Kaplan-Meier estimator is easy to calculate and well-developed, it is a step function. As a smoothed survival function is much more desirable in survival analysis, many statisticians extended the step-function to a smoothed version. Among these, Blum and Susarla (1980) and Földes et al. (1981) proposed kernel methods to smooth the estimate of survival function, Kim et al. (2003) borrowed the idea about Bezier curve smoothing from computer science, while Whittemore and Keller (1986) used the splines to get a smoothed nonparametric maximum likelihood estimator of the survival function. Some parametric methods are also involved to get the smoothed estimator of the survival function. Among these, Weibull and exponential models have been considered by Greenhouse and Silliman (1996), Gompertz model was used in Gieser et al. (1998) while Weibull, logistic and log-logistic were discussed in Hauck et al. (1997).

Density estimation is another important way to study the lifetime, both in theory and reality in statistics. Among the nonparametric estimators of the density, histogram density estimator is the simplest and first way to give the density estimation in survival analysis. Actually, based on the life time table, Gehan (1969) gave the density estimation of the survival function. Based on the Kaplan-Meier estimator, Földes et al. (1981) also gave the formula of the histogram density. Not only they gave the self-consistency property of the estimator, they gave the exact convergence rate of the histogram estimate. Burke and Horvath (1982) have considered a more general density estimator, and the histogram can be the special case of this kind of estimators. Although the histogram estimator is easy to calculate and requires few assumptions, its non-smoothed character hinders the sophisticated inference. Thereby, the kernel-type estimators have been well-developed with right-censored data after 1980. Specially, via the kernel method, Blum and Susarla (1980) have considered to maximize the deviation of the density, which has a strong powerful for goodness-of-fit test and tests for hypothesis about the density without specified form. Földes et al. (1981) showed that another kind of kernel estimator of the density enjoyed the strong convergence property, which covers the usual Parzen (1962) density estimate. As the mean integrated square error (MISE) is one of the most usual criteria to evaluate the performance of the estimates, Sánchez-

Sellero et al. (1999) talked about the convergence of MISE with censored data. Based on which, they showed the optimal bandwidth to get the minimum MISE. From the formula in Sánchez-Sellero et al. (1999), it can derive that the censoring will not affect the bias of MISE but it will affect the variance of the estimates. As competing risks is a special case of right censoring model, Burke and Horvath (1982) proposed a general estimates of the density under the senecio, which include the kernel-type estimates, series-type estimates and histogram-type estimates. McNichols and Padgett (1981) gave a weighted kernel-type estimates of the density under the proportional hazard model. The estimator is shown to be asymptotically unbiased. McNichols and Padgett (1982b) modified the weighted kernel method and showed that the new estimator enjoyed the consistency property under mild conditions. Delta sequence curve estimator, including the kernel-type estimators, is proposed by Yandell (1981). The estimator enjoys the uniform consistency and asymptotical normality properties. The nonparametric maximum likelihood estimates of the density is another method to get the estimator of the density, see McNichols and Padgett (1982a), Kooperberg and Stone (1992). Series-type estimates can be found in Kimura (1972), Tarter (1979), Tanner and Wong (1984), Antoniadis et al. (1999) and so on. Specifically, Tarter (1979) expanded the density as a linear combination of the Fourier series, and based on the maximum likelihood method, he got a consistent estimator of the density. Based on the wavelet method, Antoniadis et al. (1999) and Li (2007) gave an estimator of the density and showed the optimal rate, respectively. Splines or the log-splines are also used to estimate the density, such as in Koo et al. (1999). All the methods talked above are based on the nonparametric methods. Parallel to the survival function, there also exist some parametric density estimation methods. For example, the exponential model, the Weibull model and the log-normal are the most common model to estimate the density. The maximum likelihood method is often used to estimate the corresponding parameters in the models.

As the hazard rate can reveal the instantaneous probability that an event maybe occurs in the next instant, it can reflect much more details about the survivor time. Thereby, extensive state-of-the-art methodologies exist about the estimation of the hazard rate. Among theses, the Nelson-Aalen estimate, which was proposed by Nelson (1969) and Nelson (1972), is the simplest one to calculate. Besides, Breslow and Crowley (1974) and Aalen (1976) gave the asymptotic properties of the Nelson-Aalen estimate. As the estimate is a jump function, the smoothed version are well developed. Specifically, kernel, wavelets and the splines are the usual methods to give the smoothed estimators of the hazard. In particular, Beran (1981), Dabrowska (1987), Gray (1992), Muller and Wang (1994) and Cai (1998) have derived smooth estimators of the hazard based on the kernel methods and neighbour smoothing methods. Likelihood methods, always with a penalty to derive a more smooth estimation, are also discussed in extensive statistical literature, see Anderson and Senthilselvan (1980), O'Sullivan (1988), Antoniadis (1989) and Antoniadis and Grégoire (1990), Gu (1996), Kooperberg and Stone (1992). This penalized maximum likelihood is well-developed and the nonparametric function is often approximated by a linear combination of the splines. Cox and O'Sullivan (1990) gave the general convergence rate of this kind of estimates in a Sobolev space. Similar to the survival function, the orthogonal series is another well-developed approach to estimate the hazard (Kronmal and Tarter (1968), Tanner and Wong (1984)). Among these, the wavelet method is much more popular (Patil (1997), Antoniadis et al. (1999), Li (2002)). As most of the times, many explanatory variables have an effect on the failure time, there exists extensive literature to model the relationship between the covariates and the hazard. Among these, the relative risk model or the Cox model (Cox (1972)), the additive hazard model (Lin and Ying (1994)) and the accelerated failure time model are well known and developed. Specifically, there has numerous literature about the Cox model. Cox (1972) and Cox (1975) proposed the model and used the partial likelihood to estimate the hazard. Kalbfleisch and Prentice (1973) generalized the marginal likelihood derivation based on the Cox regression model. Other methodologies talked about the Cox model based on the maximum likelihood approach can be found in Breslow (1974), Thompson and Godambe (1974) , Jacobsen (2012) and so on. Lin and Ying (1994) gave the estimation equation about the additive hazard model. They talked about the asymptotic properties about the estimators and through modifying the nonparametric parts, they can get a positive and monotone cumulative hazard functions. But most of the above methods just can get a discontinuous estimate of the hazard as the nonparametric part is sum of the indicator function. Fosen et al. (2006) extended the kernel methods to estimate nonparametric parts of the multiplicative and additive hazard models, which can give a smoothed hazard.

1.4.2 Interval censoring

Interval-censored data arise frequently in health or medical studies when the interest is the occurrence of an event during a pre-specified period.

Nonparametric maximum likelihood estimation (NPMLE) has been discussed in extensive literature with interval censored data. The earliest work about NPMLE with current status data, namely the case-one interval censored data was proposed by Ayer et al. (1955), Eeden (1956) and Eeden (1957). The pool-adjacent-violators algorithm was discussed to compute the estimate of the distribution function. The asymptotic properties about the NPMLE were established in Groeneboom (1987). Groeneboom and Wellner (1992) have proved that the asymptotic distribution of the NPMLE was a two sided Brownian process, then based on Groeneboom and Wellner (2012), the confidence interval of the NPMLE can be derived through the estimation of the quantile. Following from the limiting theorem, the bootstrap method can also be used to get the pointwise variance estimation and the confidence interval. Without estimating the unknown parameters in the asymptotic distribution, Banerjee and Wellner (2005) established the confidence interval for case-one interval-censored data based on the likelihood ratio test. Recently, Groeneboom et al. (2015) also talked about the confidence intervals with current status data based on the likelihood ratio test with restricted and unrestricted likelihood function. Peto (1973) and Turnbull (1976) generalized the estimation to case-two interval-censored data. The Newton-Raphson algorithm and self-consistency algorithm were presented respectively in the two literature. Specifically, as there are not a closed form of the NPMLE with case-two interval censored data, various algorithms have been developed for the NPMLE with case-two interval censored data. The iterative convex minorant (ICM) algorithm was proposed by Groeneboom and Wellner (1992) and modified by Jongbloed (1998). After modification, the algorithm can make sure the increase of the objective function after every iteration and enjoys the global convergence property. The EM-ICM algorithm, which combines the EM algorithm with the ICM algorithm and was proposed by Wellner and Zhan (1997), also enjoys the global convergence property. Böhning et al. (1996) generalized the vertex-exchange or other algorithms proposed for the finite mixture model estimation to determine the NPMLE with interval censored data. Gentleman and Geyer (1994) and Li et al. (1997) also investigated the algorithm or characterized the data structure related to the NPMLE based on interval-censored data. Besides, Groeneboom and Wellner (1992) showed that the NPMLE enjoyed the uniform consistency property with case-two interval censored data. So do van de Geer (1993) and Yu et al. (1998). Further, Schick and Yu (2000) gave a strong consistency results with L_1 norm, while Yu et al. (2000) showed the self-consistency results about the NPMLE under this scenario. As pointed out by Sun (2007), the NPMLE with case-two interval censored data also has the same limiting distributing as current status data in Groeneboom and Wellner (1992). Huang (1999) showed that under a little strong assumption, both case-one and case-two interval censored data can converge to a standard normal distribution. Then the likelihood ratio statistics or the bootstrap method can be used to get the confidence interval. Further, Goodall et al. (2004) established the confidence interval for case-two interval-censored data based on three methods. The first one is based on the full information matrix, the second is based on the nonzero estimates' information matrix while the third one is relied on the likelihood ratio inference. In addition, Sun (2001) generalized the Greenwood formula to interval-censored data to estimate the variance. Similar to the Kaplan-Meier estimator, the NPMLE is also a discontinuous function. Thereby, in order to do some sophisticated inference, smoothed estimation about the survival function or density is needed. Although the complexity of the interval censored data resulting in the difficulty of smoothing the NPMLE directly, some smoothed estimation about the survival function or density estimations were proposed by Braun et al. (2005) and Pan (2000) respectively. Vandal et al. (2005) proposed the the constrained nonparametric estimation problem to estimate the distribute function with interval censored data.

As the hazard can reveal more characters of the failure time than the survival function, there exist numerous well-developed approaches to estimate the hazard with interval censored data. The earliest work is based on the nonparametric estimation without smoothing. But the hazard derived through this way is unstable and unsuitable for graphical presentation. The smoothing estimator of the hazard mainly based on the kernel methods (Lawless (2011); Tanner and Wong (1984)) the the spline methods (Kooperberg and Stone (1992), Rosenberg (1995)), and the local likelihood methods (Tibshirani and Hastie (1987)) with parametric model. Actually, the spline methods are mainly related with the likelihood methods. Specifically, in order to get a smooth estimator of the hazard, the penalized likelihood method would be introduced and the hazard or log-hazard would be modeled as a linear combination of the spline basis. Among the basis, the B-spline basis, penalized B-spline basis (Cai and Betensky (2003)), M-spline basis (Joly et al. (1998)) are used. Similar to the right censored data, there exit some regression methods to estimate the hazard. Among them, the proportional hazards model, the additive hazards model and the proportional Odds model are the most popular ones. Specifically, for the current status data, Huang et al. (1996) and Huang and Wellner (1997) gave the rigorous and detailed discussion about the use of the proportional hazards model. They gave fundamental ground work for the asymptotic study based on the maximum of the likelihood estimation and investigated the use of the proportional hazards model. The additive hazards model for current status data were investigated by Ghosh (2001), Lin et al. (1998), and Martinussen and Scheike (2002). For case-two interval-censored failure time data, Finkelstein (1986) was the first to study the use of the PH model for interval-censored data. After that, Huang and Wellner (1997) gave the asymptotic properties of the PH model with interval censored data. Alioum and Commenges (1996), Datta et al. (2000), Huber-Carol and Vonta (2004), and Pan and Chappell (2002) generalized the PH model the analysis of survival time data that involve interval censoring as well as truncation. Others that investigated the PH model include Satten (1996), Goggins et al. (1998), Betensky et al. (2002),Cai and Betensky (2003), Huber-Carol and Vonta (2004), ect. The proportional odds model involved case-two interval censoring include Huang and Rossini (1997), Rabinowitz et al. (2000), and Shen (1998). The reviews about the accelerated failure time model include Pu and Li (1999), Li and Pu (2003), Rabinowitz et al. (1995), Betensky et al. (2001), and Xue et al. (2006).

1.5 Research of outline

The remainder of the dissertation is organized as follows. Chapter 2 presents a penalized nonparametric maximum likelihood estimation of the log-hazard function with the right censored data. In particular, the log-hazard is approximated by a linear combination of the B spline basis, which derives a smoothed estimator. A reproducing kernel Hilbert space is established with a special inner product. The most appealing fact is that a functional Bahadur representation is established in the reproducing kernel space, which serves as a key technical tool for nonparametric inference of the unknown parameter/function. Both pointwise and global asymptotic properties of the resulting estimator of the unknown log-hazard function are proved. Furthermore, the local confidence interval and simultaneous confidence band of the unknown log-hazard function are provided, along with a local and global likelihood ratio tests. We also investigate issues related to the asymptotic efficiency. The fast computing algorithm is used to calculate the estimators.

As the second main part of this thesis, the aforementioned nonparametric inference approach is extended to handle interval censored data in Chapter 3. In particular, we focus on the nonparametric inference of the cumulative hazard function, instead of the log-hazard function of the interval censored data. That is because it's not so easy to take care of the inner space with log-hazard. Similarly, we have derived a functional Bahadur representation and established the asymptotic properties of the resulting estimate of the cumulative function. Particularly, the global asymptotic properties are justified under regularity conditions. A likelihood ratio test is also provided. Besides, some constrained algorithm would be used to calculate the estimator.

Some conclusions and future work would be related in Chapter 4.

Chapter 2

Nonparametric Statistical Inference for Right Censored Data Using Smoothing Splines

This chapter focuses on the statistical inference about the penalized nonparametric maximum likelihood estimation for right censored data using the smoothing splines. A functional Bahadur representation is derived firstly, which is the key technical tool of the chapter. Based on the functional Bahadur representation, we study the asymptotic properties about the estimator. Then the local confidence interval and simultaneous confidence band about the estimator is given as by products. After that, the local and global likelihood ratio test are shown. Besides, some optimal and efficiency issues are figured out. Simulation studies are carried out to verify the theories. A real data example is demonstrated.

2.1 Introduction

In survival analysis, the outcome variable of interest is the time until the occurrence of an event, such as occurrence of a disease, death, marriage, etc. The time-to-event or survival time is usually measured in days, weeks or years, which is typically positive. Censored observations, of which the survival time is incomplete, are collected frequently in medical studies, reliability and many other fields related to survival analysis. The most commonly encountered case is right censoring. To accommodate censoring, state-of-the-art statistical methodologies have been developed in past decades, including parametric, semiparametric and nonparametric methods.

Parametric approaches assume that the underlying distributions of the time-to-events are certain known probability distributions. For example, the exponential, lognormal and Weibull distributions are among those commonly used ones. Parametric methods are appealing to practitioners owing to their convenience and ease of interpretation (Johnson and Kotz (1970), Mann et al. (1974),Lawless (2011), Kalbfleisch and Prentice (2011)). The most extensively used semiparametric model for the analysis of survival data is the celebrated Cox's proportional hazards model, in which it is assumed that the hazard function of the survival time is multiplicatively related to an unknown baseline function and the covariate; see Cox (1972),Cox (1975), Cox and Oakes (1984), Lin and

Wei (1989) and Lin and Ying (1994). In contrast to parametric models, Cox's model makes no assumption on the shape of the baseline hazard function, and provides easy-to-interpret information for the relationship of the hazard function of the survival time and the covariates. The parameter regarding the covariate effect in the Cox's model is usually estimated by maximizing the partial likelihood, and its large-sample properties are beautifully justified with the martingale theory; see Andersen and Gill (1982), Kosorok (2008), and Fleming and Harrington (2011). In the analysis of survival data, an important alternative to the Cox's proportional hazards model is the accelerated failure time model (AFT), which assumes the logarithm of the survival time is linearly related to the covariates; Kalbfleisch and Prentice (1980),Cox and Oakes (1984), Wei (1992) and Zeng and Lin (2007). Intriguing semiparametric inference methods for the AFT model have been studied thoroughly in the literature, Buckley and James (1979),Prentice (1978),Ritov (1990), Tsiatis (1990), Wei et al. (1990), Lai and Ying (1991a), Lai and Ying (1991b), Lin et al. (1993) and Lin and Chen (2013).

Parametric and semiparametric methods rely very much on the distributional or model assumption. However, the underlying distribution or model is often unknown, and the inference based on the parametric and semiparametric models may suffer from possibly mis-specification. Without making assumption about the unknown distribution or an actual model form, nonparametric inference concerned about the hazard rate, survival function and density function are proposed in the literature, as hazard function is closely tied to survival function and density function through a direct relationship. Among them, the Kaplan-Meier estimator Kaplan and Meier (1958) was the nonparametric maximum likelihood estimator, which enjoys the self-consistency and asymptotic normality, see Efron (1967) and Breslow and Crowley (1974). Although the Kaplan-Meier estimator is well developed and easy to calculate, the discontinuous property would hinder the sophisticated inference. Thereby, some smoothed estimator of the hazard and density estimators are developed. For example, with censored survival data, kernel smoothing and nearest neighbor smoothing on the time axis are well-known approaches to estimate the density function or the hazard function; see Beran (1981), Dabrowska (1987),Gray (1992). In order to avoid the selection of bandwidth, ease the computation and give a smoothed estimator, penalized likelihood methods for the estimation of the hazard rate using the smoothing splines are developed in the literature; see Anderson and Senthilselvan (1980), O'Sullivan (1988) and Rosenberg (1995). It is also known that kernel estimates reflect mostly the local structure with the data, and estimates of the density function or the hazard function based on smoothing splines with a global smoothing parameter enjoy better global properties (O'Sullivan et al. (1986)). Except some consistency properties for the smoothing splines hazard estimate were reported (Cox (1972)), to the best of our knowledge, there are limited discussion on the theoretical properties of the estimate of the hazard function using smoothing splines in the literature. Moreover, the nonparametric inference for the hazard function is subject to a positivity constraint, which makes the computation complicated. In this chapter, we target at the log-hazard rate in a nonparametric framework Kooperberg et al. (1995) provide a penalized likelihood estimate using smoothing splines. Our major contribution of this chapter is to establish the local and global asymptotic properties of the proposed log-hazard estimator.

The rest of the chapter is organized as follows. Some background and preliminary knowledge are given in section 2.2. In section 2.3, we report a new functional Bahadur representation (FBR) in the Sobolev space, and investigate the local and global asymptotic properties of the resulting estimate of the log-hazard rate; We discuss the hypothesis test in section 2.4 and some simulation results are presented in section 2.5. In Section 2.6 our method is applied to a non-Hodgkin's lymphoma dataset. Section 2.7 contains some concluding remarks and further discussions. All technical proofs are deferred to the Section 2.8.

2.2 Preliminaries

2.2.1 Notation and Methodology.

We introduce notations that will be used throughout this chapter. Let T be the survival time and let C be the censoring time. We define the observation time $Y = \min(T, C)$ and $\delta = I(T \leq C)$ be the censoring indicator, where $I(\cdot)$ is the indicating function. Moreover, we denote $\lambda(t)$ as the hazard rate/function of the survival time and $g_0(t) = \log(\lambda(t))$. $\lambda(t) : \mathbb{I} \mapsto R$ is bounded away from 0 and infinity (this already is assumption). Without loss of generality, we consider $\mathbb{I} = [0, 1]$. Suppose that the observed data $(Y_i, \delta_i), i = 1, \ldots, n$, are independent and identically distributed (i.i.d) copies of (Y, δ). Then, the log-likelihood of g is:

$$l_n(g) = -\int_{\mathbb{I}} \exp\{g(t)\} S_n(t)\, dt + \frac{1}{n} \sum_{i=1}^{n} \delta_i g(Y_i),$$

where $S_n(\cdot)$ is the empirical survival function of Y; see O'Sullivan (1988). Let $l(g) \equiv E\{l_n(g)\}$. A direct calculation yields that

$$l(g) = -\int_{\mathbb{I}} \exp\{g(t)\} S(t)\, dt + \int_{\mathbb{I}} \exp\{g_0(t)\} g(t) S(t)\, dt,$$

where $S(t)$ is the survival function of Y. Throughout this chapter, we consider the true target function $g_0(t)$ belongs to the mth-order Sobolev space $\mathcal{H}^m(\mathbb{I})$ shorten as $\mathcal{H}^m$:

$$
\begin{aligned}
&\mathcal{H}^m(\mathbb{I}) \\
=&\{g : \mathbb{I} \mapsto R | g^{(j)} \text{ is absolutely continuous for } j = 0, 1, \ldots, m-1, g^{(m)} \in L_2(\mathbb{I})\}.
\end{aligned}
$$

where the constant $m > 1/2$ and is assumed to be known, $g^{(j)}$ is the jth derivative of g and $L_2(\mathbb{I})$ is the L_2 space defined in $\mathbb{I}$. Define $J(g, \tilde{g}) = \int_{\mathbb{I}} g^{(m)}(t)\tilde{g}^{(m)}(t)\,dt$. The penalized likelihood of $g(\cdot)$ is defined as:

$$
l_{n,\lambda}(g) = l_n(g) - \frac{\lambda}{2}J(g, g),
$$

where $J(g, g)$ is the roughness penalty and λ is the smoothing parameter, which converges to 0 as $n \to \infty$.

For the inference of $g_0(t)$, we propose to use B-splines to approximate g in $l_{n,\lambda}(g)$. For the finite closed interval $\mathbb{I}$, we define a partition of $\mathbb{I}$:

$$
0 = t_1 = \ldots = t_l < t_{m+1} < \ldots < t_{m_n+m} < t_{m_n+m+1} = \ldots = t_{m_n+2m} = 1,
$$

with which $[0, 1]$ is partitioned into $m_n + 1$ subintervals with knots set $\mathcal{I} \equiv \{t_i\}^{m_n+2m}$, and $m_n = o(n^v)$ for $0 < v < 1/2$. Let $\{B_{i,m}, 1 \leq i \leq q_n\}$ denote the B-spline basis functions with $q_n = m_n + m$. Let $\Psi_{m,\mathcal{I}}$ (with order m and knots $\mathcal{I}$) be the linear space spanned by the B-spline functions, that is

$$
\Psi_{m,\mathcal{I}} = \{\sum_{i=1}^{q_n} \theta_i B_{i,m} : \theta_i \in R, i = 1, \ldots, q_n\}.
$$

It follows from Schumaker (1981) that there exists a smoothing spline $g_n(t) \in \Psi_{m,\mathcal{I}}$ such that $\|g_n(t) - g_0(t)\|_\infty = O(n^{-vm})$ and $\|g(t)\|_\infty \equiv \sup_{t\in\mathbb{I}} |g(t)|$. Now, we define

$$\begin{aligned}\hat{g}_{n,\lambda} &\equiv \arg\max_{g\in\Psi_{m,\mathcal{I}}} l_{n,\lambda}(g) \\ &= \arg\max_{g\in\Psi_{m,\mathcal{I}}} \left\{ l_n(g) - \frac{\lambda}{2} J(g,g) \right\}\end{aligned}$$

as the estimator of $g_0(t)$. The numerical implementation of solving the above objective function is available in O'Sullivan (1988) with a fast computation algorithm. Moreover, a data-driven method based on AIC criterion was suggested to select the smoothing parameter λ.

2.2.2 Reproducing Kernel Hilbert Space

We now present some useful properties about the reproducing kernel Hilbert space (RKHS) as in Shang and Cheng (2013). First of all, it is known that when $m > 1/2$, $\mathcal{H}^m$ is a RKHS with the inner product $< g, \tilde{g} >_\lambda = \int_{\mathbb{I}} g(t)\tilde{g}(t)\exp\{g_0(t)\}S(t)\,dt + \lambda J(g,\tilde{g})$ and the norm $\|g\|_\lambda^2 =< g, g >_\lambda$. Furthermore, there exists a positive definite self-adjoint operator $W_\lambda : \mathcal{H}^m \mapsto \mathcal{H}^m$, which satisfies: $< W_\lambda g, \tilde{g} >_\lambda = \lambda J(g,\tilde{g})$ for any $g, \tilde{g} \in \mathcal{H}^m$ Denote $V(g,\tilde{g}) = \int_{\mathbb{I}} g(t)\tilde{g}(t)\exp\{g_0(t)\}S(t)\,dt$. Then, it follows directly that

$$< g, \tilde{g} >_\lambda = V(g,\tilde{g}) + < W_\lambda g, \tilde{g} >_\lambda .$$

Let $K(\cdot,\cdot)$ be the reproducing kernel of $\mathcal{H}^m$ defined on $\mathbb{I}\times\mathbb{I}$. Then it is known to possess the following properties:

(P_1) $K_t(\cdot) = K(t,\cdot)$ and $< K_t, g >_\lambda = g(t)$ for any g in $\mathcal{H}^m$ and any t in $\mathbb{I}$.

(P_2) There exists a constant c_m depending on m only, such that $\|K_t\|_\lambda \leq c_m h^{-1/2}$ for any $t \in \mathbb{I}$ and $h = \lambda^{1/(2m)}$. Hence, we have $\|g(t)\|_\infty \leq c_m h^{-1/2}\|g\|_\lambda$ for any $g \in \mathcal{H}^m$.

We denote two positive sequences a_n and b_n as $a_n \asymp b_n$ if they satisfy $\lim_{n\to\infty}(a_n/b_n) = c > 0$. There exists a sequence of eigenfunctions $h_j \in \mathcal{H}^m$ and eigenvalues γ_j satisfying the following properties:

(P_3) $\sup_{j\in N}\|h_j\|_\infty < \infty$, $\gamma_j \asymp j^{2m}$, where $N = \{0, 1, \ldots\}$;

(P_4) $V(h_i, h_j) = \delta_{ij}$, $J(h_i, h_j) = r_j\delta_{ij}$, where δ_{ij} is a Kronecker delta, that is $\delta_{ij} = 1$ when $i = j$; otherwise, $\delta_{ij} = 0$.

(P_5) For any $g \in \mathcal{H}^m$, we have $g = \sum_{j=0}^{\infty} V(g, h_j)h_j$ with convergence in the $\|\cdot\|_\lambda$-norm.

(P_6) For any $g \in \mathcal{H}^m$ and $t \in \mathbb{I}$, we have $\|g\|_\lambda^2 = \sum_{j=0}^{\infty} V(g, h_j)^2(1 + \lambda\gamma_j)$, $K_t(\cdot) = \sum_{j=0}^{\infty} h_j(t)h_j(\cdot)/(1 + \lambda\gamma_j)$ and $W_\lambda h_j(\cdot) = (\lambda\gamma_j)/(1 + \lambda\gamma_j)h_j(\cdot)$.

Following Shang and Cheng (2013, page 2613), the eigenvalues and eigenfunctions can be solved through the ordinary differential equations

(ODE):

$$(-1)^m h_j^{(2m)}(\cdot) = \gamma_j \exp(\hat{g}(\cdot)) S_n(\cdot) h_j(\cdot),$$
$$h_j^{(k)}(0) = h_j^{(k)}(1) = 0, \qquad k = m, m+1, \cdots, 2m-1. \tag{2.1}$$

For ease of presentation, we introduce more notations related to the Fréchet derivatives. Let $\mathcal{S}_n(g)$ and $\mathcal{S}_{n,\lambda}(g)$ be the Fréchet derivatives of $l_n(g)$ and $l_{n,\lambda}(g)$, respectively. Similarly, let $\mathcal{S}(g)$ and $\mathcal{S}_\lambda(g)$ be the Fréchet derivatives of $l(g)$ and $l_\lambda(g)$, respectively. Let D be the Fréchet derivative operator and $g_1, g_2, g_3 \in \mathcal{H}^m$ be any direction. Then, we have

$$\begin{aligned} Dl_{n,\lambda}(g)g_1 &= -\int_{\mathbb{I}} \exp\{g(t)\} g_1(t) S_n(t)\, dt + \frac{1}{n}\sum_{i=1}^{n} \delta_i g_1(Y_i) - < W_\lambda g, g_1 >_\lambda \\ &= < \mathcal{S}_n(g), g_1 >_\lambda - < W_\lambda g, g_1 >_\lambda \\ &= < \mathcal{S}_{n,\lambda}(g), g_1 >_\lambda, \end{aligned}$$

where $\mathcal{S}_n(g) = -\int_{\mathbb{I}} \exp\{g(t)\} K_t S_n(t)\, dt + n^{-1}\sum_{i=1}^{n} \delta_i K_{Y_i}$ and $\mathcal{S}_{n,\lambda}(g) = \mathcal{S}_n(g) - W_\lambda g$. Moreover,

$$D^2 l_{n,\lambda}(g) g_1 g_2 = -\int_{\mathbb{I}} \exp\{g(t)\} g_1(t) g_2(t) S_n(t)\, dt - < W_\lambda g_1, g_2 >_\lambda,$$

$$D^3 l_{n,\lambda}(g) g_1 g_2 g_3 = -\int_{\mathbb{I}} \exp\{g(t)\} g_1(t) g_2(t) g_3(t) S_n(t)\, dt.$$

Further, by a direct calculation, we can express

$$\begin{aligned}\mathcal{S}(g) &= Dl(g)\\ &= -\int_{\mathbb{I}} \exp\{g(t)\}K_t S(t)\,dt + \int_{\mathbb{I}} \exp\{g_0(t)\}K_t S(t)\,dt = E\{\mathcal{S}_n(g)\},\end{aligned}$$

as well as $\mathcal{S}_\lambda(g) = \mathcal{S}(g) - W_\lambda g$. Besides,

$$D\{\mathcal{S}(g)g_1\}g_2 = D^2 l(g)g_1 g_2 = -\int_{\mathbb{I}} \exp\{g(t)\}g_1(t)g_2(t)S(t)\,dt.$$

Hence, we have the following result:

$$\begin{aligned}< D\mathcal{S}_\lambda(g_0)f, g >_\lambda &= < D\{\mathcal{S}(g_0) - W_\lambda g_0\}f, g >_\lambda\\ &= < D\mathcal{S}(g_0)f, g >_\lambda - < W_\lambda f, g >_\lambda\\ &= < -\int_{\mathbb{I}} \exp\{g(t)\}f(t)K_t S(t)\,dt, g >_\lambda - < W_\lambda f, g >_\lambda\\ &= -\int_{\mathbb{I}} g(t)f(t)\exp\{g_0(t)\}S(t)\,dt - \lambda J(g,f)\\ &= - < f, g >_\lambda .\end{aligned}$$

Proposition 2.1. *$D\mathcal{S}_\lambda(g_0) = -id$, where id is the identity operator on $\mathcal{H}^m$.*

This proposition will be playing an important role in deriving a functional Bahadur representation (FBR) about the proposed estimator.

2.3 Functional Bahadur Representation

In this section, we derive and present the major technical tool: functional Bahadur representation (FBR), which laid down a theoretical foundation for statistical inference procedures in later sections. With the help of the FBR, we establish the asymptotic normality of the proposed smoothing spline estimate. Likelihood ratio test procedure are also justified rigorously. To begin with, we present a lemma regarding the consistency of the proposed estimate for obtaining the FBR. All theoretical conditions and proofs are deferred to Appendix.

Lemma 2.1. *(Consistency) Suppose conditions(C2.1)-(C2.3) given in Appendix hold. Then, if* $\lambda(n^{(1-v)/2} + n^{vm}) \to 0$ *as* $n \to \infty$ *for* $0 < v < 1/2$, *we have*

$$\|\hat{g}_{n,\lambda} - g_0\|_\infty = o_p(1),$$

$$J(\hat{g}_{n,\lambda} - g_0, \hat{g}_{n,\lambda} - g_0) < \tilde{C},$$

where $\tilde{C}$ *is a constant larger than* 1.

In fact, the consistency of the estimator with the infinity norm can be derived by Cox and O'Sullivan (1990). But the second theoretical result is given firstly by us.

To obtain the rate of convergence of the proposed estimator, we next drive a concentration inequality of certain empirical process. Define $\mathcal{G} =$

$\{g \in \mathcal{H}^m : \|g\|_\infty \le 1, J(g,g) \le \tilde{C}\}$, where $\tilde{C}$ is specified in Lemma 2.1. We next define

$$\mathcal{Z}_n(g) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\varphi_n(Y_i, g) - E\{\varphi_n(Y_i, g)\}],$$

where $\varphi_n(Y_i, g)$ is a real-valued function in $\mathcal{H}^m$.

Lemma 2.2. *Suppose that $\varphi_n(Y, g)$ satisfies the following condition:*

$$\|\varphi_n(Y, f) - \varphi_n(Y, g)\|_\lambda \le \|f - g\|_\infty \qquad \text{for any} \qquad f, g \in \mathcal{G}. \tag{2.2}$$

Then, we have

$$\lim_{n\to\infty} P\left[\sup_{g\in\mathcal{G}} \frac{\|\mathcal{Z}_n(g)\|_\lambda}{\|g\|_\infty^{1-1/(2m)} + n^{-1/2}} \le \{5 \log\log(n)\}^{1/2}\right] = 1.$$

By Lemma 2.1 and Lemma 2.2, we obtain the convergence rate of our estimate which is presented in the following theorem:

Theorem 2.1. *(Convergence Rate) Assume that conditions(C2.1)-(C2.3) given in Appendix are satisfied. Then, when* $\log\{\log(n)\}/(nh^2) \to 0$, $\lambda\{n^{(1-v)/2} + n^{vm}\} \to 0$ *as* $n \to \infty$, *we have*

$$\|\hat{g}_{n,\lambda} - g_0\|_\lambda = O_p\big((nh)^{-1/2} + h^m\big).$$

Remark 1. If $h \asymp n^{-1/(2m+1)}$, Theorem 2.1 suggests that $\hat{g}_{n,\lambda}$ achieves the optimal rate of convergence when we estimate $g_0 \in \mathcal{H}^m$, that is $O_p(n^{-m/(2m+1)})$.

Based on Theorem 2.1, we are ready to present the key technical tool of this chapter, a new version of functional Bahadur representation compare with that of Shang and Cheng (2013).

Theorem 2.2. *(Functional Bahadur Reprensentation) Assume that conditions(C2.1)-(C2.3) hold. Then, if* $\log\{\log(n)\}/(nh^2) \to 0$, $\lambda(n^{(1-v)/2} + n^{vm}) \to 0$ *as* $n \to \infty$, *we have*

$$\|\hat{g}_{n,\lambda} - g_0 - \mathcal{S}_{n,\lambda}(g_0)\|_\lambda = O_p(\alpha_n),$$

where $\mathcal{S}_{n,\lambda}(g_0) = n^{-1}\sum_{i=1}^n \int_{\mathbb{I}} \exp(g_0(t))K_t\, dM_i(t) - W_\lambda(g_0)$,

$$\alpha_n = n^{-1/2-vm} + n^{-vm}\{(nh)^{-1/2} + h^m\} + h^{-1/2}\{(nh)^{-1} + h^{2m}\} \\ + h^{-(6m-1)/(4m)}n^{-1/2}\{\log\log(n)\}^{1/2}\{(nh)^{-1/2} + h^m\}.$$

In fact, one of the key results leading to the FBR in Theorem 2.2 is Proposition 2.1, which can be seen from the proofs of Theorem 2.2 in Appendix. Moreover, Theorem 2.2 reveals that the "bias" of our estimate $\hat{g}_{n,\lambda}$ is approximately $\mathcal{S}_{n,\lambda}(g_0)$, a sum of certain independent and identical random variables. Applying this result, we immediately obtain the following result regarding the asymptotic normality:

Theorem 2.3. *Assume conditions (C2.1)-(C2.3) hold. For* $m > 3/4 + \sqrt{5}/4$ *and* $1/(4m) \le v \le 1/(2m)$, *suppose* $nh^{4m-1} \to 0$ *and* $nh^3 \to \infty$ *as* $n \to \infty$. *Then, we have:*

(i) For any $t_0 \in \mathbb{I}$,

$$\sqrt{nh}\{\hat{g}_{n,\lambda}(t_0) - g(t_0) - W_\lambda g_0(t_0)\} \xrightarrow{d} N(0, \sigma_{t_0}^2),$$

where $\sigma_{t_0}^2 \equiv \lim_{h\to 0} h \sum_{j=0}^{\infty} h_j^2(t_0)/(1+\lambda\gamma_j)^2$ and $\xrightarrow{d}$ means converges in distribution.

(ii) Moreover, we have $\sqrt{nh}\{\hat{g}_{n,\lambda}(t) - g_0(t) - W_\lambda g_0(t)\}$ converges weakly in $\mathbb{I}$ to a mean zero Gaussian process $Z(t)$ with the covariance function at (s,t) equals to $\Sigma(s,t)$, where

$$\Sigma(s,t) = \lim_{h\to 0} h \sum_{j=0}^{\infty} \frac{h_j(t)h_j(s)}{(1+\lambda\gamma_j)^2}.$$

Remark 2. The second part of Theorem 2.2 provides the weak convergence of our proposed estimate. To the best of our knowledge, there is no report in the literature regarding the weak convergence to a Gaussian process of a smoothing spline-type estimate.

Corollary 2.1. *Assume conditions(C2.1)-(C2.3) hold. For $m > 3/2$ and $1/(4m) \le v \le 1/(2m)$, suppose $nh^{2m} \to 0$ and $nh^3 \to \infty$ as $n \to \infty$. Then, for any $t_0 \in \mathbb{I}$, we have*

$$\sqrt{nh}\{\hat{g}_{n,\lambda}(t_0) - g_0(t_0)\} \xrightarrow{d} N(0, \sigma_{t_0}^2)$$

with $\sigma_{t_0}^2$ defined in Theorem 2.1. In addition, we show that $\sqrt{nh}\{\hat{g}_{n,\lambda}(t) - g_0(t)\}$ converges weakly in $\mathbb{I}$ to a zero-mean Gaussian process $Z(t)$ with *the covariance function at (s,t) equals to* $\Sigma(s,t)$.

Remark 3. Corollary 2.1 implies that, under certain under-smoothing conditions, the asymptotic bias for the estimation of $g_0(t_0)$ vanishes. Hence, Corollary 2.1 together with the so-called Delta-method immediately gives the pointwise confidence interval (CI) for some real-valued smooth function of $g_0(t)$ at any fixed point $t_0 \in \mathbb{I}$, denoted by $\rho\big(g_0(t_0)\big)$. Let $\dot{\rho}(\cdot)$ be the first derivative of $\rho(\cdot)$. By Corollary 2.1, for any fixed point $t_0 \in \mathbb{I}$, if $\dot{\rho}\big(g_0(t_0)\big) \neq 0$, we have

$$P\left(\rho\big(g_0(t_0)\big) \in \left[\rho\big(\hat{g}_{n,\lambda}(t_0)\big) \pm \Phi_{\frac{\alpha}{2}} \frac{\dot{\rho}\big(g_0(t_0)\big)\sigma_{t_0}}{\sqrt{nh}}\right]\right) \to 1-\alpha$$

as $n \to \infty$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function and Φ_α is the lower α-th quantitle of $\Phi(\cdot)$, that is $\Phi(\Phi_\alpha) = \alpha$.

As for the simultaneous confidence band, we employ the resampling method of Lin et al. (1993) for distributional approximation. For illustration, let $(G_1, \ldots, G_n)$ be independent standard normal random variables, independent of the data $(Y_i, \delta_i), i = 1, \ldots, n$. It can be shown that the distribution of the limiting process $Z(t)$ can be approximated by that of the following zero-mean Gaussian process

$$\hat{Z}(t) \equiv \frac{1}{\sqrt{nh^{-1}}} \sum_{i=1}^{n} \int_{\mathbb{I}} K_t(s)\, dM_i(s) G_i, \tag{2.3}$$

where $M_i(t) \equiv N_i(t) - \int_0^t I(Y_i \geq s) \exp\{g_0(s)\}\, ds$, and $N_i(t) = I(Y_i \leq$

$t, \delta_i = 1)$. It is well-known that $M_i(t)$ is a martingale. In view of this fact, we obtain a large number of realizations of $\hat{Z}(t)$ by repeatedly generating the standard normal random samples $(G_1, \ldots, G_n)$ while fixing the data. Thus, one may use the empirical distribution of these random samples to approximate the distribution of $Z(t)$. In particular, the α-percentile of $\sup_{t \in \mathbb{I}} |Z(t)|$ can be approximated by the empirical percentile of a large number of realizations of $\sup_{t \in \mathbb{I}} |\hat{Z}(t)|$, denoted by $\hat{G}_\alpha$. As a result, we can construct the global confidence band of $g_0(t)$ as follows:

$$\left(\hat{g}_{n,\lambda}(t) - \frac{1}{\sqrt{nh}} \hat{G}_\alpha, \quad \hat{g}_{n,\lambda}(t) + \frac{1}{\sqrt{nh}} \hat{G}_\alpha \right).$$

2.4 Likelihood Ratio Test

With the help of the FBR, we consider further inference of $g_0(\cdot)$ by testing local and global hypothesis. In this section, we focus on likelihood ratio tests for testing $g_0(\cdot)$.

2.4.1 Local Likelihood Ratio Test

We consider the following hypothesis for some pre-specified (t_0, ω_0):

$$H_0 : g(t_0) = \omega_0 \qquad \text{versus} \qquad H_1 : g(t_0) \neq \omega_0.$$

The penalized log-likelihood under H_0, or the "constrained" penalized log-likelihood by Shang and Cheng (2013), is defined as:

$$L_{n,\lambda}(g) = -\int_{\mathbb{I}} \exp\{g(t) + \omega_0\} S_n(t)\, dt + \frac{1}{n} \sum_{i=1}^{n} \delta_i \{g(Y_i) + \omega_0\} - \frac{\lambda}{2} J(g, g),$$

where $g \in \mathcal{H}_0 = \{g \in \mathcal{H}^m : g(t_0) = 0\}$. We consider the following likelihood ratio test (LRT) statistic:

$$LRT_{n,\lambda} = l_{n,\lambda}(\omega_0 + \hat{g}^0_{n,\lambda}) - l_{n,\lambda}(\hat{g}_{n,\lambda}),$$

where $\hat{g}^0_{n,\lambda} \equiv \arg\max_{g \in \Psi^0_{m,\mathcal{I}}} L_{n,\lambda}(g)$ is the MLE of g in

$$\Psi^0_{m,\mathcal{I}} = \{\sum_{i=1}^{q_n} \theta_i B_{i,m}, \sum_{i=1}^{q_n} \theta_i B_{i,m}(t_0) = 0\}.$$

Clearly, $\mathcal{H}_0$ is a closed subset in $\mathcal{H}^m$, and hence it is a Hilbert space endowed with the norm $\|\cdot\|_\lambda$.

The following proposition states the reproducing kernel and penalty operator of $\mathcal{H}_0$ inherited from $\mathcal{H}^m$ without proofs.

Proposition 2.2. *The reproducing kernel and penalty operator of $\mathcal{H}_0$ inherited from $\mathcal{H}^m$ satisfy the following properties:*

(a) Recall that $K(t_1, t_2)$ is the reproducing kernel for $\mathcal{H}^m$ under $<\cdot,\cdot>_\lambda$. Then, the bivariate function

$$K^*(t_1, t_2) = K(t_1, t_2) - K(t_0, t_1) K(t_0, t_2) / K(t_0, t_0)$$

is a reproducing kernel for $(\mathcal{H}_0, <\cdot,\cdot>_\lambda)$*. That is, for any* $t \in \mathbb{I}$ *and* $g \in \mathcal{H}_0$*, we have* $K_t^* \equiv K^*(t,\cdot) \in \mathcal{H}_0$ *and* $< K_t^*, g >_\lambda = g(t)$*. Moreover, we have* $\|K^*\|_\lambda \le \sqrt{2} c_m h^{-1/2}$*, where* c_m *is the same as in* P_2*.*

(b) The operator W_λ^**, defined by* $W_\lambda^* g \equiv W_\lambda g - (W_\lambda g)(t_0) K_{t_0}/K(t_0,t_0)$*, is bounded linear from* $\mathcal{H}_0$ *to* $\mathcal{H}_0$ *and satisfies* $< W_\lambda^* g, \tilde{g} >= \lambda J(g,\tilde{g})$*, for any* $g, \tilde{g} \in \mathcal{H}_0$*.*

Based on Proposition 2.2, we are in the position to derive the functional Bahadur representation (FBR) for $\hat{g}_{n,\lambda}^0$ under null hypothesis, or the so-called "restricted" FBR for $\hat{g}_{n,\lambda}^0$, which will be used to obtain the limiting distribution under null. A direct calculation yields the Fréchet derivatives of $L_{n,\lambda}$ (along directions in $\mathcal{H}_0$). Consider $g_1, g_2, g_3 \in \mathcal{H}_0$. The first-order of Fréchet derivative of $L_{n,\lambda}$, denoted by $\mathcal{S}_{n,\lambda}^0$, can be calculated as follows:

$$
\begin{aligned}
&DL_{n,\lambda}(g)g_1 \\
&= -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t)g_1(t)\,dt + \frac{1}{n}\sum_{i=1}^n \delta_i g_1(Y_i) \\
&\qquad - < W_\lambda^* g, g_1 >_\lambda \\
&= -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t) < K_t^*, g_1 >_\lambda \; dt + \frac{1}{n}\sum_{i=1}^n \delta_i < K_{Y_i}^*, g_1 >_\lambda \\
&\qquad - < W_\lambda^* g, g_1 >_\lambda \\
&=< -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t)K_t^*\,dt, g_1 >_\lambda + \frac{1}{n}\sum_{i=1}^n \delta_i < K_{Y_i}^*, g_1 >_\lambda \\
&\qquad - < W_\lambda^* g, g_1 >_\lambda \\
&=< \mathcal{S}_n^0(g), g_1 >_\lambda - < W_\lambda^* g, g_1 >_\lambda \\
&=< \mathcal{S}_{n,\lambda}^0(g), g_1 >_\lambda,
\end{aligned}
$$

where $\mathcal{S}_n^0(g) = -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t)K_t^*\,dt + n^{-1}\sum_{i=1}^n \delta_i K_{Y_i}^*$ and $\mathcal{S}_{n,\lambda}^0(g) = \mathcal{S}_n^0(g) - W_\lambda^* g$. Define $\mathcal{S}^0(g) \equiv E\{\mathcal{S}_n^0(g)\}$ and $\mathcal{S}_\lambda^0(g) \equiv \mathcal{S}^0(g) - W_\lambda^* g$. Next, we denote the second-order and the third-order Fréchet derivatives of $L_{n,\lambda}(g)$ as $D^2L_{n,\lambda}(g)g_1g_2$ and $D^3L_{n,\lambda}(g)g_1g_2g_3$ respectively. Further calculation yields that

$$
D^2L_{n,\lambda}(g)g_1g_2 = -\int_0^1 \exp\{g(t)+\omega_0\}S_n(t)g_1(t)g_2(t)\,dt - < W_\lambda^* g_2, g_1 >_\lambda,
$$

and

$$D^3 L_{n,\lambda}(g) g_1 g_2 g_3 = -\int_0^1 \exp\{g(t) + \omega_0\} S_n(t) g_1(t) g_2(t) g_3(t)\, dt.$$

We consider the derivative of $\mathcal{S}_\lambda^0(g)$ and obtain

$$D\mathcal{S}_\lambda^0(g) g_1 g_2 = -\int_0^1 \exp\{g(t) + \omega_0\} S(t) g_1(t) g_2(t)\, dt - < W_\lambda^* g_2, g_1 >_\lambda .$$

Then, by defining $g_0^0(t) = g_0(t) - \omega_0$, we got the following important equation:

$$\begin{aligned} < D\mathcal{S}_\lambda^0(g_0^0) f, g >_\lambda &= < D\{\mathcal{S}^0(g_0^0)\} f, g >_\lambda - < W_\lambda^* f, g > \\ &= -\int_0^1 \exp\{g_0^0(t) + \omega_0\} S(t) f(t) g(t)\, dt - < W_\lambda^* f, g >_\lambda \\ &= - < f, g > . \end{aligned}$$

We state this result as the next proposition.

Proposition 2.3. $D\mathcal{S}_\lambda^0(g_0^0) = -id$, *where id is the identity operator.*

Similar to Theorem 2.1 in Section 2.3, we need to prove the rate of convergence of the resulting estimator so as to obtain the FBR.

Proposition 2.4. *(Convergence Rate) Assume conditions(C2.1)-(C2.3) hold. Under H_0, if $\log\{\log(n)\}/(nh^2) \to 0$, $\lambda(n^{(1-v)/2} + n^{vm}) \to 0$ as $n \to \infty$, we have*

$$\|\hat{g}_{n,\lambda}^0 - g_0^0\|_\lambda = O_p((nh)^{-1/2} + h^m).$$

The proof of Proposition 2.4 is similar to that of Theorem 2.1. Hence it is omitted. The next theorem follows directly from Propositions 2.2-2.4.

Theorem 2.4. *(Restricted Functional Bahadur Representation) Suppose that conditions(C2.1)-(C2.3) are satisfied. Also, we assume that under* H_0, $\log\{\log(n)\}/(nh^2) \to 0$, $\lambda(n^{(1-v)/2} + n^{vm}) \to 0$ *as* $n \to \infty$*. Then, we have*

$$\|\hat{g}^0_{n,\lambda} - g^0_0 - \mathcal{S}^0_{n,\lambda}(g^0_0)\|_\lambda = O_p(\alpha_n),$$

where α_n *is defined in Theorem 2.2.*

Our main result on the local likelihood ratio test follows immediately from Theorem 2.4 and is presented below.

Theorem 2.5. *(Local Likelihood Ratio Test) Assume conditions(C2.1)-(C2.3) hold. For* $m > (5 + \sqrt{21})/4$ *and* $1/(4m) \le v \le 1/(2m)$*, suppose that* $nh^{2m} \to 0$ *and* $nh^4 \to \infty$ *as* $n \to \infty$*. Furthermore, for any* $t_0 \in \mathbb{I}$*, if* $\sigma_{t_0} \neq 0$*, let* $c_{t_0} = \lim_{h\to 0} V(K_{t_0}, K_{t_0})/\|K_{t_0}\|^2_\lambda \in (0, 1]$*. Then, under* H_0*, we have:*

(i) $\|\hat{g}_{n,\lambda} - \hat{g}^0_{n,\lambda} - \omega_0\|_\lambda = O_p(n^{-1/2})$;

(ii) $-2nLRT_{n,\lambda} = n\|\hat{g}_{n,\lambda} - \hat{g}^0_{n,\lambda} - \omega_0\|^2_\lambda + o_p(1)$;

(iii) $-2nLRT_{n,\lambda} \xrightarrow{d} c_{t_0}\chi^2_1$.

Remark 4. The central Chi-square limiting distribution in part (iii) above is established under those under-smoothing assumptions in Theorem 2.5. One may also relax those conditions for h at the price of obtaining a noncentral Chi-square limiting distribution. We also note that the convergence rate stated in Theorem 2.5 is reasonable under local restriction.

2.4.2 Global Likelihood Ratio Test

It is of paramount importance to study the global behavior of a smooth function. In this section, we consider the following "global" hypothesis:

$$H_0^{global} : g = g_0 \qquad \text{versus} \qquad H_1 : g \neq g_0,$$

where $g_0 \in \mathcal{H}^m$ can be either known or with unknown. The penalized likelihood ratio rest (PLRT) statistic is defined as:

$$\text{PLRT}_{n,\lambda} \equiv l_{n,\lambda}(g_0) - l_{n,\lambda}(\hat{g}_{n,\lambda}).$$

We next derive the null limiting distribution of $\text{PLRT}_{n,\lambda}$.

Theorem 2.6. *Assume conditions(C2.1)-(C2.3) hold. For* $m > (3+\sqrt{5})/4$ *and* $1/(4m) \leq v \leq 1/(2m)$*, suppose that* $nh^{2m+1} = O(1)$ *and* $nh^3 \to \infty$ *as* $n \to \infty$*. Define* $\sigma_\lambda^2 \equiv \sum_{j=0}^{\infty} h/(1+\lambda\gamma_j)$*,* $\rho_\lambda^2 \equiv \sum_{j=0}^{\infty} h/(1+\lambda\gamma_j)^2$*,* $\gamma_\lambda \equiv \sigma_\lambda^2/\rho_\lambda^2$*,* $\nu_\lambda \equiv h^{-1}\sigma_\lambda^4/\rho_\lambda^2$*. Under* H_0^{global}*, we have*

$$(2\nu_\lambda)^{-1/2}(-2n\gamma_\lambda PLRT_{n,\lambda} - n\gamma_\lambda\|W_\lambda g_0(t)\|_\lambda^2 - \nu_\lambda) \xrightarrow{d} N(0,1).$$

In fact, the null limiting distribution above remains unchanged when g_0 in the null hypothesis is unknown. Moreover, it can be easily verified that $h \asymp n^{-d}$ with $1/(2m+1) \leq d < 1/3$ satisfies those conditions in Theorem 2.6. We can also show that $n\|W_\lambda g_0\|^2 = o(h^{-1}) = o(\nu_\lambda)$. Thus, $-2n\gamma_\lambda PLRT_{n,\lambda}$ is asymptotically $N(\nu_\lambda, 2\nu_\lambda)$, which approaches $\chi^2_{\nu_\lambda}$ as n goes to infinity. In other words, we have

$$-2n\gamma_\lambda PLRT_{n,\lambda} \sim \chi^2_{\nu_\lambda},$$

suggesting the Wilks phenomenon holds for the PLRT.

Lastly, to conclude this section, we show that the PLRT achieves the optimal minimax rate of testing given by Ingster (1993) based on the uniform version of the FBR. To this end, we consider the alternative hypothesis $H_{1n} : g = g_{n_0}$, where $g_{n_0} = g_0 + g_n$, $g_0 \in \mathcal{H}^m$ and g_n belongs to the alternatives value set $\mathcal{A} = \{g \in \mathcal{H}^m, \exp\{g_n(t)\} \leq \zeta, J(g,g) \leq \zeta\}$ for some constant $\zeta > 0$.

Theorem 2.7. *Assume that conditions(C2.1)-(C2.3) are satisfied. For $m > (3+\sqrt{5})/4$ and $1/(4m) \leq v \leq 1/(2m)$, suppose that $h \asymp n^{-d}$ for $1/(2m+1) \leq d < 1/3$ and uniformly over $g_n \in \mathcal{A}$, $\|\hat{g}_{n,\lambda} - g_{n_0}\|_\lambda = O_p\big((nh)^{-1/2} + h^m\big)$ holds under $H_{1n} : g = g_{n_0}$. Then, for any $\delta \in (0,1)$, there always exist positive constants C and N such that*

$$\inf_{n \geq N} \inf_{g_n \in \mathcal{A}, \|g_n\|_\lambda \geq C\eta_n} P(\textit{reject} \quad H_0^{global} | H_{1n} \ \textit{is} \ \ \textit{true}) \geq 1 - \delta,$$

where $\eta_n \geq \sqrt{h^{2m} + (nh^{1/2})^{-1}}$. *Moreover, the minimal lower bound of* η_n *is* $n^{-2m/(4m+1)}$, *which can be achieved when* $h = h^{**} = n^{-2/(4m+1)}$.

Importantly, when $h = h^{**} = n^{-2/(4m+1)}$, Theorem 2.7 proves that the PLRT can detect any local alternatives with separation rate no faster than $n^{-2m/(4m+1)}$, which is actually the minimax rate of hypothesis testing in the sense of Ingster (1993).

2.5 Simulation Results

To verify the theoretical results, we present three simulated examples in this section. In the simulation studies, we set $\nu = 1/5$ and the number of knots is $[3 \times n^{1/5}]$, where $[x]$ is the integer part of x. Fore ease of presentation, more notations are needed. We define

$$H \equiv \int_0^1 \exp\{g(t)\}B(s)B(s)^\top S_n(s)\,ds,$$

$$\Omega_{lk} \equiv \int_0^1 \ddot{B}_{l,m}(s)\ddot{B}_{k,m}(s)\,ds, \qquad l, k = 1, 2, \ldots, q_n,$$

$\Omega = (\Omega_{lk})$ which is the matrix with the (l, k) element being Ω_{lk} and $\ddot{B}_{l,m}(s)$ is the second derivative of $B_{l,m}(s)$. The following AIC criterion proposed by O'Sullivan (1988) is used to select the smoothing parameter λ:

$$AIC(\lambda) = -l_n(\hat{g}_{n,\lambda}) + \frac{trace[(\hat{H} + \lambda\Omega)^{-1}\hat{H}]}{n},$$

In linear algebra, the trace of an n-by-n square matrix A is defined to be the sum of the elements on the main diagonal (the diagonal from the upper left to the lower right) of A.

Example 1: In this example, the failure time follows Beta$(1, \beta_1)$ and the censoring time follows Beta$(1, 1-\beta_1)$, where β_1 is chosen to yield 20% or 40% censoring rate. The sample size $n = 250, 500$ and the replication time is 500.

To examine the performance of the pointwise confidence interval and global simultaneous confidence band, we compare our method with the Bayesian confidence interval proposed by Wahba (1983), denoted by B-CI. And its corresponding coverage probability is denoted by BCP. We refer our proposed pointwise (local) confidence interval and its coverage probability as LCI and LCP, while we refer our proposed simultaneous (global) confidence band as GCI. Through the simulations, we find that $\exp(\hat{g}(t))S_n(t)$ is nearly flat in the interval [0.2, 0.8], we set the eigenfunctions as the trigonometric series. To be specific, we let r be the mean value of $[\{\exp(\hat{g}(t))S_n(t)\}]^{1/2}, t \in [0.2, 0.8]$. Thus, the eigenfunctions and eigenvalues are given as the following:

$$h_j(t) = \begin{cases} 1/r, & j = 0; \\ \sqrt{2}\sin(2\pi k)/r, & j = 2k-1, k = 1, 2, \ldots; \\ \sqrt{2}\cos(2\pi k)/r, & j = 2k, l = 1, 2, \ldots; \end{cases}$$

$$\gamma_j(t) = \begin{cases} 0, & j = 0; \\ \frac{(2\pi k)^{2m}}{r^2}, & j = 2k-1, k = 1, 2, \ldots; \\ \frac{(2\pi k)^{2m}}{r^2}, & j = 2k, l = 1, 2, \ldots; \end{cases}$$

Then following from the method in Shang and Cheng (2013) on page 2631, we can get the estimate of σ_{t_0}. Specifically, we can estimate σ_{t_0} by $I_2 r^{-(2-1/m)}/\pi$, where $I_2 = \int_0^\infty (1 + x^{2m})^{-2}\, dx$. By plugging in the eigenfunctions and eigenvalues obtained previously into 2.3 to get $\hat{G}_\alpha$, we can compute the global coverage probability (CP). The simulation results are presented in Figures 1-2. We observe that the length of our proposed local confidence interval (LCI) is shorter than that of Wahba's method (1983), which is consistent with that of Shang and Cheng (2013). The LCP is close to 95% for $t \in [0.1, 0.8]$ while the BCP is almost 1 due to over-estimation in the variance. Besides, table 1 gives the Global CP at different intervals. From the table, we can get that the global confidence band is reasonable.

To show the power of the test, we considered the test functions: $g(t) = \log(\beta_1) - \log(1-x) + cx$, with $c = 0, 0.5, 1, 1.5$. As $J(g, g) = \infty$, we modify $PLRT_{n,\lambda} \asymp l_n(g_0) - l_n(\hat{g}_{n,\lambda})$. The results about the global likelihood ratio tests are listed in Table 2. From the table, we can find that when $c = 0$, the power is around 5% or less 5%, and when $c > 0$, the power increases to 1 with the samples or c increasing. This means that the modified

$PLRT_{n,\lambda}$ can also do some likelihood ratio test when $J(g,g) = \infty$, namely the true function do not belong to the $\mathcal{H}^m$.

Example 2: The failure time is generated from the truncated Weibull distribution on $[0.1, \infty]$ with density function

$$f(x) \propto \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} \exp(-(\frac{x}{\lambda})^k), \qquad x \in [0.1, \infty],$$

with $k = 2.8$ and $\lambda = 0.8$. We generate the censoring time from the truncated Weibull distribution on $[0.1, 1.1]$ with $\lambda = 5$ and k is chosen to yield 20% and 40% censoring rate. In this example, we set m=2. We are able to obtain the eigenvalues and eigenfunctions of the Hilbert space by solving the ODE functions (2.1) numerically, which therein are used in the estimation of σ_{t_0} and $\hat{G}_\alpha$ similar to Example 1. The simulation results of the pointwise and global confidence interval and coverage probability are reported in Figures 3-4. We observe that the local and global CI and CP exhibit similar patterns as in example 1, for example, LCP is close to 95% for [0.2, 0.9] and the global confidence band works reasonably well.

To show the power of the test, we consider the test functions: $g(t) = \log(k) + (k-1)\log(t) - k\log(\lambda) + cx$, with $c = 0, 0.5, 1, 1.5$. Still following from the ODE function (2.1), we can get the eigenvalues of the different g are $\gamma_j \approx (\alpha j)^{2m}$, with $\alpha = 2.9162, 2.4562, 2.6543, 2.2476$ and alpha $= 3.1169, 2.9013, 2.6673, 2.4820$ with the censoring rate being 20%, n=250 and n=500, respectively. Then we can get the $\gamma_\lambda = 1.333$ while $h\nu_\lambda =$

$0.5078, 0.5579, 0.6029, 0.6589$ or $h\nu_\lambda = 0.4751, 0.5104, 0.5552, 0.5967$, respectively. The results about the global likelihood ratio tests are listed in Table 4. The table shows that when $c = 0$, the power is around 5%, and when $c > 0$, the power increases to 1 with the samples or c increasing.

Example 3: In our simulation studies, the failure time follows from the exponential distribution with the corresponding mean parameter $1/2$, while the censoring time follows from the uniform distribution $Unif(0, L)$ and the end of the study time is 1. L is chosen to yield the censoring rate being 20% or 40%. We considered the sample sizes $n = 250, 500$ while the replication times are 500.

To show the power of $PLRT_{n,\lambda}$, we considered the test functions: $g(t) = \log(2) + cx$, with $c = 0, 0.5, 1, 1.5$. Following from the ODE function (2.1), we can get the eigenvalues of the different g are $\gamma_j \approx (\alpha j)^{2m}$, with $\alpha = 1.887, 1.7584, 1.6431, 1.5446$ and $\alpha = 1.8727, 1.7665, 1.6507, 1.5572$ with the censoring rate being 20%, n=250 and n=500, respectively. Then we can get the $\gamma_\lambda = 1.333$ while $h\nu_\lambda = 0.7883, 0.8488, 0.9013, 0.9588$ or $h\nu_\lambda = 0.7908, 0.8384, 0.8972, 0.9510$, respectively. Similarly, we can also get the $h\nu_\lambda$ when the censoring rate is 40% and n=250, 500 respectively. The results about the global likelihood ratio tests are listed in Table 5. The table shows that when $c = 0$, the power is around 5%, and when $c > 0$, the power increases to 1 with the samples or c increasing.

2.6 Application

For illustration, we apply the proposal to analyze the study of non-Hodgkin's lymphoma Dave et al. (2004). The goal of the experiment is to detect the effect of follicular lymphoma on the patients' survival time. The data were obtained from seven institutions from 1974 to 2001. The samples are from 191 patients with untreated follicular lymphoma, who are diagnosed at the ages from 23 to 81 years (median 51). The follow-up times are ranging from 1.0 to 28.2 years (median 6.6). After removing 4 samples with missing censoring indicator and observation time, we have $n = 187$ samples and around 50% censoring rate. As suggested by Iglewicz and Hoaglin (1993), we also calculate an outlier statistic: $Z_i = 0.6745|Y_i - \text{median}(Y)|/\text{mad}(Y)$, where i refers to the ith subject, median(Y) and mad(Y) are the median and median absolute deviation of the 187 observation times, respectively. According to Iglewicz and Hoaglin (1993), an observation is an outlier if $Z > 3.5$. In this analysis, we observe that the 170th subject is the outlier. Then we would clean it out and use the left samples to do the data analysis. We standardize the survival times to range from 0 to 1. The results are summarized in Figure 5.

For comparison, we also compute the Kaplan-Meier estimate, the smooth Kaplan-Meier estimate and our proposed method about the cumulative hazard function. Besides, we give the lower and up bound of the

95% CI of cumulative hazard with the method of Kaplan-Meier estimator. The results are shown in the left panel of Figure 5. From the figure, we can get that our method can give an appropriate estimation, which is very close to the other classical methods. The right panel in figure 5 gives the estimation of the log-hazard of our method, the LCI and BCI. From the figure, we can get that the pointwise interval is shorter than that given by Wahba (1983), which are accordance to the simulation results.

2.7 Appendix

The following regularity conditions are assumed to prove the main results.

(C2.1) The probability $P(\min(T, C) \geq 1) > 0$.

(C2.2) The censoring time C and the survival time T are independent.

(C2.3) The hazard function of T, $\lambda(t)$ is bounded away from 0 and ∞, that is, there exist constants $C_1 > 0$ and $C_2 < \infty$ such that $C_1 \leq \lambda(t) \leq C_2$.

To prove Lemma 2.1, we define the inner product $< \cdot, \cdot >_1$ as a special case of $< \cdot, \cdot >_\lambda$ when $\lambda = 1$ and the corresponding norm $\|g\|_1^2 =< g, g >_1$.

Proof of Lemma 2.1. Let $g_n(t)$ be the B-spline function satisfying $\|g_n - g_0\|_\infty = O(n^{-vm})$. Since any two norms in a finite dimensional Hilbert space is equivalent, we choose $h_n \in \Psi_{m,\mathcal{I}}$ satisfying $\|h_n\|_2 = O(n^{-(1-v)/2} + n^{-vm})$ and $\|h_n\|_\infty = O(n^{-(1-v)/2} + n^{-vm})$. For some $\alpha \in R$, write

$$\begin{aligned} H_n(\alpha) &= l_{n,\lambda}(g_n + \alpha h_n) \\ &= -\int_{\mathbb{I}} \exp\{g_n(t) + \alpha h_n(t)\} S_n(t)\,dt + \frac{1}{n}\sum_{i=1}^{n} \delta_i (g_n + \alpha h_n)(Y_i) \\ &\qquad -\frac{1}{2} < W_\lambda(g_n + \alpha h_n), g_n + \alpha h_n >_\lambda . \end{aligned}$$

The derivative of $H_n(\alpha)$ with respect to α is

$$\begin{aligned} & H_n'(\alpha) \\ &= -\int_{\mathbb{I}} \exp\{g_n(t) + \alpha h_n(t)\} h_n(t) S_n(t)\,dt + \frac{1}{n}\sum_{i=1}^{n} \delta_i h_n(Y_i) \\ &\qquad - \alpha < W_\lambda h_n, h_n >_\lambda - < W_\lambda g_n, h_n >_\lambda \\ &= -\int_{\mathbb{I}} \big[\exp\big(g_0(t)\big) + \exp\big(g_0(t)\big)\{g_n - g_0 + \alpha h_n(t)\}\{1 + o(1)\}\big] h_n(t) \\ &\times S_n(t)\,dt - \alpha < W_\lambda h_n, h_n >_\lambda - < W_\lambda g_n, h_n >_\lambda + \frac{1}{n}\sum_{i=1}^{n} \delta_i h_n(Y_i) \\ &= -\alpha \int_{\mathbb{I}} h_n^2(t)\{1 + o(1)\} \exp\big(g_0(t)\big) S_n(t)\,dt - \alpha < W_\lambda h_n, h_n >_\lambda \\ &\qquad - \left[\int_{\mathbb{I}} \exp\big(g_0(t)\big) h_n(t) S_n(t)\,dt - \frac{1}{n}\sum_{i=1}^{n} \delta_i h_n(Y_i)\right] \\ &\qquad - \int_{\mathbb{I}} (g_n - g_0)\{1 + o(1)\} h_n(t) \exp\big(g_0(t)\big) S_n(t)\,dt - < W_\lambda g_n, h_n >_\lambda \\ &\equiv -\alpha I_1 - \alpha I_2 + I_3 + I_4 + I_5. \end{aligned}$$

Because $S_n(t)$ is a Donsker-Class,we have that $\|S_n(t)-S(t)\|_\infty = O_p(n^{-1/2})$. Then,

$$
\begin{aligned}
|I_1| &= \left|\int_{\mathbb{I}} h_n^2(t)\{1+o(1)\}\exp\{g_0(t)\}S_n(t)\,dt\right| \\
&\geq O_p(1)P(Y \geq 1)C_1\|h_n\|_2^2 \\
&= O_p(n^{-(1-v)} + n^{-2vm}).
\end{aligned}
$$

Next, we consider I_3. In view of the fact that $M_i(t)$ is a martingale, we have

$$
\begin{aligned}
E\{|I_3|^2\} &= \frac{1}{n}E\left[\delta_i h_n(Y_i) - \int_{\mathbb{I}} \exp\{g_0(t)\}h_n(t)S_n(t)\,dt\right]^2 \\
&= \frac{1}{n}E\left\{\int_{\mathbb{I}} h_n(t)\,dM_i(t)\right\}^2 \\
&= \frac{1}{n}\int_{\mathbb{I}} h_n^2 \exp\{g_0(t)\}S(t)\,dt \\
&= O(n^{-(2-v)} + n^{-2vm-1}).
\end{aligned}
$$

Thereby, we have $I_3 = O_p(n^{-\frac{(2-v)}{2}} + n^{-vm-1/2})$. On the other hand, from $\|g_n - g_0\|_\infty = O(n^{-vm})$, we get

$$\begin{aligned}
|I_4| &= \left| -\int_{\mathbb{I}} (g_n - g_0)\{1 + o(1)\} h_n(t) \exp\{g_0(t)\} S_n(t)\, dt \right| \\
&\leq \left| -\int_{\mathbb{I}} (g_n - g_0)\{1 + o(1)\} h_n(t) \exp\{g_0(t)\} \{S_n(t) - S(t)\}\, dt \right| \\
&\qquad + \left| \int_{\mathbb{I}} (g_n - g_0)\{1 + o(1)\} h_n(t) \exp\{g_0(t)\} S(t)\, dt \right| \\
&\leq O_p(n^{-\frac{2-v}{2}-vm} + n^{-2vm-1/2}) + O(n^{-\frac{1-v}{2}-vm} + n^{-2vm}).
\end{aligned}$$

Lastly, it follows from the property of B-spline that $\|g_n^{(m)}(t)\|_{L^2} \leq C_0$, for some constant C_0 depending on $\|g_0^{(m)}(t)\|_{L^2}$ and m, where $\| \cdot \|_{L^2}$ is the L_2 norm, which is the integral norm. Thus, we have

$$\begin{aligned}
|I_5| &= |< W_\lambda g_n, h_n >_\lambda| \\
&\leq \lambda \|g_n^{(m)}\|_{L^2} \|h_n\|_1 \\
&= \lambda O_p(n^{-\frac{1-v}{2}} + n^{-vm}) = o_p(n^{-(1-v)} + n^{-2vm}).
\end{aligned}$$

As a result, we can conclude that $\alpha H_n'(\alpha) < 0$. Further, it is not hard to see that

$$H_n''(\alpha) = -\int_{\mathbb{I}} \exp\{g_n(t) + \alpha h_n(t)\} h_n^2(t) S_n(t)\, dt - < W_\lambda h_n, h_n >_\lambda \ \leq 0,$$

which implies $H_n'(\alpha)$ is a nonincreasing function. Hence, $\hat{g}_{n,\lambda} \in [g_n - \alpha h_n, g_n + \alpha h_n]$. Note that

$$\begin{aligned}
\|\hat{g}_{n,\lambda} - g_0\|_\infty &\leq \|\hat{g}_{n,\lambda} - g_n\|_\infty + \|\hat{g}_n - g_0\|_\infty \\
&\leq \alpha \|h_n\|_\infty + O(n^{-vm}) = O(n^{-\frac{1-v}{2}} + n^{-vm}),
\end{aligned}$$

which goes to zero as $n \to \infty$. Recall that any two norms in the finite dimension Hilbert Space are equivalent. Then, $\|\hat{g}_{n,\lambda} - g_n\|_1 = O(\|\hat{g}_{n,\lambda} - g_n\|_\infty) = O(n^{-\frac{1-v}{2}} + n^{-vm})$. Therefore, we have

$$\begin{aligned} \|\hat{g}_{n,\lambda} - g_0\|_1 &\leq \|\hat{g}_{n,\lambda} - g_n\|_1 + \|g_n - g_0\|_1 \\ &\leq O(n^{-\frac{1-v}{2}} + n^{-vm}) + \|g_n - g_0\|_1. \end{aligned}$$

Using $\|g_n - g_0\|_\infty = O(n^{-vm})$, $\|g_n^{(m)}(t)\|_{L^2} \leq C_0$ and $g_0 \in \mathcal{H}^m$, we have

$$\|\hat{g}_{n,\lambda} - g_0\|_1 < \tilde{C},$$

and

$$J(\hat{g}_{n,\lambda} - g_0, \hat{g}_{n,\lambda} - g_0) < \tilde{C},$$

where $\tilde{C}$ only depends on g_0 and m. The proof of Lemma 2.1 is complete.

Proof of Lemma 2.2. Following from equation (2.2) and Theorem 2 of Hoeffding (1963), we have

$$P(\|\mathcal{Z}_n(g) - \mathcal{Z}_n(f)\|_\lambda \geq t) \leq 2 \exp\left(\frac{-t^2}{8\|f - g\|_\infty} \right).$$

Together with Lemma 2.2.1 of Van Der Vaart and Wellner (1996), we have

$$\big\| \|\mathcal{Z}_n(f) - \mathcal{Z}_n(g)\|_\lambda \big\|_{\psi_2} \leq 8\|f - g\|_\infty,$$

where $\|\cdot\|_{\psi_2}$ denotes the orlicz norm associated with $\psi_2(s) = \exp(s^2) - 1$. Applying Theorem 2.2.4 of Van Der Vaart and Wellner (1996), for any $\delta > 0$, we have

$$
\begin{aligned}
& \left\| \sup_{f,g\in\mathcal{G},\|f-g\|_\infty\le\delta} \|\mathcal{Z}_n(g)-\mathcal{Z}_n(f)\|_\lambda \right\|_{\psi_2} \\
\le\ & C'\left(\int_0^\delta \sqrt{\log\left\{1+N(\delta,\mathcal{G},\|\cdot\|_\infty)\right\}} + \delta\sqrt{\log\left[1+\{N(\delta,\mathcal{G},\|\cdot\|_\infty)\}^2\right]}\right) \\
\approx\ & \delta^{1-\frac{1}{2m}},
\end{aligned}
$$

where $N(\delta,\mathcal{G},\|\cdot\|_\infty)$ is the covering number, which means the minimum number of $\|\cdot\|_\infty$ δ-balls needed to cover $\mathcal{G}$. Thus, we have

$$
P\left(\sup_{g\in\mathcal{G},\|g\|_\infty\le\delta}\|\mathcal{Z}_n(g)\|_\lambda \ge t\right) \le 2\exp(\delta^{-2+1/m}t^2).
$$

For brevity, we denote $\gamma \equiv 1-1/(2m)$, $T_n \equiv \{5\log\log(n)\}^{1/2}$, $b_n = \sqrt{n}$, $\epsilon = b_n^{-1}$, and $Q_\epsilon = [-\log(\epsilon)-1]$. Write

$$
\begin{aligned}
& P\left(\sup_{g\in\mathcal{G}} \frac{b_n\|\mathcal{Z}_n(g)\|_\lambda}{b_n\|g\|_\infty^\gamma+1} \geq T_n\right) \\
\leq\ & P\left(\sup_{g\in\mathcal{G},\|g\|_\infty\leq\epsilon^{1/\gamma}} \frac{b_n\|\mathcal{Z}_n(g)\|_\lambda}{b_n\|g\|_\infty^\gamma+1} \geq T_n\right) \\
& +\sum_{l=0}^{Q_\epsilon} P\left(\sup_{g\in\mathcal{G},(e^l\epsilon)^{1/\gamma}\leq\|g\|_\infty\leq(e^{(l+1)}\epsilon)^{1/\gamma}} \frac{b_n\|\mathcal{Z}_n(g)\|_\lambda}{b_n\|g\|_\infty^\gamma+1} \geq T_n\right) \\
\leq\ & P\left(\sup_{g\in\mathcal{G},\|g\|_\infty\leq\epsilon^{1/\gamma}} b_n\|\mathcal{Z}_n(g)\|_\lambda \geq T_n\right) \\
& +\sum_{l=0}^{Q_\epsilon} P\left(\sup_{g\in\mathcal{G},\|g\|_\infty\leq\{e^{(l+1)}\epsilon\}^{1/\gamma}} \frac{b_n\|\mathcal{Z}_n(g)\|_\lambda}{b_n\|g\|_\infty^\gamma+1} \geq T_n\right) \\
\leq\ & 2\exp\big\{-(\epsilon^{1/\gamma})^{-2+1/m}T_n^2/n\big\} \\
& +2\sum_{l=1}^{Q_\epsilon} \exp\left(-\big[\{e^{(l+1)}\epsilon\}^{1/\gamma}\big]^{-2+1/m}T_n^2(e^l+1)^2/n\right) \\
=\ & 2\exp(-T_n^2)+2\sum_{l=1}^{Q_\epsilon} 2\exp\{-e^{-2(l+1)T_n^2}(e^l+1)^2\} \\
\leq\ & 2(Q_\epsilon+2)\exp(-T_n^2/4) \\
=\ & C\log(n)\{\log(n)\}^{-5/4}\to 0
\end{aligned}
$$

as $n\to\infty$. We complete the proof of Lemma 2.2.

Proof of Theorem 2.1. Denote $g=\hat{g}_{n,\lambda}-g_0$. By Lemma 2.1, it is clear that $g\in\mathcal{G}$. Write

$$
\begin{aligned}
l_{n,\lambda}(g_0+g)-l_{n,\lambda}(g_0) &= \mathcal{S}_{n,\lambda}(g_0)g+\frac{1}{2}D\mathcal{S}_{n,\lambda}(g_0)gg+\frac{1}{6}D^2\mathcal{S}_{n,\lambda}(g^*)ggg \\
&\equiv I_1+I_2+I_3, \qquad (2.4)
\end{aligned}
$$

where $g^* = g_0 + \alpha_1 g$ and $\alpha_1 \in [0, 1]$. We will next discuss the order of each term in (2.4). From the definition of $\hat{g}_{n,\lambda}$, we can get that $l_{n,\lambda}(g_0 + g) - l_{n,\lambda}(g_0) \geq 0$.

First, it follows from $g \in \mathcal{G}$ that there exists a constant $\tilde{c}$ such that $\exp\{|g(t)|\} \leq \tilde{c}$. Then, we have

$$\begin{aligned}
|6I_3| &= |D^2\mathcal{S}_{n,\lambda}(g^*)ggg| \\
&= \left|\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}}\exp\{g_0(t) + \alpha_1 g(t)\}g^3(t)I(Y_i \geq t)\,dt\right| \\
&\leq \|g\|_\infty \left|\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}}\exp\{g_0(t) + \alpha_1 g(t)\}g^2(t)I(Y_i \geq t)\,dt\right| \\
&\leq \tilde{c}\|g\|_\infty \left|\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}}\exp\{g_0(t)\}g^2(t)I(Y_i \geq t)\,dt\right| \\
&\leq \frac{\tilde{c}\|g\|_\infty}{n}\left|\sum_{i=1}^{n}\int_{\mathbb{I}}\exp\{g_0(t)\}g^2(t)I(Y_i \geq t)\,dt - n\int_{\mathbb{I}}\exp\{g_0(t)\}g^2(t)S(t)\,dt\right| \\
&\quad + \tilde{c}\|g\|_\infty \int_{\mathbb{I}}\exp\{g_0(t)\}g^2(t)S(t)\,dt. \qquad (2.5)
\end{aligned}$$

Denote $\psi(Y; g) = \int_{\mathbb{I}}\exp\{g_0(t)\}g(t)I(Y \geq t)K_t\,dt$ and $\tilde{\psi}(Y; g) = C_2^{-1}c_m^{-1}h^{1/2}\psi(Y; g)$. Then, we have

$$\begin{aligned}
&\left|\sum_{i=1}^{n}\int_{\mathbb{I}}\exp\{g_0(t)\}g^2(t)I(Y_i \geq t)\,dt - n\int_{\mathbb{I}}\exp\{g_0(t)\}g^2(t)S(t)\,dt\right| \\
&= C_2^{-1}c_m^{-1}h^{1/2}\left|\sum_{i=1}^{n} < \tilde{\psi}(Y_i; g), g >_\lambda - < E\tilde{\psi}(Y_i; g), g >_\lambda\right|. \qquad (2.6)
\end{aligned}$$

Under condition(C2.2), we have

$$
\begin{aligned}
& \|\tilde{\psi}(Y;g)-\tilde{\psi}(Y;f)\|_\lambda \\
= \;& C_2^{-1}c_m^{-1}h^{1/2}\|\psi(Y;g)-\psi(Y;f)\|_\lambda \\
= \;& C_2^{-1}c_m^{-1}h^{1/2}\left\|\int_{\mathbb{I}}\exp\{g_0(t)\}\{f(t)-g(t)\}I(Y\geq t)K_t\,dt\right\|_\lambda \\
\leq \;& C_2^{-1}c_m^{-1}h^{1/2}\int_{\mathbb{I}}\exp\{g_0(t)\}\,dt\|f-g\|_\infty\|K_t\|_\lambda \\
\leq \;& C_2^{-1}c_m^{-1}h^{1/2}C_2c_mh^{-1/2}\|f-g\|_\infty \\
= \;& \|f-g\|_\infty.
\end{aligned}
$$

Together with Lemma 2.2, the following inequality hold with probability one

$$
\left\|\sum_{i=1}^{n}\tilde{\psi}(Y_i;g)-E\tilde{\psi}(Y_i;g)\right\|_\lambda \leq \sqrt{n}\left\{\|g\|_\infty^{1-1/(2m)}+1\right\}\{5\log\log(n)\}^{1/2}. \quad (2.7)
$$

The order of the first term in equation (2.4), directly derived from (2.5) and (2.6)

$$
\begin{aligned}
& \frac{\tilde{c}\|g\|_\infty}{n}\left|\sum_{i=1}^{n}\int_{\mathbb{I}}\exp\{g_0(t)\}g^2(t)I(Y_i\geq t)\,dt-n\int_{\mathbb{I}}\exp\{g_0(t)\}g^2(t)S(t)\,dt\right| \\
\leq \;& \frac{\tilde{c}\|g\|_\infty}{n}C_2^{-1}c_m^{-1}h^{1/2}\|g\|_\lambda\left\|\sum_{i=1}^{n}\tilde{\psi}(Y_i;g)-E\tilde{\psi}(Y_i;g)\right\|_\lambda \\
= \;& O_p[\{5\log\log(n)\}^{1/2}n^{-1/2}h^{-1}]\|g\|_\lambda^2.
\end{aligned}
$$

As $\log\{\log(n)\}/nh^2 \to 0$, we have

$$\begin{aligned}
& \frac{\tilde{c}\|g\|_\infty}{n}\left|\sum_{i=1}^n \int_I \exp\{g_0(t)\}g^2(t)I(Y_i \geq t)\,dt - n\int_I \exp\{g_0(t)\}g^2(t)S(t)\,dt\right| \\
= \; & o_p(1)\|g\|_\lambda^2.
\end{aligned}$$

For the second term in (2.4), we have

$$\begin{aligned}
\tilde{c}\|g\|_\infty\left|\int_{\mathbb{I}} \exp\{g_0(t)\}g^2(t)S(t)\,dt\right| &= \tilde{c}\|g\|_\infty V(g,g) \\
&\leq \tilde{c}\|g\|_\infty\|g\|_\lambda^2.
\end{aligned}$$

Thus, we have $|6I_3| = o_p(1)\|g\|_\lambda^2$. It then follows from the Cauthy-Schwarz inequality that

$$|I_1| = |\mathcal{S}_{n,\lambda}(g_0)g| \leq \|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda\|g\|_\lambda.$$

For $\mathcal{S}_{n,\lambda}(g_0)$, we have

$$\begin{aligned}
\|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda &= \left\|-\int_{\mathbb{I}} \exp\{g_0(t)\}K_t S_n(t)\,dt + \frac{1}{n}\sum_{i=1}^n \delta_i K_{Y_i} - W_\lambda g_0\right\|_\lambda \\
&\leq \frac{1}{n}\left\|\sum_{i=1}^n \int_{\mathbb{I}} K_t\,dM_i(t)\right\|_\lambda + \|W_\lambda g_0\|_\lambda = O_P((nh)^{-1/2} + \lambda^{1/2}).
\end{aligned}$$

Regarding I_2, we have

$$\begin{aligned}
2I_2 &= D\mathcal{S}_{n,\lambda}(g_0)gg \\
&= D\mathcal{S}_{n,\lambda}(g_0)gg - D\mathcal{S}_\lambda(g_0)gg + D\mathcal{S}_\lambda(g_0)gg \\
&= -\|g\|_\lambda^2 + D\mathcal{S}_{n,\lambda}(g_0)gg - D\mathcal{S}_\lambda(g_0)gg.
\end{aligned}$$

This is because

$$
\begin{aligned}
& |D\mathcal{S}_{n,\lambda}(g_0)gg - D\mathcal{S}_\lambda(g_0)gg| \\
= & \left|\frac{1}{n}\sum_{i=1}^{n}\left[\int_{\mathbb{I}}\exp\{g_0(t)\}I(Y_i \geq t)g^2(t)\,dt - \int_{\mathbb{I}}\exp\{g_0(t)\}S(t)g^2(t)\,dt\right]\right| \\
\leq & \|g\|_\infty(\int_{\mathbb{I}}\exp\{g_0(t)\}| < K_t, g >_\lambda ||S_n(t) - S(t)|)\,dt) \\
= & \|g\|_\infty\|K_t\|_\lambda\|g\|_\lambda\left(\int_{\mathbb{I}}\exp\{g_0(t)\}\|S_n(t) - S(t)\|_\infty\,dt\right) \\
= & o_p(1)O_p\{(nh)^{-1/2}\}\|g\|_\lambda.
\end{aligned}
$$

Therefore, we have

$$
\|g\|_\lambda^2\{1 + o_p(1)\} \leq \left[O_p\{(nh)^{-1/2} + \lambda^{1/2}\} + o_p\{(nh)^{-1/2}\}\right]\|g\|_\lambda,
$$

which leads to $\|g\|_\lambda = O_p((nh)^{-1/2} + h^m)$.

Proof of Theorem 2.2. For brevity, we denote $g = \hat{g}_{n,\lambda} - g_0$, $r_n = M\{(nh)^{-1/2}+h^m\}$, $\tilde{g} = d_n^{-1}g$ and $d_n = c_m r_n h^{-1/2}$. Recall $\|g\|_\lambda = O_p((nh)^{-1/2}+h^m)$ in Theorem 2.1. Then, there exists a constant M such that the event $B_n = \{\|g\|_\lambda \leq r_n\}$ happens with large probability. Since $h = o(1)$ and $\log\{\log(n)\}/(nh^2) \to 0$ as $n \to \infty$, it is easy to see that $d_n = o(1)$. On the other hand, when event B_n happens, one can get $\|\tilde{g}\|_\infty \leq 1$ and

$$
J(\tilde{g}, \tilde{g}) = d_n^{-2}\lambda^{-1}\lambda J(g, g) \leq d_n^{-2}\lambda^{-1}\|g\|_\lambda^2 = d_n^{-2}\lambda^{-1}r_n^2 \leq c_m^{-2}h\lambda^{-1}.
$$

It then follows directly that $\tilde{g} \in \mathcal{F}$, where $\mathcal{F} = \{g : \|g\|_\infty \leq 1, J(g, g) \leq c_m^{-2}h\lambda^{-1}\}$ when the event B_n happens. Next, by a Taylor expansion, we have

$$
\begin{aligned}
& \mathcal{S}_n(\hat{g}_{n,\lambda}) - \mathcal{S}_n(g_0) - \{\mathcal{S}(\hat{g}_{n,\lambda}) - \mathcal{S}(g_0)\} \\
= \;& -\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}} \exp\{g_0(t)+g(t)\} I(Y_i \geq t) K_t \, dt + \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}} \exp\{g_0(t)\} I(Y_i \geq t) K_t \, dt \\
& - \left[-\int_{\mathbb{I}} \exp\{g_0(t)+g(t)\} S(t) K_t \, dt + \int_{\mathbb{I}} \exp\{g_0(t)\} S(t) K_t \, dt \right] \\
= \;& -\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}} \exp\{g_0(t)\}[\exp\{g(t)\}-1] I(Y_i \geq t) K_t \, dt \\
& \qquad\qquad + \int_{\mathbb{I}} \exp\{g_0(t)\}[\exp\{g(t)\}-1] S(t) K_t \, dt \\
= \;& -\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}} \exp\{g_0(t)\} \left[g(t) + \frac{g(t)^2}{2}\{1+o_p(1)\} \right] \{I(Y_i \geq t) - S(t)\} K_t \, dt \\
= \;& -\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}} \exp\{g_0(t)\} g(t) \{I(Y_i \geq t) - S(t)\} K_t \, dt \\
& -\frac{1}{2n}\sum_{i=1}^{n}\int_{\mathbb{I}} \exp\{g_0(t)\} g(t)^2 \{I(Y_i \geq t) - S(t)\} K_t \, dt \{1+o_p(1)\} \\
\equiv \;& I_1 + I_2.
\end{aligned}
$$

Observe that

$$
\begin{aligned}
\|2I_2\|_\lambda &= \left\| -\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}} \exp\{g_0(t)\}g^2(t)\{I(Y_i \ge t) - S(t)\}K_t\,dt\{1+o_p(1)\}\right\|_\lambda \\
&= \left\|\int_{\mathbb{I}} \exp\{g_0(t)\}g^2(t)\{S_n(t) - S(t)\}K_t\,dt\{1+o_p(1)\}\right\|_\lambda \\
&\le \left\|\int_{\mathbb{I}} \exp\{g_0(t)\}\,dt\right\|_\lambda \|S_n(t)-S(t)\|_\infty \|g\|_\infty^2 \|K_t\|_\lambda \\
&\le O_p(n^{-1/2})(c_m h^{-1/2})^2\|g\|_\lambda^2 c_m h^{-1/2} \\
&= O_p\{(n^{1/2}h)^{-1}\}\|g\|_\lambda^2 c_m h^{-1/2}.
\end{aligned}
$$

The fact $\log(\log(n))/(nh^2) \to 0$ imples $nh^2 \to \infty$ as $n \to \infty$. Thus, the term

$$
I_2 = o_p(1)c_m h^{-1/2}\|g\|_\lambda^2 = o_p(1)c_m h^{-1/2}\{(nh)^{-1/2} + h^m\}^2. \tag{2.8}
$$

We next consider I_1. Write

$$
\begin{aligned}
-I_1 &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}} \exp\{g_0(t)\}g(t)\{I(Y_i \ge t) - S(t)\}K_t\,dt \\
&= \frac{1}{n}\sum_{i=1}^{n}\phi(Y_i, g) - E\{\phi(Y_i, g)\},
\end{aligned}
$$

where $\phi(Y, g) = \int_{\mathbb{I}} \exp\{g_0(t)\}g(t)I(Y \ge t)K_t\,dt$. We denote:

$$
\tilde{\phi}(Y; \tilde{g}) = C_2^{-1}(c_m h^{-1/2})^{-1} d_n^{-1}\phi(Y, d_n^{-1}\tilde{g}).
$$

By a careful evaluation, we can show that

$$
\|\tilde{\phi}(Y; \tilde{g}) - \tilde{\phi}(Y; \tilde{f})\|_\lambda \le \|\tilde{f} - \tilde{g}\|_\infty.
$$

Define

$$\mathcal{Z}_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi(Y_i, g) - E\phi(Y_i, g),$$

In light of Lemma $S.1$ in Shang and Cheng (2013) and the proof of Lemma 2.2, we have

$$\lim_{n\to\infty} P\left[\sup_{g\in\mathcal{F}} \frac{\|\mathcal{Z}_n(g)\|_\lambda}{h^{-(2m-1)/4m}\|g\|_\infty^{1-1/(2m)} + n^{-1/2}} \le \{5\log\log(n)\}^{1/2}\right] = 1.$$

Then, the term

$$\begin{aligned}
\|I_1\|_\lambda &= C_2(c_m h^{-1/2})d_n \left\| \frac{1}{n}\sum_{i=1}^{n} \tilde{\phi}(Y_i, \tilde{g}) - E\{\tilde{\phi}(Y_i, \tilde{g})\} \right\|_\lambda \\
&= \frac{C_2(c_m h^{-1/2})d_n}{n} \left\{\sqrt{n}\|\tilde{g}\|_\infty^{1-1/(2m)} h^{-(2m-1)/(4m)} + 1\right\} \{5\log\log(n)\}^{1/2}.
\end{aligned}$$

In view of the fact that $m > 1/2$ and $\|\tilde{g}\|_\infty \le 1$, we get

$$\begin{aligned}
&\frac{C_2(c_m h^{-1/2})d_n}{n} \left\{\sqrt{n}\|\tilde{g}\|_\infty^{1-1/(2m)} h^{-(2m-1)/(4m)} + 1\right\} \{5\log\log(n)\}^{1/2} \\
&= O(h^{-(6m-1)/(4m)}\{n^{-1/2} + h^{(2m-1)/(4m)}\}\{5\log\log(n)\}^{1/2}\{(nh)^{-1/2} + h^m\}) \\
&= O(h^{-(6m-1)/(4m)} n^{-1/2}\{\log\log(n)\}^{1/2}\{(nh)^{-1/2} + h^m\}). \qquad (2.9)
\end{aligned}$$

Hence, combing (2.8) and (2.9), we have

$$\mathcal{S}_n(\hat{g}_{n,\lambda}) - \mathcal{S}_n(g_0) - (\mathcal{S}(\hat{g}_{n,\lambda}) - \mathcal{S}(g_0))$$

$$= O_p(h^{-(6m-1)/(4m)} n^{-1/2}\{\log\log(n)\}^{1/2}\{(nh)^{-1/2} + h^m\} + h^{-1/2}\{(nh)^{-1} + h^{2m}\}).$$

On the other hand, we note that

$$\begin{aligned}
& \mathcal{S}_n(\hat{g}_{n,\lambda}) - \mathcal{S}_n(g_0) - \{\mathcal{S}(\hat{g}_{n,\lambda}) - \mathcal{S}(g_0)\} \\
= \; & \mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda}) - \mathcal{S}_{n,\lambda}(g_0) - \{\mathcal{S}_\lambda(\hat{g}_{n,\lambda}) - \mathcal{S}_\lambda(g_0)\} \\
= \; & \mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda}) - \mathcal{S}_{n,\lambda}(g_0) - \left\{ D\mathcal{S}_\lambda(g_0)g + \int_{\mathbb{I}}\int_{\mathbb{I}} sD^2\mathcal{S}_\lambda(g_0 + s'sg)g^2 \, ds \, ds' \right\} \\
= \; & g - \mathcal{S}_{n,\lambda}(g_0) - \int_{\mathbb{I}}\int_{\mathbb{I}} sD^2\mathcal{S}_\lambda(g_0 + s'sg)g^2 \, ds \, ds' + \mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda}).
\end{aligned}$$

For any $h \in \mathcal{H}^m$, there exists $h_n \in \Psi_{m,\mathcal{I}}$ such that $\|h - h_n\|_\infty = O(n^{-vm})$. Furthermore, by the definition of $\mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda})$, we have $\mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda})h_n = 0$. Then, we further write

$$\begin{aligned}
& \mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda})h \\
& = \mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda})(h - h_n) \\
& = \mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda})(h - h_n) - \mathcal{S}_n(g_0)(h - h_n) + \mathcal{S}_n(g_0)(h - h_n) \\
& = \left[- \int_{\mathbb{I}} \exp\{g_0(t)\}g(t)\{1 + o_p(1)\}\{h(t) - h_n(t)\}S_n(t) \, dt - < W_\lambda g, h - h_n >_\lambda \right] \\
& - \int_{\mathbb{I}} \exp\{g_0(t)\}\{h(t) - h_n(t)\}S_n(t) \, dt + \frac{1}{n}\sum_{i=1}^{n} \delta_i(h - h_n)(Y_i) - < W_\lambda g_0, h - h_n >_\lambda \\
& \equiv L_1 + L_2 + L_3.
\end{aligned}$$

First, we consider L_1. Write

$$|L_1| = \left| \int_{\mathbb{I}} \exp\{g_0(t)\} g(t)\{1+o(1)\}\{h(t)-h_n(t)\} S_n(t)\, dt + < W_\lambda g, h-h_n >_\lambda \right|$$

$$\leq \left| \left[\int_{\mathbb{I}} \exp\{g_0(t)\} g^2(t) S_n(t)\, dt \right]^{1/2} \left[\int_{\mathbb{I}} \exp\{g_0(t)\}\{h(t)-h_n(t)\}^2 S_n(t)\, dt \right]^{1/2} \right|$$

$$\times \{1+o(1)\} + | < W_\lambda g, h-h_n >_\lambda |$$

$$\leq O_p(\|g\|_\lambda \|h-h_n\|_\infty) + o_p(\lambda n^{-vm})$$

$$= O_p(n^{-vm}\{(nh)^{-1/2} + h^m\}) + o_p(h^{2m} n^{-vm}).$$

Next, we consider L_2 and get

$$\begin{aligned} E\{|L_2|^2\} &= E\left[\left| \int_{\mathbb{I}} \exp\{g_0(t)\}\{h(t)-h_n(t)\} S_n(t)\, dt - \frac{1}{n}\sum_{i=1}^n \delta_i \{h-h_n(t)\} \right|^2 \right] \\ &= \frac{1}{n} E\left[\left| \int_I \exp\{g_0(t)\}\{h(t)-h_n(t)\}\, dM_i(t) \right|^2 \right] \\ &\leq \frac{C_4}{n} \|h-h_n\|_\infty^2 \\ &= O(n^{-1-2vm}), \end{aligned}$$

which implies $|L_2| = O_p(n^{-1/2-vm})$. Third, it is not hard to verify that $|L_3| = O(\lambda n^{-vm})$. Combining the asymptotic order of L_1, L_2 and L_3, we have $\|\mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda})\|_\lambda = O_p(n^{-1/2-vm} + h^{2m} n^{-vm} + n^{-vm}\{(nh)^{-1/2} + h^m\})$. Lastly, we observe

$$\left\| \int_{\mathbb{I}} \int_{\mathbb{I}} s D^2 \mathcal{S}_\lambda(g_0 + s' s g) g^2\, ds\, ds' \right\|_\lambda \leq \int_{\mathbb{I}} \int_{\mathbb{I}} \left\| D^2 \mathcal{S}_\lambda(g_0 + s' s g) g^2 \right\|_\lambda ds\, ds'.$$

In particular,

$$\begin{aligned} \|D^2\mathcal{S}_\lambda(g_0+s'sg)g^2\|_\lambda &= \left\| \int_{\mathbb{I}} \exp\{g_0(t)+ss'g(t)\}S(t)g^2(t)K_t\,dt \right\|_\lambda \\ &\le \tilde{c}(c_m h^{-1/2})\|g\|_\lambda^2. \end{aligned}$$

Finally, we get

$$\|g-\mathcal{S}_{n,\lambda}(g_0)\|_\lambda = O_p(\alpha_n),$$

where

$$\begin{aligned} \alpha_n = n^{-1/2-vm} + n^{-vm}((nh)^{-1/2}+h^m) + h^{-1/2}((nh)^{-1}+h^{2m}) \\ +h^{-(6m-1)/(4m)}n^{-1/2}(\log\log(n))^{1/2}((nh)^{-1/2}+h^m). \end{aligned}$$

The proof of Theorem 2.2 is complete.

Proof of Theorem 2.3. For ease of presentation, we denote $R_n = \hat{g}_{n,\lambda} - g^* - \mathcal{S}_n(g_0)$ and $g^* = (id - W_\lambda)g_0$. It follows from Theorem 2.2 directly that $\|R_n\|_\lambda = O_p(\alpha_n) = o_p(n^{-1/2})$. It can be checked that $\|\mathcal{S}_n\|_\lambda = O_p((nh)^{-1/2})$. Hence, R_n is asymptotically negligible compare with $\mathcal{S}_n$. In the following, we shall derive the asymptotic distribution of $(nh)^{-1/2}\{\hat{g}_{n,\lambda}(t_0) - g^*(t_0)\}$. Recall that for any $t \in \mathbb{I}$ and $g \in \mathcal{H}^m$, we have $< K_t, g >_\lambda = g(t)$. Then, we have

$$\begin{aligned} |(nh)^{1/2} < K_{t_0}, \hat{g}_{n,\lambda} - g^* - \mathcal{S}_n(g_0) >_\lambda| &\le \|K_{t_0}\|_\lambda \|R_n\|_\lambda (nh)^{1/2} \\ &\le c_m h^{-1/2}(nh)^{1/2} o_p(n^{-1/2}) \\ &= o_p(1). \end{aligned}$$

Next, we write

$$
\begin{aligned}
& -(nh)^{1/2} < K_{t_0}, \mathcal{S}_n(g_0) >_\lambda \\
= \ & (nh)^{1/2} \int_{\mathbb{I}} \exp\{g_0(t)\} S_n(t) K_t(t_0)\, dt - \frac{1}{n}\sum_{i=1}^{n} \delta_i K_{Y_i}(t_0) \\
= \ & (nh)^{1/2} \int_{\mathbb{I}} \exp\{g_0(t)\} S_n(t) K_{t_0}(t)\, dt - \frac{1}{n}\sum_{i=1}^{n} \delta_i K_{t_0}(Y_i) \\
= \ & (nh)^{1/2} \frac{1}{n}\sum_{i=1}^{n} \int_{\mathbb{I}} K_{t_0}(t)\, dM_i(t) \\
= \ & \frac{1}{\sqrt{nh^{-1}}} \sum_{i=1}^{n} \int_{\mathbb{I}} K_{t_0}(t)\, dM_i(t).
\end{aligned}
$$

Observe that

$$
Var\left\{\int_{\mathbb{I}} K_{t_0}(t)\, dM_i(t)\right\} = \int_{\mathbb{I}} K_{t_0}^2(t) \exp\{g_0(t)\} S(t)\, dt = V(K_{t_0}, K_{t_0}).
$$

Invoking $hV(K_{t_0}, K_{t_0}) < h\|K\|_\lambda^2 < c_m^2$ and $hV(K_{t_0}, K_{t_0}) \to \sigma_{t_0}^2$ as $n \to \infty$, we have

$$
(nh)^{1/2} < K_{t_0}, \mathcal{S}_n(g_0) >_\lambda \xrightarrow{d} N(0, \sigma_{t_0}^2)
$$

as $n \to \infty$. By the multivariate central limit theorem, $(nh)^{1/2}\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}} K_s(t)\, dM_i(t)$ converges to a zero-mean Gaussian distribution with covariance function $\Sigma(s,t)$.

Moreover, since $K_t(s) = \sum_{j=0}^{\infty} \frac{h_j(t)}{1+\lambda\gamma_j} h_j(s)$, it can be shown that

$$\begin{aligned}
(nh)^{1/2}\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}} K_t(s)\,dM_i(s) &= (nh)^{1/2}\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}}\sum_{j=0}^{\infty}\frac{h_j(t)}{1+\lambda\gamma_j}h_j(s)\,dM_i(s) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=0}^{\infty}\frac{h^{1/2}h_j(t)}{1+\lambda\gamma_j}\int_{\mathbb{I}} h_j(s)\,dM_i(s) \\
&= \sum_{j=0}^{\infty}\frac{h^{1/2}h_j(t)}{1+\lambda\gamma_j}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\mathbb{I}} h_j(s)\,dM_i(s) \\
&\equiv Z_n(t).
\end{aligned}$$

Note that $h^{1/2}h_j(t)/(1+\lambda\gamma_j)$ is a bounded deterministic function and

$$\sum_{i=1}^{n}(1/\sqrt{n})\int_{\mathbb{I}} h_j(s)\,dM_i(s)$$

is tight. Then,

$$\frac{h^{1/2}h_j(t)}{\sqrt{n}(1+\lambda\gamma_j)}\sum_{i=1}^{n}\int_{\mathbb{I}} h_j(s)\,dM_i(s)$$

is also tight. By Theorem 2.1 of Kosorok (2008) and $\sum_{i=1}^{n}(1/\sqrt{n})\int_{\mathbb{I}} h_j(s)\,dM_i(s)$ is integral of martingale, we can show

$$\frac{h^{1/2}h_j(t)}{1+\lambda\gamma_j}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\mathbb{I}} h_j(s)\,dM_i(s)$$

is a Donsker-Class. Also, for any integer M, it can verified by Corollary 9.32 of Kosorok (2008) that

$$Z_{n,M}(t) \equiv \sum_{j=1}^{M}\frac{h^{1/2}h_j(t)}{1+\lambda\gamma_j}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\mathbb{I}} h_j(s)\,dM_i(t)$$

is a Donsker-Class. Hence, it follows from Theorem 2.1 of Kosorok (2008) that there exists a semimetric ρ for which I is totally bounded and

$$\lim_{\delta\downarrow 0}\lim_{n\to\infty} P(\sup_{s,t\in\mathbb{I}\ \text{with}\rho(s,t)<\delta} |Z_{n,M}(t) - Z_{n,M}(s)| > \epsilon) = 0$$

for all $\epsilon > 0$. Moreover, it can be shown similarly that $Z_{n,M}(t)$ converges uniformly in $\mathbb{I}$ to $Z_n(t)$ as $M \to \infty$. So far, we have shown, for any $\epsilon > 0$, there exists a M such that $|Z_n(t) - Z_{n,M}(t)| < \epsilon/4$ for all $t \in \mathbb{I}$. Consequently,

$$\begin{aligned}
& \lim_{\delta\downarrow 0}\lim_{n\to\infty} P(\sup_{s,t\in\mathbb{I}\ \text{with}\rho(s,t)<\delta} |Z_n(t) - Z_n(s)| > \epsilon) \\
\leq\ & \lim_{\delta\downarrow 0}\lim_{n\to\infty} P(\sup_{s,t\in I \text{with}\rho(s,t)<\delta} |Z_{n,M}(t) - Z_{n,M}(s)| + |Z_{n,M}(t) - Z_n(t)| \\
+\ & |Z_{n,M}(s) - Z_n(s)| > \epsilon) \\
\leq\ & \lim_{\delta\downarrow 0}\lim_{n\to\infty} P(\sup_{s,t\in I \text{with}\rho(s,t)<\delta} |Z_{n,M}(t) - Z_{n,M}(s)| > \epsilon/2) = 0,
\end{aligned}$$

implying $Z_n(t)$ is tight. Finally, for any finite-dimension $(t_1, t_2, \cdots, t_k)$, $(Z_n(t_1), Z_n(t_2), \cdots, Z_n(t_k)) \xrightarrow{d} (Z(t_1), Z(t_2), \cdots, Z(t_k))$ as $n \to \infty$ can imply that $Z_n(t) \xrightarrow{d} Z(t)$ uniformly in $\mathbb{I}$. Hence, we have shown $\sqrt{nh}\{\hat{g}_{n,\lambda}(t) - g^*(t)\}$ converges weakly in $\mathbb{I}$ to a mean zero Gaussian process $Z(t)$ with covariance function at (s,t) being $\Sigma(s,t)$. The proof of Theorem 2.3 is complete.

Proof of Corollary 2.1. First, for any t,

$$W_\lambda g_0(t) =< W_\lambda g_0, K_t >_\lambda = \sum_{j=0}^{\infty} \frac{\lambda\gamma_j}{1+\lambda\gamma_j} h_j(t) V(g_0, h_j).$$

By the Cauchy-Schwarz's inequality,

$$\begin{aligned}\sum_{j=0}^{\infty}\frac{\lambda\gamma_j}{1+\lambda\gamma_j}h_j(t)V(g_0,h_j) &\leq \left\{\sum_{j=0}^{\infty}\lambda\gamma_j V^2(g_0,h_j)\right\}^{1/2}\left\{\sum_{j=0}^{\infty}\frac{\lambda\gamma_j}{(1+\lambda\gamma_j)^2}h_j^2(t)\right\}^{1/2}\\ &\leq h^m\sup_{j\in N}\|h_j\|_\infty\sqrt{J(g_0,g_0)}\left\{\sum_{j=0}^{\infty}\frac{\lambda\gamma_j}{(1+\lambda\gamma_j)^2}\right\}^{1/2}.\end{aligned}$$

Invoking $g_0 \in \mathcal{H}^m$ and $\gamma_j \approx j^{2m}$, $W_\lambda g_0(t) = O(h^{m-1/2})$. Hence,

$$\sqrt{nh}W_\lambda g_0(t) = O(n^{1/2}h^m) = o(1).$$

It follows directly from Theorem 2.3 that the results of Corollary 2.1 hold.

Proof of Theorem 2.5 (ii). For notational convenience, we denote $\hat{g} = \hat{g}_{n,\lambda}$, $\hat{g}^0 = \hat{g}^0_{n,\lambda}$, $g = \hat{g}^0 + \omega_0 - \hat{g}$. By Theorem 2.4,

$$\|g\|_\lambda = \|\hat{g}^0 + \omega_0 - \hat{g}\|_\lambda \leq \|\hat{g}^0 + \omega_0 - g_0\|_\lambda + \|\hat{g} - g_0\|_\lambda = O_p(r_n),$$

where $r_n = (nh)^{-1/2} + h^m$. Applying Taylor expansion,

$$\begin{aligned}\mathrm{LRT}_{n,\lambda} &= l_{n,\lambda}(\omega_0 + \hat{g}^0) - l_{n,\lambda}(\hat{g})\\ &= \mathcal{S}_{n,\lambda}(\hat{g})(\omega_0 + \hat{g}^0 - \hat{g}) + \int_{\mathbb{I}}\int_{\mathbb{I}} sD\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg\,ds\,ds'.\end{aligned}$$

It follows from the definition of $\mathcal{S}_{n,\lambda}(\hat{g})$ that $\mathcal{S}_{n,\lambda}(\hat{g})(\hat{g}^0 + \omega_0 - \hat{g}) = 0$. Hence,

$$
\begin{aligned}
&\mathrm{LRT}_{n,\lambda} \\
&= \int_{\mathbb{I}}\int_{\mathbb{I}} sD\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg\, ds\, ds' \\
&= \int_{\mathbb{I}}\int_{\mathbb{I}} s\left\{D\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg - D\mathcal{S}_{n,\lambda}(g_0)gg\right\} ds\, ds' + \int_{\mathbb{I}}\int_{\mathbb{I}} sD\mathcal{S}_{n,\lambda}(g_0)gg\, ds\, ds' \\
&= \int_{\mathbb{I}}\int_{\mathbb{I}} s\{D\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg - D\mathcal{S}_{n,\lambda}(g_0)gg\}\, ds\, ds' + \frac{1}{2}\{D\mathcal{S}_{n,\lambda}(g_0)gg - D\mathcal{S}_{\lambda}(g_0)gg\} \\
&+ \frac{1}{2}D\mathcal{S}_{\lambda}(g_0)gg \\
&\equiv I_1 + I_2 + I_3.
\end{aligned}
$$

We first consider I_1. Denote $\tilde{g} = \hat{g} + ss'g - g_0$ for any $0 \le s, s' \le 1$. Then, $\|\tilde{g}\|_\lambda = O_p(r_n)$ and

$$
\begin{aligned}
D\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg &= D\mathcal{S}_{n,\lambda}(\tilde{g} + g_0)gg \\
&= -\int_{\mathbb{I}} \exp\{g_0(t) + \tilde{g}(t)\}g(t)g(t)S_n(t)\, dt - < W_\lambda g, g >_\lambda .
\end{aligned}
$$

Thus, we can write

$$
\begin{aligned}
& |D\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg - D\mathcal{S}_{n,\lambda}(g_0)gg| \\
= & \left| -\int_{\mathbb{I}} [\exp\{g_0(t) + \hat{g}(t)\} - \exp\{g_0(t)\}]g(t)g(t)S_n(t)\, dt \right| \\
= & \left| \int_{\mathbb{I}} \exp\{g_0(t)\}\tilde{g}(t)\{1 + o_p(1)\}g^2(t)S_n(t)\, dt \right| \\
\le & \|\tilde{g}\|_\infty \left\| \int_{\mathbb{I}} \exp\{g_0(t)\}g^2(t)S_n(t)\, dt \right\| \\
\le & \|\tilde{g}\|_\infty \int_{\mathbb{I}} \exp\{g_0(t)\}g^2(t)|S_n(t) - S(t)|\, dt + \|\tilde{g}\|_\infty \left\| \int_{\mathbb{I}} \exp\{g_0(t)\}g^2(t)S(t)\, dt \right\| .
\end{aligned}
$$

Under condition(C2.3) and the assumption that $nh^4 \to \infty$ as $n \to \infty$,

$$\begin{aligned}
&\|\tilde{g}\|_\infty \int_{\mathbb{I}} \exp\{g_0(t)\} g^2(t) |S_n(t) - S(t)|\, dt \\
= &\ \|\tilde{g}\|_\infty \int_{\mathbb{I}} \exp\{g_0(t)\} g(t) < K_t, g >_\lambda |S_n(t) - S(t)|\, dt \\
\leq &\ \|\tilde{g}\|_\infty \|K_t\|_\lambda \|g\|_\lambda \|S_n(t) - S(t)\|_\infty \|g(t)\|_\infty \int_{\mathbb{I}} \exp\{g_0(t)\}\, dt \\
= &\ O_p(n^{-1/2}h^{-1}) \|g\|_\lambda^2 \|\tilde{g}\|_\infty \\
= &\ o_p(1) \|g\|_\lambda^2 \|\tilde{g}\|_\infty.
\end{aligned}$$

Moreover, note that

$$\|\tilde{g}\|_\infty \left\| \int_{\mathbb{I}} \exp\{g_0(t)\} g^2(t) S(t)\, dt \right\| \leq \|\tilde{g}\|_\infty \|g\|_\lambda^2.$$

which gives that $|I_1| = O_p(1)\|\tilde{g}\|_\infty \|g\|_\lambda^2 = O_p(h^{-1/2} r_n^3)$.

We next consider I_2. Write

$$\begin{aligned}
2|I_2| &= |D\mathcal{S}_{n,\lambda}(g_0)gg - D\mathcal{S}_\lambda(g_0)gg| \\
&= \frac{1}{n} \left| \sum_{i=1}^n \int_{\mathbb{I}} \exp\{g_0(t)\} g^2(t) \{I(Y_i \geq t) - S(t)\}\, dt \right| \\
&= \frac{1}{n} \left| \left\langle \sum_{i=1}^n \int_{\mathbb{I}} \exp\{g_0(t)\} g(t) K_t \{I(Y_i \geq t) - S(t)\}\, dt, g \right\rangle_\lambda \right|.
\end{aligned}$$

We can show along the same lines of Theorem 2.2 that $|I_2| = O_p(r_n a_n')$ and

$$a_n' = h^{-(6m-1)/(4m)} n^{-1/2} \{\log\log(n)\}^{1/2} \{(nh)^{-1/2} + h^m\}.$$

Lastly, we consider I_3. Applying the fact that $I_3 = -\|g\|_\lambda^2/2$ and combining the previous arguments, we have

$$\mathrm{LRT}_{n,\lambda} \;=\; -\frac{\|g\|_\lambda^2}{2} + O_p(h^{-1/2}r_n^3 + r_n a_n').$$

Recall that $nh^{2m} \to 0$. Hence, $nh^{2m+1} \to 0$ as $n \to \infty$. Together with $nh^4 \to \infty$, we have shown $h^{-1/2}r_n^3 + r_n a_n' = o(n^{-1})$. As a result,

$$-2n\mathrm{LRT}_{n,\lambda} = n\|\hat{g}^0 + \omega_0 - \hat{g}\|_\lambda^2 + o_p(1).$$

Proof of Theorem 2.5 (iii). In view of $-2n\mathrm{LRT}_{n,\lambda} = n\|\hat{g}^0+\omega_0-\hat{g}\|_\lambda^2+o_p(1)$ in Theorem 2.5 part(ii), to show part (iii), it suffices to derive the asymptotic properties of $n\|\hat{g}^0 + \omega_0 - \hat{g}\|_\lambda^2$. It is not hard to see that

$$\begin{aligned}
& n^{1/2}\left\|\hat{g}^0 + \omega_0 - \hat{g} - \mathcal{S}_{n,\lambda}^0(g_0^0) + \mathcal{S}_{n,\lambda}(g_0)\right\|_\lambda \\
\leq\; & n^{1/2}\|\hat{g}^0 + \omega_0 - \mathcal{S}_{n,\lambda}^0(g_0^0)\|_\lambda + n^{1/2}\|\hat{g} - \mathcal{S}_{n,\lambda}(g_0)\|_\lambda \\
=\; & O_p(n^{1/2}a_n) = o_p(1).
\end{aligned}$$

Thus, we only need to focus on $n^{1/2}\{\mathcal{S}_{n,\lambda}^0(g_0^0) - \mathcal{S}_{n,\lambda}(g_0)\}$. Recall that

$$\begin{aligned}
&\mathcal{S}_{n,\lambda}^0(g_0^0) \\
&= -\int_{\mathbb{I}} \exp\{g_0(t)\}S_n(t)K_t^*\,dt + \frac{1}{n}\sum_{i=1}^n \delta_i K_{Y_i}^* - W_\lambda^* g_0^0 \\
&= -\int_{\mathbb{I}} \exp\{g_0(t)\}S_n(t)\left\{K_t - \frac{K_{t_0}(t)K_{t_0}}{K(t_0,t_0)}\right\}dt + \frac{1}{n}\sum_{i=1}^n \delta_i\left\{K_{Y_i} - \frac{K_{t_0}(Y_i)K_{t_0}}{K(t_0,t_0)}\right\} \\
&\quad - \left\{W_\lambda g_0 - \frac{(W_\lambda g_0)(t_0)}{K(t_0,t_0)}K_{t_0}\right\}.
\end{aligned}$$

Then,

$$\mathcal{S}^0_{n,\lambda}(g^0_0) - \mathcal{S}_{n,\lambda}(g_0) = \frac{K_{t_0}}{K(t_0,t_0)}\big[\frac{1}{n}\sum_{i=1}^{n}\int_I K_{t_0}(t)\,dM_i(t) + (W_\lambda g_0)(t_0)\big],$$

and

$$\sqrt{n}\|\mathcal{S}^0_{n,\lambda}(g^0_0) - \mathcal{S}_{n,\lambda}(g_0)\|_\lambda = \left|\frac{1}{\sqrt{K(t_0,t_0)}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\mathbb{I}} K_{t_0}(t)\,dM_i(t) + \sqrt{n}(W_\lambda g_0)(t_0)\right]\right|.$$

Applying $nh^{2m} \to 0$, we get

$$\frac{\sqrt{n}(W_\lambda g_0)(t_0)}{\|K_{t_0}\|_\lambda} \le \frac{\sqrt{nh}(W_\lambda g_0)(t_0)}{h^{1/2}\|V^{1/2}(K_{t_0},K_{t_0})\|_\lambda} = O(1)\frac{\sqrt{nh}(W_\lambda g_0)(t_0)}{\sigma_{t_0}} = O(\sqrt{n}h^m) = o(1).$$

Combining these gives

$$\frac{1}{\sqrt{K(t_0,t_0)}}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\mathbb{I}} K_{t_0}(t)\,dM_i(t) + \sqrt{n}(W_\lambda g_0)(t_0)\right\} \xrightarrow{d} N(0, c_{t_0})$$

where

$$c_{t_0} = \lim_{h\to 0}\frac{V(K_{t_0},K_{t_0})}{\|K_{t_0}\|^2} \in (0,1].$$

As a result, it follows immediately that $-2n\mathrm{LRT}_{n,\lambda} \xrightarrow{d} c_{t_0}\chi^2_1$, which implies that $\|\hat{g}^0 + \omega_0 - \hat{g}\|_\lambda = O_p(n^{-1/2})$. Thereby, we prove the first part of Theorem 2.5. The proof of Theorem 2.5 is complete.

Proof of Theorem 2.6. For simplicity, we denote $g = g_0 - \hat{g}_{n,\lambda}$ and $r_n = (nh)^{-1/2} + h^m$. By a Taylor expansion,

$$\begin{aligned} \mathrm{PLRT}_{n,\lambda} &= l_{n,\lambda}(g_0) - l_{n,\lambda}(\hat{g}_{n,\lambda}) \\ &= \mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda})(g_0 - \hat{g}_{n,\lambda}) + \int_{\mathbb{I}}\int_{\mathbb{I}} sD\mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda} + ss'g)\,ds\,ds' \\ &\equiv I_1 + I_2. \end{aligned}$$

We first consider I_1. Along similar lines of the proof of Theorem 2.2, we have

$$\begin{aligned} |I_1| &= |\mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda})g| \\ &\leq \|\mathcal{S}_{n,\lambda}\|\|g\|_\lambda \\ &= O_p[n^{-1/2-vm} + h^{2m}v^{-vm} + n^{-vm}\{(nh)^{-1/2} + h^m\}]\|g\| \\ &= O_p(r_n[n^{-1/2-vm} + n^{-vm}\{(nh)^{-1/2} + h^m\}]). \end{aligned}$$

Similar to the proof of Theorem 2.5(ii), it can be easily verified that

$$|I_2| = -\frac{\|g\|_\lambda^2}{2} + O_p(h^{-1/2}r_n^3 + r_n\alpha_n'),$$

where $\alpha_n' = h^{-(6m-1)/(4m)}n^{-1/2}\{\log\log(n)\}^{1/2}r_n$. Thus,

$$\mathrm{PLRT}_{n,\lambda} = -\frac{\|g\|_\lambda^2}{2} + O_p(h^{-1/2}r_n^3 + r_n\alpha_n''),$$

where $\alpha_n'' = \alpha_n' + n^{-1/2-vm} + n^{-vm}\{(nh)^{-1/2} + h^m\}$. Under the conditions that $m > (3+\sqrt{5})/4$, $1/(4m) \leq v \leq 1/(2m)$, $nh^{2m+1} = O(1)$ and $nh^3 \to \infty$,

$$-2n\mathrm{PLRT}_{n,\lambda} = n\|g\|_\lambda^2 + o_p(h^{-1/2}).$$

On the other hand, under H_0^{global}, g_0 is true function. Then, Theorem 2.2 gives $\|\hat{g}_{n,\lambda} - g_0 - \mathcal{S}_{n,\lambda}(g_0)\| = O_p(\alpha_n)$ and Theorem 2.3 suggests $n^{1/2}\alpha_n = o(1)$. Combining these gives

$$n^{1/2}\|g\|_\lambda = n^{1/2}\|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda + o_p(1).$$

Next, we consider $\|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda$. Through direct calculation,

$$n\|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda^2 = n^{-1}\left\|\sum_{i=1}^n \int_{\mathbb{I}} K_t \, dM_i(t)\right\|_\lambda^2 + 2 < \sum_{i=1}^n \int_{\mathbb{I}} K_t \, dM_i(t), W_\lambda g_0 > + n\|W_\lambda g_0\|_\lambda^2.$$

We first approximate $\|W_\lambda g_0\|_\lambda$. To this end, we define $m_\lambda(j) \equiv |V(g_0, h_j)|^2 \gamma_j \frac{\lambda\gamma_j}{1+\lambda\gamma_j}$ and $m(j) \equiv |V(g_0, h_j)|^2\gamma_j$, $j = 0, 1, 2, \cdots$. Note that $|m_\lambda(j)|$ is a sequence of functions satisfying $|m_\lambda(j)| \le m(j)$. Since $g_0 \in \mathcal{H}^m$,

$$|V(g_0, h_j)|^2\gamma_j = \int_N m(j)\, d\mu(j) = J(g_0, g_0) < \infty,$$

where $\mu(\cdot)$ is the counting measure. Invoking $\lim_{\lambda\to 0} m_\lambda(j) = 0$,

$$\sum_j |V(g_0, h_j)|^2 \frac{\lambda\gamma_j^2}{1+\lambda\gamma_j} = \int_N m_\lambda(j)\, dm(j) \to 0$$

as $\lambda \to 0$ by the Lebesgue dominated convergence theorem. That is,

$$\|W_\lambda g_0\|_\lambda^2 = \sum_j |V(g_0, h_j)|^2 \frac{\lambda^2\gamma_j^2}{1+\lambda\gamma_j} = o(\lambda).$$

Using this fact, we have

$$E\left|< \sum_{i=1}^{n}\int_{\mathbb{I}} K_t\, dM_i(t), W_\lambda g_0 >\right|^2$$

$$= \left|\sum_{i=1}^{n}\int_{\mathbb{I}} W_\lambda g_0(t)\, dM_i(t)\right|^2$$

$$= n\int_{\mathbb{I}} \exp\{g_0(t)\}S(t)\,\{W_\lambda(g_0(t))\}^2\ dt$$

$$\leq n\|W_\lambda(g_0(t))\|_\lambda^2 = o(n\lambda).$$

Together with $nh^{2m+1} = O(1)$, it follows that

$$< \sum_{i=1}^{n}\int_{\mathbb{I}} K_t\, dM_i(t), W_\lambda g_0 >= o_p\{(n\lambda)^{1/2}\} = o_p(n^{1/2}h^m) = o_p(h^{-1/2}).$$

So far, we have shown that

$$n\|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda^2 = n^{-1}\left\|\sum_{i=1}^{n}\int_{\mathbb{I}} K_t\, dM_i(t)\right\|_\lambda^2 + o_p(h^{-1}).$$

In what follows, we shall derive the limiting distribution of $n^{-1}\|\sum_{i=1}^{n}\int_{\mathbb{I}} K_t\, dM_i(t)\|_\lambda^2$. A direct calculation yields that

$$\frac{1}{n}\left\|\sum_{i=1}^{n}\int_{\mathbb{I}} K_t\, dM_i(t)\right\|_\lambda^2 = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}}\int_{\mathbb{I}} < K_t, K_s >\ dM_i(t)\, dM_i(s) + \frac{1}{n}W_n,$$

where $W_n = \sum_{i\neq j}\int_{\mathbb{I}}\int_{\mathbb{I}} < K_t, K_s >\ dM_i(t)\, dM_j(s)$. Denoting $W_{ij} = 2\int_I\int_I <$ $K_t, K_s >\ dM_i(t)\, dM_j(s)$, one can write $W_n = \sum_{1\leq i<j\leq n} W_{ij}$. So, W_n is clean (de1987central). Next, we aim to derive the limiting distribution of W_n. Let

$\sigma_n^2 = Var(W_n)$. Write

$$\begin{aligned}
\sigma_n^2 &= \frac{n(n-1)}{2} E(W_{ij}^2) \\
&= 2n(n-1) E\left\{ \int_{\mathbb{I}} \int_{\mathbb{I}} < K_t, K_s > \, dM_i(t) \, dM_j(s) \right\}^2 \\
&= 2n(n-1) \sum_{l=0}^{\infty} \frac{1}{(1+\lambda\gamma_l)^2}.
\end{aligned}$$

More notations are needed here. Define G_1, G_2 and G_4 as follows:

$$G_1 \equiv \sum_{i<j} E(W_{ij}^4),$$

$$G_2 \equiv \sum_{i<j<k} \left\{ E(W_{ij}^2 W_{ik}^2) + E(W_{ji}^2 W_{jk}^2) + E(W_{ki}^2 W_{kj}^2) \right\},$$

$$G_4 \equiv \sum_{i<j<k<l} \{ E(W_{ij} W_{ik} W_{lj} W_{lk}) + E(W_{ij} W_{il} W_{kj} W_{kl}) + E(W_{ik} W_{il} W_{jk} W_{jl}) \}.$$

By Proposition 3.2 of Jong (1987), if G_1, G_2, G_3 are all of lower order than σ_n^4, $\sigma_n^{-1} W_n$ converges weakly to the standard normal distribution. Now, we study the order of each $G_i, i = 1, 2, 3$. First, observe that

$$
\begin{aligned}
&E\{W_{ij}^4\} \\
&= 16E\left\{\int_{\mathbb{I}}\int_{\mathbb{I}} < K_t, K_s > \, dM_i(t)\, dM_j(s)\right\}^4 \\
&= 16\int_{\mathbb{I}}\int_{\mathbb{I}}\int_{\mathbb{I}}\int_{\mathbb{I}}\int_{\mathbb{I}}\int_{\mathbb{I}}\int_{\mathbb{I}}\int_{\mathbb{I}} < K_{t_1}, K_{s_1} >< K_{t_2}, K_{s_2} >< K_{t_3}, K_{s_3} >< K_{t_4}, K_{s_4} > \\
&E\Big\{\, dM_i(t_1)\, dM_j(s_1)\, dM_i(t_2)\, dM_j(s_2)\, dM_i(t_3)\, dM_j(s_3)\, dM_i(t_4)\, dM_j(s_4)\Big\} \\
&\leq 16\|K_{t_1}\|^4\|K_{s_1}\|^4 \int_{\mathbb{I}}[E\{dM_i(t_1)\}]^4 \int_I [E\{dM_j(s_1)\}]^4 \\
&= O(h^{-4}),
\end{aligned}
$$

which implying $G_1 = O(n^2h^{-4})$. Next, by Cauchy-Schwarz inequity,

$$E\{W_{ij}^2W_{ik}^2\} \leq [E\{W_{ij}^4\}]^{1/2}[E\{W_{ik}^4\}]^{1/2} = O(h^{-4}),$$

which gives $G_2 = O(n^3h^{-4})$. A straightforward calculation yields that

$$E(W_{ij}W_{ik}W_{lj}W_{lk}) = 16\sum_{j=0}^{\infty}\frac{1}{(1+\lambda\gamma_j)^4} = O(h^{-1}).$$

Therefore, $G_4 = O(n^4h^{-1})$. Combining the fact that $\sigma_n^4 = (\sigma_n^2)^2 = O(n^4h^{-2})$ and the assumptions that $nh^3 \to \infty$ and $h = o(1)$, G_1, G_2, G_4 are of lower order than that of σ_n^4. Hence, by Jong (1987),

$$\sigma_n^{-1}W_n \xrightarrow{d} N(0,1)$$

as $n \to \infty$. Recall that $\rho_\lambda^2 = \sum_{j=0}^{\infty}\frac{h}{(1+\lambda\gamma_j)^2}$. We have

$$\frac{1}{\sqrt{2h^{-1}}n\rho_\lambda}W_n \xrightarrow{d} N(0,1). \tag{2.10}$$

Lastly, we consider $\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}}\int_{\mathbb{I}} < K_t, K_s > \, dM_i(t)\, dM_i(s)$. By a direct calculation,

$$E\left\{\int_{\mathbb{I}}\int_{\mathbb{I}} < K_t, K_s > \, dM_i(t)\, dM_i(s)\right\}^2 = O(\|K_t\|_\lambda^4) = O(h^{-2}).$$

Then,

$$\begin{aligned} & E\left\{\sum_{i=1}^{n}\int_{\mathbb{I}}\int_{\mathbb{I}} < K_t, K_s > \, dM_i(t)\, dM_i(s) - h^{-1}\sigma_\lambda^2\right\}^2 \\ \leq \; & nE\big\{\int_{\mathbb{I}}\int_{\mathbb{I}} < K_t, K_s > \, dM_i(t)\, dM_i(s)\big\}^2 = O(nh^{-2}), \end{aligned}$$

where $\sigma_\lambda^2 = \sum_{j=0}^{\infty}\frac{h}{1+\lambda\gamma_j}$. Combining these gives

$$\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{I}}\int_{\mathbb{I}} < K_t, K_s > \, dM_i(t)\, dM_i(s) = h^{-1}\sigma_\lambda^2 + O_p\{(n^{1/2}h)^{-1}\}. \quad (2.11)$$

Combining (2.10) and (2.11), we have $n\|\mathcal{S}_{n,\lambda}\|_\lambda^2 = O_p(h^{-1})$ and therefore $n^{1/2}\|\mathcal{S}_{n,\lambda}\|_\lambda = O_p(h^{-1/2})$. As a result,

$$\begin{aligned} -2n\mathrm{PLRT}_{n,\lambda} &= \{n^{1/2}\|\mathcal{S}_{n,\lambda}\|_\lambda + o_p(1)\}^2 + o_p(h^{-1/2}) \\ &= n\|\mathcal{S}_{n,\lambda}\|_\lambda^2 + o_p(h^{-1/2}). \end{aligned} \quad (2.12)$$

In view of (2.10), (2.11) and (2.12), we have that as $n \to \infty$,

$$(2h^{-1}\sigma_\lambda^4/\rho_\lambda^2)^{-1/2}\left\{-2n\gamma_\lambda PLRT_{n,\lambda} - n\gamma_\lambda\|W_\lambda g_0(t)\|_\lambda^2 - h^{-1}\sigma_\lambda^4/\rho_\lambda^2\right\} \xrightarrow{d} N(0,1).$$

Proof of Theorem 2.7. First, it can be easily verified that $m > (3+\sqrt{5})/4$, $1/(4m) \leq v \leq 1/(2m)$ and $h \asymp n^{-d}$ with $1/(2m+1) \leq d < 1/3$ satisfy those conditions in Theorem 2.6. Throughout this proof, we only consider

$g_{n0} = g_0 + g_n$ for $g_n \in \mathcal{A}$ in H_1. To prove Theorem 2.7, we write

$$-2n \cdot \mathrm{PLRT}_{n,\lambda} = -2n\{l_{n,\lambda}(g_0) - l_{n,\lambda}(g_{n0})\} - 2n\{l_{n,\lambda}(g_{n0}) - l_{n,\lambda}(\hat{g}_{n,\lambda})\}$$

$$\equiv I_1 + I_2. \tag{2.13}$$

We first consider I_1. For simplicity, we denote

R_i

$$= \left[-\int_{\mathbb{I}} \exp\{g_0(t)\} I(Y_i \geq t)\, dt + \delta_i g_0(Y_i)\right] - \left[-\int_{\mathbb{I}} \exp\{g_{n0}(t)\} I(Y_i \geq t)\, dt + \delta_i g_{n0}(Y_i)\right]$$

$$= -\int_{\mathbb{I}} g_n(t)\, dM_i(t) - \int_{\mathbb{I}}\int_{\mathbb{I}} \exp\{g_{n0}(t) - s g_n(t)\} g_n^2(t) I(Y_i \geq t)\, dt\, ds.$$

Then,

$E\{R_i^2\}$

$$\leq 2\int_{\mathbb{I}} g_n^2(t) S(t) \exp\{g_{n0}(t)\}\, dt + 2E\left[\int_{\mathbb{I}}\int_{\mathbb{I}} \exp\{g_{n0}(t) - s g_n(t)\} g_n^2(t) I(Y_i \geq t)\, dt\, ds\right]^2$$

$$= O(\|g_n\|_\lambda^2 + \|g_n\|_\lambda^4).$$

Therefore, we can get

$$E\left\{\left|\sum_{i=1}^n (R_i - ER_i)\right|^2\right\} \leq nE\{R_i^2\} = (n\|g_n\|_\lambda^2 + n\|g_n\|_\lambda^4).$$

Combining these gives

$$n\left[l_{n,\lambda}(g_0) - l_{n,\lambda}(g_{n0}) - E\{l_{n,\lambda}(g_0) - l_{n,\lambda}(g_{n0})\}\right] = O_p(n^{1/2}\|g_n\|_\lambda + n^{1/2}\|g_n\|_\lambda^2).$$

On the other hand, invoking $D\mathcal{S}_\lambda(g)g_ng_n < 0$ for any $g \in \mathcal{H}^m$, there exists constant $c' > 0$ satisfies that

$$
\begin{aligned}
E\{D\mathcal{S}_{n,\lambda}(g^*_{n0})g_ng_n\} &\le c'E\{D\mathcal{S}_{n,\lambda}(g_{n0})g_ng_n\} \\
&= -c'\|g_n\|_\lambda^2.
\end{aligned}
$$

Then, we can write

$$
\begin{aligned}
E\{l_{n,\lambda}(g_0) - l_{n,\lambda}(g_{n0})\} &= E\left\{\mathcal{S}_{n,\lambda}(g_{n0})(-g_n) + \frac{1}{2}D\mathcal{S}_{n,\lambda}(g^*_{n0})g_ng_n\right\} \\
&\le \lambda J(g_{n0}, g_n) - \frac{c'\|g_n\|_\lambda^2}{2} \\
&\le \{J(g_n, g_n) + J(g_0, g_n)\} - \frac{c'\|g_n\|_\lambda^2}{2} \\
&\le \{J(g_n, g_n) + J(g_0, g_0)^{1/2}J(g_n, g_n)^{1/2}\} - \frac{c'\|g_n\|_\lambda^2}{2} \\
&= O(\lambda) - \frac{c'\|g_n\|_\lambda^2}{2}.
\end{aligned}
$$

It then follows that

$$
\begin{aligned}
I_1 &\ge n\|g_n\|_\lambda^2 + O_p(n\lambda + n^{1/2}\|g_n\|_\lambda + n^{1/2}\|g_n\|_\lambda^2) \\
&= n\|g_n\|_\lambda^2\{1 + O_p(\lambda\|g_n\|_\lambda^{-2} + n^{-1/2}\|g_n\|_\lambda^{-1} + n^{-1/2})\}. \qquad (2.14)
\end{aligned}
$$

Second, we consider I_2. Wnder H_{1n}, note that $\|\hat{g}_{n,\lambda} - g_{n0}\| = O_p\{(nh)^{-1/2} + h^m\}$. It then follows by the FBR in Theorem 2.2 that

$$
\inf_{n\ge N}\inf_{g_n\in\mathcal{A}} P_{g_{n0}}\left(\|\hat{g}_{n,\lambda} - g_{n0} - S_{n,\lambda}(g_{n,0})\|_\lambda \le Mr_n\right), \qquad (2.15)
$$

where $r_n = (nh)^{-1/2} + h^m$, $P_{g_{n0}}$ means the probablity relies on g_{n0}. Along the lines of Theorem 2.6, we can show I_2 has the same limiting distribution as in Theorem 2.6, uniformly for any $g_n \in \mathcal{A}$. In other words, uniformly over all

$g_n \in \mathcal{A}$,

$$(2\nu_{n_0})^{-1/2}(I_2 - n\|W_\lambda g_{n_0}\|_\lambda^2 - h^{-1}\sigma_{n_0,\lambda}^2) = O_p(1), \tag{2.16}$$

where $\nu_{n_0} = h^{-1}\sigma_{n_0,\lambda}^4/\rho_{n_0,\lambda}^2$, $\sigma_{n_0,\lambda}^2$ and $\rho_{n_0,\lambda}^2$ are defined the same as σ_λ^2 and ρ_λ^2 but with eigenvalues and eigenvectors obtained under g_{n_0}. Next, let V_{n_0} and V_0 be similar functions defined like V as in section 2. Thus, for any $f \in \mathcal{H}^m$,

$$\begin{aligned}|V_{n_0}(f,f) - V_0(f,f)| &= \left|\int_{\mathbb{I}} \big[\exp\{g_{n_0}(t)\} - \exp\{g_0(t)\}\big] S(t) f^2(t)\,dt\right| \\ &\le \|\exp\{g_n(t)\}\|_\infty V_0(f,f)\|g_n\|_\infty \\ &= \zeta V_0(f,f)\|g_n\|_\infty.\end{aligned}$$

It follows from the Shang and Cheng (2013) in the supplementary on page 56 that

$$\sigma_{n_0,\lambda}^2 - \sigma_\lambda^2 = O(h^{-1/2}\|g_n\|_\lambda). \tag{2.17}$$

Combining (2.14), (2.16) and (2.17) gives

$$\begin{aligned}&(2\nu_n)^{-1/2}(-2nr_\lambda \mathrm{PLRT}_{n,\lambda} - \nu_n) \\ =&(2\nu_n)^{-1/2}\{-r_\lambda(I_1 + I_2) - \nu_n\} \\ =&(2\nu_n)^{-1/2} r_\lambda (I_2 - n\|W_\lambda g_{n_0}\|_\lambda^2 - h^{-1}\sigma_{n_0,\lambda}^2) + (2\nu_n)^{-1/2} r_\lambda n \|W_\lambda g_{n_0}\|_\lambda^2 \\ &\qquad + (2\nu_n)^{-1/2} r_\lambda I_1 + (2\nu_n)^{-1/2} r_\lambda h^{-1}(\sigma_{n_0,\lambda}^2 - \sigma_\lambda^2) \\ \ge& O_p(1) + (2\nu_n)^{-1/2} r_\lambda n \|g_n\|_\lambda^2 \{1 + O_p(\lambda\|g_n\|_\lambda^{-2} \\ &+ n^{-1/2}\|g_n\|_\lambda^{-1} + n^{-1/2})\} + O(h^{-1}\|g_n\|_\lambda),\end{aligned}$$

where $O_p(1)$ holds uniformly in $\mathcal{A}$, $\nu_n = h^{-1}\sigma_\lambda^4/\rho_\lambda^2$ and r_λ is defined in Theorem 2.6. Let $\lambda\|g_n\|_\lambda^{-2} \le 1/C$, $n^{-1/2}\|g_n\|_\lambda^{-1} \le 1/C$, $Ch^{-1}\|g_n\|_\lambda \le (nh^{1/2})\|g_n\|_\lambda^2$, and $\|g_n\|_\lambda^2 \geq C(nh^{1/2})^{-1}$ for some sufficiently small constant C. In other words,

$$|(2\nu_n)^{-1/2}(-2nr_\lambda \text{PLRT}_{n,\lambda} - \nu_n)| \geq c_\alpha,$$

where c_α is the critical value (based on $N(0,1)$) to H_0^{global} at nominal level α. This leads to

$$\|g_n\|_\lambda^2 \geq C\{h^{2m} + (nh^{1/2})^{-1}\}. \tag{2.18}$$

Combine (2.15) and (2.18), the proof of Theorem 2.7 is complete.

Table 2.1: The estimated size and power of PLRT for example 1, where the test function is $g(t) = g_0(t) + ct$ with various c values and the nominal significance level is 95%.

Censoring rate	n	$c = 0$	$c = 0.5$	$c = 1$	$c = 1.5$
20%	250	0.016	0.604	1	1
	500	0	0.912	1	1
40%	250	0.056	0.634	1	1
	500	0.022	0.906	1	1

Table 2.2: Estimated global coverage probability with the nominal coverage probability being 95% for example 1.

Censoring rate	n	[0,0.96]	[0,0.97]	[0,0.98]	[0,0.982]
20%	250	98.8%	98.2%	93.2%	88.8%
	500	100%	100%	99.4%	93.6%
40%	250	95.2%	95.2%		
	500	99.8%	99.6%	96.4%	91.6%

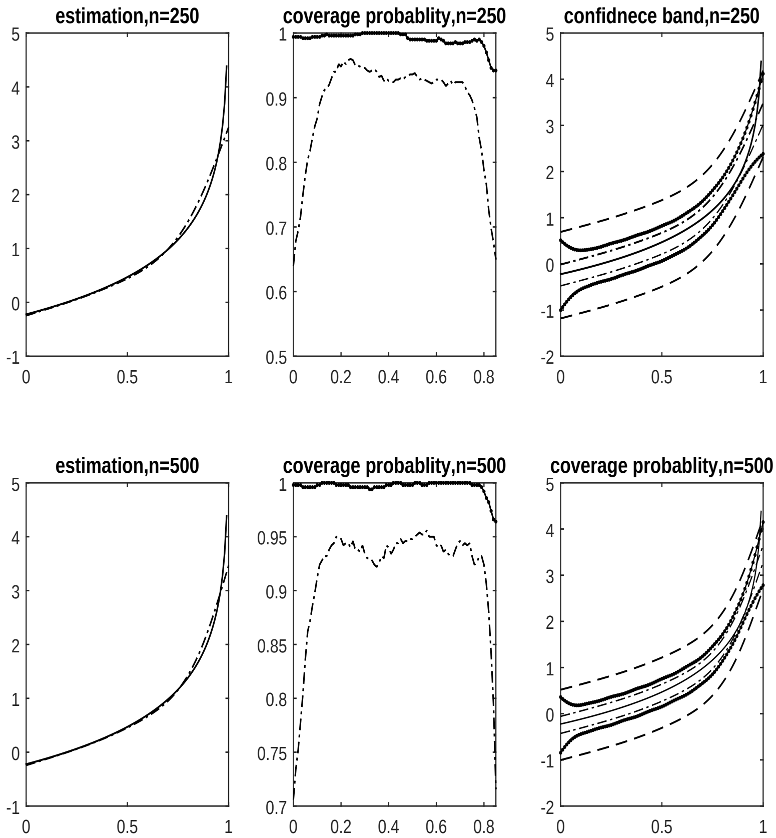

Figure 2.1: The pictures show the simulation results with censoring rate being 20%. The first panel and fourth panel is the estimation (marked with $-.$) and the true function (solid line), respectively. The second and fifth panels display the estimate coverage probabilities given by LCP (marked with $-.$) and BCP (marked with $\cdot$), while the third and last panels show the estimate pointwise confidence intervals given by LCI (marked with $-.$) and BCI (marked with $\cdot$).

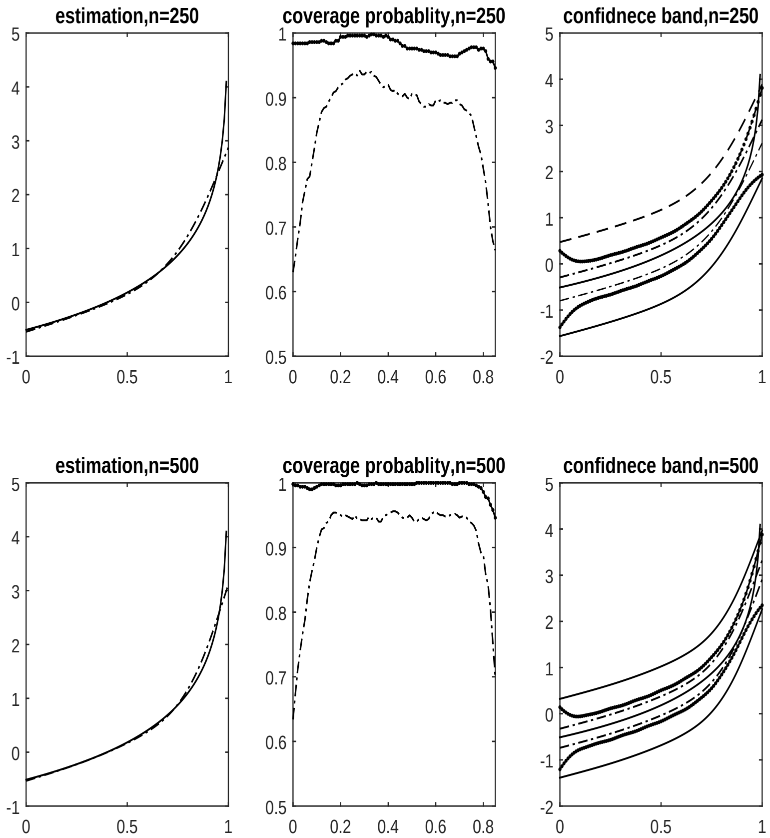

Figure 2.2: The pictures show the simulation results with censoring rate being 40%. The first panel and fourth panel is the estimation (marked with −.) and the true function (solid line), respectively. The second and fifth panels display the estimate coverage probabilities given by LCP (marked with −.) and BCP (marked with ·), while the third and last panels show the estimate pointwise confidence intervals given by LCI (marked with −.) and BCI (marked with ·).

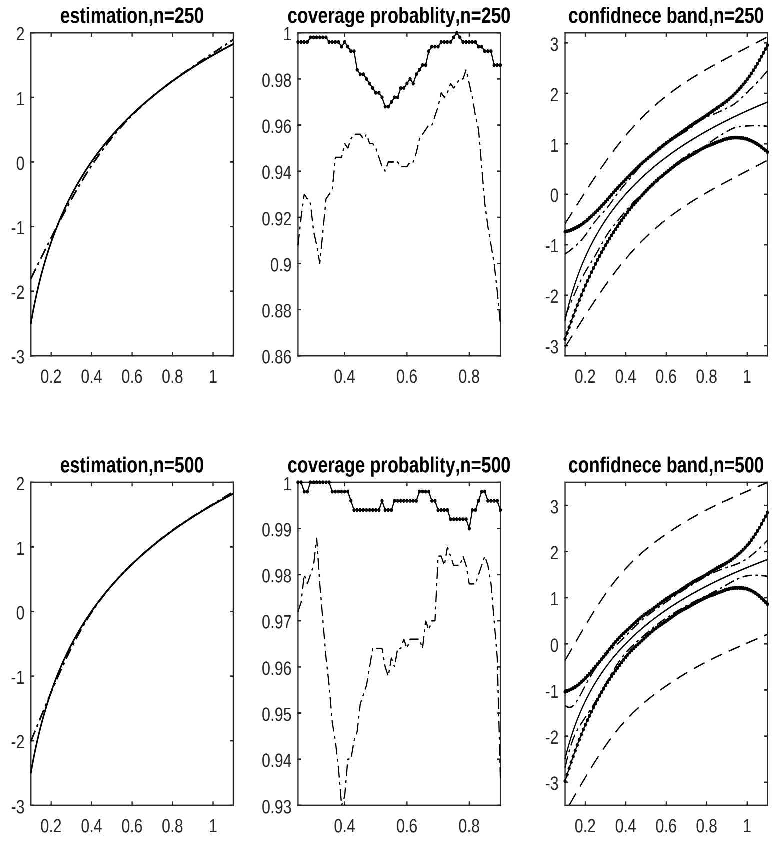

Figure 2.3: The pictures show the simulation results with censoring rate being 20%. The first panel and fourth panel is the estimation (marked with $-.$) and the true function (solid line), respectively. The second and fifth panels display the estimate coverage probabilities given by LCP (marked with $-.$) and BCP (marked with $\cdot$), while the third and last panels show the estimate pointwise confidence intervals given by LCI (marked with $-.$) and BCI (marked with $\cdot$).

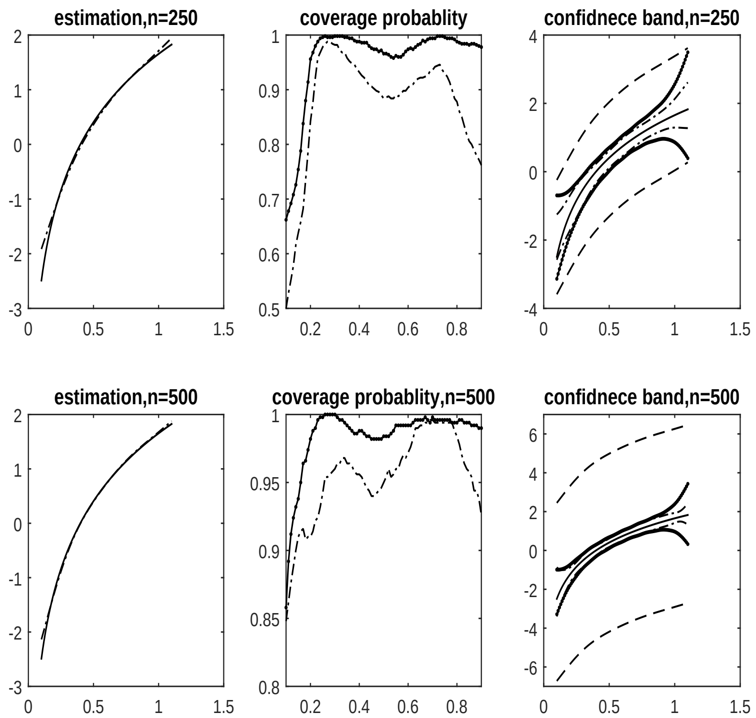

Figure 2.4: The pictures show the simulation results with censoring rate being 40%. The first panel and fourth panel is the estimation (marked with $-.$) and the true function (solid line), respectively. The second and fifth panels display the estimate coverage probabilities given by LCP (marked with $-.$) and BCP (marked with $\cdot$), while the third and last panels show the estimate pointwise confidence intervals given by LCI (marked with $-.$) and BCI (marked with $\cdot$).

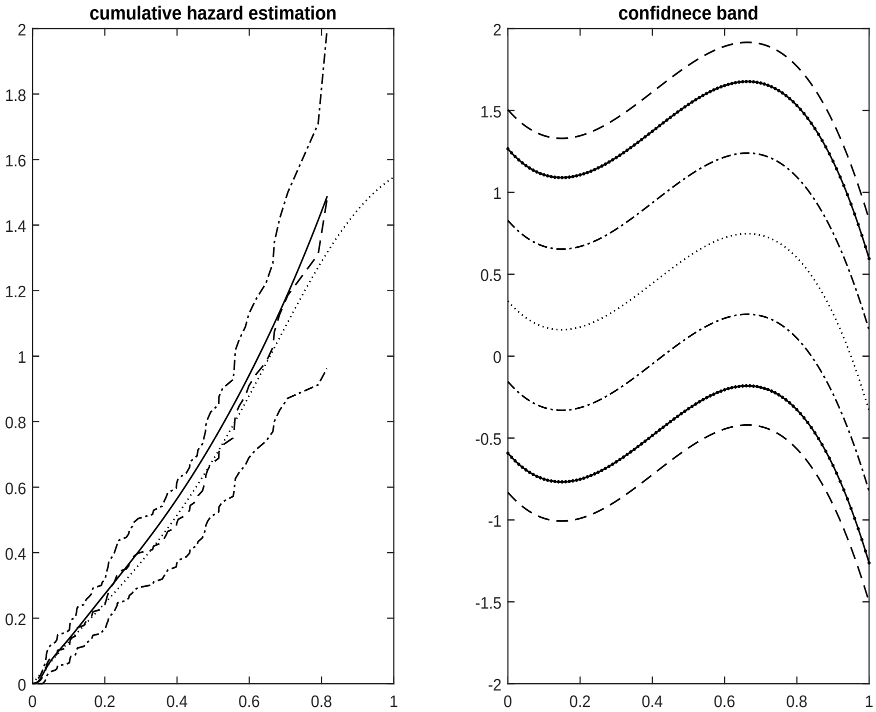

Figure 2.5: The pictures display the real data analysis. The first panel displays the cumulative hazard estimation: the kernel methods (solid line), the Kaplan-Meier estimator (marked with $--$) and our estimation (the dash line), and the confidence band is given by the Kaplan-Meier estimation marked with $-.$ line. The second panel shows the log-hazard estimation and its confidence band various the three methods: the solid line is the log-hazard estimation, the $--$ line the simultaneous confidence band, the $-.$ line is the local pointwise confidence interval while the line marked with $\cdot$ is the confidence interval given by Wahba.

Table 2.3: The estimated size and power of PLRT for example 2, where the test function is $g(t) = g_0(t)+ct$ with various c values and the nominal significance level is 95%.

Censoring rate	n	$c = 0$	$c = 0.5$	$c = 1$	$c = 1.5$
20%	250	0.064	1	1	1
	500	0.044	1	1	1
40%	250	0.052	0.982	1	1
	500	0.052	1	1	1

Table 2.4: Estimated global coverage probability with the nominal coverage probability being 95% for example 1.

Censoring rate	n	[0.18,1.1]	[0.17,1.1]	[0.16,1.1]	[0.1,1.1]
20%	250	95.6%	92.8%	90.0%	81.6%
	500	98.2%	98.2%	98.0%	95.6%
40%	250	98.2%	98.2%	98.2%	71.0%
	500	96.6%	96.2%	96.2%	96.0%

Table 2.5: The estimated size and power of PLRT for example 3, where the test function is $g(t) = g_0(t)+ct$ with various c values and the nominal significance level is 95%.

Censoring rate	n	$c = 0$	$c = 0.5$	$c = 1$	$c = 1.5$
20%	250	0.058	0.790	1	1
	500	0.052	0.954	1	1
40%	250	0.070	0.492	1	1
	500	0.060	0.716	1	1

Chapter 3

Nonparametric Statistical Inference for Case One Interval Censored Data

In this chapter, we consider the nonparametric statistical inference with case-one interval censored data, namely the current status data.

3.1 Introduction

When analyzing the survivor data, interval censoring arises frequently in medical and public health examples. In particular, interval censored data occur when the exact failure time data could not be observed, instead we just know that it lies within an interval or not. Among these data, due to the constraints, costs, character of interest events and many other difficulties, one extreme form is that the failure time is just be known before the examination time or not. This kind of data is called the case-one interval censored data or current status data (Groeneboom and Wellner (1992)).

Nonparametric maximum likelihood, as a widely used method in survival analysis, has been derived for the survival functions with current status data. Specifically, Ayer et al. (1955) and Eeden (1956) gave the nonparametric maximum likelihood estimators of current status data, while Banerjee and Wellner (2005) established the estimator's self-consistency property and the confidence interval. However, as the estimator is discrete, it is not suitable to study the density and the hazard. Groeneboom et al. (2010) introduced the smoothed estimator for the survival function with current status data based on the kernel method.

As hazard can give more insight about the event of interest than the survival function, numerous articles have been developed based on the hazard. While the nonparametric estimation without smoothing is not stable, some smooth estimation approaches have been developed. Similar to the right censored data, the kernel-based approaches were given in Eubank (1999) and Groeneboom et al. (2010). In order to avoid selecting the sensitive bandwidth in estimation, spline methods have also been extended to the interval censored data. Specifically, Joly et al. (1998) used the M-spline to model the log-hazard, Rosenberg (1995) suggested using the non-negative coefficients to model the hazard, while Cai and Betensky (2003) developed the penalized B-spline basis to model the hazard. However, all the method with the splines have not talked about the asymptotic properties about the estimators.

In this chapter, we also use the penalized likelihood method to get the estimator of the cumulative hazard function. Specifically, a functional Bahadur representation would be established. Based on the technical tool, we show the estimator enjoys the pointwise asymptotic normality and global asymptotical gaussian process. Further more, the likelihood ratio test is shown, which reveals some efficient properties of the test.

The setup of this chapter is organized as follows: In section 3.2, we give some preliminary knowledge. Specifically, we talk about the statistical model and some estimating procedures; then we present how to construct the Sobolev space with a special inner product; Section 3.3 develops a new functional Bahadur representation (FBR) in the space and gives the local and global asymptotic property of the estimators of the cumulative hazard function; In Section 3.4, we propose the hypothesis test; In Section 3.5, some simulation results are presented; In Section 3.6, the real example analysis is given; All technical proofs are deferred to the 3.7.

3.2 Methodology

Denote U as an "examination" or "observation" time, T as the failure time. Then under the scenario of current status data, the observation consists of the random vector $X = (\delta, U)$ where $\delta = 1(T \leq U)$. Through the whole chapter, we assume that the examination time is independent of the failure time. Let $X_i = (\delta_i, U_i), i = 1, 2, \ldots, n$ are the i.i.d copies of $X = (\delta, U)$. Under this condition, the likelihood function can be denoted as:

$$L_n(F) = \prod_{i=1}^{n} F(U_i)^{\delta_i}(1 - F(U_i))^{1-\delta_i} h(U_i),$$

where F is the cumulative distribute function of the failure time T and that h is the density of the examination time. As U is independent of T, the log-likelihood of the failure time without an additive term not involving F is:

$$l_n(F) = \frac{1}{n}\sum_{i=1}^{n} \delta_i \log(F(U_i)) + (1-\delta_i)\log(1-F(U_i)).$$

Assume that g_0 is the true cumulative-hazard function of the failure time, and $g_0(t) : I \mapsto R$ is bounded away from 0 and infinity, without generality, we can assume that $I = [0,1]$. Then the log-likelihood of g is:

$$\begin{aligned} l_n(g) &= \frac{1}{n}\sum_{i=1}^{n} \delta_i \log[1-\exp\{-g(U_i)\}] + (1-\delta_i)\log[\exp\{-g(U_i)\}] \\ &= \frac{1}{n}\sum_{i=1}^{n} \delta_i \log[1-\exp\{-g(U_i)\}] - (1-\delta_i)g(U_i). \end{aligned}$$

Define $l(g) = El_n(g)$, and

$$\begin{aligned} &\mathcal{H}^m(I) \\ =&\{g : I \mapsto R | g^{(j)} \text{is absolutely continuous for} \quad j = 0, 1, \cdots, m-1, g^{(m)} \in L_2(I)\}, \end{aligned}$$

where $m > 1/2$ and is assumed to be known. Define $J(g, \tilde{g}) = \int_I g^{(m)}(t)\tilde{g}^{(m)}(t)\,dt$. To make an inference about $g_0(t)$, the penalized log-likelihood of g is:

$$\begin{aligned} l_{n,\lambda}(g) &= \frac{1}{n}\sum_{i=1}^{n} \delta_i \log[1-\exp\{-g(U_i)\}] - (1-\delta_i)g(U_i) - \frac{\lambda}{2}\int_0^1 \{g^{(m)}(t)\}^2\,dt \\ &= l_n(g) - \frac{\lambda}{2}J(g,g). \end{aligned}$$

Define the inner product in the space $\mathcal{H}^m$ is:

$$< g, h >_\lambda = E_U \frac{h(U)g(U)\exp\{-g_0(U)\}}{1-\exp\{-g_0(U)\}} + \lambda \int_0^1 g^{(m)}(t)h^{(m)}(t)\,dt,$$

where E_U is the expectation regarding to U. As $\exp\{-g_0(U)\} \leq C < 1$, we have that under the inner product, $\mathcal{H}^m$ is the reproducing kernel Hilbert space (RKHS) with the norm $\|g\|_\lambda^2 = < g, g >_\lambda$. Besides, there exists a positive self-adjoint operator:

$$W_\lambda : \mathcal{H}^m \to \mathcal{H}^m,$$

which satisfies: $< W_\lambda f, g > = \lambda J(f, g)$ for any $f, g \in \mathcal{H}^m$. Define:

$$V(f, g) = E_U \frac{f(U)g(U)\exp\{-g_0(U)\}}{1-\exp\{-g_0(U)\}},$$

we have that

$$< f, g >_\lambda = V(f, g) + < W_\lambda f, g >_\lambda .$$

Further, the reproducing kernel $K(\cdot, \cdot)$ of $\mathcal{H}^m$ defined on $I \times I$ satisfies the following properties:

(P_1) $K_t(\cdot) = K(t, \cdot)$ and $< K_t, g >_\lambda = g(t)$ for any g in $\mathcal{H}^m$ and any t in I.

(P_2) There exists a constant c_m which only depends on m s.t. $\|K_t\|_\lambda \leq c_m h^{-1/2}$ for $\forall t \in I$, where $h = \lambda^{1/2m}$. Thereby, for any $g \in \mathcal{H}^m$, we have $\|g\|_\infty \leq c_m h^{-1/2} \|g\|_\lambda$.

Denote positive sequences a_i and b_i as $a_i \asymp b_i$ if $\lim_{i\to\infty}(a_i/b_i) = c > 0$, and when c=1, we write $a_i \sim b_i$. There exist a sequence of eigenfunctions $h_j \in \mathcal{H}^m$ and eigenvalues γ_j satisfying the following properties:

(P_3) $\sup_{j\in N} \|h_j\|_\infty < \infty$, $\gamma_j \asymp j^{2m}$;

(P_4) $V(h_i, h_j) = \delta_{ij}$, $J(h_i, h_j) = r_j\delta_{ij}$, where δ_{ij} is a Kronecker's delta, which means that when $i = j$, $\delta_{ij} = 1$; otherwise, it's zero.

(P_5) For any $g \in \mathcal{H}^m$, we have $g = \sum_j V(g, h_j)h_j$ with an convergence in the $\|\cdot\|_\lambda$-norm.

(P_6) For any $g \in \mathcal{H}^m$ and $t \in I$, we have $\|g\|_\lambda^2 = \sum_j V(g, h_j)^2(1 + \lambda\gamma_j)$, $K_t(\cdot) = \sum_j h_j(t)h_j(\cdot)/(1 + \lambda\gamma_j)$ and $W_\lambda h_j(\cdot) = (\lambda\gamma_j)/(1 + \lambda\gamma_j)h_j(\cdot)$.

It follows from Shang and Cheng (2013) that, the underling eigensystem satisfies the following ODE functions:

$$
\begin{aligned}
(-1)^m h_j^{(2m)}(\cdot) &= \gamma_j \frac{\exp\{-g_0(\cdot)\}}{1 - \exp\{-g_0(\cdot)\}}\pi(\cdot)h_j(\cdot), \\
h_j^{(k)}(0) = h_j^{(k)}(1) &= 0, k = m, m+1, \ldots, 2m-1, \qquad (3.1)
\end{aligned}
$$

where $\pi(\cdot)$ is the density of U.

Denote $\mathcal{S}_n(g)\big(\mathcal{S}_{n,\lambda}(g)\big)$ and $\mathcal{S}(g)\big(\mathcal{S}_\lambda(g)\big)$ as the Fréchet derivatives of $l_n(g)\big(l_{n,\lambda}(g)\big)$ and $l(g)\big(l_\lambda(g)\big)$ respectively, D be the Fréchet derivative operator, $g_1, g_2, g_3 \in$

$\mathcal{H}^m$ be any direction, then we have:

$$
\begin{aligned}
Dl_{n,\lambda}(g)g_1 &= \frac{1}{n}\sum_{i=1}^{n} -g_1(U_i) + \delta_i \frac{g_1(U_i)}{1-\exp\{-g(U_i)\}} - \lambda\int_0^1 g^{(m)}(t)g_1^{(m)}(t)\,dt \\
&= \frac{1}{n}\sum_{i=1}^{n} < -K_{U_i}, g_1 >_\lambda + < \delta_i \frac{K_{U_i}}{1-\exp\{-g(U_i)\}}, g_1 >_\lambda - < W_\lambda g, g_1 >_\lambda \\
&= < \mathcal{S}_n(g), g_1 >_\lambda - < W_\lambda g, g_1 >_\lambda,
\end{aligned}
$$

where

$$\mathcal{S}_n(g) = \frac{1}{n}\sum_{i=1}^{n} -K_{U_i} + \delta_i \frac{K_{U_i}}{1-\exp\{-g(U_i)\}},$$

$$\mathcal{S}_{n,\lambda}(g) = \mathcal{S}_n(g) - W_\lambda g.$$

$$D^2 l_{n,\lambda}(g)g_1 g_2 = -\frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{g_1(U_i)g_2(U_i)\exp\{-g(U_i)\}}{[1-\exp\{-g(U_i)\}]^2} - \lambda\int_0^1 g_1^{(m)}(t)g_2^{(m)}(t)\,dt.$$

$$D^3 l_{n,\lambda}(g)g_1 g_2 g_3 = \frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{g_1(U_i)g_2(U_i)g_3(U_i)\exp\{-g(U_i)\}[1+\exp\{-g(U_i)\}]}{[1-\exp\{-g(U_i)\}]^3}.$$

Further, we have

$$\mathcal{S}(g) = Dl(g) = EK_U \frac{\exp\{-g(U)\} - \exp\{-g_0(U)\}}{1-\exp\{-g(u)\}},$$

$$\mathcal{S}_\lambda(g) = \mathcal{S}(g) - W_\lambda g.$$

Besides, we have that

$$D\{\mathcal{S}(g)g_1\}g_2 = D^2 l(g)g_1 g_2 = -E_U \frac{[1-\exp\{-g_0(U_i)\}]g_1(U_i)g_2(U_i)\exp\{-g(U_i)\}}{[1-\exp\{-g(U_i)\}]^2}.$$

Thereby, we have that

$$\begin{aligned} < D\mathcal{S}_\lambda(g_0)f, g > &= -E_U \frac{f(U_i)g(U_i)\exp\{-g_0(U_i)\}}{1-\exp\{-g(U_i)\}} - < W_\lambda f, g >_\lambda \\ &= - < f, g >_\lambda . \end{aligned}$$

Proposition 3.1. $D\mathcal{S}_\lambda(g_0) = -id$, *where id is the identity operator.*

This proposition would paly a key role in the FRB, as following from it, the first term of the taylor expansion of $\mathcal{S}_{n,\lambda}(g)$ at g_0 can be approximated by $-id(g-g_0)$. This would result in that we may have a sum of the independent and identically distributed random variables.

3.3 Functional Bahadur Representation

The Functional Bahadur Representation (FBR), which is a key technique in the whole chapter is established in this chapter, and we then give the asymptotic normality of the estimators through the straightforward application of the FBR. The following lemma shows that the estimator is consistent in the $\|\cdot\|_\infty$ and $\|\cdot\|_1$, which denotes that $\|\cdot\|_\lambda$ with $\lambda = 1$.

$$\hat{g}_{n,\lambda} = \arg\max_{g\in\mathcal{H}^m} l_{n,\lambda}(g).$$

Lemma 3.1. *(Consistency) Assume that Conditions* $(C3.1) \sim (C3.2)$ *are* *satisfied. If* $\lambda n^{1-2\mu} \to 0$ *as* $n \to \infty$ *for any* $\mu > 0$*, we have* $\|\hat{g}_{n,\lambda} - g_0\|_\infty = o_p(1)$*,* $J(\hat{g}_{n,\lambda} - g_0, \hat{g}_{n,\lambda} - g_0) = o_p(1)$*, which means that* $\|\hat{g}_{n,\lambda} - g_0\|_1 = o_p(1)$*.*

Based on Lemma 3.1, we can get the exact rate of convergence.

Theorem 3.1. *(The Rate of Convergence) If* $\log\{\log(n)\}/(nh^2) \to 0$*,* $\lambda n^{1-2\mu} \to 0$ *as* $n \to \infty$ *for any* $\mu > 0$*, we have* $\|\hat{g}_{n,\lambda} - g_0\|_\lambda = O_p((nh)^{-1/2} + h^m)$*.*

Given the exact rate of convergence, we derive a new version of FBR.

Theorem 3.2. *(Functional Bahadur Reprensentation) Assume that Conditions* $(C3.1) \sim (C3.2)$ *are satisfied. If* $\log\{\log(n)\}/(nh^2) \to 0$*,* $\lambda n^{1-2\mu} \to 0$ *as* $n \to \infty$ *for any* $\mu > 0$*, we have* $\|\hat{g}_{n,\lambda} - g_0 - \mathcal{S}_{n,\lambda}(g_0)\|_\lambda = O_p(\alpha_n)$*, where*

$$\alpha_n = h^{-1/2}\{(nh)^{-1}+h^{2m}\}+h^{-(6m-1)/(4m)}n^{-1/2}[\log\{\log(n)\}]^{1/2}\{(nh)^{-1/2}+h^m\}.$$

From Theorem 3.2, we can find that the 'bias' of the estimator is very close to a sum of some independently and identically distributed random variables, which is very helpful to study the asymptotic normality.

Theorem 3.3. *(Asymptotic Normality) Assume that Conditions* $(C3.1) \sim (C3.2)$ *are satisfied. If* $m > 3/4 + \sqrt{5}/4$*,* $nh^{4m-1} \to 0$ *and* $nh^3 \to \infty$ *as* $n \to \infty$*. For* $\forall t_0 \in I$*, define* $\sigma^2_{t_0} = \lim_{h\to 0} h \sum_{j=0}^{\infty} h_j^2(t_0)/(1 + \lambda\gamma_j)^2$*. Let* $g^* = (id - W_\lambda)g_0$ *be the biased 'true parameter', then we have*

$$\sqrt{nh}\{\hat{g}_{n,\lambda}(t_0) - g^*(t_0)\} \xrightarrow{d} N(0, \sigma^2_{t_0}).$$

Besides, we have $\sqrt{nh}\{\hat{g}_{n,\lambda}(t) - g^*(t)\}$ *converges weakly in* I *to a zero mean*

Gaussian process with the covariance function at (s,t) equals to $\Sigma(s,t)$*, where*

$$\Sigma(s,t) = \lim_{h\to 0} h \sum_{j=0}^{\infty} \frac{h_j(t)h_j(s)}{(1+\lambda\gamma_j)^2}.$$

Corollary 3.1. *Assume Conditions* $(C3.1) \sim (C3.2)$ *hold. If* $m > 3/2$, $nh^{2m} \to 0$ *and* $nh^3 \to \infty$ *as* $n \to \infty$*, we have*

$$\sqrt{nh}\{\hat{g}_{n,\lambda}(t_0) - g_0(t_0)\} \xrightarrow{d} N(0, \sigma_{t_0}^2).$$

Besides, we have $\sqrt{nh}\{\hat{g}_{n,\lambda}(t) - g_0(t)\}$ *converges weakly in* I *to a zero mean Gaussian process* $Z(t)$ *with the covariance function at (s,t) equals to* $\Sigma(s,t)$*.*

Remark: Corollary 3.1 together with Delta-method immediately gives the the pointwise CI for some real-valued smooth function of $g_0(t_0)$ at any fixed point $t_0 \in I$, denoted as $\rho(g_0(t))$. Let $\dot{\rho}(\cdot)$ be the first derivative of $\rho(\cdot)$. If $\dot{\rho}(g_0(t_0)) \neq 0$, we have

$$P(\rho\{g_0(t_0)\} \in \big[\rho\{\hat{g}_{n,\lambda}(t_0)\} \pm \Phi(\alpha/2)\frac{\dot{\rho}\{g_0(t_0)\}\sigma_{t_0}}{\sqrt{nh}}]) \to 1-\alpha,$$

where Φ is the standard normal distribution function and $\Phi(\alpha)$ is the lower αth quantile of Φ.

Regarding to the confidence band, it is not so easy to get as $Z(t)$ is not so easy to sample. To overcome this difficulty, we appeal to the resampling approach (e.g., Lin et al. (1993)). Let $(G_1, \ldots, G_n)$ be independent standard normal random variables independent of the data. It can be shown following from Lin et al. (1993) that the distribution of the process $Z(t)$ can be approximated by that of the zero-mean Gaussian process

$$\hat{Z}(t) = \sqrt{\frac{h}{n}} \sum_{i=1}^{n} K_t(U_i)\big[1 - \frac{\delta_i}{1 - \exp\{-\hat{g}_{n,\lambda}(U_i)\}}\big] G_i.$$

Thus, we can first obtain a large number of realizations of $\hat{Z}(t)$ by repeatedly generating the standard normal random sample $(G_1, \ldots, G_n)$ while fixing the data at their observed values, and then use the empirical distribution of these realizations to approximate the distribution of $Z(t)$. More specifically, the α-percentile of $\sup_{t \in I} |Z(t)|$ can be obtained by the empirical percentile of a large number of realizations from $\sup_{t \in I} |\hat{Z}(t)|$, denoted as $\hat{Z}_\alpha$. Then the global confidence band of $g_0(t)$ is:

$$\left(\hat{g}_{n,\lambda}(t) - \frac{1}{\sqrt{nh}} \hat{Z}_\alpha, \hat{g}_{n,\lambda}(t) + \frac{1}{\sqrt{nh}} \hat{Z}_\alpha \right).$$

3.4 Likelihood Ratio Test

Based on the likelihood ratio test, we give the local and global hypothesis test about g_0.

3.4.1 Local Likelihood Ratio Test

For some prespecified point (t_0, ω_0), we consider the following hypothesis:

$$H_0 : g(t_0) = \omega_0 \qquad V.S. \qquad H_1 : g(t_0) \neq \omega_0.$$

The "constrained" penalized log-likelihood is defined as:

$$L_{n,\lambda}(g) = \frac{1}{n}\sum_{i=1}^{n} \delta_i \log[1 - \exp\{-g(U_i) - \omega_0\}] - (1 - \delta_i)\{\omega_0 + g(U_i)\} - \frac{\lambda}{2}J(g, g),$$

where $g \in \mathcal{H}_0 = \{g \in \mathcal{H}^m : g(t_0) = 0\}$. We consider the likelihood ratio test (LRT) statistic:

$$LRT_{n,\lambda} = L_{n,\lambda}(\omega_0 + \hat{g}^0_{n,\lambda}) - L_{n,\lambda}(\hat{g}_{n,\lambda}),$$

where $\hat{g}^0_{n,\lambda} = \arg\max_{g \in \mathcal{H}_0} L_{n,\lambda}(g)$.

Endowed with the norm $\|\cdot\|_\lambda$, H_0 is a closed subset in $\mathcal{H}^m$, and thus a Hilbert space. The following proposition says that $\mathcal{H}_0$ also inherits the reproducing kernel and penalty operator from the space $\mathcal{H}^m$. The proof is trivial thus omitted.

Proposition 3.2. *(a) Recall that* $K(t_1, t_2)$ *is the reproducing kernel for* $\mathcal{H}^m$ *under* $< \cdot, \cdot >_\lambda$*, the bivariate function* $K^*(t_1, t_2) = K(t_1, t_2) - K(t_0, t_1)K(t_0, t_2)/K(t_0, t_0)$ *is a reproducing kernel for* $(\mathcal{H}_0, < \cdot, \cdot >_\lambda)$*. That is, for any* $t' \in I$ *and* $g \in \mathcal{H}_0$*, we have* $K^*_{t'} \equiv K^*(t', \cdot) \in \mathcal{H}_0$ *and* $< K^*_{t'}, g >_\lambda = g(t')$*. Besides,* $\|K^*\|_\lambda \leq \sqrt{2}c_m h^{-1/2}$*.*

(b) The operator W^*_λ *defined by* $W^*_\lambda g = W_\lambda g - (W_\lambda g)(t_0)K_{t_0}/K(t_0, t_0)$ *is bounded linear from* $\mathcal{H}_0$ *to* $\mathcal{H}_0$ *and satisfies* $< W^*_\lambda g, \tilde{g} >= \lambda J(g, \tilde{g})$*.*

On the basis of the proposition 3.2, we denote the restricted FBR for $\hat{g}^0_{n,\lambda}$, which will be used to obtain the null limiting distribution. By straightforward calculation we can find the Fréchet derivatives of $L_{n,\lambda}$ (under $\mathcal{H}_0$). Let $g_1, g_2, g_3 \in \mathcal{H}_0$, the first-order of Fréchet derivative of $L_{n,\lambda}$ (L_n) is denoted as

$\mathcal{S}^0_{n,\lambda}$ ($\mathcal{S}^0_n$), then we have

$$
\begin{aligned}
&DL_{n,\lambda}(g)g_1 \\
&= \frac{1}{n}\sum_{i=1}^n \delta_i \frac{\exp\{-g(U_i)-\omega_0\}g_1(U_i)}{1-\exp\{-g(U_i)-\omega_0\}} - (1-\delta_i)g_1(U_i) - \lambda \int_0^1 g^{(m)}(t)g_1^{(m)}(t)\,dt \\
&= \frac{1}{n}\sum_{i=1}^n \frac{\delta_i g_1(U_i)}{1-\exp\{-g(U_i)-\omega_0\}} - g_1(U_i) - \lambda \int_0^1 g^{(m)}(t)g_1^{(m)}(t)\,dt \\
&=< \frac{1}{n}\sum_{i=1}^n \frac{\delta_i K^*_{U_i}}{1-\exp\{-g(U_i)-\omega_0\}} - K^*_{U_i}, g_1 >_\lambda - < W^*_\lambda g, g_1 >_\lambda \\
&=< \mathcal{S}^0_n(g), g_1 >_\lambda - < W^*_\lambda g, g_1 >_\lambda \\
&=< \mathcal{S}^0_{n,\lambda}(g), g_1 >_\lambda .
\end{aligned}
$$

where

$$\mathcal{S}^0_n(g) = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i K^*_{U_i}}{1-\exp\{-g(U_i)-\omega_0\}} - K^*_{U_i},$$

$$\mathcal{S}_{n,\lambda}(g) = \mathcal{S}^0_n(g) - W^*_\lambda g.$$

Define

$$\mathcal{S}^0(g) = E\{\mathcal{S}^0_n(g)\}$$

$$\mathcal{S}^0_\lambda(g) = \mathcal{S}^0(g) - W^*_\lambda g.$$

Then the second-order and the third-order of the Fréchet derivatives of $L_{n,\lambda}(g)$ are denoted as $D^2L_{n,\lambda}(g)g_1g_2$ and $D^3L_{n,\lambda}(g)g_1g_2g_3$ respectively. Besides, we have

$$D^2L_{n,\lambda}(g)g_1g_2 = -\frac{1}{n}\sum_{i=1}^n \frac{\delta_i \exp\{-g(U_i)-\omega_0\}g_1(U_i)g_2(U_i)}{[1-\exp\{-g(U_i)-\omega_0\}]^2} - < W^*_\lambda g_2, g_1 >_\lambda,$$

$$D^3 L_{n,\lambda}(g) g_1 g_2 g_3$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i g_1(U_i) g_2(U_i) g_3(U_i) \exp\{-g(U_i)-\omega_0\}[1+\exp\{-g(U_i)-\omega_0\}]}{[1-\exp\{-g(U_i)-\omega_0\}]^3}.$$

Considering the derivative of $\mathcal{S}^0_\lambda(g)$, we have

$$D\mathcal{S}^0_\lambda(g) g_1 g_2 = -E\frac{\delta_i \exp\{-g(U_i)-\omega_0\} g_1(U_i) g_2(U_i)}{[1-\exp\{-g(U_i)-\omega_0\}]^2} - < W^*_\lambda g_2, g_1 >_\lambda,$$

Define $g^0_0(t) = g_0(t) - \omega_0$, then we have

$$\begin{aligned} < D\mathcal{S}^0_\lambda(g^0_0) f, g >_\lambda &= < D\{\mathcal{S}^0(g^0_0)\} f, g >_\lambda - < W^*_\lambda f, g > \\ &= -E\frac{\exp\{-g_0(U_i)\} g_1(U_i) g_2(U_i)}{1-\exp\{-g_0(U_i)\}} - < W^*_\lambda f, g >_\lambda \\ &= - < f, g > . \end{aligned}$$

Following from the equation, we have the next proposition.

Proposition 3.3. $D\mathcal{S}^0_\lambda(g^0_0) = -id$, *where id is the identity operator.*

Proposition 3.4. *(The Rate of Convergence) Assume that Conditions* $(C3.1) \sim (C3.2)$ *are satisfied. Under* H_0, *if* $(\log\log(n))/(nh^2) \to 0$, $\lambda(n^{1/2-\mu}) \to 0$ *as* $n \to \infty$ *for* $\mu > 0$, *we have* $\|\hat{g}^0_{n,\lambda} - g^0_0\|_\lambda = O_p((nh)^{-1/2} + h^m)$.

Proposition 3.4 is very similar to the proof of theorem 3.1.

Theorem 3.4. *(Restricted FBR) Assume that Conditions* $(C3.1) \sim (C3.2)$ *are satisfied. Under* H_0, *suppose that* $(\log\log(n))/(nh^2) \to 0$, $\lambda(n^{1/2-\mu}) \to 0$ *as* $n \to \infty$ *for any* $\mu > 0$ *hold, we have* $\|\hat{g}^0_{n,\lambda} - g^0_0 - \mathcal{S}^0_{n,\lambda}(g^0_0)\|_\lambda = O_p(\alpha_n)$, *where* α_n *is defined as in theorem 3.2.*

Our main result follows immediately from the Restricted FBR.

Theorem 3.5. *(Local Likelihood Ratio Test) Assume that Conditions* $(C3.1) \sim (C3.2)$ *are satisfied. If* $m > (5+\sqrt{21})/4$, $nh^{2m} \to 0$ *and* $nh^4 \to \infty$ *as* $n \to \infty$. *Furthermore, for* $\forall t_0 \in I$, *if* $\sigma_{t_0} \neq 0$, $c_{t_0} = \lim_{h\to 0} V(K_{t_0}, K_{t_0})/\|K_{t_0}\|_\lambda^2 \in (0, 1]$, *Under* H_0, *we have that: (i)* $\|\hat{g}_{n,\lambda} - \hat{g}^0_{n,\lambda} - \omega_0\|_\lambda = O_p(n^{-1/2})$; *(ii)*$-2nLRT_{n,\lambda} = n\|\hat{g}_{n,\lambda} - \hat{g}^0_{n,\lambda} - \omega_0\|_\lambda^2 + o_p(1)$; *(iii)*$-2nLRT_{n,\lambda} \xrightarrow{d} c_{t_0}\chi_1^2$.

Note that the parametric convergence rate stated in theorem 3.5 is reasonable since the restriction is local.

3.4.2 Global Likelihood Ratio Test

Consider the following "global" hypothesis:

$$H_0^{global} : g = g_0 \qquad V.S. \qquad H_1 : g \neq g_0,$$

where $g_0 \in \mathcal{H}$ can be either known or unknown. The PLRT statistic is defined as:

$$PLRT_{n,\lambda} = l_{n,\lambda}(g_0) - l_{n,\lambda}(\hat{g}_{n,\lambda}).$$

Theorem 3.6. *Assume that Conditions* $(C3.1) \sim (C3.2)$ *are satisfied. If* $m > (3+\sqrt{5})/4$, $nh^{2m+1} = O(1)$, $nh^3 \to \infty$ *as* $n \to \infty$. *Define* $\sigma_\lambda^2 = \sum_{j=0}^\infty h/(1+\lambda\gamma_j)$, $\rho_\lambda^2 = \sum_{j=0}^\infty h/(1+\lambda\gamma_j)^2$, $\gamma_\lambda = \sigma_\lambda^2/\rho_\lambda^2$, $\nu_\lambda = h^{-1}\sigma_\lambda^4/\rho_\lambda^2$. *Under* H_0^{global}, *we have*

$$(2\nu_\lambda)^{-1/2}(-2n\gamma_\lambda PLRT_{n,\lambda} - n\gamma_\lambda\|W_\lambda g_0(t)\|_\lambda^2 - \nu_\lambda) \xrightarrow{d} N(0, 1).$$

A direct examination reveals that $h \asymp n^{-d}$, where $1/(2m+1) \leq d < 1/3$ satisfies the conditions required by theorem 3.6. As we can show that $n\|W_\lambda g_0\|^2 = o(h^{-1}) = o(\nu_\lambda)$, we have $-2n\gamma_\lambda PLRT_{n,\lambda}$ is asymptotically $N(\nu_\lambda, 2\nu_\lambda)$. As $N(\nu_\lambda, 2\nu_\lambda)$ is asymptotically distributed as $\chi^2_{\nu_\lambda}$, denoted

$$-2n\gamma_\lambda PLRT_{n,\lambda} \sim \chi^2_{\nu_\lambda}.$$

That shows that the Wilks phenomenon holds for the PLRT.

To conclude this section, we show that the PLRT achieves the optimal minimax rate of testing specified in Ingster (1993) based on the uniform version of the FBR. Write $H_1 : g = g_{n_0}$, where $g_{n_0} = g_0 + g_n$, where $g_0 \in \mathcal{H}^m$ and g_n belongs to the alternatives value set $\mathcal{A} = \{g \in \mathcal{H}^m, \exp(g_n(t)) \leq \zeta, J(g,g) \leq \zeta\}$ for some constant $\zeta > 0$.

Theorem 3.7. *Assume that Conditions* $(C3.1) \sim (C3.2)$ *are satisfied. If* $m > (3+\sqrt{5})/4$, $h \asymp n^{-d}$, *where* $1/(2m+1) \leq d < 1/3$. *Suppose that uniformly over* $g_n \in \mathcal{A}$, $\|\hat{g}_{n,\lambda} - g_{n_0}\|_\lambda = O_p((nh)^{-1/2} + h^m)$ *holds under* $H_{1n} : g = g_{n_0}$. *Then for any* $\delta \in (0,1)$, *there exist positive constants* C *and* N *that*

$$\inf_{n \geq N} \inf_{g_n \in \mathcal{A}, \|g_n\|_\lambda \geq C\eta_n} P(reject \quad H_0^{global} | H_{1n} \quad is \quad true) \geq 1 - \delta,$$

where $\eta_n \geq \sqrt{h^{2m} + (nh^{1/2})^{-1}}$. *The minimal lower bound of* η_n, *that is* , $n^{-2m/(4m+1)}$, *is achieved when* $h = h^** = n^{-2/(4m+1)}$.

Theorem 3.7 presents that, $h = h^** = n^{-2/(4m+1)}$, the PLRT can detect any local alternatives with separation rates no faster than $n^{-2m/(4m+1)}$, which turns out to be the minimax rate of testing in the sense of Ingster (1993).

3.5 Simulation

In order to get the estimator of the cumulative hazard function, we would use the B-spline functions to approximate the true function. Specifically, we have $g(t) \approx B'(t)\theta$, where $B(t) = (B_1(t), B_2(t), \ldots, B_q(t))^\top$ is the B spline basis and q is the number of basis.

The objective function to be minimized for the B-spline coefficients θ is:

$$l_{n,\lambda}(\theta) = \frac{1}{n}\sum_{i=1}^{n} -\delta_i \log(1 - \exp(-B^\top(t)\theta) + (1-\delta_i)B^\top(U_i)\theta + \frac{\lambda}{2}\theta^\top \Omega\theta,$$

where $\Omega_{lm} \equiv \int_0^1 \ddot{B}_l(s)\ddot{B}_m(s)\,ds, l, m = 1, 2, \ldots, q$, $\Omega = (\Omega_{lm})$. If we let $\theta^{(k)}$ be the kth approximation to the minimizer, then $\theta^{(k+1)}$ is:

$$\theta^{(k+1)} = \Pi\big[\theta^{(k)} - \tau \cdot [H^{(k)} + \lambda\Omega]^{-1}\big],$$

where τ is the value in the sequence $\{1, 1/2, 1/4, 1/8, 1/16, 0\}$ to result in a reduction in the objective function, $\Pi[x] = arg\min_{z\in\mathcal{C}}$, where $\mathcal{C} = \{z \in R^q : 0 = z_1 \le z_2 \le \ldots \le z_q\}$.

$$H_{lm}^{(k)} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{B_l(U_i)B_m(U_i)\exp(-\sum_{i=1}^{q} B^\top(U_i)\theta)}{(1-\exp(-\sum_{i=1}^{n} B^\top(U_i)\theta)^2},$$

$$d_j^{(k)} = \frac{1}{n}\sum_{i=1}^{n} -\delta_i \frac{\exp(-\sum_{i=1}^{q} B^\top(U_i)\theta)B_j(U_i)}{1-\exp(-\sum_{i=1}^{q} B^\top(U_i)\theta)} + (1-\delta_i)B_j(U_i) + \lambda[\Omega\theta^{(k)}]_j,$$

where The notation $[\Omega\theta^{(k)}]_j$ stands for the jth component of the vector $\Omega\theta^{(k)}$. In order to verify the theoretical results, we present some simulated examples.

Example 1: In this example, the failure time follows from Weibull distribution:

$$f(x) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}\exp(-(\frac{x}{\lambda})^k), x \in [0, \infty],$$

with the corresponding parameter $k = 0.3, \lambda = 0.2$, while the censoring time follows from the truncated exponential distribution from 0 to 3, with mean of the exponential distribute function is chosen to yield 20% and 30% censoring rate. In order to get the estimate, we use the cubic spline to estimate of the true function, and the number of knots is at the order of $q_n = [4.7n^{1/4}]$. In this section, we use the AIC proposed by O'Sullivan (1988) to select the parameter λ. Specifically, define

$$H = \int_0^1 \frac{\exp\{g(t)\}}{1 - \exp(g(t))} B(s)B(s)^\top \, ds,$$

$$\Omega_{lm} = \int_0^1 \ddot{B}_l(s)\ddot{B}_m(s)\, ds, \qquad l, m = 1, 2, \ldots, q_n,$$

and $\Omega = (\Omega_{lm})$, $\ddot{B}_l(s)$ is the second derivative of $B_l(s)$, $B = \{B_l(s), l = 1, 2 \ldots, q_n\}$ is the B-spline basis, then the AIC is:

$$AIC = -l_n + \frac{trace[(\hat{H} + \lambda\Omega)^{-1}\hat{H}]}{n},$$

In linear algebra, the trace of an n-by-n square matrix A is defined to be the sum of the elements on the main diagonal (the diagonal from the upper left to the lower right) of A.

To check the pointwise confidence interval and global simultaneous confidence band, we compare the method proposed by Wahba (1983). Specially, we denote the coverage probability derived from Wahba (1983) as W_CP, while that derived from our method is $local\_CP$. In this example, we set m=2, and following from the ODE functions (3.1), we can get the eigenvalues and eigenfunctions of the Hilbert space. Then plug in the estimation of σ_{t_0} and $\hat{G}_\alpha$, we can get the pointwise and global CI. The simulation results are reported in Figures 3.1-3.3 and Table 3.1.

From the figures, we can get that the local confidence interval got from our method is reasonable, that is accordance to the results in Shang and Cheng (2013). Besides, from Figures 3.1-3.3, we can get that the W_CP is almost around 1, that is because the variance is too large, while the $local\_CP$ is nearest to 95% at [0.5, 2.5] when the censoring rate being 20%. While the censoring rate is 30%, the W_CP is also almost around 1, while the $local\_CP$ is nearest to 95% at [0.5, 2]. Further, Table 3.1 gives the Global CP at different intervals. From the table, we can get that the global confidence band is reasonable.

3.6 Application

We apply our proposed method to the analysis of one dataset of heart disease (Detrano et al. (1989)). Actually, there were total 200 patients undergoing the heart disease at the Veterans Administration Medical Center in Long Beach, California, from 1984 to 1987. If we deem the age of patients as the observation time and the heart attack as the interest event, at the end of the study, there were 75% patients died of the heart disease, with the other 25% patients were subjected to case-one interval censored. Besides, we would scale the data to make the survival time ranging from 0 to 1. For comparison, we would give the confidence interval through various methods: the method given by Wahba (1983), our pointwise confidence interval and the global confidence interval. The results are shown in Figure 3.4. From the figure, we can get that the pointwise interval is shortest among the three methods, which are accordance to the simulation results.

3.7 Appendix

In order to study the asymptotic properties of the proposed estimators, we need the following regularity conditions.

(C3.1) $\exp\{-g_0(U)\} \le C < 1$ almost surely.

(C3.2) U and T are independent.

Proof of Lemma 3.1: Let $h_n(t) \in \mathcal{H}^m$ which satisfies that $\|h_n\|_\infty = O(n^{-1/2+\mu})$. Then we have that:

$$H_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \delta_i \log[1 - \exp\{-g_0(U_i) - \alpha h_n(U_i)\}] - (1-\delta_i)\{g_0(U_i) + \alpha h_n(U_i)\}$$

$$- \frac{\lambda}{2}\int_0^1 \{g_0^{(m)}(t) + \alpha h_n^{(m)}(t)\}^2\, dt$$

Then the derivative of $H_n(\alpha)$ is:

$$H_n'(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\delta_i \frac{h_n(U_i)}{1-\exp\{-g_0(U_i)-\alpha h_n(U_i)\}} - h_n(U_i)$$

$$-\lambda \int_0^1 g_0^{(m)}(t)h_n^{(m)}(t)\,dt - \alpha\lambda\int_0^1 (h_n^{(m)})^2\,dt$$

$$= \frac{1}{n}\sum_{i=1}^{n}\delta_i h_n(U_i)\big[\frac{1}{1-\exp\{-g_0(U_i)-\alpha h_n(U_i)\}} - \frac{1}{1-\exp\{-g_0(U_i)\}}\big] + h_n(U_i)$$

$$\times\big[\delta_i \frac{1}{1-\exp\{-g_0(U_i)\}} - 1\big] - \lambda\int_0^1 g_0^{(m)}(t)h_n^{(m)}(t)\,dt - \alpha\lambda\int_0^1 (h_n^{(m)}(t))^2\,dt$$

$$= -\alpha[\frac{1}{n}\sum_{i=1}^{n}\delta_i \frac{h_n^2(U_i)\exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]^2} + \lambda\int_0^1 h_n^{(m)}(t)^2\,dt]$$

$$+\frac{1}{n}\sum_{i=1}^{n}h_n(U_i)\big[\delta_i\frac{1}{1-\exp\{-g_0(U_i)\}} - 1\big] - \lambda\int_0^1 g_0^{(m)}(t)h_n^{(m)}(t)\,dt,$$

As $\|h_n\|_\infty = O(n^{-1/2+\mu})$, $\|h_n^{(m)}\|_\infty = O(n^{-1/2+\mu})$, $\lambda n^{1/2-\mu} \to 0$, we have that $H_n'(\alpha)\alpha < 0$. Besides, as

$$H_n''(\alpha) = -[\frac{1}{n}\sum_{i=1}^{n}\delta_i \frac{h_n^2(U_i)\exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]^2} + \lambda\int_0^1 h_n^{(m)}(t)^2\,dt],$$

we have that $H_n'(\alpha)$ is a nonincreasing function, so $\hat{g}_{n,\lambda}(t) \in [g_0(t)-\alpha h_n(t), g_0(t)+\alpha h_n(t)]$. Then we have that $\|\hat{g}_{n,\lambda} - g_0\|_\infty \le \alpha\|h_n\|_\infty \to 0$. Following from Schumaker (1981), there exist two B-spline functions $\hat{h}_{m,\lambda}, \hat{h}_{m,0}$ that we have $\|\hat{g}_{n,\lambda} - \hat{h}_{m,\lambda}\|_\infty = O(n^{-\nu m})$, $\|\hat{h}_{m,0} - g_0\|_\infty = O(n^{-\nu m})$, where $0 < \nu < 1/2$

and the number of knots have the same order of (n^{ν}). Then we have that

$$\begin{aligned} & \|\hat{h}_{m,\lambda} - \hat{h}_{m,0}\|_\infty \\ \leq\ & \|\hat{g}_{n,\lambda} - \hat{h}_{m,\lambda}\|_\infty + \|\hat{h}_{m,0} - g_0\|_\infty + \|\hat{g}_{n,\lambda} - g_0\|_\infty \\ =\ & O_p(n^{-\nu m} + n^{-1/2+\mu}). \end{aligned}$$

Then as $\hat{h}_{m,\lambda}, \hat{h}_{m,0}$ belong to the same finite space, we have that $\|\hat{h}_{m,\lambda}^{(m)} - \hat{h}_{m,0}^{(m)}\|_\infty$ has the same order of $\|\hat{h}_{m,0} - \hat{h}_{m,\lambda}\|_\infty$, namely $\|\hat{h}_{m,\lambda}^{(m)} - \hat{h}_{m,0}^{(m)}\|_\infty = O_p(n^{-\nu m} + n^{-1/2+\mu})$. Thereby, we have

$$\|\hat{g}_{n,\lambda}^{(m)} - g_0^{(m)}\|_\infty \leq \|\hat{g}_{n,\lambda}^{(m)} - \hat{h}_{m,\lambda}^{(m)}\|_\infty + \|\hat{g}_{n,\lambda} - \hat{h}_{m,\lambda}^{(m)}\|_\infty + \|g_0^{(m)} - \hat{h}_{m,0}^{(m)}\|_\infty.$$

As Schumaker(1981) says, the derivative of the spline approximate the derivatives of the function it approximates, so the first term and the last term in the above format is o(1), thereby,$\|\hat{g}_{n,\lambda}^{(m)} - g_0^{(m)}\|_\infty \to 0$.Then we have that $J(\hat{g}_{n,\lambda}^{(m)} - g_0^{(m)}, \hat{g}_{n,\lambda}^{(m)} - g_0^{(m)}) = o_p(1)$. Then the conclusions derived.

Proof of Theorem 3.1: Denote $g = \hat{g}_{n,\lambda} - g_0$, then

$$\begin{aligned} l_{n,\lambda}(g + g_0) - l_{n,\lambda}(g_0) &= \mathcal{S}_{n,\lambda}(g_0)g + \frac{1}{2}D\mathcal{S}_{n,\lambda}(g_0)gg + \frac{1}{6}D^2\mathcal{S}_{n,\lambda}(g^*)ggg \\ &\equiv I_1 + I_2 + I_3, \end{aligned}$$

where $g^* = g_0 + \alpha g$,with $\alpha \in [0,1]$.

$$\begin{aligned}
|6I_3| &= |D^2\mathcal{S}_{n,\lambda}(g^*)ggg| \\
&= \Big|\frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{[1+\exp\{-g^*(U_i)\}]\exp\{-g^*(U_i)\}g^3(U_i)}{[1-\exp\{-g^*(U_i)\}]^3}\Big| \\
&\leq \|g\|_\infty\Big|\frac{1}{n}\delta_i\frac{[1+\exp\{-g^*(U_i)\}]\exp\{-g^*(U_i)\}g^2(U_i)}{[1-\exp\{-g^*(U_i)\}]^3}\Big|
\end{aligned}$$

As $\|g\|_\infty = o(1)$, then when n is large enough, we have $\|\exp\{-g(U_i)\}\|_\infty \approx 1$. Besides, as $\exp\{-g_0(U_i)\} \leq C < 1$, we have

$$\begin{aligned}
|6I_3| &\approx \|g\|_\infty\frac{2}{1-C}\Big|\frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{\exp\{-g_0(U_i)\}g^2(U_i)}{[1-\exp\{-g_0(U_i)\}]^2}\Big| \\
&\leq \frac{2\|g\|_\infty}{n(1-C)}\Big|\sum_{i=1}^{n}\delta_i\frac{\exp\{-g_0(U_i)\}g^2(U_i)}{[1-\exp\{-g_0(U_i)\}]^2} - E\big(\frac{\exp\{-g_0(U_i)\}g^2(U_i)}{[1-\exp\{-g_0(U_i)\}]^2}\big)\Big| \\
&+ \|g\|_\infty\frac{2}{(1-C)}E\big(\frac{\exp\{-g_0(U_i)\}g^2(U_i)}{[1-\exp\{-g_0(U_i)\}]^2}\big) \\
&= \frac{2\|g\|_\infty}{n(1-C)}\Big| < \sum_{i=1}^{n}\psi(\delta_i,U_i,g)K_{U_i} - E[\psi(\delta,U,g)K_U], g >_\lambda \Big| \\
&+ \frac{2\|g\|_\infty}{1-C}E\big(\frac{\exp\{-g_0(U_i)\}g^2(U_i)}{[1-\exp\{-g_0(U_i)\}]^2}\big).
\end{aligned}$$

Define $\tilde{\psi}(\delta_i,U_i,g) = \frac{(1-C)^2}{C}c_m^{-1}h^{1/2}\psi(\delta_i,U_i,g)$, $i = 1,2,\ldots,n$, then

$$\begin{aligned}
&|\tilde{\psi}(\delta_i,U_i,g) - \tilde{\psi}(\delta_i,U_i,f)| \\
&= \frac{(1-C)^2}{C}c_m^{-1}h^{1/2}\frac{\delta_i\exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]^2}|f(U_i)-g(U_i)| \\
&\leq c_m^{-1}h^{1/2}\|f-g\|_\infty.
\end{aligned}$$

Then it follows from Shang and Cheng (2013) that

$$\|\sum_{i=1}^{n}[\tilde{\psi}(\delta_i, U_i, g)K_{U_i} - E\{\tilde{\psi}(\delta_i, U_i, g)K_{U_i}\}]\|_\lambda$$

$$\leq (n^{1/2}\|g\|_\infty^{1-1/(2m)} + 1)\{5\log\log(n)\}^{1/2}.$$

Thereby, we have that

$$\frac{2\|g\|_\infty}{n(1-C)}| < \sum_{i=1}^{n}\psi(\delta_i, U_i, g)K_{U_i} - E[\psi(\delta, U, g)K_{U_i}], g >_\lambda |$$

$$\leq \quad c_m h^{-1/2}\frac{C}{(1-C)^2}\frac{\|g\|_\infty\|g\|_\lambda}{n}(n^{1/2}\|g\|_\infty^{1-1/(2m)} + 1)\{5\log\log(n)\}^{1/2}.$$

As

$$\|g\|_\infty\frac{2}{1-C}E\big[\frac{\exp\{-g_0(U_i)\}g^2(U_i)}{1-\exp\{-g_0(U_i)\}}\big] \quad \leq \quad \|g\|_\infty\frac{2}{(1-C)^2}\|g\|_\lambda^2,$$

we have that

$$\begin{aligned}
&|6I_3| \\
&\leq \frac{C}{(1-C)^2}c_m h^{-1/2}\frac{\|g\|_\infty\|g\|_\lambda}{n}(n^{1/2}\|g\|_\infty^{1-1/(2m)} + 1)\{5\log\log(n)\}^{1/2} \\
&+ \frac{2}{(1-C)^2}\|g\|_\infty\|g\|_\lambda^2 \\
&\leq \frac{2}{n^{1/2}h(1-C)^3}c_m^2\{\log\log(n)\}^{1/2}\|g\|_\lambda + \frac{2}{(1-C)^2}\|g\|_\infty\|g\|_\lambda.
\end{aligned}$$

Thereby, following $(n^{1/2}h)^{-1}\{\log\log(n)\}^{1/2} = o(1)$, we have $|6I_3| = o_p(1)\|g\|_\lambda^2$.

$$|I_1| = |\mathcal{S}_{n,\lambda}(g_0)g| \leq \|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda\|g\|_\lambda = O_p((nh)^{-1/2} + \lambda^{1/2})\|g\|_\lambda.$$

Regarding to I_2, we have

$$
\begin{aligned}
2I_2 &= D\mathcal{S}_{n,\lambda}(g_0)gg \\
&= \{D\mathcal{S}_{n,\lambda}(g_0)gg - ED\mathcal{S}_{n,\lambda}(g_0)gg\} + ED\mathcal{S}_{n,\lambda}(g_0)gg \\
&= -\|g\|_\lambda^2 + \{D\mathcal{S}_{n,\lambda}(g_0)gg - ED\mathcal{S}_{n,\lambda}(g_0)gg\} \\
&= -\|g\|_\lambda^2 + \frac{1}{n}\sum_{i=1}^{n}\Big(\delta_i \frac{g^2(U_i)\exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]^2} - E\frac{g^2(U_i)\exp\{-g_0(U_i)\}}{1-\exp\{-g_0(U_i)\}}\Big).
\end{aligned}
$$

It follows from the proof of I_3, we have that

$$
\begin{aligned}
&\Big|\frac{1}{n}\sum_{i=1}^{n}\delta_i \frac{g^2(U_i)\exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]^2} - E\big[\frac{g^2(U_i)\exp\{-g_0(U_i)\}}{1-\exp\{-g_0(U_i)\}}\big]\Big| \\
\leq\ & \frac{1}{(1-C)^2}c_m h^{-1/2}\|g\|_\lambda\big(\frac{1}{\sqrt{n}}\|g\|_\infty^{1-1/(2m)} + \frac{1}{n}\big)\{5\log\log(n)\}^{1/2} \\
=\ & \frac{1}{\sqrt{nh}(1-C)^2}c_m\|g\|_\lambda\|g\|_\infty^{1-1/(2m)}\{5\log\log(n)\}^{1/2} \\
+\ & \frac{1}{\sqrt{h}n(1-C)^2}c_m\|g\|_\lambda\{5\log\log(n)\}^{1/2}.
\end{aligned}
$$

Then, we have that

$$
\begin{aligned}
2I_2 &= -\|g\|_\lambda + \frac{1}{\sqrt{nh}(1-C)^2}c_m\|g\|_\lambda\|g\|_\infty^{1-1/(2m)}\{5\log\log(n)\}^{1/2} \\
&+ \frac{1}{n\sqrt{h}(1-C)^2}c_m\|g\|_\lambda\{5\log\log(n)\}^{1/2}.
\end{aligned}
$$

Thereby, we have that

$$\begin{aligned}
&\|g\|_\lambda(1+o_p(1))\\
&\le \|g\|_\lambda^{2-1/(2m)} n^{-1/2} c_m^{2-1/(2m)} h^{-1+1/(4m)} \{5\log\log(n)\}^{1/2}\\
&+\frac{1}{nh^{1/2}(1-C)^2}\|g\|_\lambda\{5\log\log(n)\}^{1/2}+O_p((nh)^{-1/2}+\lambda^{1/2})\|g\|_\lambda.
\end{aligned}$$

$$\begin{aligned}
\|g\|_\lambda \le ((nh)^{-1/2}+\lambda^{1/2}) + \|g\|_\lambda^{1-1/(2m)} n^{-1/2} c_m^{2-1/(2m)} h^{-1+1/(4m)}\{5\log\log(n)\}^{1/2}\\
+\frac{1}{nh^{1/2}(1-C)^2}\{5\log\log(n)\}^{1/2}.
\end{aligned}$$

As $(nh^{1/2})^{-1}\{5\log\log(n)\}^{1/2} = o((nh)^{-1/2})$, we have that

$$\|g\|_{n,\lambda} \le (nh)^{-1/2}+\lambda^{1/2}+\|g\|_\infty^{1-1/(2m)} n^{-1/2}\{5\log\log(n)\}^{1/2} c_m h^{-1/2}.$$

Besides, as $\|g\|_\infty = o_p(1)$, $(nh)^{-1/2}\{5\log\log(n)\}^{1/2} = o(1)$, we have that $\|g\|_\lambda \le (nh)^{-1/2}+h^m$.

Proof of Theorem 3.2: Denote $g=\hat{g}_{n,\lambda}-g_0$, following from theorem 3.1, we have that $\|g\|_\lambda = O_p((nh)^{-1/2}+h^m)$. Thereby, there exists a constant M, s.t. $B_n=\{\|g\|_\lambda \le r_n \equiv M((nh)^{-1/2}+h^m)\}$ has large probability. Define $\tilde{g}=d_n^{-1}g$, where $d_n = c_m r_n h^{-1/2}$, since $h=o(1)$, and $\{\log\log(n)\}(nh^2)^{-1}\to 0$, we have that $d_n=o(1)$. Besides, on B_n, we have $\|\tilde{g}\|_\infty \le 1$ and $J(\tilde{g},\tilde{g}) = d_n^{-2}\lambda^{-1}(\lambda J(g,g)) \le d_n^{-2}\lambda^{-1}\|g\|_\lambda^2 = c_m^{-2}\lambda^{-1}h^{-1}$. Thus, when the event B_n holds, we have $\tilde{g}\in\mathcal{F}$, where $\mathcal{F}=\{g:\|g\|_\infty\le 1, J(g,g)\le c_m^{-2}h\lambda^{-1}\}$. By the

Taylor expansion, we have that

$$
\begin{aligned}
& \mathcal{S}_n(\hat{g}_{n,\lambda}) - \mathcal{S}_n(g_0) - (\mathcal{S}(\hat{g}_{n,\lambda}) - \mathcal{S}(g_0)) \\
= & -\frac{1}{n}\sum_{i=1}^{n} K_{U_i} + \delta_i \frac{K_{U_i}}{1-\exp\{-\hat{g}_{n,\lambda}(U_i)\}} + \frac{1}{n}\sum_{i=1}^{n}(K_{U_i} - \delta_i \frac{K_{U_i}}{1-\exp\{-g_0(U_i)\}}) \\
- & E[\delta \frac{K_U}{1-\exp\{-\hat{g}_{n,\lambda}(U)\}} - \delta \frac{K_U}{1-\exp\{-g_0(U)\}}] \\
= & \frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{K_{U_i}}{1-\exp\{-\hat{g}_{n,\lambda}(U_i)\}} - \delta_i \frac{K_{U_i}}{1-\exp\{-g_0(U_i)\}} \\
- & E[\delta \frac{K_U}{1-\exp\{-\hat{g}_{n,\lambda}(U)\}} - \delta \frac{K_U}{1-\exp\{-g_0(U)\}}] \\
= & -\frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{K_{U_i}\exp\{-g_0(U_i)\}g(U_i)}{[1-\exp\{-g_0(U_i)\}]^2} \\
- & E\delta_i \frac{K_{U_i}\exp\{-g_0(U_i)\}g(U_i)}{[1-\exp\{-g_0(U_i)\}]^2} \\
- & \frac{1}{n}\sum_{i=1}^{n} \Big(\delta_i \frac{\exp\{-g_0(U_i)\}K_{U_i}g^2(U_i)[1+\exp\{-g_0(U_i)\}]}{[1-\exp\{-g_0(U_i)\}]^3} \\
- & E\delta_i \frac{\exp\{-g_0(U_i)\}K_{U_i}g^2(U_i)[1+\exp\{-g_0(U_i)\}]}{[1-\exp\{-g_0(U_i)\}]^3}\Big)(1+o_p(1)) \\
\equiv & I_1 + I_2.
\end{aligned}
$$

$$
\begin{aligned}
I_1 &= -\frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{K_{U_i}\exp\{-g_0(U_i)\}g(U_i)}{[1-\exp\{-g_0(U_i)\}]^2} - E\delta_i \frac{K_{U_i}\exp\{-g_0(U_i)\}g(U_i)}{[1-\exp\{-g_0(U_i)\}]^2} \\
&= -\frac{1}{n}\sum_{i=1}^{n} \phi(\delta_i, U_i, g)K_{U_i} - E\phi(\delta_i, U_i, g)K_{U_i},
\end{aligned}
$$

where

$$\phi(\delta_i, U_i, g) = \delta_i \frac{\exp\{-g_0(U_i)\}g(U_i)}{[1-\exp\{-g_0(U_i)\}]^2},$$

define

$$\tilde{\phi}(\delta_i, U_i, g) = \frac{1-C}{C} d_n^{-1} \phi(\delta_i, U_i, d_n \tilde{g}) c_m^{-1} h^{1/2},$$

then we have

$$\begin{aligned}
& |\tilde{\phi}(\delta_i, U_i, \tilde{g}) - \tilde{\phi}(\delta_i, U_i, \tilde{f})| \\
\leq \quad & \frac{1-C}{C} d_n^{-1} c_m^{-1} h^{1/2} |\phi(\delta_i, U_i, d_n \tilde{g}) - \phi(\delta_i, U_i, d_n \tilde{f})| \\
= \quad & \frac{1-C}{C} d_n^{-1} c_m^{-1} h^{1/2} \Big| \frac{\delta_i \exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]^2} d_n(\tilde{g} - \tilde{f}) \Big| \\
\leq \quad & c_m^{-1} h^{1/2} \|\tilde{f} - \tilde{g}\|_\infty,
\end{aligned}$$

Then we have

$$\begin{aligned}
& \| \sum_{i=1}^n \tilde{\phi}(\delta_i, U_i, \tilde{g}) K_{U_i} - E\tilde{\phi}(\delta_i, U_i, \tilde{g}) K_{U_i} \|_\lambda \\
\leq & (n^{-1/2} h^{-(2m-1)/(4m)} \|\tilde{g}\|_\lambda^{1-1/(2m)} + 1)\{5 \log\log(n)\}^{1/2}.
\end{aligned}$$

Thereby, we have that

$$\begin{aligned}
& \|I_1\|_\lambda \\
& = \frac{1}{n} \| \sum_{i=1}^n \phi(\delta_i, U_i, g) - E\phi(\delta_i, U_i, g) \|_\lambda \\
& \leq \frac{1}{n} (n^{1/2} h^{-(2m-1)/(4m)} \|\tilde{g}\|_\infty^{1-1/(2m)} + 1)\{5 \log\log(n)\}^{1/2} \frac{C}{1-C} d_n c_m h^{-1/2} \\
& \leq (n^{-1/2} h^{-(2m-1)/(4m)} \|\tilde{g}\|_\infty^{1-1/(2m)} + n^{-1})\{5 \log\log(n)\}^{1/2} \frac{C}{1-C} d_n c_m h^{-1/2}.
\end{aligned}$$

As $\|\tilde{g}\|_\infty \leq 1$, we have

$$\begin{aligned}\|I_1\|_\lambda &\leq (n^{-1/2}h^{-(6m-1)/(4m)} + n^{-1}h^{-1})\{5\log\log(n)\}^{1/2}\frac{C}{1-C}c_m^2 r_n \\ &= O_p(n^{-1/2}h^{-(6m-1)/(4m)}\{5\log\log(n)\}^{1/2}\{(nh)^{-1/2} + h^m\}).\end{aligned}$$

Regarding to I_2, we have that

$$\begin{aligned}I_2 &= -\frac{1}{n}\sum_{i=1}^n\Big(\delta_i\frac{\exp\{-g_0(U_i)\}K_{U_i}g^2(U_i)[1+\exp\{-g_0(U_i)\}]}{[1-\exp\{-g_0(U_i)\}]^3} \\ &- E\delta_i\frac{\exp\{-g_0(U_i)\}K_{U_i}g^2(U_i)[1+\exp\{-g_0(U_i)\}]}{[1-\exp\{-g_0(U_i)\}]^3}\Big),\end{aligned}$$

Thus, we have

$$\|I_2\|_\lambda \leq \frac{2\|g\|_\infty}{1-C}\|I_1\|_\lambda = o_p(\|I_1\|_\lambda).$$

Thereby, it is easy to check that

$$\begin{aligned}&\|\mathcal{S}_n(\hat{g}_{n,\lambda}) - \mathcal{S}_n(g_0) - (\mathcal{S}(\hat{g}_{n,\lambda}) - \mathcal{S}(g_0))\|_\lambda \\ =\ & O_p(n^{-1/2}h^{-(6m-1)/(4m)}\{\log\log(n)\}^{1/2}\{(nh)^{-1/2} + h^m\}).\end{aligned}$$

On another hand, we have

$$\begin{aligned}&\mathcal{S}_n(\hat{g}_{n,\lambda}) - \mathcal{S}_n(g_0) - (\mathcal{S}(\hat{g}_{n,\lambda}) - \mathcal{S}(g_0)) \\ =\ & \mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda}) - \mathcal{S}_{n,\lambda}(g_0) - (\mathcal{S}_\lambda(\hat{g}_{n,\lambda}) - \mathcal{S}_\lambda(g_0)) \\ =\ & -\mathcal{S}_{n,\lambda}(g_0) - (\mathcal{S}_\lambda(\hat{g}_{n,\lambda}) - \mathcal{S}_\lambda(g_0)) \\ =\ & g - \mathcal{S}_{n,\lambda}(g_0) - \int_0^1\int_0^1 sD^2\mathcal{S}_\lambda(g_0 + ss'g)g^2\,ds\,ds',\end{aligned}$$

As $\| \int_0^1 \int_0^1 sD^2\mathcal{S}_\lambda(g_0 + ss'g)g^2 \, ds \, ds' \|_\lambda \le \int_0^1 \int_0^1 \|D^2\mathcal{S}_\lambda(g_0 + ss'g)g^2\|_\lambda \, ds \, ds'$, and

$$\|D^2\mathcal{S}_\lambda(g_0 + ss'g)g^2\|_\lambda = O_p(h^{-1/2}\{(nh)^{-1/2} + h^m\}^2),$$

we have that $\|g - \mathcal{S}_{n,\lambda}(g_0)\|_\lambda = O_p(\alpha_n)$, where

$$\alpha_n = h^{-1/2}\{(nh)^{-1} + h^{2m}\} + n^{-1/2}h^{-(6m-1)/(4m)}\{\log\log(n)\}^{1/2}\{(nh)^{-1/2} + h^m\}.$$

Proof of Theorem 3.3: Define $Rem_n = \hat{g}_{n,\lambda} - g^* - \mathcal{S}_n(g_0)$,then it follows from the Functional Bahadur representation, we have that $\|Rem_n\|_\lambda = O_p(\alpha_n)$. As $nh^3 \to \infty$, $nh^{4m-1} \to 0$, $m > (3 + \sqrt{5})/4$, we have that $\alpha_n = o_p(n^{-1/2})$. Since $\|\mathcal{S}_n(g_0)\|_\lambda = O_p((nh)^{-1/2})$, thus Rem_n is negligible compared with $\mathcal{S}_n(g_0)$. Next, we would show the asymptotic distribution of $(nh)^{-1/2}\{\hat{g}_{n,\lambda}(t_0) - g^*(t_0)\}$. We would use the fact that for any $t \in I$, and any $g \in \mathcal{H}^m$, we have $< K_t, g >_\lambda = g(t)$. Thereby, for any fixed $t_0 \in I$, we have

$$|(nh)^{-1/2} < K_{t_0}, \hat{g}_{n,\lambda} - g^* - \mathcal{S}_n(g_0) >_\lambda | \le \|K_{t_0}\|_\lambda \|\hat{g}_{n,\lambda} - g^* - \mathcal{S}_n(g_0)\|_\lambda (nh)^{1/2}$$
$$= o_p(1),$$

As

$$-(nh)^{1/2} < K_{t_0}, \mathcal{S}_n(g_0) >_\lambda = (nh)^{1/2}\frac{1}{n}\sum_{i=1}^n K_{t_0}(U_i)\big(1 - \frac{\delta_i}{1 - \exp\{-g_0(U_i)\}}\big),$$

$Var\Big(K_{t_0}(U_i)\big(1-\frac{\delta_i}{1-\exp\{-g_0(U_i)\}}\big)\Big) = V(K_{t_0},K_{t_0}),\ hV(K_{t_0},K_{t_0}) \to \sigma_{t_0}^2 < c_m^2,$

we have

$$(nh)^{1/2} < K_{t_0}, \mathcal{S}_n(g_0) >_\lambda \xrightarrow{d} N(0, \sigma_{t_0}^2).$$

Then the conclusion follows.

It follows from the multivariate central limit theorem that

$$(n^{-1}h)^{1/2}\sum_{i=1}^n K_{t_0}(U_i)\big(1-\delta_i[1-\exp\{-g_0(U_i)\}^{-1}]\big)$$

converges in finite dimensional distributions to a zero-mean Gaussian distribution with the covariance function is $\Sigma(s,t)$. Besides, As

$$K_t(s) = \sum_{j=0}^\infty \frac{h_j(t)}{1+\lambda\gamma_j}h_j(s),$$

it's easy to get that

$$\begin{aligned}
& (nh)^{1/2}\frac{1}{n}\sum_{i=1}^n K_t(U_i)\big[1-\frac{\delta_i}{1-\exp\{-g_0(U_i)\}}\big] \\
= \quad & (nh)^{1/2}\frac{1}{n}\sum_{i=1}^n\sum_{j=0}^\infty \frac{h_j(t)}{1+\lambda\gamma_j}h_j(U_i)\Big[1-\frac{\delta_i}{1-\exp\{-g_0(U_i)\}}\Big] \\
= \quad & \sum_{j=0}^\infty \frac{h^{1/2}h_j(t)}{1+\lambda\gamma_j}\frac{1}{\sqrt{n}}\sum_{i=1}^n h_j(U_i)\Big[1-\frac{\delta_i}{1-\exp\{-g_0(U_i)\}}\Big]
\end{aligned}$$

Denote

$$Z_n(t) = \sum_{j=0}^\infty \frac{h^{1/2}h_j(t)}{1+\lambda\gamma_j}\frac{1}{\sqrt{n}}\sum_{i=1}^n h_j(U_i)\big(1-\delta_i[1-\exp\{-g_0(U_i)\}]^{-1}\big).$$

As $h^{1/2}h_j(t)/(1+\lambda\gamma_j)$ is a bounded deterministic function, $n^{-1/2}\sum_{i=1}^n h_j(U_i)\big(1-\delta_i[1-\exp\{-g_0(U_i)\}]^{-1}\big)$ is tight, $h^{1/2}h_j(t)/\{\sqrt{n}(1+\lambda\gamma_j)\}\sum_{i=1}^n h_j(U_i)\big(1-\delta_i[1-\exp\{-g_0(U_i)\}]^{-1}\big)$ is also tight. Then it follows from Kosorok (2008) that, we can get $h^{1/2}h_j(t)/\{\sqrt{n}(1+\lambda\gamma_j)\}\sum_{i=1}^n h_j(U_i)\big(1-\delta_i[1-\exp\{-g_0(U_i)\}]^{-1}\big)$ is a Donsker-Class. Again, following form Kosorok (2008), we have for any inter M, $\sum_{j=0}^M h^{1/2}h_j(t)/\{\sqrt{n}(1+\lambda\gamma_j)\}\sum_{i=1}^n h_j(U_i)\big(1-\delta_i[1-\exp\{-g_0(U_i)\}]^{-1}\big)$ is a Donsker-Class, denoted as $Z_{n,M}(t)$. Thereby, it follows from Kosorok (2008) that there exists a semimetric ρ for which I is totally bounded and

$$\lim_{\delta\downarrow 0}\lim P(\sup_{s,t\in I \text{ with} \rho(s,t)<\delta}|Z_{n,M}(t)-Z_{n,M}(s)|>\epsilon)=0$$

for all $\epsilon>0$. As it's easy to prove that $Z_{n,M}(t)$ uniformly in I converges to $Z_n(t)$ as $M\to\infty$. We have that for any $\epsilon>0$, there exists a M, s.t. $|Z_n(t)-Z_{n,M}(t)|<\epsilon/4$ for all $t\in I$. Then we have

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\lim P(\sup_{s,t\in I \text{ with} \rho(s,t)<\delta}|Z_n(t)-Z_n(s)|>\epsilon)\\
\le\ &\lim_{\delta\downarrow 0}\lim P(\sup_{s,t\in I \text{with} \rho(s,t)<\delta}|Z_{n,M}(t)-Z_{n,M}(s)|+|Z_{n,M}(t)-Z_n(t)|\\
&+|Z_{n,M}(s)-Z_n(s)|>\epsilon)\\
\le\ &\lim_{\delta\downarrow 0}\lim P(\sup_{s,t\in I \text{with} \rho(s,t)<\delta}|Z_{n,M}(t)-Z_{n,M}(s)|>\epsilon/2)=0
\end{aligned}$$

thereby, $Z_n(t)$ is tight. As for any finite-dimension $(t_1,t_2,\cdots,t_k)$, $Z_n(t_1,t_2,\cdots,t_k)\xrightarrow{d} Z(t_1,t_2,\cdots,t_k)$ when $n\to\infty$, we can get $Z_n(t)\xrightarrow{d} Z(t)$ uniformly in I. Thereby, $\sqrt{nh}\{\hat{g}_{n,\lambda}(t)-g^*(t)\}$ converges weakly in I to a mean zero Gaussian process $Z(t)$ with the covariance function at (s,t) equal to $\Sigma(s,t)$.

Proof of theorem 3.5(ii): For notational convenience, denote $\hat{g} = \hat{g}_{n,\lambda}$, $\hat{g}^0 = \hat{g}^0_{n,\lambda}$, $g = \hat{g}^0 + \omega_0 - \hat{g}$. By theorem 3.4, we have that $\|g\|_\lambda = \|\hat{g}^0 + \omega_0 - \hat{g}\|_\lambda \leq \|\hat{g}^0 + \omega_0 - g_0\|_\lambda + \|\hat{g} - g_0\|_\lambda = O_p(r_n)$, where $r_n = (nh)^{-1/2} + h^m$. By Taylor expansion, we have

$$\begin{aligned} LRT_{n,\lambda} &= L_{n,\lambda}(\omega_0 + \hat{g}^0) - L_{n,\lambda}(\hat{g}) \\ &= \mathcal{S}_{n,\lambda}(\hat{g})(\omega_0 + \hat{g}^0 - \hat{g}) + \int_0^1 \int_0^1 sD\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg\, ds\, ds' \end{aligned}$$

It follows from the definition of $\mathcal{S}_{n,\lambda}(\hat{g}) = 0$, we have $\mathcal{S}_{n,\lambda}(\hat{g})(\omega_0 + \hat{g}^0 - \hat{g}) = 0$. Then

$$\begin{aligned} &LRT_{n,\lambda} \\ &= \int_0^1 \int_0^1 sD\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg\, ds\, ds' \\ &= \int_0^1 \int_0^1 s[D\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg - D\mathcal{S}_{n,\lambda}(g_0)gg]\, ds\, ds' + \int_0^1 \int_0^1 sD\mathcal{S}_{n,\lambda}(g_0)gg\, ds\, ds' \\ &= \frac{1}{2}[D\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg - D\mathcal{S}_{n,\lambda}(g_0)gg] + \frac{1}{2}[D\mathcal{S}_{n,\lambda}(g_0)gg - D\mathcal{S}_{\lambda}(g_0)gg] + \frac{1}{2}D\mathcal{S}_{\lambda}(g_0)gg \\ &\equiv I_1 + I_2 + I_3. \end{aligned}$$

Define $\tilde{g} = \hat{g} + ss'g - g_0$, for any $0 \leq s \leq s' \leq 1$, $\|\tilde{g}\|_\lambda = \|\hat{g} - g_0 + ss'g\|_\lambda \leq \|\hat{g} - g_0\|_\lambda + \|g\|_\lambda = O_p(r_n)$.Then we have

$$\begin{aligned} D\mathcal{S}_{n,\lambda}(\hat{g} + ss'g)gg &= D\mathcal{S}_{n,\lambda}(\tilde{g} + g_0)gg \\ &= -\frac{1}{n}\sum_{i=1}^n \delta_i \frac{g^2(U_i)\exp\{-\tilde{g}(U_i) - g_0(U_i)\}}{[1 - \exp\{-\tilde{g}(U_i) - g_0(U_i)\}]^2} - \lambda \int_0^1 \{g^{(m)}(t)\}^2\, dt. \end{aligned}$$

Thereby, we have

$$
\begin{aligned}
&|D\mathcal{S}_{n,\lambda}(\hat{g}+ss'g)gg - D\mathcal{S}_{n,\lambda}(g_0)gg| \\
&= |\frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{g^2(U_i)\exp\{-\tilde{g}(U_i)-g_0(U_i)\}}{[1-\exp\{-\tilde{g}(U_i)-g_0(U_i)\}]^2} - \frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{g^2(U_i)\exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]^2}| \\
&\approx |\frac{1}{n}\sum_{i=1}^{n}\delta_i g^2(U_i)\frac{\exp\{-g_0(U_i)\}[1+\exp\{-g_0(U_i)\}]\tilde{g}(U_i)}{[1-\exp\{-g_0(U_i)\}]^3}| \\
&\leq \frac{2\|\tilde{g}_\infty\|}{1-C}|\frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{g^2(U_i)\exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]^2} - E\frac{g^2(U_i)\exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]}| \\
&+ \frac{2\|\tilde{g}_\infty\|}{1-C}E\frac{g^2(U_i)\exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]} \\
&\leq \frac{2\|\tilde{g}_\infty\|}{1-C}|\frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{g^2(U_i)\exp\{-g_0(U_i)\}}{[1-\exp\{-g_0(U_i)\}]^2} - E\frac{g^2(U_i)\exp\{-g_0(U_i)\}}{1-\exp\{-g_0(U_i)\}}| + \frac{2\|\tilde{g}_\infty\|}{1-C}\|g\|_\lambda^2 \\
&\equiv I_{11} + I_{12}.
\end{aligned}
$$

Following from the proof of theorem 3.2, we have that

$$I_{11} = \|\tilde{g}\|_\infty O_p(r_n\alpha_n'),$$

where $\alpha_n' = n^{-1/2}\{(nh)^{-1/2} + h^m\}h^{-(6m-1)/(4m)}\{\log\log(n)\}^{1/2}$. Then we have that

$$|I_1| = \|\tilde{g}\|_\infty O_p(r_n\alpha_n') + \|\tilde{g}\|_\infty O_p(r_n^2).$$

Following from the condition of λ, we have $n^{-1/2}h^{-(6m-1)/(4m)}\{\log\log(n)\}^{1/2} = o_p(1)$, thereby, we have $\alpha_n' = o_p(r_n)$, so we have

$$|2I_1| = \|\tilde{g}\|_\infty O_p(r_n^2) \leq h^{-1/2}r_n O_p(r_n^2) = O_p(h^{-1/2}r_n^3).$$

Further, we can easy get that

$$
\begin{aligned}
|2I_2| &= |D\mathcal{S}_{n,\lambda}(g_0)gg - D\mathcal{S}_\lambda(g_0)gg| \\
&= O_p(r_n\alpha_n').
\end{aligned}
$$

As $I_3 = -\|g\|_\lambda^2/2$, we have

$$
LRT_{n,\lambda} = -\frac{\|g\|_\lambda^2}{2} + O_p(h^{-1/2}r_n^3 + r_n\alpha_n').
$$

Following from $nh^{2m} \to c_0$, then we have $nh^{2m+1} \to 0$, along with $nh^4 \to \infty$, we have $h^{-1/2}r_n^3 + r_n\alpha_n' = o(n^{-1})$. Thereby, we have that

$$
-2nLRT_{n,\lambda} = n\|\hat{g}^0 + \omega_0 - \hat{g}\|_\lambda^2 + o_p(1).
$$

Then the first part of the proof is completed.

Proof of theorem 3.5(iii):As $-2nLRT_{n,\lambda} = n\|\hat{g}^0 + \omega_0 - \hat{g}\|_\lambda^2 + o_p(1)$, it's sufficient to give the asymptotic property of $n\|\hat{g}^0 + \omega_0 - \hat{g}\|_\lambda^2$. As

$$
\begin{aligned}
& n^{1/2}\|\hat{g}^0 + \omega_0 - \hat{g} - \mathcal{S}_{n,\lambda}^0(g_0^0) + \mathcal{S}_{n,\lambda}(g_0)\|_\lambda \\
\leq \; & n^{1/2}\|\hat{g}^0 + \omega_0 - \mathcal{S}_{n,\lambda}^0(g_0^0)\|_\lambda + n^{1/2}\|\hat{g} - \mathcal{S}_{n,\lambda}(g_0)\|_\lambda \\
= \; & O_p(n^{1/2}a_n) = o_p(1).
\end{aligned}
$$

Thereby, we just have to focus on $n^{1/2}\{\mathcal{S}^0_{n,\lambda}(g^0_0) - \mathcal{S}_{n,\lambda}(g_0)\}$. Recall that

$$\begin{aligned}
\mathcal{S}^0_{n,\lambda}(g^0_0) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K^*_{U_i}}{1-\exp\{-g^0_0(U_i)-\omega_0\}} - K^*_{U_i} - W^*_\lambda g^0_0 \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{1-\exp\{-g^0_0(U_i)-\omega_0\}}\big(K_{U_i} - \frac{K_{U_i}(t_0)K_{t_0}}{K(t_0,t_0)}\big) \\
&- \big(K_{U_i} - \frac{K_{U_i}(t_0)K_{t_0}}{(K(t_0,t_0))}\big) - \big(W_\lambda g_0 - \frac{W_\lambda(g_0)(t_0)K_{t_0}}{K(t_0,t_0)}\big).
\end{aligned}$$

Thus,

$$\mathcal{S}^0_{n,\lambda}(g^0_0) - \mathcal{S}_{n,\lambda}(g_0) = \frac{K_{t_0}}{K(t_0,t_0)}\big[\frac{1}{n}\sum_{i=1}^{n}\frac{-\delta_i K_{U_i}(t_0)}{1-\exp\{-g_0(U_i)\}} + K_{U_i}(t_0) + (W_\lambda g_0)(t_0)\big].$$

Thereby,

$$\begin{aligned}
&n^{1/2}\|\mathcal{S}^0_{n,\lambda}(g^0_0) - \mathcal{S}_{n,\lambda}(g_0)\|_\lambda \\
&= \Big|\frac{1}{\sqrt{K(t_0,t_0)}}\big[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{-\delta_i K_{U_i}(t_0)}{1-\exp\{-g_0(U_i)\}} + K_{U_i}(t_0) + (W_\lambda g_0)(t_0)\big]\Big|.
\end{aligned}$$

As $nh^{2m} \to 0$, we have

$$\frac{\sqrt{n}(W_\lambda g_0)(t_0)}{\|K_{t_0}\|_\lambda} \leq \frac{\sqrt{nh}(W_\lambda g_0)(t_0)}{h^{1/2}\|V^{1/2}(K_{t_0},K_{t_0})\|_\lambda} = O(1)\frac{\sqrt{nh}(W_\lambda g_0)(t_0)}{\sigma_{t_0}} = O(\sqrt{n}h^m) = o(1).$$

Thereby, we have

$$\frac{1}{\sqrt{K(t_0,t_0)}}\big[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{-\delta_i K_{U_i}(t_0)}{1-\exp(-g_0(U_i))} + K_{U_i}(t_0) + (W_\lambda g_0)(t_0)\big] \xrightarrow{d} N(0, c_{t_0}),$$

where

$$c_{t_0} = \lim_{h\to 0}\frac{V(K_{t_0},K_{t_0})}{\|K_{t_0}\|^2} \in (0,1].$$

Thereby, we have $-2nLRT_{n,\lambda} \xrightarrow{d} c_{t_0}\chi_1^2$. It follows immediately that $\|\hat{g}^0 + \omega_0 - \hat{g}\|_\lambda = O_p(n^{-1/2})$.

Proof of theorem 3.6: For simplify, denote $g = g_0 - \hat{g}_{n,\lambda}$, $r_n = (nh)^{-1/2} + h^m$. Using the Taylor expansion, we have

$$
\begin{aligned}
PLRT_{n,\lambda} &= l_{n,\lambda}(g_0) - l_{n,\lambda}(\hat{g}_{n,\lambda}) \\
&= \mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda})(g_0 - \hat{g}_{n,\lambda}) + \int_I \int_I s D\mathcal{S}_{n,\lambda}(\hat{g}_{n,\lambda} + ss'g)\, ds\, ds' \\
&\equiv I_1 + I_2.
\end{aligned}
$$

It follows from the the definition of $\mathcal{S}_{n,\lambda}$, we have that

$$|I_1| = 0.$$

using the argument very similar to the proof of theorem 5(ii), we have

$$|I_2| = -\frac{\|g\|_\lambda^2}{2} + O_p(h^{-1/2}r_n^3 + r_n\alpha_n'),$$

where $\alpha_n' = h^{-(6m-1)/(4m)}n^{-1/2}(\log\log(n))^{1/2}r_n$. Thus,

$$PLRT_{n,\lambda} = -\frac{\|g\|_\lambda^2}{2} + O_p(h^{-1/2}r_n^3 + r_n\alpha_n').$$

Following from the conditions that $m > (3+\sqrt{5})/4$, $1/(4m) \le v \le 1/(2m)$, $nh^{2m+1} = O(1)$, $nh^3 \to \infty$, we have

$$-2nPLRT_{n,\lambda} = n\|g\|_\lambda^2 + o_p(h^{-1/2}).$$

Under the hypothesis H_0^{global} that g_0 is "true" parameter, by theorem 2, we have $\|\hat{g}_{n,\lambda} - g_0 - \mathcal{S}_{n,\lambda}(g_0)\| = O_p(\alpha_n)$. Following from the theorem 3, we have $n^{1/2}\alpha_n = o(1)$, then we have

$$n^{1/2}\|g\|_\lambda = n^{1/2}\|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda + o_p(1).$$

Next, we study the leading term $\|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda$. Through direct computation, we have

$$\begin{aligned} n\|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda^2 &= n\|\frac{1}{n}\sum_{i=1}^{n} -K_{U_i} + \delta_i \frac{K_{U_i}}{1-\exp\{-g_0(U_i)\}} - W_\lambda g_0\|_\lambda^2 \\ &= \frac{1}{n}\|\sum_{i=1}^{n} -K_{U_i} + \delta_i \frac{K_{U_i}}{1-\exp\{-g_0(U_i)\}}\|_\lambda^2 \\ &- 2 < \sum_{i=1}^{n} -K_{U_i} + \delta_i \frac{K_{U_i}}{1-\exp\{-g_0(U_i)\}}, W_\lambda g_0 >_\lambda + n\|W_\lambda g_0\|_\lambda^2 \end{aligned}$$

We first approximate $\|W_\lambda g_0\|_\lambda$.Firstly, define

$$m_\lambda(j) = |V(g_0, h_j)|^2 \gamma_j \frac{\lambda\gamma_j}{1+\lambda\gamma_j}, \text{for} \qquad j = 0, 1, 2, \ldots.$$

Then $|m_\lambda(j)|$ is a sequence of functions satisfying that $|m_\lambda(j)| \leq |V(g_0, h_j)|^2\gamma_j \equiv m(j)$. From $g_0 \in \mathcal{H}^m$, we have that $|V(g_0, h_j)|^2\gamma_j = \int_N m(j)\, d\mu(j) = J(g_0, g_0) < \infty$, where $\mu(\dot{})$is the counting measure. As

$$\lim_{\lambda\to 0} m_\lambda(j) = 0$$

, we have

$$\sum_j |V(g_0, h_j)|^2 \frac{\lambda\gamma_j^2}{1+\lambda\gamma_j} = \int_N m_\lambda(j)\, dm(j) \to 0$$

based on the Lebesgue dominated convergence theorem. That is,

$$\|W_\lambda g_0\|_\lambda^2 = \sum_j |V(g_0, h_j)|^2 \frac{\lambda^2\gamma_j^2}{1+\lambda\gamma_j} = o(\lambda).$$

Following from this, we have

$$\begin{aligned} & E| < \sum_{i=1}^n -K_{U_i} + \delta_i \frac{K_{U_i}}{1-\exp\{-g_0(U_i)\}}, W_\lambda g_0 > |^2 \\ = \;& E|\sum_{i=1}^n \big(\frac{\delta_i}{1-\exp\{-g_0(U_i)\}} - 1\big) W_\lambda g_0|^2 \\ = \;& nE\big[\frac{\exp\{-g_0(U_i)\}}{1-\exp\{-g_0(U_i)\}}(W_\lambda g_0)^2)\big] \\ \leq \;& n\|W_\lambda(g_0(t))\|_\lambda^2 = o(n\lambda) \end{aligned}$$

Thereby, it follows from $nh^{2m+1} = O(1)$ that

$$\begin{aligned} & < \sum_{i=1}^n -K_{U_i} + \delta_i \frac{K_{U_i}}{1-\exp\{-g_0(U_i)\}}, W_\lambda g_0 > \\ & = o_p((n\lambda)^{1/2}) = o_p(n^{1/2}h^m) = o_p(h^{-1/2}). \end{aligned}$$

Thereby, we have

$$n\|\mathcal{S}_{n,\lambda}(g_0)\|_\lambda^2 = \frac{1}{n}\|\sum_{i=1}^n -K_{U_i} + \delta_i \frac{K_{U_i}}{1-\exp\{-g_0(U_i)\}}\|_\lambda^2 + o_p(h^{-1}).$$

In what follows, we study the limiting property of $n^{-1}\| \sum_{i=1}^{n} -K_{U_i} + \delta_i K_{U_i}[1-\exp\{-g_0(U_i)\}]^{-1}\|_\lambda^2$. Through directly computing, we have

$$\frac{1}{n}\| \sum_{i=1}^{n} -K_{U_i} + \delta_i \frac{K_{U_i}}{1-\exp\{-g_0(U_i)\}} \|_\lambda^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} \big[\frac{\delta_i}{1-\exp\{-g_0(U_i)\}} - 1\big]^2 < K_{U_i}, K_{U_i} > + \frac{1}{n} W_n,$$

where

$$W_n = \sum_{i\neq j} \big(\frac{\delta_i}{1-\exp\{-g_0(U_i)\}} - 1\big)\big(\frac{\delta_j}{1-\exp\{-g_0(U_j)\}} - 1\big) < K_{U_i}, K_{U_j} > .$$

Denote

$$W_{ij} = 2\big(\frac{\delta_i}{1-\exp\{-g_0(U_i)\}} - 1\big)\big(\frac{\delta_j}{1-\exp\{-g_0(U_j)\}} - 1\big) < K_{U_i}, K_{U_j} >,$$

then we can write $W_n = \sum_{1\le i<j\le n} W_{ij}$, so that W_n is clean (Jong (1987)). Next, we will derive the limiting distribution of W_n. Let $\sigma_n^2 = var(W_n)$, then through direct computation, we have

$$\sigma_n^2$$

$$= \frac{n(n-1)}{2} E(W_{ij}^2)$$

$$= 2n(n-1)E\left(\big(\frac{\delta_i}{1-\exp\{-g_0(U_i)\}} - 1\big)\big(\frac{\delta_j}{1-\exp\{-g_0(U_j)\}} - 1\big) < K_{U_i}, K_{U_j} >\right)^2$$

$$= 2n(n-1)\sum_{l=0}^{\infty} \frac{1}{(1+\lambda\gamma_l)^2}.$$

Let G_1, G_2, G_4 be:

$$G_1 = \sum_{i<j} E(W_{ij}^4),$$

$$G_2 = \sum_{i<j<k} E\{W_{ij}^2 W_{ik}^2\} + E\{W_{ji}^2 W_{jk}^2\} + E\{W_{ki}^2 W_{kj}^2\},$$

$$G_4 = \sum_{i<j<k<l} E\{W_{ij}W_{ik}W_{lj}W_{lk}\} + E\{W_{ij}W_{il}W_{kj}W_{kl}\} + E\{W_{ik}W_{il}W_{jk}W_{jl}\}.$$

It follows from the proposition 3.2 of Jong (1987) that, if G_1, G_2, G_3 are all of the lower order than σ_n^4, we will have $\sigma_n^{-1}W_n$ converge weekly to the standard normal distribution.

$$\begin{aligned}
&E\{W_{ij}^4\} \\
&= 16E\left\{\Big(\frac{\delta_i}{1-\exp\{-g_0(U_i)\}} - 1\Big)\Big(\frac{\delta_j}{1-\exp\{-g_0(U_j)\}} - 1\Big) < K_{U_i}, K_{U_j} >\right\}^4 \\
&= O(h^{-4}).
\end{aligned}$$

Thereby, we have $G_1 = O(n^2h^{-4})$. It follows from Cauchy-Schwarz inequity, we have $EW_{ij}^2W_{ik}^2 \leq (EW_{ij}^4)^{1/2}(EW_{ik}^4)^{1/2} = O(h^{-4})$. Thereby, $G_2 = O(n^3h^{-4})$. Straight forward calculation yields that

$$E\{W_{ij}W_{ik}W_{lj}W_{lk}\} = 16\sum_{j=0}^{\infty} \frac{1}{(1+\lambda\gamma_j)^4} = O(h^{-1}).$$

Therefore, we have $G_4 = O(n^4h^{-1})$. As $\sigma_n^4 = (\sigma_n^2)^2 = O(n^4h^{-2})$ and $nh^3 \to \infty$, $h = o(1)$, we have G_1, G_2, G_4 are of lower order than that of σ_n^4.So we have

$$\sigma_n^{-1}W_n \xrightarrow{d} N(0,1).$$

Following from $\rho_\lambda^2 = \sum_{j=0}^{\infty} h/(1+\lambda\gamma_j)^2$, we have

$$\frac{1}{\sqrt{2h^{-1}}n\rho_\lambda} W_n \xrightarrow{d} N(0,1). \tag{3.2}$$

Now, consider

$$\frac{1}{n}\sum_{i=1}^{n}\big[\frac{\delta_i}{1-\exp\{-g_0(U_i)\}}-1\big]^2 < K_{U_i}, K_{U_i} > .$$

Through straight forward calculation, we have

$$E\big\{\big(\frac{\delta_i}{1-\exp\{-g_0(U_i)\}}-1\big)^2 < K_{U_i}, K_{U_i} > \big\}^2 = O(\|K_U\|_\lambda^4) = O(h^{-2}).$$

Therefore, a direct calculation leads to

$$\begin{aligned} & E\big\{\sum_{i=1}^{n}\big[\frac{\delta_i}{1-\exp\{-g_0(U_i)\}}-1\big]^2 < K_{U_i}, K_{U_i} > -h^{-1}\sigma_\lambda^2\big\}^2 \\ \leq \; & nE\big\{\big(\frac{\delta_i}{1-\exp(-g_0(U_i))}-1\big)^2 < K_{U_i}, K_{U_i} > \big\}^2 = O(nh^{-2}), \end{aligned}$$

where $\sigma_\lambda^2 = \sum_{j=0}^{\infty} h/(1+\lambda\gamma_j)$. Thereby, we have

$$\frac{1}{n}\sum_{i=1}^{n}\big[\frac{\delta_i}{1-\exp\{-g_0(U_i)\}}-1\big]^2 < K_{U_i}, K_{U_i} >= h^{-1}\sigma_\lambda^2 + O_p((n^{1/2}h)^{-1}). \tag{3.3}$$

It follows from (3.2) and (3.3), we have $n\|\mathcal{S}_{n,\lambda}\|_\lambda^2 = O_p(h^{-1})$. Hence, we have

$$n^{1/2}\|\mathcal{S}_{n,\lambda}\|_\lambda = O_p(h^{-1/2}).$$

Thus, we have

$$\begin{aligned} -2nPLRT_{n,\lambda} &= \{n^{1/2}\|\mathcal{S}_{n,\lambda}\|_\lambda + o_p(1)\}^2 + o_p(h^{-1/2}) \\ &= n\|\mathcal{S}_{n,\lambda}\|_\lambda^2 + o_p(h^{-1/2}). \end{aligned} \tag{3.4}$$

It follows from (3.2),(3.3)and (3.4) that,

$$(2h^{-1}\sigma_\lambda^4/\rho_\lambda^2)^{-1/2}(-2n\gamma_\lambda PLRT_{n,\lambda} - n\gamma_\lambda\|W_\lambda g_0(t)\|_\lambda^2 - h^{-1}\sigma_\lambda^4/\rho_\lambda^2) \overset{d}{\longrightarrow} N(0,1).$$

Proof of theorem 3.7: Firstly, through straightforward calculation, we can verify that $m > (3+\sqrt{5})/4$, $1/(4m) \le v \le 1/(2m)$, $h \asymp n^{-d}$, where $1/(2m+1) \le d < 1/3$ satisfy the conditions in theorem 6. Throughout, we can only consider $g_{n_0} = g_0 + g_n$ for $g_n \in \mathcal{A}$. In order to prove the theorem, rewrite

$$\begin{aligned} -2n \cdot PLRT_{n,\lambda} &= -2n(l_{n,\lambda}(g_0) - l_{n,\lambda}(g_{n_0})) - 2n(l_{n,\lambda}(g_{n_0}) - l_{n,\lambda}(\hat{g}_{n,\lambda})) \\ &= I_1 + I_2. \end{aligned} \tag{3.5}$$

In order to complete the proof, it's sufficient to derive the order of the two parts respectively. Firstly, we would see the first part I_1. For simplicity, define

$$\begin{aligned} &R_i \\ &= (\delta_i \log[1-\exp\{-g_0(U_i)\}] - (1-\delta_i)g_0(U_i)) - (\delta_i \log[1-\exp\{-g_{n_0}(U_i)\}] \\ &-(1-\delta_i)g_{n_0}(U_i)) \\ &= \delta_i\big(\log[1-\exp\{-g_0(U_i)\}] - \log[1-\exp\{-g_{n_0}(U_i)\}]\big) - (1-\delta_i)g_n(U_i). \end{aligned}$$

It is easy to calculate that

$$ER_i^2 = O(\|g_n\|_\lambda^2).$$

Thereby, we have

$$E\{|\sum_{i=1}^{n}(R_i - ER_i)|^2\} \leq nER_i^2 = (n\|g_n\|_\lambda^2).$$

Thus, we have that $n\,(l_{n,\lambda}(g_0) - l_{n,\lambda}(g_{n_0}) - E(l_{n,\lambda}(g_0) - l_{n,\lambda}(g_{n_0}))) = O_p(n^{1/2}\|g_n\|_\lambda)$. On the other hand, follows from $D\mathcal{S}_\lambda(g)g_ng_n < 0$, for any $g \in \mathcal{H}^m$, there exists constant $c' > 0$ satisfies that

$$\begin{aligned}E\{D\mathcal{S}_{n,\lambda}(g_{n_0}^*)g_ng_n\} &\leq c'E\{D\mathcal{S}_{n,\lambda}(g_{n_0})g_ng_n\} \\ &= \frac{-c'\|g_n\|_\lambda^2}{2}.\end{aligned}$$

Thereby, we have that

$$\begin{aligned}E\{l_{n,\lambda}(g_0) - l_{n,\lambda}(g_{n_0})\} = E\{\mathcal{S}_{n,\lambda}(g_{n_0})(-g_n) + \frac{1}{2}D\mathcal{S}_{n,\lambda}(g_{n_0}^*)g_ng_n\} \\ \lambda J(g_{n_0}, g_n) - \frac{c'\|g_n\|_\lambda^2}{2} \\ \leq \{J(g_n, g_n) + J(g_0, g_n)\} - \frac{c'\|g_n\|_\lambda^2}{2} \\ \leq \{J(g_n, g_n) + J(g_0, g_0)^{1/2}J(g_n, g_n)^{1/2}\} - \frac{c'\|g_n\|_\lambda^2}{2} \\ = O(\lambda) - \frac{c'\|g_n\|_\lambda^2}{2}.\end{aligned}$$

Following from the above, we can get

$$\begin{aligned}I_1 &\geq n\|g_n\|_\lambda^2 + O_p(n\lambda + n^{1/2}\|g_n\|_\lambda + n^{1/2}\|g_n\|_\lambda^2) \\ &= n\|g_n\|_\lambda^2(1 + O_p(\lambda\|g_n\|_\lambda^{-2} + n^{-1/2}\|g_n\|_\lambda^{-1} + n^{-1/2})). \qquad (3.6)\end{aligned}$$

Secondly, let us see the second part I_2. It's easy to note that under H_{1n}, following from $\|\hat{g}_{n,\lambda} - g_{n0}\| = O_p((nh)^{-1/2} + h^m)$ and the FBR, we can get that

$$\inf_{n\geq N}\inf_{g_n\in\mathcal{A}} P_{g_{n0}}\left(\|\hat{g}_{n,\lambda}-g_{n0}-S_{n,\lambda}(g_{n,0})\|_\lambda \leq Mr_n\right), \tag{3.7}$$

where $r_n = (nh)^{-1/2} + h^m$. Similar to theorem 6, we can get I_2 has the same distribution as in theorem 6, but uniformly for $\forall g_n \in \mathcal{A}$. That's to say,

$$(2\nu_{n_0})^{-1/2}(I_2 - n\|W_\lambda g_{n_0}\|_\lambda^2 - h^{-1}\sigma_{n_0,\lambda}^2) = O_p(1), \tag{3.8}$$

uniformly for $\forall g_n \in \mathcal{A}$, where $\nu_{n_0} = h^{-1}\sigma_{n_0,\lambda}^4/\rho_{n_0,\lambda}^2$ with $\sigma_{n_0,\lambda}^2$, $\rho_{n_0,\lambda}^2$ are in the form of σ_λ^2, ρ_λ^2 but the eigenvalues and eigenvectors are derived under g_{n_0}. Let V_{n_0}, V_0 be the V functions as defined in section 2, then $\forall f \in \mathcal{H}^m$, we have

$$|V_{n_0}(f,f) - V_0(f,f)| = \zeta V_0(f,f)\|g_n\|_\infty.$$

It follows from shang and Cheng (2013) that

$$\sigma_{n_0,\lambda}^2 - \sigma_\lambda^2 = O(h^{-1/2}\|g_n\|_\lambda). \tag{3.9}$$

Thereby, it follows from (3.6), (3.8) and (3.9), we have

$$\begin{aligned}
&(2\nu_n)^{-1/2}(-2nr_\lambda PLRT - \nu_n)\\
=&(2\nu_n)^{-1/2}(-r_\lambda(I_1+I_2) - \nu_n)\\
=&(2\nu_n)^{-1/2}r_\lambda(I_2 - n\|W_\lambda g_{n_0}\|_\lambda^2 - h^{-1}\sigma_{n_0,\lambda}^2) + (2\nu_n)^{-1/2}r_\lambda n\|W_\lambda g_{n_0}\|_\lambda^2\\
&+ (2\nu_n)^{-1/2}r_\lambda I_1 + (2\nu_n)^{-1/2}r_\lambda h^{-1}(\sigma_{n_0,\lambda}^2 - \sigma_\lambda^2)\\
\geq& O_p(1) + (2\nu_n)^{-1/2}r_\lambda n\|g_n\|_\lambda^2(1 + O_p(\lambda\|g_n\|_\lambda^{-2} + n^{-1/2}\|g_n\|_\lambda^{-1} + n^{-1/2}))\\
&+ O(h^{-1}\|g_n\|_\lambda).
\end{aligned}$$

where $O_p(1)$ holds uniformly in $\mathcal{A}$, $\nu_n = h^{-1}\sigma_\lambda^4/\rho_\lambda^2$, and r_λ is defined in theorem 6. Let $\lambda\|g_n\|_\lambda^{-2} \leq 1/C$, $n^{-1/2}\|g_n\|_\lambda^{-1} \leq 1/C$, $Ch^{-1}\|g_n\|_\lambda \leq (nh^{1/2})\|g_n\|_\lambda^2$, and $\|g_n\|_\lambda^2 \geq C(nh^{1/2})^{-1}$ for some sufficiently large constant C. This means that $|(2\nu_n)^{-1/2}(-2nr_\lambda PLRT - \nu_n)| \geq c_\alpha$, where c_α is the cutoff value of (based on $N(0,1)$) for rejecting H_0^{global} at normal level α. This means that

$$\|g_n\|_\lambda^2 \geq C(h^{2m} + (nh^{1/2})^{-1}). \tag{3.10}$$

Combine (3.7) and (3.10), we can get the conclusion.

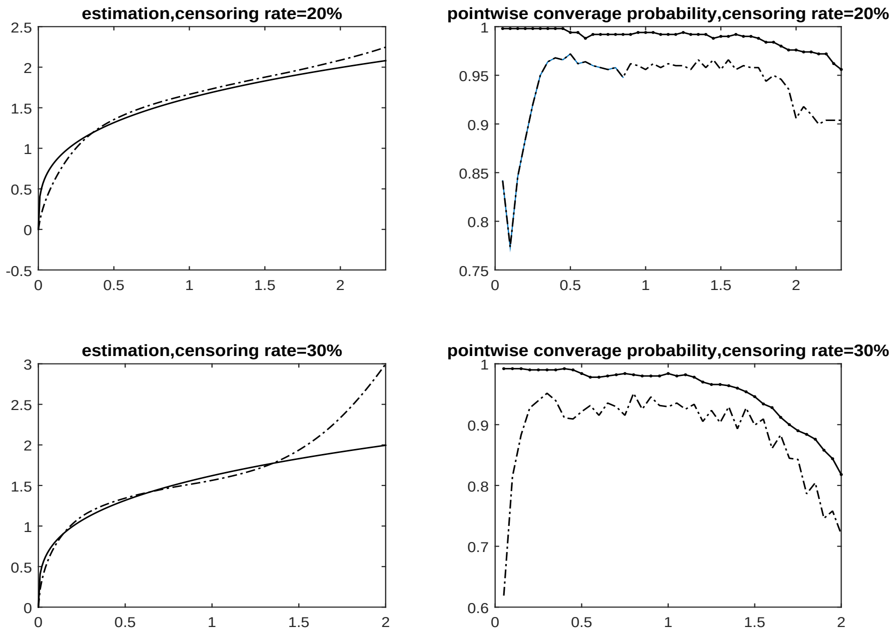

Figure 3.1: The picture gives the estimation and local CI of the example 1 with n=100. It shows the cumulative hazard estimation with censoring rate being 20% and 30% and the coverage probabilities with the censoring rate being 20% and 30%, respectively. Specifically, the solid line is the true cumulative hazard estimation, the $-.$ lines are that according to the censoring rate being 20% and 30%, respectively. The solid lines marked with $\cdot$ are the CI according to the method of Wahba.

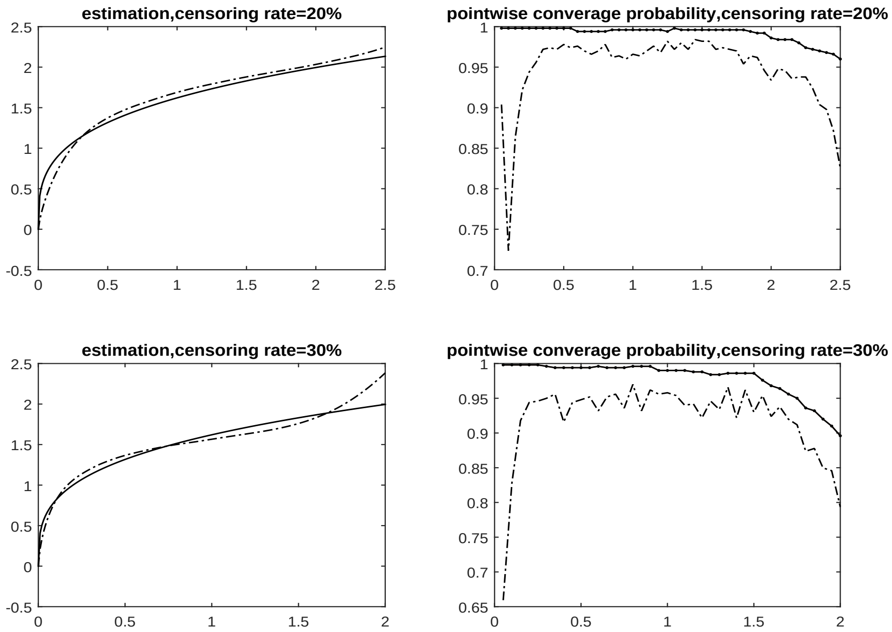

Figure 3.2: The picture gives the estimation and local CI of the example 1 with n=200. It shows the cumulative hazard estimation with censoring rate being 20% and 30% and the coverage probabilities with the censoring rate being 20% and 30%, respectively. Specifically, the solid line is the true cumulative hazard estimation, the $-.$ lines are that according to the censoring rate being 20% and 30%, respectively. The solid lines marked with $\cdot$ are the CI according to the method of Wahba.

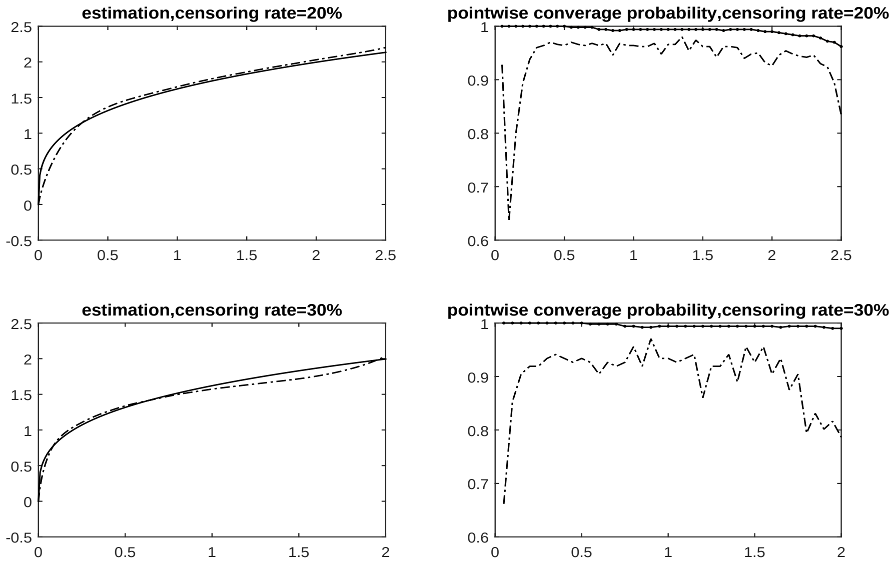

Figure 3.3: The picture gives the estimation and local CI of the example 1 with n=300. It shows the cumulative hazard estimation with censoring rate being 20% and 30% and the coverage probabilities with the censoring rate being 20% and 30%, respectively. Specifically, the solid line is the true cumulative hazard estimation, the −. lines are that according to the censoring rate being 20% and 30%, respectively. The solid lines marked with · are the CI according to the method of Wahba.

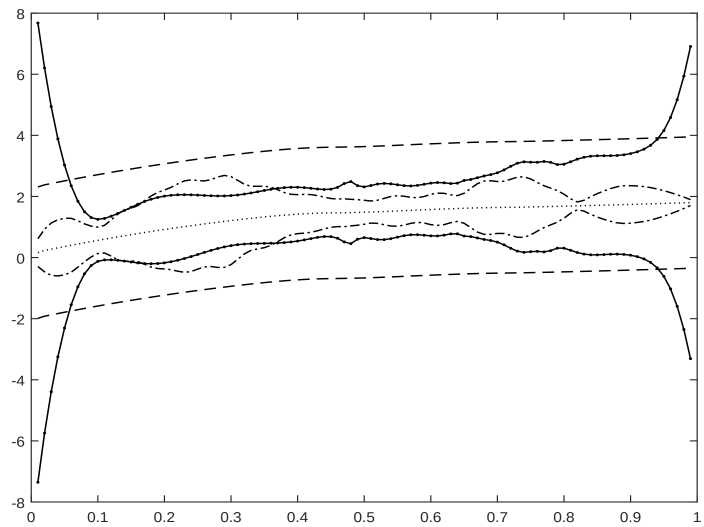

Figure 3.4: The pictures gives the real data analysis. It shows the cumulative hazard estimation and its confidence band various the three methods: the dash line is the cumulative hazard estimation, the $--$ line the simultaneous confidence band, the $-.$ line is the local pointwise confidence interval while the line marked with $\cdot$ is the confidence interval given by Wahba.

Table 3.1: Estimated global coverage probability with the nominal coverage probability being 95% for example 1.

interval	censoringrate=20%			censoringrate=30%		
	n=100	n=200	n=300	n=100	n=200	n=300
$[0.05, 0.95]$	95%	97%	97%	99%	99%	98%
$[0.05, 1.05]$	95%	97%	97%	98%	99%	98%
$[0.05, 1.15]$	95%	97%	97%	98%	99%	98%
$[0.05, 1.25]$	95%	97%	97%	97%	99%	98%
$[0.05, 1.35]$	95%	97%	97%	97%	99%	98%
$[0.05, 1.45]$	95%	97%	97%	96%	98%	98%
$[0.05, 1.55]$	95%	97%	97%	95%	98%	98%
$[0.05, 1.65]$	95%	97%	97%	94%	98%	98%
$[0.05, 1.75]$	95%	97%	97%	93%	98%	98%
$[0.05, 1.85]$	95%	97%	96%	91%	97%	97%
$[0.05, 1.95]$	95%	97%	96%	89%	96%	96%
$[0.05, 2.05]$	95%	97%	96%	89%	95%	95%
$[0.05, 2.15]$	95%	97%	96%	87%	94%	94%
$[0.05, 2.25]$	95%	97%	96%	85%	92%	93%
$[0.05, 2.35]$	94%	97%	96%	82%	90%	90%
$[0.05, 2.45]$	94%	97%	96%	80%	88%	88%
$[0.05, 2.55]$	93%	96%	96%	78%	85%	85%
$[0.05, 2.65]$	93%	96%	96%	75%	82%	83%
$[0.05, 2.75]$	92%	96%	95%	71%	78%	79%
$[0.05, 2.85]$	91%	95%	95%	68%	74%	75%
$[0.05, 2.95]$	87%	92%	94%	64%	70%	71%

Chapter 4

Conclusions and future work

This chapter draws conclusions on the thesis, and points out some possible research directions related to the work done in this thesis.

4.1 Conclusions

This thesis focuses on the nonparametric statistical inference of censored data. In particular, the right censored data and the current status data are studied in great detail.

1. In Chapter 2, the nonparametric inference focuses on the log-hazard function. The major advantage of doing so is that there is no constraint on the target function, and hence it simplifies the computation. Since the penalized nonparametric maximum likelihood estimation is quite useful to balance the smoothness and goodness-of-fit of the resulting estimator, we adopt the method here to estimate the log-hazard function with right censored data. On the other hand, the idea of smoothing B-spline can be also found in Schumaker (1981) for a smooth estima-

tion. The most appealing finding of the chapter is that a functional Bahadur representation is established in the Sobolev space $\mathcal{H}^m$ with a proper inner product, which serves as a key tool for nonparametric inference of the unknown parameter/function. Asymptotic properties of the resulting estimate of the unknown log-hazard function are justified. Furthermore, the local confidence interval and simultaneous confidence band of the unknown log-hazard function are provided, along with a local and global likelihood ratio tests. We also investigate issues related to the asymptotic efficiency. Extensive simulations have been conducted to verify the theories.

2. In Chapter 3, the nonparametric inference approach in Chapter 2 is extended to handle interval censored data. Chapter 3 mainly focuses on the current status data, that is the case-one interval censored data. What difference in this chapter is that we target at the cumulative hazard function, instead of the log-hazard function. One key step is to derive an appropriate inner product. With the inner product defined satisfactorily, we derive a functional Bahadur representation and establish the asymptotic properties of the resulting estimate of the cumulative hazard function. In particular, the global asymptotic properties of the resulting estimator are shown under certain regularity conditions. A likelihood ratio test is also provided. Numerous simulations have been conducted to verify the theories.

4.2 Future Work

We may pursue along the following directions for future work.

1. Case-two interval censored data are quite common but much more complex than the current status data in survival analysis. To the best of our knowledge, there is limited report on the asymptotic properties of the smoothed estimation of the survival function or the hazard function with case-two interval censored data. In this regard, we shall strive to generalize the approach for estimating the survival function or the hazard function with case-two interval censored data, and carry out corresponding statistical inference for the target function.

2. Functional data regression is fast developing. Proportional hazard functional regression model is one of the most widely used models, in which the response refers to the time-to-event. For such a model, the penalized B-spline method has been used to estimate the functional coefficients To the best of our knowledge, there is no results reported in the literature on the theoretical properties of the estimate of the entire function. Hence, we consider to extend our approach to this kind of semiparametric models. Moreover, we also consider to adopt some penalty function to select important non-functional covariates. In the meanwhile, we may apply the penalized B-spline method to control the roughness of the functional coefficients. As in some case, the functional coefficients may remain a constant in certain interval. Another penalty function may also be im-

posed on the B-spline coefficients to control the shape of the resulting function.

3. Nowadays, ultrahigh-dimensional data, in which the number of candidate predictors or parameters may be at the exponential rate of the sample size, arise in many fields of modern science. As the big data brings to us its own features, such as heterogeneity, spurious correlation, noise accumulation etc, traditional penalized method may not be suitable both in theory and in computation. Hence, one need to reduce the high dimensionality to a moderate scale by some screening techniques. However, statistical inference for high dimensional survival data after screening remains challenging. We shall work along this direction for future extension and refinement .

Bibliography

Aalen, O. (1976), "Nonparametric inference in connection with multiple decrement models," *Scandinavian Journal of Statistics*, 3, 15–27.

Alioum, A. and Commenges, D. (1996), "A proportional hazards model for arbitrarily censored and truncated data," *Biometrics*, 52, 512–524.

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.

Anderson, J. and Senthilselvan, A. (1980), "Smooth estimates for the hazard function," *Journal of the Royal Statistical Society. Series B (Methodological)*, 42, 322–327.

Antoniadis, A. (1989), "A penalty method for nonparametric estimation of the intensity function of a counting process," *Annals of the Institute of Statistical Mathematics*, 41, 781–807.

Antoniadis, A. and Grégoire, G. (1990), "Penalized likelihood estimation for rates with censored survival data," *Scandinavian Journal of Statistics*, 17, 43–63.

Antoniadis, A., Grégoire, G., and Nason, G. (1999), "Density and hazard rate estimation for right-censored data by using wavelet methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 63–84.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W., Silverman, E., et al. (1955), "An empirical distribution function for sampling with incomplete information," *The Annals of Mathematical Statistics*, 26, 641–647.

Banerjee, M. and Wellner, J. A. (2005), "Confidence intervals for current status data," *Scandinavian Journal of Statistics*, 32, 405–424.

Beadle, G., Harris, J., Comes, S., Henderson, C., Silver, B., and Hellman, S. (1984), "The effect of adjuvant chemotherapy on the cosmetic results after primary radiation treatment for early breast cancer: a preliminary analysis," *International Journal of Radiation Oncology* Biology* Physics*, 10, 2131–2137.

Beran, R. (1981), "Nonparametric regression with randomly censored survival data." *Technical Report, Univ. California, Berkeley.*

Berkson, J. and Gage, R. P. (1952), "Survival curve for cancer patients following treatment," *Journal of the American Statistical Association*, 47, 501–515.

Betensky, R. A., Rabinowitz, D., and Tsiatis, A. A. (2001), "Computationally simple accelerated failure time regression for interval censored data," *Biometrika*, 88, 703–711.

Betensky, R. A., Lindsey, J. C., Ryan, L. M., and Wand, M. (2002), "A local likelihood proportional hazards model for interval censored data," *Statistics in Medicine*, 21, 263–275.

Bitouze, D., Laurent, B., and Massart, P. (1999), "A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator," *In Annales de l'IHP Probabilites et statistiques*, 35, 735–763.

Blum, J. and Susarla, V. (1980), "Maximal deviation theory of density and failure rate function estimates based on censored data," *Journal of Multivariate Analysis*, 5, 213–222.

Böhmer, P. E. (1912), "Theorie der unabhängigen Wahrscheinlichkeiten," *In Rapports Memoires et Proces verbaux de Septieme Congres International dActuaires Amsterdam*, 2, 327–343.

Böhning, D., Schlattmann, P., and Dietz, E. (1996), "Interval censored data: A note on the nonparametric maximum likelihood estimator of the distribution function," *Biometrika*, 83, 462–466.

Braun, J., Duchesne, T., and Stafford, J. E. (2005), "Local likelihood density estimation for interval censored data," *Canadian Journal of Statistics*, 33, 39–60.

Breslow, N. (1974), "Covariance analysis of censored survival data," *Biometrics*, 30, 89–99.

Breslow, N. and Crowley, J. (1974), "A large sample study of the life table and product limit estimates under random censorship," *The Annals of Statistics*, 2, 437–453.

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

Burke, M. D. and Horvath, L. (1982), "Density and failure rate estimation in a competing risks model," *Sankhya:The Indian Journal of Statistics, Series A*, 46, 135–154.

Cai, T. and Betensky, R. A. (2003), "Hazard Regression for Interval-Censored Data with Penalized Spline," *Biometrics*, 59, 570–579.

Cai, Z. (1998), "Kernel density and hazard rate estimation for censored dependent data," *Journal of Multivariate Analysis*, 67, 23–34.

Cox, D. D. and O'Sullivan, F. (1990), "Asymptotic analysis of penalized likelihood and related estimators," *The Annals of Statistics*, 18, 1676–1695.

Cox, D. R. (1972), "Regression models and life tables (with discussion)," *Journal of the Royal Statistical Society*, 34, 187–220.

Cox, D. R. (1975), "Partial likelihood," *Biometrika*, 62, 269–276.

Cox, D. R. and Oakes, D. (1984), *Analysis of Survival Data*, vol. 21, CRC Press.

Csörgö, S. and Horváth, L. (1980), "Random censorship from the left," *Studia Sc. Math. Hung. 15, 397*, 491.

Cutler, S. J. and Ederer, F. (1958), "Maximum utilization of the life table method in analyzing survival," *Journal of Chronic Diseases*, 8, 699–712.

Dabrowska, D. M. (1987), "Non-parametric Regression with Censored Survival Time Data," *Scandinavian Journal of Statistics*, 14, 181–197.

Dabrowska, D. M. (1989), "Uniform consistency of the kernel conditional Kaplan-Meier estimate," *The Annals of Statistics*, 17, 1157–1167.

Datta, S., Satten, G. A., and Williamson, J. M. (2000), "Consistency and asymptotic normality of estimators in a proportional hazards model with interval censoring and left truncation," *Annals of the Institute of Statistical Mathematics*, 52, 160–172.

Dave, S. S., Wright, G., Tan, B., Rosenwald, A., Gascoyne, R. D., Chan, W. C., Fisher, R. I., Braziel, R. M., Rimsza, L. M., Grogan, T. M., et al. (2004), "Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells," *New England Journal of Medicine*, 351, 2159–2169.

De Gruttola, V. and Lagakos, S. W. (1989), "Analysis of doubly-censored survival data, with application to AIDS," *Biometrics*, 45, 1–11.

Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J.-J., Sandhu, S., Guppy, K. H., Lee, S., and Froelicher, V. (1989), "International application of a new probability algorithm for the diagnosis of coronary artery disease," *The American journal of cardiology*, 64, 304–310.

Eeden, V. (1956), "Maximum likelihood estimation of ordered probabilities," 59, 444–455.

Eeden, V. (1957), "Maximum likelihood estimation of partially or completely ordered parameters. I," 60, 128–136.

Efron, B. (1967), "The two sample problem with censored data," *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 4, 831–853.

Eubank, R. L. (1999), *Nonparametric regression and spline smoothing*, CRC press.

Finkelstein, D. M. (1986), "A proportional hazards model for interval-censored failure time data," *Biometrics*, 42, 845–854.

Fleming, T. R. and Harrington, D. P. (2011), *Counting processes and survival analysis*, John Wiley and Sons.

Földes, A., Rejtö, L., and Winter, B. (1981), "Strong consistency properties of nonparametric estimators for randomly censored data, II: Estimation of density and failure rate," *Periodica Mathematica Hungarica*, 12, 15–29.

Fosen, J., Borgan, Ø., Weedon-Fekjær, H., and Aalen, O. O. (2006), "Dynamic analysis of recurrent event data using the additive hazard model," *Biometrical Journal*, 48, 381–398.

Galbraith, S., Daniel, J. A., and Vissel, B. (2010), "A study of clustered data and approaches to its analysis," *The Journal of Neuroscience*, 30, 10601–10608.

Gehan, E. A. (1969), "Estimating survival functions from the life table," *Journal of Chronic Diseases*, 21, 629–644.

Gentleman, R. and Geyer, C. J. (1994), "Maximum likelihood for interval censored data: Consistency and computation," *Biometrika*, 81, 618–623.

Ghosh, D. (2001), "Efficiency considerations in the additive hazards model with current status data," *Statistica Neerlandica*, 55, 367–376.

Gieser, P. W., Chang, M. N., Rao, P. V., Shuster, J. J., and Pullen, J. (1998), "Modelling cure rates using the Gompertz model with covariate information," *Statistics in Medicine*, 17, 831–839.

Goggins, W. B., Finkelstein, D. M., Schoenfeld, D. A., and Zaslavsky, A. M. (1998), "A Markov chain Monte Carlo EM algorithm for analyzing interval-censored data under the Cox proportional hazards model," *Biometrics*, 54, 1498–1507.

Gomez, G., Julià, O., Utzet, F., and Moeschberger, M. L. (1992), "Survival analysis for left censored data," 211, 269–288.

Gómez, G., Julià, O., and Utzet, F. (1994), "Asymptotic properties of the left Kaplan-Meier estimator," *Communications in Statistics-Theory and Methods*, 23, 123–135.

Goodall, R. L., Dunn, D. T., and Babiker, A. G. (2004), "Interval-censored survival time data: confidence intervals for the non-parametric survivor function," *Statistics in Medicine*, 23, 1131–1145.

Gray, R. J. (1992), "Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis," *Journal of the American Statistical Association*, 87, 942–951.

Greenhouse, J. B. and Silliman, N. P. (1996), "Applications of a mixture survival model with covariates to the analysis of a depression prevention trial," *Statistics in Medicine*, 15, 2077–2094.

Groeneboom, P. (1987), "Asymptotics for interval censored observations," *Report*, 87, 18.

Groeneboom, P. and Wellner, J. A. (1992), *Information bounds and nonparametric maximum likelihood estimation*, vol. 19, Springer Science and Business Media.

Groeneboom, P. and Wellner, J. A. (2012), "Computing Chernoff's distribution," *Journal of Computational and Graphical Statistics*.

Groeneboom, P., Jongbloed, G., Witte, B. I., et al. (2010), "Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model," *The Annals of Statistics*, 38, 352–387.

Groeneboom, P., Jongbloed, G., et al. (2015), "Nonparametric confidence intervals for monotone functions," *The Annals of Statistics*, 43, 2019–2054.

Gu, C. (1996), "Penalized likelihood hazard estimation: a general procedure," *Statistica Sinica*, 6, 861–876.

Hauck, W. W., McKee, L. J., and Turner, B. J. (1997), "Two-part survival models applied to administrative data for determining rate of and predictors for maternal-child transmission of HIV," *Statistics in Medicine*, 16, 1683–1694.

Hoeffding, W. (1963), "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, 58, 13–30.

Hoel, D. G. and Walberg, H. E. (1972), "Statistical analysis of survival experiments," *Journal of the National Cancer Institute*, 49, 361–372.

Huang, J. (1999), "Asymptotic properties of nonparametric estimation based on partly interval-censored data," *Statistica Sinica*, 9, 501–519.

Huang, J. et al. (1996), "Efficient estimation for the proportional hazards model with interval censoring," *The Annals of Statistics*, 24, 540–568.

Huang, J. and Rossini, A. (1997), "Sieve estimation for the proportional-odds failure-time regression model with interval censoring," *Journal of the American Statistical Association*, 92, 960–967.

Huang, J. and Wellner, J. A. (1997), "Interval censored survival data: a review of recent progress," 163, 123–169.

Huber-Carol, C. and Vonta, I. (2004), "Frailty models for arbitrarily censored and truncated data," *Lifetime Data Analysis*, 10, 369–388.

Iglewicz, B. and Hoaglin, D. C. (1993), *How to detect and handle outliers*, vol. 16, Asq Press.

Ingster, Y. I. (1993), "Asymptotically minimax hypothesis testing for non-parametric alternatives. I, II, III," *Math. Methods Statist*, 2, 85–114.

Jacobsen, M. (2012), *Statistical analysis of counting processes*, vol. 12, Springer Science and Business Media.

Jewell, N. P. and Van Der Laan, M. (2004), "Case-control current status data," *Biometrika*, 91, 529–541.

Johnson, N. L. and Kotz, S. (1970), *Distributions in Statistics: Continuous Univariate Distributions*, vol. 2, Houghton Mifflin.

Joly, P., Commenges, D., and Letenneur, L. (1998), "A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia," *Biometrics*, 54, 185–194.

Jong, D. P. (1987), "A central limit theorem for generalized quadratic forms," *Probability Theory and Related Fields*, 75, 261–277.

Jongbloed, G. (1998), "The iterative convex minorant algorithm for nonparametric estimation," *Journal of Computational and Graphical Statistics*, 7, 310–321.

Kalbfleisch, J. and Prentice, R. (1980), "The Statistical Analysis of Failure Time Data. New York: John Wiley and Sons," *KalbfleischThe Statistical Analysis of Failure Time Data1980.*

Kalbfleisch, J. D. and Prentice, R. L. (1973), "Marginal likelihoods based on Cox's regression and life model," *Biometrika*, 60, 267–278.

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

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

Kim, C., Park, B. U., Kim, W., and Lim, C. (2003), "Bezier curve smoothing of the Kaplan-Meier estimator," *Annals of the Institute of Statistical Mathematics*, 55, 359–367.

Kimura, D. K. (1972), *Fourier Series Methods for Censored Data*, Ph.D. Dissertation, University of LJashington.

Klein, J. P. and Moeschberger, M. L. (1997), "Statistics for biology and health," *Survival Analysis: Techniques for Censored and Truncated data.*

Koo, J.-Y., Kooperberg, C., and Park, J. (1999), "Logspline density estimation under censoring and truncation," *Scandinavian Journal of Statistics*, 56, 87–105.

Kooperberg, C. and Stone, C. J. (1992), "Logspline density estimation for censored data," *Journal of Computational and Graphical Statistics*, 1, 301–328.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995), "Hazard regression," *Journal of the American Statistical Association*, 90, 78–94.

Kosorok, M. (2008), "Introduction to Empirical Processes and Semiparametric Inference Springer: New York," .

Kronmal, R. and Tarter, M. (1968), "The estimation of probability densities and cumulatives by Fourier series methods," *Journal of the American Statistical Association*, 63, 925–952.

Lai, T. L. and Ying, Z. (1991a), "Large sample theory of a modified Buckley-James estimator for regression analysis with censored data," *The Annals of Statistics*, 19, 1370–1402.

Lai, T. L. and Ying, Z. (1991b), "Rank regression methods for left-truncated and right-censored data," *The Annals of Statistics*, 19, 531–556.

Lawless, J. F. (2011), *Statistical models and methods for lifetime data*, vol. 362, John Wiley and Sons.

Li, L. (2002), "Hazard rate estimation for censored data by wavelet methods," *Communications in Statistics-Theory and Methods*, 31, 943–960.

Li, L. (2007), "On the minimax optimality of wavelet estimators with censored data," *Journal of Statistical Planning and Inference*, 137, 1138–1150.

Li, L. and Pu, Z. (2003), "Rank estimation of log-linear regression with interval-censored data," *Lifetime Data Analysis*, 9, 57–70.

Li, L., Watkins, T., and Yu, Q. (1997), "An EM Algorithm for Smoothing the Self-consistent Estimator of Survival Functions with Interval-censored Data," *Scandinavian Journal of Statistics*, 24, 531–542.

Lin, D. and Ying, Z. (1994), "Semiparametric analysis of the additive risk model," *Biometrika*, 81, 61–71.

Lin, D., Oakes, D., and Ying, Z. (1998), "Additive hazards regression with current status data," *Biometrika*, 85, 289–298.

Lin, D. Y. and Wei, L.-J. (1989), "The robust inference for the Cox proportional hazards model," *Journal of the American Statistical Association*, 84, 1074–1078.

Lin, D. Y., Wei, L.-J., and Ying, Z. (1993), "Checking the Cox model with cumulative sums of martingale-based residuals," *Biometrika*, 80, 557–572.

Lin, Y. and Chen, K. (2013), "Efficient estimation of the censored linear regression model," *Biometrika*, 100, 525–530.

Mann, N. R., Singpurwalla, N. D., and Schafer, R. E. (1974), *Methods for statistical analysis of reliability and life data*, Wiley.

Martinussen, T. and Scheike, T. H. (2002), "Efficient estimation in additive hazards regression with current status data," *Biometrika*, 89, 649–658.

McNichols, D. T. and Padgett, W. J. (1981), "Kernel density estimation under random censorship," *Statistics Technical Report*, 74.

McNichols, D. T. and Padgett, W. J. (1982a), "Maximum likelihood estimation of un1modal and decreasing densities based on arbitrarily right-censored data," *Communications in Statistics-Theory and Methods*, 11, 2259–2270.

McNichols, D. T. and Padgett, W. J. (1982b), "A modified kernel density estimator for randomly right-censored data," *Statistics Technical Report*, 90.

Muller, H. G. and Wang, J. L. (1994), "Hazard rate estimation under random censoring with varying kernels and bandwidths," *Biometrics*, 50, 61–76.

Nelson, W. (1969), "azard plotting for incomplete failure data," *Journal of Quality Technology*, 1, 27–52.

Nelson, W. (1972), "Theory and applications of hazard plotting for censored failure data," *Technometrics*, 42, 12–25.

O'Sullivan, F. (1988), "Fast computation of fully automated log-density and log-hazard estimators," *SIAM Journal on Scientific and Statistical Computing*, 9, 363–379.

O'Sullivan, F., Yandell, B. S., and Raynor Jr, W. J. (1986), "Automatic smoothing of regression functions in generalized linear models," *Journal of the American Statistical Association*, 81, 96–103.

Pan, W. (2000), "Smooth estimation of the survival function for interval censored data," *Statistics in Medicine*, 19, 2611–2624.

Pan, W. and Chappell, R. (2002), "Estimation in the Cox Proportional Hazards Model with Left-Truncated and Interval-Censored Data," *Biometrics*, 58, 64–70.

Parzen, E. (1962), "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, 33, 1065–1076.

Patil, P. (1997), "Nonparametric hazard rate estimation by orthogonal wavelet methods," *Journal of Statistical Planning and Inference*, 60, 153–168.

Peto, R. (1973), "Experimental survival curves for interval-censored data," *Applied Statistics*, 22, 86–91.

Prentice, R. L. (1978), "Linear rank tests with right censored data," *Biometrika*, 65, 167–179.

Pu, Z. and Li, L. (1999), "Regression models with arbitrarily interval-censored observations," *Communications in Statistics-Theory and Methods*, 28, 1547–1563.

Rabinowitz, D., Tsiatis, A., and Aragon, J. (1995), "Regression with interval-censored data," *Biometrika*, 82, 501–513.

Rabinowitz, D., Betensky, R. A., and Tsiatis, A. A. (2000), "Using conditional logistic regression to fit proportional odds models to interval censored data," *Biometrics*, 56, 511–518.

Ritov, Y. (1990), "Estimation in a linear regression model with censored data," *The Annals of Statistics*, 18, 303–328.

Rosenberg, P. S. (1995), "Hazard function estimation using B-splines," *Biometrics*, 51, 874–887.

Sánchez-Sellero, C., González-Manteiga, W., and Cao, R. (1999), "Bandwidth selection in density estimation with truncated and censored data," *Annals of the Institute of Statistical Mathematics*, 51, 51–70.

Satten, G. A. (1996), "Rank-based inference in the proportional hazards model for interval censored data," *Biometrika*, 83, 355–370.

Schick, A. and Yu, Q. (2000), "Consistency of the GMLE with Mixed Case Interval-Censored Data," *Scandinavian Journal of Statistics*, 27, 45–55.

Schumaker, L. (1981), *Spline functions: basic theory. 1981.*

Shang, Z. and Cheng, G. (2013), "Local and global asymptotic inference in smoothing spline models," *The Annals of Statistics*, 41, 2608–2638.

Shen, X. (1998), "Propotional odds regression and sieve maximum likelihood estimation," *Biometrika*, 85, 165–177.

Spierdijk, L. (2008), "Nonparametric conditional hazard rate estimation: a local linear approach," *Computational Statistics and Data Analysis*, 52, 2419–2434.

Sun, J. (2001), "Variance estimation of a survival function for interval-censored survival data," *Statistics in Medicine*, 20, 1249–1257.

Sun, J. (2007), *The statistical analysis of interval-censored failure time data*, Springer Science and Business Media.

Tanner, M. A. and Wong, W. H. (1984), "Data-based nonparametric estimation of the hazard function with applications to model diagnostics and exploratory analysis," *Journal of the American Statistical Association*, 79, 174–182.

Tarter, M. E. (1979), "Trigonometric maximum likelihood estimation and application to the analysis of incomplete survival information," *Journal of the American Statistical Association*, 74, 132–139.

Thompson, M. and Godambe, V. (1974), "Likelihood Ratio vs. Most Powerful Rank Test a Two-Sample Problem in Hazard Analysis," *Sankhyā: The Indian Journal of Statistics, Series A*, 36, 13–40.

Tibshirani, R. and Hastie, T. (1987), "Local likelihood estimation," *Journal of the American Statistical Association*, 82, 559–567.

Tsiatis, A. A. (1990), "Estimating regression parameters using linear rank tests for censored data," *The Annals of Statistics*, 18, 354–372.

Turnbull, B. W. (1976), "The empirical distribution function with arbitrarily grouped, censored and truncated data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 38, 290–295.

van de Geer, S. (1993), "Hellinger-consistency of certain nonparametric maximum likelihood estimators," *The Annals of Statistics*, 21, 14–44.

Van Der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence*, Springer.

Vandal, A. C., Gentleman, R., and Liu, X. (2005), "Constrained estimation and likelihood intervals for censored data," *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 33, 71–83.

Wahba, G. (1983), "Bayesian" confidence intervals" for the cross-validated smoothing spline," *Journal of the Royal Statistical Society. Series B (Methodological)*, 45, 133–150.

Wang, W. and Ding, A. A. (2000), "On assessing the association for bivariate current status data," *Biometrika*, 87, 879–893.

Ware, J. H. and Demets, D. L. (1976), "Reanalysis of some baboon descent data," *Biometrics*, 32, 459–463.

Wei, L. (1992), "The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis," *Statistics in Medicine*, 11, 1871–1879.

Wei, L., Ying, Z., and Lin, D. (1990), "Linear regression analysis of censored survival data based on rank tests," *Biometrika*, 77, 845–851.

Wellner, J. A. (2007), "On an exponential bound for the Kaplan-Meier estimator," *Lifetime Data Analysis*, 13, 481–496.

Wellner, J. A. and Zhan, Y. (1997), "A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data," *Journal of the American Statistical Association*, 92, 945–959.

Whittemore, A. S. and Keller, J. B. (1986), "Survival estimation using splines," *Biometrics*, 42, 495–506.

Xue, H., Lam, K., Cowling, B. J., and de Wolf, F. (2006), "Semi-parametric accelerated failure time regression analysis with application to interval-censored HIV/AIDS data," *Statistics in Medicine*, 25, 3850–3863.

Yandell, B. S. (1981), *Non-parametric inference for rates and densities with censored serial data*, Doctoral dissertation, University of California, Berkeley.

Yu, Q., Schick, A., Li, L., and Wong, G. Y. (1998), "Asymptotic properties of the GMLE with case 2 interval-censored data," *Statistics and Probability Letters*, 37, 223–228.

Yu, Q., Li, L., and Wong, G. Y. (2000), “On Consistency of the Self-Consistent Estimator of Survival Functions with Interval-Censored Data,” *Scandinavian Journal of Statistics*, 27, 35–44.

Zeng, D. and Lin, D. (2007), “Efficient estimation for the accelerated failure time model,” *Journal of the American Statistical Association*, 102, 1387–1396.